Journal of Accounting and Economics 75 (2023) 101526

Contents lists available at ScienceDirect

Journal of Accounting and Economics

journal homepage: www.journals.elsevier.com/journal-of-accounting-and-economics

Predictive analytics and centralization of authority[☆]

Eva Labro [a], Mark Lang [a,*], James D. Omartian [b]

[a] Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Campus Box 3490, McColl Building, Chapel Hill, NC 27599, USA

[b] Ross School of Business, University of Michigan, 701 Tappan Avenue, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 12 March 2019
Received in revised form 6 July 2022
Accepted 25 July 2022
Available online 24 August 2022

JEL classification:
D23
L60
M52

Keywords:
Authority in organizations
Managerial incentives
Delegation

ABSTRACT

We examine the relation between plant-level predictive analytics use and centralization of authority for more than 25,000 manufacturing plants using proprietary US Census data. We focus on headquarters' authority over plants through delegation of decision-making and design of performance-based incentives. We find that increased predictive analytics use is associated with reduced delegation of decision-rights to local managers, increased centralization of control over data gathering and reduced plant managerial payrolls. In terms of incentives, predictive analytics use is associated with more accurate targets and tighter linkages between rewards to workers (performance-based bonuses, promotions and firings) and measured performance. Overall, our findings suggest that predictive analytics use is associated with increased centralization of authority in headquarters.

1. Introduction

One of the most substantive recent innovations in manufacturing has been the expansion in the use of predictive analytics, driven by an increased ability to generate real-time data, a reduction in the cost of computing power necessary to store and process large datasets, and the ready availability of high quality software to conduct statistical analyses and machine learning (Laaper et al., 2018). Although there is no formal generally accepted definition, SAS, one of the leading providers, defines predictive analytics as "the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data."[1] Predictive analytics permit plants to convert large quantities of location-specific

☆ This research was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Triangle Census Research Data Center and the Michigan Census Research Data Center. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1532. All results have been reviewed to ensure that no confidential information is disclosed. We thank Wayne Guay (editor), Brian Miller (referee), John Bai (discussant), Susanna Gallani (discussant), Stephen Glaeser, Frank Verbeeten (discussant), Joanna Ho, Raffi Indjejikian, Venky Nagar, participants at the 2019 CETAFE conference at USC, the CAED conference in Michigan, the 2019 Harvard IMO conference, the 2019 GMARS symposium, the 2019 EIASM conference on Performance Measurement and Control in Nice, the 2019 Rotman CPA Ontario Centre for Accounting Innovation Research Conference and workshop participants at Carnegie Mellon University, Katholieke Universiteit Leuven, Maastricht University, the University of California Irvine, the University of Colorado, the University of Michigan, the University of North Carolina at Chapel Hill, and Virginia Tech for helpful comments.

* Corresponding author.

E-mail addresses: eva_labro@unc.edu (E. Labro), mark_lang@unc.edu (M. Lang), omartian@umich.edu (J.D. Omartian).

[1] In our primary data source, the Management Organization and Practices Survey (MOPS) from the US Census Bureau, predictive analytics are defined as, "statistical models that provide forecasts in areas such as demand, production and human resources."

https://doi.org/10.1016/j.jacceco.2022.101526
0165-4101/

data into real-time forecasts that can be used to optimize production efficiency, labor utilization, product mix, maintenance and quality control.[2]

While the practitioner literature motivates the adoption of predictive analytics largely in terms of direct efficiency and productivity gains (Dilda et al., 2017; Swaminathan, 2018; Trentmann, 2019), use of predictive analytics has significant implications for the organizational architecture of the firm through its effects on available information. Agency theory is predicated on information asymmetry between the principal (headquarters in our setting) and the agent (plant management). Headquarters exerts authority over the activities of plant managers either by directly assuming decision-making responsibility, or by guiding choices through performance-based incentives. By changing the information available across the organization, the use of predictive analytics has the potential to alter the power balance between headquarters and plant management—through changes in the delegation of decision-making and the design of performance-based incentive plans for plant managers. Our goal is to explore the relation between predictive analytics use and these two components of organizational architecture, building on the literature that establishes it is important to consider the specific character of the information provided (Garicano, 2000; Bloom et al., 2014).[3] In particular, we focus on what is the relation between predictive analytics use and (1) delegation of decision-making, and (2) performance-based incentive design.

In terms of delegation, headquarters entrusts some decisions to plant managers (e.g., hiring, marketing, sourcing) because local plant managers have an informational advantage. All else equal, the greater the local information advantage, the more decisions will be delegated. While use of predictive analytics likely changes the local information advantage, the direction of the shift is unclear. Predictive analytics results in large amounts of granular local information, potentially increasing the amount of information available to local managers (Abernethy et al., 2004; Bushman et al., 2000). However, unlike traditional local information which is more subjective and requires the local manager's expertise for interpretation, local information generated by predictive analytics tends to be more objective, quantitative, and easier to transmit to headquarters. This difference potentially reduces the need for delegation (Garicano, 2000), if headquarters can easily control, access, and interpret the output of predictive analytics. As a result, the relation between the use of predictive analytics and shifts in decision-making delegation is an empirical question.

In addition to directly intervening in decision-making, headquarters may indirectly influence plant manager choices through local incentives—motivating actions by plant managers that are consistent with headquarters' objectives. Prior literature on the controllability principle suggests that, when delegation is reduced, performance-based incentives become less effective because they increase managers' risk exposure to factors outside their control (Merchant, 1998). However, by enriching the information available centrally, predictive analytics permit headquarters to better evaluate local managers' performance and compare it to explicit and accurate targets, allowing headquarters to exert authority over plant managers by strengthening the link between measured performance and outcomes (e.g., bonuses, promotions, firings). As with the delegation of decision-making authority, the direction of the relation between predictive analytics and the extent to which incentives are linked to measured performance is an empirical question.

Confidential data from the Census allow us to examine these research questions for a sample of more than 25,000 manufacturing plants covering a representative swath of US manufacturing activity. For the first time in 2015, the Census' Management Organization and Practices Survey (MOPS) queried this sample of plants on their reliance on predictive analytics in 2010 and 2015. As a result, we can observe intertemporal changes in the use of predictive analytics across plants within a given firm, as well as detailed changes in centralization of both decision-making and performance-based incentive practices.

The advantages of using these Census data for our study are threefold. First, the sampling is randomized and stratified to ensure that it covers a representative spectrum of plants not limited by industry, geography or size, and includes both publicly- and privately-held firms. Response rates are high because compliance is required by law and census personnel follow up on non-responses repeatedly. Because responses are confidential (even from other branches of government) and there are strict penalties for unauthorized disclosure, there should be few reasons to misreport.

Second, since many plants in the sample are part of multi-plant firms, we can conduct analyses using firm- and plant-level controls, along with an extensive fixed effects structure. We use firm fixed effects (along with plant-industry and county fixed effects), to compare changes in predictive analytics among plants within a given firm, controlling for location and industry effects, which allows us to rule out a wide range of potential omitted variables such as firm-level changes, local economics and industry-wide shocks. Third, coupled with the Census' Annual Survey of Manufacturers (ASM) and the Census of Manufacturers (CMF), we can observe changes in managerial payroll, along with a wide range of control variables, such as capital expenditures, plant capacity, product mix and asset composition. Given the richness of the Census data, we can also employ causal regression forests (Wager and Athey, 2018; Athey et al., 2019), using machine learning to incorporate hundreds of potential covariates, along with interactions and nonlinearities.

[2] We include controls for a wide range of other information-related variables, including data driven decision making (Brynjolfsson and McElheran (2016a; 2016b)), capital information technology stock, instrumentation and managerial feedback. As discussed later, results are robust to a range of alternate measures and expanding our definition of predictive analytics to include a broader set of related information sources.

[3] The decisions to adopt predictive analytics and increase headquarters' authority are likely made simultaneously. Our interest is in understanding whether predictive analytics and headquarters authority work as complements (whether predictive analytics use facilitates headquarters authority) rather than whether predictive analytics use directly causes increased authority.

We provide consistent evidence that predictive analytics and decision-making centralization act as complements: greater use of predictive analytics is associated with increased centralization of decision-making, particularly over marketing (product mix, pricing and advertising) and human relations (hiring, bonuses and firing of plant employees). Also consistent with reduced delegation, primary control over specific plant-level data to be gathered is increasingly centralized in headquarters. Increases in predictive analytics usage are associated with decreases in plant management (but not staff) payrolls, consistent with increased centralization reducing the need for local managerial decision-making.

In addition to a link between predictive analytics usage and centralizing decision-making authority directly, we also observe increases in headquarters' indirect influence on local managers' decision-making through incentives. Plants adopting predictive analytics are more likely to initiate use of performance targets, and existing targets become more accurate and are communicated more saliently throughout the plant. Increases in predictive analytics usage are also associated with increases in the extent to which incentives for plant management (e.g., bonuses, promotions and terminations) are explicitly tied to measured performance, consistent with more accurate measurement permitting more effective incentive design.

We make several contributions to the literature. Most directly, we provide evidence of an association between predictive analytics and the authority headquarters exerts over plant management. Contrary to much of the prior literature which suggests that local information production is associated with greater delegation (Abernethy et al., 2004; Bushman et al., 2000), our results suggest that predictive analytics use is associated with centralization of authority in headquarters. The difference likely results from the fact that information from traditional sources (e.g., personal observation by the manager), was "softer" and more subjective. However, predictive analytics uses large datasets with sophisticated processing to help headquarters interpret and contextualize local information, reducing the advantage of local insight. As predictive analytics increase in sophistication and pervasiveness, our results suggest a likely increase in centralized authority.[4]

Second, our results suggest a complementary relationship between centralization of decision-making and performance-based incentives for local plant management. Taken on its face, our finding that predictive analytics use is associated with reduced delegation would suggest that incentives for local managers should be less tied to local performance because those managers face increased exposure to risks outside their control (Merchant, 1998). However, our results suggest that any increased risk associated with loss of control is limited by the improved ability of headquarters to objectively parse out the effect of managers' actions relative to factors outside their control.

Finally, prior literature suggests that improved information technology can facilitate the oversight and automation of routine tasks such as warehouse, clerical and customer service functions (Bresnahan et al., 2002). However, the implications for more complex managerial employment are less clear. Our results indicate that predictive analytics use is associated with reductions in agency and employment for local managers, suggesting implications even for workers in higher-skilled managerial positions.

2. Background and research questions

Predictive analytics use statistical methods such as machine learning to observe key patterns within data with the goal of forecasting likely outcomes. Predictive analytics are particularly suited to manufacturing because the plant environment facilitates the collection of large amounts of real-time data and the processes naturally lend themselves to rapid adjustment as conditions change.[5] The recent rise in predictive analytics use reflects a combination of an increased ability to track real-time data, a reduction in the cost of computing power necessary to store and process large datasets, and the ready availability of high quality software to conduct statistical analyses and machine learning. Beyond improving efficiency, changes in the information environment likely also affect the allocation of authority within the firm. Our study focuses on the authority headquarters exerts over plant managers through decision-making delegation and performance-based incentives.

We begin by providing descriptive evidence on the types of plants, firms and industries that significantly expanded their use of predictive analytics between 2010 and 2015. While not a primary focus, this analysis allows us to understand the drivers of predictive analytics use for our later analyses. Our results build on prior work (Brynjolfsson and McElheran, 2016a,b) suggesting that adoption of IT systems more generally is likely to rely on implementation by sophisticated employees, so we expect greater use by plants with a highly educated workforce. We expect greater use of predictive analytics by firms that are larger and not family-owned since those firms are more likely to have the resources required for effective implementation. Because predictive analytics use historical data to predict outcomes, they are likely to be more effective for firms in stable industries.

Our primary research questions focus on the relation between predictive analytics use and the exertion of authority by headquarters over local plant management. Information from predictive analytics differs from many other sources in that it is

[4] Our results are based on manufacturing firms and, therefore, may not generalize to specialized service sectors (e.g., medicine, counseling and education) where context-specific "soft" information may be more difficult to communicate.

[5] For example, predictive analytics can forecast short-term fluctuations in demand and potential bottlenecks. As conditions change, labor and machine use can be optimized, product mix can be shifted, maintenance and repair can be anticipated and incorporated, and anomalies can be detected and mitigated.

both local to conditions in the specific plant (potentially increasing local information advantage) but also easily transmitted and interpreted by headquarters (potentially reducing local information advantage). We focus on two attributes of headquarters' authority over plant managers: (1) centralization of decision making, and (2) incentive design.[6]

Research Question 1. *Is predictive analytics use associated with the extent of delegation over decision-making from headquarters to local managers?*

The direction of the association between predictive analytics and the delegation of decision-making authority from headquarters to local managers is not obvious a priori. The traditional argument is that enhanced local information is associated with increased delegation.[7] However, central to that argument is the notion that local information is difficult to effectively transmit, so it increases the information advantage of local managers. Garicano (2000) develops a theoretical model in which cheaper acquisition of knowledge by local employees facilitates delegation but cheaper transmission of knowledge facilitates centralization of decision-making. Bloom et al. (2014) examine those predictions empirically using adoption of ERP and CAD/CAM systems as proxies for technologies that facilitate local data acquisition, and adoption of intranets as a proxy for technology that facilitates communication. Consistent with predictions, ERP and CAD/CAM adoption is associated with delegation of decision-making and intranets are associated with centralization of decision-making.

Predictive analytics use is characterized by both increasing local information while also providing data that are easily transmitted to headquarters. As a result, we view the direction of the relation between predictive analytics and delegation of decision-making authority to be an open empirical question. In our tests, we examine the relation between changes in predictive analytics use and changes in the delegation of decision making to plants, the choice of whether the data needed to run the plant is controlled centrally, and the size of the managerial payroll at the plant.

Research Question 2. *Is predictive analytics use associated with changes in incentive design and performance targets for plant managers?*

While centralizing decision rights in headquarters is one way for headquarters to exercise authority over plant managers, an alternate mechanism is to align plant managers' incentives more closely to headquarters' objectives. Headquarters structures incentives to reward preferred, and penalize undesired, outcomes.

Several characteristics of predictive analytics likely alter the ability of headquarters to implement effective incentives, both in terms of rewarding measured performance and in terms of setting performance targets. First, use of statistical modeling and measurement underlying predictive analytics permit headquarters to better link plant managers' actions to realized outcomes, improving insight into managerial performance. One of the primary challenges in a principal/agent setting is accurately disentangling the effect of managerial action from the effects of outside forces, reducing noise in performance measures. Predictive analytics are designed to help identify causal links and abstract from the effects of idiosyncratic noise. As a result, predictive analytics permit headquarters to more accurately link managerial actions to outcomes and, therefore, to more tightly link monetary bonuses and career-based incentives (promotions and terminations) for local managers to measured performance.[8]

In addition to improving ex post performance measures, the forward-looking nature of predictive analytics also facilitates setting effective performance targets. In many settings, it is difficult to set effective targets because the information necessary to set optimal targets resides with local managers who have incentives to bias targets or build in slack (Baiman and Evans, 1983; Merchant et al., 2018). Because the information from predictive analytics is objective, easily transmissible and forward looking, we expect predictive analytics usage to be associated with greater use of performance targets, more accurate targets and greater awareness of targets throughout the plant hierarchy.

3. Empirical approach

We draw on confidential microdata gathered by the US Census Bureau to measure predictive analytics usage and other plant and firm characteristics. While known primarily for its decennial tabulation of individuals, Census also gathers highly detailed data on US economic activity. These data are published in aggregate form for the benefit of policy makers and businesses making operational decisions. Researchers on approved projects may apply for access to the underlying microdata. We use a panel of manufacturing plant-level observations to explore the relations between changes in predictive analytics use and changes in exerting authority in organizations.

[6] These two components are analogous to "decision-facilitating" and "decision-influencing" in the Demski and Feltham (1976) framework.

[7] For example, the accounting literature suggests that decision-making delegation will be greater when plants have greater access to local information, offering them the opportunity to react nimbly to changes in their local environment (Abernethy et al., 2004; Bushman et al., 2000).

[8] The prediction is not without tension. As noted earlier, prior literature on the controllability principle (Merchant, 1998) suggests that when local managers control fewer aspects of running the plant (e.g. because headquarters assumes decision-making or automates tasks that were previously under the manager's control), performance-based incentives increase managers' exposure to forces outside their control.(Moers, 2006; Nagar, 2002; Prendergast, 2002).

3.1. Data sources

Census gathers detailed information about domestic manufacturing establishments—both privately and publicly held.[9] A key feature of these data is that they are collected at the "establishment" (i.e. plant) level. Every five years the Census Bureau conducts a full Census of Manufactures (CMF), collecting operational data including sales, employees, payroll, inventories, capital expenditures, and operating expenses from all but the smallest plants. In other years, a stratified random sample of plants is surveyed (Annual Survey of Manufactures, or ASM). The survey population is stable for a five-year window, and the populations have significant overlap across windows, allowing us to construct a panel of plant-level data.

In 2015, Census sent the Management and Organizational Practices Survey (MOPS) with the ASM.[10] MOPS includes detailed questions on management practices, organizational hierarchy, data and decision-making, uncertainty, and background information. We rely on MOPS responses to measure the use of predictive analytics at the plant, as well as organizational authority.[11] For each question, respondents provide answers for 2010 and 2015, typically selected from a menu of choices. About 51,000 plants were surveyed and 71% provided valid responses. While this survey provides unparalleled detail for such a large and broad sample of plants, data for predictive analytics are only available for 2010 and 2015, and the 2010 data are "recall" data.[12] For inclusion in our main sample, we require valid MOPS responses and plant data from the 2010 and 2015 ASM. These restrictions leave us with roughly 26,500 individual plants across roughly 15,000 firms. Compared to other studies in the managerial accounting literature, our sample is one of the most representative and free of selection bias. Table 1 presents average values for operational characteristics of plants in our sample, and Appendix B provides specifics on the data collection methodology.

3.2. Research design

We employ a plant-level changes analysis to explore the relation between predictive analytics and the allocation of authority in organizations. The rich Census data allow us to control for a wide range of potential confounds, and having observations from multiple plants for the same firm allows us to conduct within-firm analyses. To address potential confounds, we use OLS regressions (with and without fixed effects) and causal random forests as our primary specifications; we assess alternate measures in our robustness analyses.

3.2.1. OLS regressions

Our baseline models regress changes in features of the plant's organizational authority on changes in predictive analytics usage.

$$\Delta OAFeature_p = \alpha_1 \Delta PredAnal_p + \beta Controls_p + \gamma Controls_f + \delta Controls_i + \varepsilon_p \tag{1}$$

Table 1
Descriptive Statistics of ASM/CMF Data. Average values for plant-level variables from sample plants over 2008–2016. All variables are defined in Appendix C.

Variable	Mean
ln ($TVS_{p,t}$)	10.460
ln ($NumProds_{p,t}$)	1.180
$MgtComp_{p,t}$	0.071
$StaffComp_{p,t}$	0.105

[9] These data underpin economic indicators such as the Federal Reserve's Industrial Production index, the Bureau of Labor Statistics' producer price index, and the Bureau of Economic Analysis's input-output tables.

[10] The MOPS was intended to be completed by the local plant manager (rather than by headquarters) to measure variation across plants within a firm. To that end, the survey was addressed to the "Plant Manager" and mailed to the physical address of the plant (rather than to a "business address" that might be located at headquarters). While it is possible that the survey might have been forwarded to headquarters, that seems unlikely given that the information required should have been readily available at the plant and completing the survey was not onerous. If surveys were forwarded to headquarters, it should not bias our results (other than perhaps understating within-firm variation).

[11] The MOPS survey instrument can be found here: https://www2.census.gov/programs-surveys/mops/technical-documentation/questionnaires/ma-10002_15_final_3-2-16.pdf.

[12] There are advantages and disadvantages associated with recall data. The fact that data for both 2015 and 2010 are reported on the same form, and likely by the same employee at the plant, increases the probability that definitions for constructs such as "predictive data analytics" are applied consistently over time. However, given "anchoring bias" (a tendency to rely too heavily on more recent salient data in recalling past events (Furnham and Boo, 2011)), our results may understate changes in predictive analytics use.

where p indexes plant, f indexes firm, and i indexes industry. $\Delta OAFeature_p$ is one of a number of ways in which headquarters exerts authority over plant managers, such as delegation of decision-making or tightening of performance incentives. Appendix C provides precise definitions of each feature we analyze. A detailed correspondence between the MOPS survey questions (with screenshots of the survey instrument) and our variables is included in Internet Appendix E.

We operationalize three measures of predictive analytics changes ($\Delta PredAnal_p$) based on the MOPS question: "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 select one frequency—never, yearly, monthly, weekly, or daily—individually for 2010 and 2015 which we transform in three ways.[13] $\Delta PAScore_p$ scales these frequencies from zero to one for each year and takes the difference from 2010 to 2015, thus capturing intensity changes in predictive analytics usage. Because the frequency choices on the Census form are non-linear, and the intensity may have non-linear effects as well, we also use an indicator $PAIncrease_p$ which takes a value of 1 if $\Delta PAScore_p$ is positive and 0 otherwise. To isolate changes on the extensive margin, we measure $StartPA_p$ as an indicator if the respondent from plant p checked that the plant did not use predictive analytics in 2010, but did (in any frequency) in 2015.[14]

We include a battery of plant-, firm-, and industry-level controls in our analyses. To control for operational differences and changes explaining the shifts in organizational authority, we include plant-level production variables. Specifically, we include the 2010 values of the natural log of total value of shipments, $ln\,(TVS_{p,2010})$, capital expenditure intensity ($TCE_{p,2010}/TVS_{p,2010}$), cost of materials ($CM_{p,2010}/TVS_{p,2010}$), and inventory ($TIE_{p,2010}/TVS_{p,2010}$). We also include percentage changes of these variables (without scaling by TVS) from 2010 to 2015. To ensure results are not driven by changes in the workforce, we include changes in management and staff education ($\Delta MgtEducation_p$; $\Delta StaffEducation_p$) and percentage changes in firm and plant payroll ($\frac{PlantPayroll_{p,2015}-PlantPayroll_{p,2010}}{PlantPayroll_{p,2010}}$; $\frac{FirmPayroll_{f,2015}-FirmPayroll_{f,2010}}{FirmPayroll_{f,2010}}$). To control for changes in growth opportunities, we include aggregate industry sales growth on a percentage basis in the plant's industry ($IndGrowth_p$).

To ensure that our results reflect changes in predictive analytics and not information technology in general, we control for the change in IT capital stock at the plant ($\Delta\ln(ITCapitalStock_p)$). To isolate the link with predictive analytics specifically and further rule out the effects of changes in data usage more broadly, we control for changes in data-driven decision-making, ΔDDD_p (Brynjolfsson and McElheran, 2016b), and the intensity of information from performance indicators from production technology or instruments, $\Delta ProdPerfIndScore_p$. Lastly, because we are interested in shifts in the hierarchical level of management authority and not changes in managerial intervention in general, we control for changes in the reliance on formal or informal feedback from managers $\Delta MgrFdbckScore_p$. For parsimony and because of disclosure restrictions from Census, we do not tabulate the coefficient estimates for our control variables, but they are included in the estimation of each model. For all specifications, we winsorize all ratios at 1% and 99%, and cluster standard errors at the firm level.

The high number of observations allows us to include fixed effects along multiple dimensions. We also estimate models in the form:

$$\Delta OAFeature_p = \alpha_1 \Delta PredAnal_p + \beta Controls_p + \zeta_f + \eta_i + \theta_c + \varepsilon_p \tag{2}$$

where c indexes county and all other subscripts remain the same. ζ_f, η_i, and θ_c correspond with firm, industry (4-digit NAICS), and county fixed effects. In most studies using firm fixed effects, identification comes from inter-temporal, intra-firm variation. However, because our unit of observation is the plant, in equation (2) identification comes from differences in changes in predictive analytics across plants within the same company. Coupled with a changes design, this approach is very powerful at removing potential confounds because the firm fixed effects absorb any changes in $\Delta OAFeature_p$ coming from firm-level shifts.[15] For example, changes in cost of capital, corporate strategy, C-suite management, and company-wide technology improvements are all captured and controlled for by the firm fixed effects. Industry fixed effects mitigate concerns about general changes in sector-level economic conditions driving our results.[16] County fixed effects mitigate potential concerns that changes in economic conditions in specific locations (e.g., economic growth or unemployment rates) could affect organizational authority. Because of collinearity, when we include fixed effects we drop firm-level and industry-level control variables from our models.

3.2.2. Causal forests

We are careful not to make strong causal statements because we recognize that predictive analytics and variables capturing delegation and incentive design are likely to coevolve over time. Our analysis suggests characteristics of plants

[13] As discussed in Section 4.4, we also conduct analyses based on changes in the breadth of data use ($\Delta Breadth_p$).

[14] A very small number of plants report using predictive analytics during 2010 but not in 2015, or a decline in predictive analytics usage. Our results are virtually identical if we construct $PAIncrease_p[StartPA_p]$ as the difference of two binary variables (1 = increase [start], 0 = no change, −1 = decrease [end] using predictive analytics). We do not tabulate these analyses because they increase the risk that specific individual responses could be inferred.

[15] To our knowledge Campbell et al. (2009), is the only other study in the accounting literature on delegation that has been able to focus on within-firm variation (although they are unable to observe changes). They use data from convenience store chains to understand why some stores are franchised while others are owned outright.

[16] Because industry is a plant-level variable and a firm may have plants operating in multiple industries, firm fixed effects do not subsume industry fixed effects.

increasing predictive analytics use, such as availability of resources and an educated workforce, and we control for those variables, along with an extensive fixed effects structure and a range of other factors suggested by prior literature and theory. However, we recognize that our OLS regressions may exclude potential determinants. There are hundreds of plant-specific characteristics in the Census data that we could have included in the analysis, but we only included those that we thought were most likely to affect our variables of interest to reduce the likelihood of overfitting. Additionally, OLS only permits linear effects, whereas the literature posits many organizational features have interactive or nonlinear effects (Prendergast, 2002).

To address the possibility that we may have excluded important controls or that interactive or nonlinear effects are important, we also employ a non-parametric, data driven, machine learning approach to model selection. Machine learning algorithms have successfully been used to develop prediction models, and recently have been adapted to econometric problems of causal inference (Athey and Imbens, 2017). Causal forests are an adaptation of the popular random forest algorithm (Breiman, 2001). Because causal forests are new to the accounting literature, we provide a detailed description of the methodology in Appendix D. Loosely speaking, we provide the algorithm with all of the available plant-level variables for 2010 and 2015 from the ASM and for 2010 from the MOPS (other than our variables of interest), and the causal forest algorithm infers the important correlates, their functional form and interactions based on out-of-sample prediction to ensure that we have efficiently controlled for any important plant characteristics without overfitting the control variables. Recent work has shown that the resulting estimator from the causal forest procedure is unbiased and asymptotically normal (Wager and Athey, 2018; Athey et al., 2019).

Specifically, we follow the methodology in Athey and Wager (2019) to estimate average treatment effects of our three operationalizations of $\Delta PredAnal_p$ on the various definitions of $\Delta OAFeature_p$. This approach has several attractive features. The non-parametric nature of causal forests allows us to incorporate non-linearities and interactions in both our outcome and treatment assignment estimations. Because regression forests are resistant to overfitting (because prediction is out of sample), we can provide all available covariates to the estimation procedure and let the data speak on which to include in the model and in what form.[17] The sample splitting and cross-validation allow us to make use of the full dataset, but be "honest" (i.e. using different subsamples to build a tree than to estimate predictions from it). This approach also allows us to adjust clustering effects (which are likely present at the firm level in our setting) into our standard error estimates. Practically speaking, we make 323 variables available to the causal forest model as outlined in Internet Appendix F. In essence, this controls for a massive number of potential confounds, including non-linear transformations and interactions among available variables.

While we are careful not to claim causality, the causal forest approach has the advantage of incorporating all available covariates in a parsimonious model. We acknowledge that there may be unavailable covariates leading some plants to adopt predictive analytics while others do not (e.g., unobservable idiosyncratic characteristics of plant managers). However, we are unaware of any clearly missing variables. Further, the robustness and conceptual consistency across specifications provides reassurance that results are not spurious.[18]

4. Results

All results using confidential Census data undergo review by Census disclosure personnel to insure there is no risk of inferring individual responses. Because each reported number is vetted to ensure confidentiality, we minimize nonessential quantitative data tabulated, limiting reporting of, for example, coefficients on control variables.

Table 2 presents descriptive statistics for key variables of interest. In 2015, 72.4% of manufacturing plants reported using predictive analytics (*UsePA*) and, in terms of intensity, the average plant relied on predictive analytics on slightly longer than a monthly frequency (*PAScore* = 0.478). These levels represent a substantial increase relative to 2010—on the extensive margin 7.6% of plants report commencing using predictive analytics, and looking at *PAScore*, which captures both changes in the extensive and intensive margins, we see an increase of 0.087 from 0.391, equating to the average plant roughly moving from a quarterly to a monthly usage frequency. Among all plants in the sample, 23.1% report heavier reliance on predictive analytics in 2015 relative to 2010 ($PAIncrease_p$), and 30.1% of firms report an increase for at least one plant in its *PAScore*.

4.1. Determinants of predictive analytics

We begin by presenting evidence on determinants of the use of predictive analytics. While this analysis is descriptive, there is not, to our knowledge, other broad sample evidence on use of predictive analytics. We report results for both levels and changes but, in subsequent analyses, focus on changes to better assist with identification. In Table 3 we report 2015 levels and changes analyses with *PAScore* as the dependent variable (results for *StartPA* and *PAIncrease* are very similar). In terms of

[17] As a result, the unconfoundedness assumption of Rosenbaum and Rubin (1983)—that after conditioning on observed confounders, treatment assignment is as good as random—is much more likely to hold.

[18] Another potential source of bias is survivorship. While we cannot completely rule this out, we model the decision to close a plant using plants that filled out the 2010 MOPS form. Using fitted values from the model, we select plants that filled out both the 2010 and 2015 MOPS forms that are most similar to the plants that close during that window. We overweight each of these matched surviving plants in our regressions for each corresponding plant that falls out of our sample, in effect adding back closed plants into our sample with similar surviving plants as a substitute. Our results are nearly identical under this re-weighting (untabulated).

Table 2
Descriptive Statistics of MOPS Data. Average values for 2015 levels and 2010 to 2015 changes in variables from MOPS. Levels are tabulated for the key constructs of interest and for the target accuracy variables. All variables are defined in Appendix C.

Variable	2015 Level	2010 to 2015 Change
$UsePA_p$	0.724	0.073
$StartPA_p$	0.076	
$PAScore_p$	0.478	0.087
$PAIncrease_p$	0.231	
$Breadth_p$		0.393
$ProdPerfIndScore_p$	0.751	0.062
$MgrFdbckScore_p$	0.801	0.050
$ChooseInfoLocal_p$	0.838	−0.003
$ChooseInfoHQ_p$	0.538	0.039
$DelegationScore_p$	0.397	−0.019
$StartTarget_p$	0.028	
$TargetAwareness_p$	0.698	0.166
$TargetEffortLittle_p$	0.030	−0.010
$TargetEffortSome_p$	0.105	−0.022
$TargetEffortNormal_p$	0.419	−0.007
$TargetEffortMoreThanNormal_p$	0.331	0.057
$TargetEffortExtreme_p$	0.052	−0.016
$Achieve_p$	0.706	0.002
$StartBonus_p$	0.054	
$Termination_p$	0.515	0.082
$Promotion_p$	0.821	0.028

firm-level determinants, results in Column (1) suggest that plants are more likely to rely on predictive analytics if the parent firm is larger (i.e., larger payroll, multinational or multi-plant firm) and younger. Family-run firms rely less on predictive analytics, consistent with lower managerial sophistication and fewer resources. There is a negative relation between use of predictive analytics and both industry growth and technological intensity, suggesting that predictive analytics use is less common in industries that are innovating and changing rapidly. Similar to the firm-level relation, larger plants are more likely to rely on predictive analytics, as are plants with higher-educated workforces.

Column (2) reports results with firm, industry, and county fixed effects, focusing on variation within firm and controlling for differences in industry and local economic conditions. Results are consistent in suggesting that, within firms, the plants that are likelier to employ predictive analytics are larger with better-educated staff (although manager education is insignificant).

Panel B reports results for changes in predictive analytics use based on changes in plant characteristics, both with and without fixed effects. Results are generally consistent with the results for levels. Use of predictive analytics increased for firms that increased in size and for firms that increased education levels of both managers and staff. While coefficient estimates are significant, the explanatory power of the plant-level variables is relatively low, suggesting that there is a significant idiosyncratic component to predictive analytics use.

Finally, to investigate the relation between predictive analytics and other information sources, we regress our predictive analytics variables on measures of the use of data from production technology and managerial feedback, along with the control variables from Table 3. Our concern is that the predictive analytics measures may simply reflect the effects of other information sources. Results, tabulated in Table IA.1 of the internet appendix, indicate a positive relation between use of predictive analytics and use of both performance indicators from production technology and managerial feedback.[19] However, the coefficient estimates on changes in other data sources are substantially less than one (ranging from 0.180 to 0.238) and the explanatory power of the regressions excluding fixed effects (but including controls) is 0.077 and 0.156, suggesting that the proportion of shared variation between predictive analytics and other information measures is low. To ensure that our predictive analytics measures are not capturing other aspects of data gathering, we include controls for the other information variables (*ProdPerfIndScore* and *MgrFdbckScore*) in all of our empirical analyses. In addition, we control for both the change in the logarithm of capital IT stock *ITCapitalStock* as defined by Brynjolfsson and McElheran (2016a) and the data-driven decision making measure *DDD* used by Brynjolfsson and McElheran (2016b) and Brynjolfsson and McElheran (2016a). Our results on the effects of predictive analytics are very similar irrespective of whether we control for other information, suggesting that predictive analytics use can be separated empirically from other sources of information.[20]

[19] A potential concern is whether it is possible to separate predictive analytics from the underlying instrumentation that permits measurement. To assess robustness, we also compute a variable combining use of predictive analytics with "performance indicators and instrumentation." Results are very similar to those reported in our primary analyses.

[20] In addition, we control for a wide range of other variables, including capital expenditures, plant capacity, product mix and asset composition to ensure that results are not driven by, for example, automation or significant plant renovation.

Table 3
Determinants of Predictive Analytics. This table presents regressions of predictive analytics on firm, plant and industry characteristics. Panel A tabulates 2015 levels analyses and panel B tabulates changes from 2010 to 2015. In both panels column (2) includes fixed effects for the plant's firm, 4-digit NAICS, and county. Standard errors are clustered at the firm level. All variables are defined in Appendix C.

Panel A: 2015 Levels

	$PAScore_{i,2015}$	
	(1)	(2)
ln ($FirmPayroll_{f,2015}$)	0.024***	
	(0.003)	
$MultinationalFirm_{f,2015}$	0.028***	
	(0.008)	
$FamilyCEO_{f,2015}$	−0.017***	
	(0.006)	
ln ($FirmAge_{f,2015}$)	−0.021**	
	(0.009)	
$MultiplantFirm_{f,2015}$	0.042***	
	(0.008)	
$IndGrowth_{i,2015}$	−0.080***	
	(0.013)	
$TechInd_i$	−0.066***	
	(0.007)	
$RemotePlant_p$	0.054***	0.015
	(0.006)	(0.013)
ln ($PlantAge_{p,2015}$)	−0.008	−0.007
	(0.006)	(0.011)
ln ($PlantPayroll_{p,2015}$)	0.036***	0.032***
	(0.003)	(0.006)
$MgtEducation_{p,2015}$	0.018**	−0.001
	(0.008)	(0.017)
$StaffEducation_{p,2015}$	0.043***	0.038***
	(0.009)	(0.021)
Constant	−0.025	
	(0.026)	
N	26,500	26,500
R^2	0.149	0.718
Fixed Effects:		
Firm	No	Yes
Industry	No	Yes
County	No	Yes

Panel B: 2010 to 2015 Changes

	$\Delta PAScore_i$	
	(1)	(2)
$GrowthFirmPayroll_f$	0.012***	
	(0.003)	
$\Delta IndGrowth_i$	−0.015***	
	(0.003)	
$GrowthPlantPayroll_p$	0.004	0.003
	(0.003)	(0.007)
$\Delta MgtEducation_p$	0.200***	0.163***
	(0.011)	(0.023)
$\Delta StaffEducation_p$	0.158***	0.152***
	(0.012)	(0.028)
Constant	0.692***	
	(0.015)	
N	26,500	26,500
R^2	0.053	0.597
Fixed Effects:		
Firm	No	Yes
Industry	No	Yes
County	No	Yes

4.2. Delegation of decision-making authority

4.2.1. Managerial decision-making

In our first main research question, we consider the relation between predictive analytics use and the delegation of formal decision-making rights from headquarters to plant management. Even if predictive analytics are primarily used to increase efficiency, the availability of the resulting data has the potential to change the power balance between management at the plant and headquarters. Based on the model in Garicano (2000), the direction of the effect depends largely on whether use of predictive analytics primarily promotes acquisition of local knowledge by plant managers (facilitating delegation of decision-making) or transmission of knowledge to headquarters (facilitating centralization).

Table 4 reports results relating the change in predictive analytics to changes in the delegation of decision-rights ($\Delta DelegationScore_p$). Delegation is measured based on the plant's discretionary authority over decisions including human relations (hiring, promotion and pay), marketing (advertising, pricing and new product introduction) and capital expenditures. Because plants that are co-located with headquarters do not respond to the delegation questions, the sample size is smaller and limited to remote plants.

Results in Table 4 provide consistent evidence that predictive analytics use is associated with reductions in decision-making delegation to local plant managers. The coefficient on predictive analytics is negative and significant regardless of the operationalization of predictive analytics, model type (OLS or causal forest), or inclusion of fixed effects. The fact that results for predictive analytics are robust to inclusion of firm (as well as industry and county) fixed effects is important because it isolates within-firm variation in the use of predictive analytics across plants controlling for other changes in underlying economics. Within the firm, reductions in delegation of decision-making are disproportionately concentrated in plants that have increased use of predictive analytics. Further, robustness to causal forest estimation using hundreds of plant- and firm-level variables reduces the likelihood of missing covariates that explain both delegation of decision-making and predictive analytics.

Based on the earlier discussion, these results indicate that predictive analytics act as a complement to centralized decision-making. Interpreted through the lens of the Garicano (2000) model, the results suggest that, because the information is easily transmitted to headquarters, data from predictive analytics primarily substitute for the private informational advantage of plant managers, reducing the need for headquarters to delegate decision-making.[21]

For parsimony, we use a unidimensional factor in our primary analysis to capture the extent of delegation over a range of decisions. However, the nature of the decision also likely affects the extent of delegation. In terms of components (not tabulated), the relation between changes in predictive analytics and delegation is driven by marketing (advertising, pricing and new product introduction) and human resources (hiring, promotion and pay), as opposed to capital expenditures. The result for marketing is consistent with the notion that predictive analytics permit headquarters to better forecast demand and adjust advertising, pricing and product mix without relying on local expertise. Similarly, the result for human relations suggests that predictive analytics facilitate headquarters' ability to anticipate local staffing needs, and centralize hiring, promotion and pay.

The preceding analysis suggests that predictive analytics use is positively associated with centralization of decision-making authority. To the extent that headquarters uses data from predictive analytics to increase control over local decisions, we also expect headquarters to increasingly dictate the specific information to be gathered. In Table 5, we explore the relation between predictive analytics use and this dimension of authority using responses to the question "who chose what type of data to gather at this establishment." In odd (even) numbered columns, the dependent variable represents the 2010 to 2015 change in the extent to which respondents indicate that local management (headquarters) chose the information to be

Table 4
Delegation of decision-making rights. This table presents OLS regressions and causal forests of changes in the amount of decision-making delegated to local plant management on changes in predictive analytics. Columns (1)–(6) include control variables: ln ($TVS_{p,2010}$); $TCE_{p,2010}$, $TIE_{p,2010}$, and $CM_{p,2010}$ each scaled by $TVS_{p,2010}$; the 2010 to 2015 percentage changes in TVS_p, TCE_p, TIE_p, and CM_p; $\Delta MgtEducation_p$, $\Delta StaffEducation_p$, Δ ln ($ITCapitalStock_p$), ΔDDD_p, $\Delta ProdPerfIndScore_p$, $\Delta MgrFdbckScore_p$, and the percentage change in $PlantPayroll_p$. Columns (1)–(3) also include the percentage change in $FirmPayroll_f$ and $IndGrowth_p$. Available controls to the causal forest estimation procedure are listed in the internet appendix. Standard errors are clustered at the firm level. All variables are defined in Appendix C.

	$\Delta DelegationScore_p$								
Model Type:	OLS						Causal Forests		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\Delta PAScore_p$	−0.029***			−0.023*			−0.022***		
	(0.007)			(0.013)			(0.008)		
$StartPA_p$		−0.020***			−0.022**			−0.010*	
		(0.006)			(0.010)			(0.006)	
$PAIncrease_p$			−0.011***			−0.009*			−0.012***
			(0.003)			(0.005)			(0.003)
Fixed effects:	No	No	No	Yes	Yes	Yes			
R^2	0.016	0.016	0.016	0.624	0.625	0.623			
N	11,500	11,500	11,500	11,500	11,500	11,500	11,500	11,500	11,500

[21] This result contrasts with the Bloom et al. (2014) finding that ERP systems are associated with increased delegation. An interpretation of the two sets of results is that ERP systems serve as tools to facilitate local managers' decision-making by complementing local knowledge, while data from predictive analytics are easily transmitted and interpreted by headquarters, reducing the local manager's information advantage.

Table 5
Choice of information collection. This table presents estimates from regressing changes in who chooses the information collected at the plant on changes in predictive analytics usage. The dependent variable in columns (1), (3), and (5), is the change in local plant management's choosing which data to collect, and the dependent variable in columns (2), (4), and (6) is the change in headquarters' choosing which data to collect. Panel A tabulates OLS regressions without fixed effects, whereas Panel B includes firm, county, and industry fixed effects. Panel A and Panel B include control variables: ln ($TVS_{p,2010}$); $TCE_{p,2010}$, $TIE_{p,2010}$, and $CM_{p,2010}$ each scaled by $TVS_{p,2010}$; the 2010 to 2015 percentage changes in TVS_p, TCE_p, TIE_p, and CM_p; $\Delta MgtEducation_p$, $\Delta StaffEducation_p$, Δ ln ($ITCapitalStock_p$), ΔDDD_p, $\Delta ProdPerfIndScore_p$, $\Delta MgrFdbckScore_p$, and the percentage change in $PlantPayroll_p$. Panel A also includes the percentage change in $FirmPayroll_f$ and $IndGrowth_p$. Available controls to the causal forest estimation procedure used in Panel C are listed in the internet appendix. For all panels, standard errors are clustered at the firm level. All variables are defined in Appendix C.

	$\Delta ChooseInfo$					
	$Local_p$	HQ_p	$Local_p$	HQ_p	$Local_p$	HQ_p
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS regressions, no fixed effects						
$\Delta PAScore_p$	0.030***	0.089***				
	(0.012)	(0.013)				
$StartPA_p$			0.015*	0.047***		
			(0.008)	(0.009)		
$PAIncrease_p$					0.005	0.039***
					(0.004)	(0.005)
Fixed effects:	No	No	No	No	No	No
R^2	0.004	0.031	0.035	0.024	0.002	0.032
N	26,500	26,500	26,500	26,500	26,500	26,500
Panel B: OLS regressions, fixed effects						
$\Delta PAScore_p$	0.025	0.107***				
	(0.028)	(0.033)				
$StartPA_p$			−0.007	0.069***		
			(0.022)	(0.027)		
$PAIncrease_p$					0.007	0.038***
					(0.012)	(0.013)
Fixed effects:	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.556	0.566	0.555	0.564	0.555	0.564
N	26,500	26,500	26,500	26,500	26,500	26,500
Panel C: Causal forests						
$\Delta PAScore_p$	0.011	0.088***				
	(0.011)	(0.012)				
$StartPA_p$			0.008	0.04***		
			(0.008)	(0.008)		
$PAIncrease_p$					0.003	0.043***
					(0.004)	(0.005)
N	26,500	26,500	26,500	26,500	26,500	26,500

collected.[22] Across all three measures, there is a significant positive relation between predictive analytics use and the extent to which headquarters chose the data to be collected at the plant. In contrast, there is relatively little evidence that predictive analytics use is associated with local plant management choosing data to be collected. In most cases, the relation between predictive analytics and local managers choosing data to be collected is insignificant and in all cases the effect is much smaller than for headquarters. Robustness to inclusion of firm fixed effects suggests that the increase in data gathering by headquarters is not firm-wide but, rather, is concentrated in plants using predictive analytics.

4.2.2. Managerial payroll

To the extent that predictive analytics use is associated with increased headquarters authority over plants, we expect reduced reliance on local plant manager decision-making and, therefore, a reduction in managerial payroll at the plant. While we cannot observe the number of plant managers or the pay of individual managers, we can observe the total compensation of plant managers, as well as the total compensation of lower-level staff.

We change our research design for this analysis because we can use annual data on salaries and wages from the ASM. As a result, for Table 6 only, we build a panel of annual data for each plant for 2008–2016, and conduct a difference-in-differences analysis, using increased reliance on predictive analytics as the treatment. We estimate models of the following structure:

$$Workforce_{p,t} = \alpha_1 (Post_t \times \Delta PAScore_p) + B \cdot (Post_t \times \Delta MOPSControls_p) + \Gamma \cdot ASMControls_{p,t} + \zeta_p + \eta_{i,t} + \varepsilon_{p,t} \quad (3)$$

where p indexes plant, i indexes the plant's 4-digit NAICS industry, and t indexes year. ζ_p is a time-invariant plant fixed effect (which subsumes firm fixed effects) and $\eta_{i,t}$ is a time-varying industry fixed effect controlling for economic changes to the

[22] Respondents were directed to check all that apply, so a change indicating that both local management and headquarters select the data gathered at the plant is a valid response.

Table 6
Workforce composition. This table presents estimates of regressing workforce characteristics on predictive analytics usage based on yearly observations from the ASM dataset from 2008 to 2016. We include as controls $Post_t \times \Delta ProdPerfIndScore_p$, $Post_t \times \Delta MgrFdbckScore_p$, $Post_t \times \Delta \ln(ITCaptialStock_p)$, $Post_t \times \Delta DDD_p$, $Post_t \times \Delta MgtEducation_p$, and $Post_t \times \Delta StaffEducation_p$. We also include $TCE_{p,t}$ scaled by $TVS_{p,t}$, and $\ln(NumProds_{p,t})$. Standard errors are clustered at the firm level. All variables are defined in Appendix C.

	$MgtComp_{p,t}$	$StaffComp_{p,t}$
	(1)	(2)
$Post_t \times \Delta PAScore_p$	−0.002*	−0.00002
	(0.001)	(0.001)
Fixed Effects:		
Plant	Yes	Yes
Industry × Year	Yes	Yes
N	180,000	180,000
R^2	0.855	0.878

industry. We employ two measures of the workforce as dependent variables: $StaffComp_{p,t}$, which is production workers' salaries and wages scaled by total value of shipments, and $MgtComp_{p,t}$, which is salaries and wages other than from production workers, also scaled by total value of shipments. $Post_t$ is an indicator that takes on a value of one after 2010.

Because MOPS data are measured only in 2010 and 2015, we do not know precisely when the plant increased its use of predictive analytics. We adopt a lower-bound estimate and consider 2011−2016 to be the post period. α_1 is the difference-in-difference estimator for the effect of predictive analytics on payroll. $\Delta MOPSControls_p$ is a vector of controls from the MOPS and $ASMControls_{p,t}$ is a vector of controls from the ASM. Specifically, from MOPS we interact the control variables for alternative information sources ($\Delta ProdPerfIndScore_p$ and $\Delta MgrFdbckScore_p$), IT investment in general ($\Delta \ln(ITCaptialStock_p)$), data-driven decision-making (ΔDDD_p), and education ($\Delta MgtEducation_p$ and $\Delta StaffEducation_p$) each with our $Post_t$ indicators (the main effects are subsumed by the plant fixed effects). To control for changes in the production process, we also include $TCE_{p,t}$ scaled by $TVS_{p,t}$, and $\ln(NumProds_{p,t})$. Because predictive analytics may be affected by firm-level decisions, we cluster standard errors at the firm level.

Consistent with a reduction in the importance of local plant management's decision-making, the results reported in Table 6, Column (1), indicate that increased predictive analytics use is associated with reductions in managerial payroll.[23] However, we observe no significant reduction for total production worker compensation in Column (2). These results reinforce the conclusion that predictive analytics use is associated with increased centralization of decision-making and not simply a general increase in productive efficiency.

4.3. Exerting authority over plant managers through incentive alignment

4.3.1. Performance-based incentives

The second important dimension of authority is the ability of top management to design compensation to align the incentives of plant managers with the headquarters' objectives. In the standard principal-agent framework, a challenge with performance-based incentives is that, if the outcome measure captures the agent's actions with noise, the incentive effect is reduced and the agent must be compensated for taking on additional risk. By providing more informative outcome measures, predictive analytics enable headquarters to rely more heavily on performance incentives to reduce agency problems and more tightly control managerial actions.[24] Results reported in Table 7, Columns (1)−(3), indicate that the initiation of formal performance-based bonuses is positively associated with increases in all three measures of predictive analytics use, suggesting that predictive analytics facilitate implementation of performance-based compensation.

In addition to compensation, headquarters can also create incentives by linking career-based outcomes such as firings and promotions to measured performance. Firings and promotions are particularly important because they represent major events in a manager's career trajectory and are understudied in the literature (Campbell, 2008). To the extent that predictive analytics use increases headquarters ability to accurately evaluate plant managers' performance, we expect firing and promotion decisions to be more heavily based on measured performance as opposed to seniority or more subjective criteria.

Table 7, Columns (4)−(6), reports results linking predictive analytics use to how quickly under-performing manager are dismissed or reassigned. Across all three specifications, increased predictive analytics use is associated with more rapid dismissal or reassignment for under-performance. Similarly, Columns (7)−(9) provide evidence linking predictive analytics use to promotions. Across all three specifications, results indicate a significant increase in the link between predictive analytics use and promotions based on performance, as opposed to promotions based on, “factors other than performance and ability (for example, tenure or family connections).” Results are consistent including firm fixed effects, indicating that within-firm variation drives the relation between predictive analytics and the use of formal incentives linked to

[23] This contrasts with the Bresnahan et al. (2002) finding that increases in IT more generally tend to increase the need for managers, and highlights the importance of considering potential heterogeneity in the components of overall informational infrastructure.

[24] As noted earlier, in these analyses we return to the specifications in equations (1) and (2) because we again rely on data comparing 2015 to 2010.

Table 7
Formal performance-based managerial incentives. This table presents estimates from regressing changes in managerial incentives (monetary bonuses, termination speed, and promotions based on merit) on changes in predictive analytics usage. Panel A tabulates OLS regressions without fixed effects, whereas Panel B includes firm, county, and industry fixed effects. Panel A and Panel B include control variables: ln ($TVS_{p,2010}$); $TCE_{p,2010}$, $TIE_{p,2010}$, and $CM_{p,2010}$ each scaled by $TVS_{p,2010}$; the 2010 to 2015 percentage changes in TVS_p, TCE_p, TIE_p, and CM_p; $\Delta MgtEducation_p$, $\Delta StaffEducation_p$, Δ ln ($ITCapitalStock_p$), ΔDDD_p, $\Delta ProdPerfIndScore_p$, $\Delta MgrFdbckScore_p$, and the percentage change in $PlantPayroll_p$. Panel A also includes the percentage change in $FirmPayroll_f$ and $IndGrowth_p$. Available controls to the causal forest estimation procedure used in Panel C are listed in the internet appendix. For all panels, standard errors are clustered at the firm level. All variables are defined in Appendix C.

	$StartBonus_p$			$\Delta Termination_p$			$\Delta Promotion_p$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS regressions, no fixed effects									
$\Delta PAScore_p$	0.091***			0.208***			0.084***		
	(0.011)			(0.013)			(0.008)		
$StartPA_p$		0.078***			0.134***			0.064***	
		(0.008)			(0.009)			(0.006)	
$PAIncrease_p$			0.031***			0.076***			0.027***
			(0.004)			(0.005)			(0.003)
Fixed effects:	No	No	No	No	No	No	No	No	No
R^2	0.031	0.030	0.029	0.114	0.079	0.109	0.074	0.068	0.062
N	26,500	26,500	26,500	26,500	26,500	26,500	26,500	26,500	26,500
Panel B: OLS regressions, fixed effects									
$\Delta PAScore_p$	0.078***			0.196***			0.076***		
	(0.026)			(0.031)			(0.018)		
$StartPA_p$		0.062***			0.136***			0.056***	
		(0.018)			(0.024)			(0.014)	
$PAIncrease_p$			0.022**			0.066***			0.022***
			(0.011)			(0.013)			(0.007)
Fixed effects:	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.722	0.722	0.722	0.668	0.656	0.665	0.700	0.698	0.696
N	26,500	26,500	26,500	26,500	26,500	26,500	26,500	26,500	26,500
Panel C: Causal forests									
$\Delta PAScore_p$	0.040***			0.202***			0.096***		
	(0.008)			(0.015)			(0.009)		
$StartPA_p$		0.028***			0.086***			0.045***	
		(0.005)			(0.010)			(0.006)	
$PAIncrease_p$			0.024***			0.093***			0.043***
			(0.003)			(0.006)			(0.003)
N	26,500	26,500	26,500	26,500	26,500	26,500	26,500	26,500	26,500

performance. Overall, results in Table 7 suggest that predictive analytics use is associated with greater headquarters authority over plant managers as reflected in closer links between measured performance and bonuses, firings and promotions.

4.3.2. Targets

The preceding evidence suggests a positive relation between predictive analytics use and stronger performance-based incentives. While the MOPS only provides one question each on performance-based bonuses, firings and promotions, it provides more detail on explicit performance targets. Such explicit performance targets are not only used in the implementation of performance-based incentives but also to establish clear goals that are shared across employees (Matějka, 2018). However, for targets to be effective, they need to be accurately set and communicated throughout the organization. We expect predictive analytics to provide information that facilitates greater use of explicit targets that are communicated deeper into the organization, and are more accurately tied to attainable performance.

We begin by examining the relation between predictive analytics and use of targets within the plant. Table 8, Columns (1)–(3), report results on the relation between predictive analytics and the initiation of targets. The significant positive coefficients on $StartTarget_p$ across all specifications provide consistent evidence that increased use of predictive analytics is associated with increased target initiation.

MOPS also provides data on “who was aware of the production targets at this establishment.” To the extent that targets become more important, we expect them to be pushed deeper into the organization to facilitate headquarters’ influence and authority over all employee levels. The significant positive coefficients on $\Delta TargetAwareness_p$ in columns (4)–(6) provide strong evidence that predictive analytics use is associated with target awareness deeper into the organization, consistent with targets increasing in saliency for the subset of plants that adopt predictive analytics.

In addition to predictive analytics increasing the use and salience of targets, we expect an increase in target accuracy and effectiveness. Prior research (Eyring and Narayanan, 2018; Webb et al., 2013) suggests that targets are most effective when they are challenging but achievable with a reasonable amount of effort (i.e., neither too easy nor too difficult to achieve). MOPS includes questions on the level of effort required to achieve targets (“without much effort,” “with some effort,” “with

Table 8

Target usage. This table presents estimates from regressing an indicator of plants starting to use performance targets and changes in target awareness at the plant on changes in predictive analytics usage. Panel A tabulates OLS regressions without fixed effects, whereas Panel B includes firm, county, and industry fixed effects. Panel A and Panel B include control variables: ln ($TVS_{p,2010}$); $TCE_{p,2010}$, $TIE_{p,2010}$, and $CM_{p,2010}$ each scaled by $TVS_{p,2010}$; the 2010 to 2015 percentage changes in TVS_p, TCE_p, TIE_p, and CM_p; $\Delta MgtEducation_p$, $\Delta StaffEducation_p$, Δ ln ($ITCapitalStock_p$), ΔDDD_p, $\Delta ProdPerfIndScore_p$, $\Delta MgrFdbckScore_p$, and the percentage change in $PlantPayroll_p$. Panel A also includes the percentage change in $FirmPayroll_f$ and $IndGrowth_p$. Available controls to the causal forest estimation procedure used in Panel C are listed in the internet appendix. For all panels, standard errors are clustered at the firm level. All variables are defined in Appendix C.

	$StartTarget_p$			$\Delta TargetAwareness_p$		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS regressions, no fixed effects						
$\Delta PAScore_p$	0.091***			0.374***		
	(0.010)			(0.013)		
$StartPA_p$		0.086***			0.190***	
		(0.007)			(0.009)	
$PAIncrease_p$			0.024***			0.146***
			(0.003)			(0.005)
Fixed Effects:	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.063	0.081	0.048	0.266	0.186	0.278
N	26,500	26,500	26,500	26,500	26,500	26,500
Panel B: OLS regressions, fixed effects						
$\Delta PAScore_p$	0.074***			0.333***		
	(0.022)			(0.032)		
$StartPA_p$		0.060***			0.168***	
		(0.016)			(0.021)	
$PAIncrease_p$			0.016**			0.131***
			(0.007)			(0.013)
Fixed Effects:	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.747	0.751	0.742	0.736	0.711	0.738
N	26,500	26,500	26,500	26,500	26,500	26,500
Panel C: Causal forests						
$\Delta PAScore_p$	0.027***			0.314***		
	(0.004)			(0.013)		
$StartPA_p$		0.013***			0.119***	
		(0.004)			(0.008)	
$PAIncrease_p$			0.015***			0.140***
			(0.002)			(0.005)
N	26,500	26,500	26,500	26,500	26,500	26,500

normal amount of effort," "with more than normal effort," and "with extraordinary effort"), as well as how frequently targets were met or beaten. If predictive analytics use helps in formulating effective targets, we expect fewer responses in the extremes and more toward the middle (particularly "normal" and "more than normal" effort). In addition, if targets are effective in motivating action, we expect plants to more frequently be able to meet or exceed targets.

Results reported in Table 9 are consistent with predictive analytics helping to set accurate targets. For these analyses we exclude firm fixed effects (results are insignificant with their inclusion, which is consistent with target-setting practices moving to a top-down, company-wide approach). In columns (1) through (5), we observe significant negative coefficients for targets that are in the extremes ("without much effort," or "only with extraordinary effort"), and significant positive coefficients on targets that require "more than normal" but not extreme effort (i.e., realistic targets), suggesting that predictive analytics are associated with more accurate targets. The significant negative coefficients for both the "little" and "extreme" effort levels suggest that the targets did not simply become easier but, rather, became more accurate.[25] The fact that targets requiring "more than normal effort" increase most significantly is consistent with the use of "stretch" targets that are challenging but attainable.

Column (6) reports results relating predictive analytics use to the ability of managers to meet or exceed their targets. Consistent with the preceding results, targets are, on average, more accurate in that managers are better able to reach them given increased use of predictive analytics. Overall, results in Table 9 suggest that predictive analytics use is associated with increased influence of headquarters through more frequent, more salient and more accurate use of targets.

4.4. Robustness analysis: alternate measures of predictive analytics

Results to this point provide consistent evidence that predictive analytics use is associated with increased headquarters authority both in terms of reduced delegation of decision-making to plant management and greater use of performance-

[25] The fact that the coefficient on "without much effort" is significantly negative (i.e. fewer plants report easy targets) is also inconsistent with the notion that plants simply became more efficient and therefore targets were easier to meet.

Table 9

Target accuracy. This table presents causal forest estimates from regressing changes in performance target accuracy on changes in predictive analytics usage. Panel A tabulates OLS regressions with county and industry fixed effects that include control variables: ln ($TVS_{p,2010}$); $TCE_{p,2010}$, $TIE_{p,2010}$, and $CM_{p,2010}$ each scaled by $TVS_{p,2010}$; the 2010 to 2015 percentage changes in TVS_p, TCE_p, TIE_p, and CM_p; $\Delta MgtEducation_p$, $\Delta StaffEducation_p$, $\Delta \ln (ITCapitalStock_p)$, ΔDDD_p, $\Delta ProdPerfIndScore_p$, $\Delta MgrFdbckScore_p$, and the percentage change in $PlantPayroll_p$ and $FirmPayroll_f$. Available controls to the causal forest estimation procedure used in Panel B are listed in the internet appendix. For all panels, standard errors are clustered at the firm level. All variables are defined in Appendix C.

Panel A: OLS Regressions, industry and county fixed effects

	$\Delta TargetEffort$					$\Delta Achieve_p$
	$Little_p$	$Some_p$	$Normal_p$	$> Normal_p$	$Extreme_p$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta PAScore_p$	−0.034***	−0.031*	0.108***	0.078***	−0.114***	0.019*
	(0.010)	(0.018)	(0.023)	(0.024)	(0.014)	(0.010)
R^2	0.104	0.100	0.106	0.100	0.120	0.105
N	26,500	26,500	26,500	26,500	26,500	26,500
$StartPA_p$	−0.025***	−0.003	0.088***	0.034**	−0.090***	0.008
	(0.007)	(0.012)	(0.016)	(0.016)	(0.010)	(0.007)
R^2	0.105	0.099	0.106	0.099	0.115	0.105
N	26,500	26,500	26,500	26,500	26,500	26,500
$PAIncrease_p$	−0.010**	−0.014*	0.022**	0.032***	−0.028***	0.006
	(0.004)	(0.008)	(0.011)	(0.011)	(0.006)	(0.004)
R^2	0.103	0.100	0.106	0.100	0.114	0.103
N	26,500	26,500	26,500	26,500	26,500	26,500

Panel B: Causal forests

	$\Delta TargetEffort$					$\Delta Achieve_p$
	$Little_p$ (1)	$Some_p$ (2)	$Normal_p$ (3)	$> Normal_p$ (4)	$Extreme_p$ (5)	(6)
$\Delta PAScore_p$	−0.011	0.018	−0.026	0.080***	−0.034***	0.023**
	(0.009)	(0.014)	(0.023)	(0.022)	(0.012)	(0.010)
N	26,500	26,500	26,500	26,500	26,500	26,500
$StartPA_p$	−0.016***	0.012	0.004	0.042***	−0.020***	0.006
	(0.005)	(0.010)	(0.014)	(0.014)	(0.006)	(0.006)
N	26,500	26,500	26,500	26,500	26,500	26,500
$PAIncrease_p$	−0.009**	0.008	−0.010	0.038***	−0.012***	0.008**
	(0.003)	(0.006)	(0.009)	(0.009)	(0.004)	(0.003)
N	26,500	26,500	26,500	26,500	26,500	26,500

based incentives. In this section, we explore robustness to refinements of our measures of predictive analytics. To this point our primary measures of predictive analytics rely on responses to direct questions about predictive analytics usage. An alternative approach is to measure the breadth of activities supported by data analysis, under the assumption that predictive analytics will play a stronger role in plants that use data analytics for a broader set of activities in general. This analysis ensures that results are not sensitive to responses to particular questions or our measurement approach.

MOPS asks respondents about the extent to which data analysis affected three activities: new product design, demand forecasting, and supply chain management. From these responses, we construct a variable, $\Delta Breadth$, which measures the change in the number of activities supported by data analysis at the plant. Table IA.2 in the internet appendix reports abbreviated results substituting $\Delta Breadth$ as our independent variable of interest across all of our primary tests. Results across specifications are consistent with the expectation that application breadth matters in terms of the impact of data analytics. In particular, increases in the breadth of application are significantly negatively associated with delegation of decision-making to the plant and positively associated with the extent to which headquarters controls data gathering. Similarly, increases in $\Delta Breadth$ are positively associated with increases in formal performance-based incentives, target usage and target accuracy.

5. Conclusions

We provide large-scale evidence on the relation between predictive analytics use and centralization of authority in headquarters for a representative sample of US manufacturing plants. Our analysis suggests that predictive analytics use has become increasingly pervasive in manufacturing, particularly for well-resourced, young firms with educated workforces in relatively stable industries.

In terms of our primary research questions, we focus on two channels through which headquarters can exert authority over plants: delegation of decision-making and performance incentive design. Overall, we provide consistent evidence that predictive analytics use is associated with increased authority by headquarters over plant operations. In terms of delegation, we find that predictive analytics use is associated with increased headquarters centralization of decision-making, particularly over marketing (e.g., product mix, pricing and advertising) and human resources (e.g., hiring, bonuses and firing). Primary control over specific plant-level data to be gathered is increasingly centralized in headquarters. Consistent with reduced reliance on plant managers, predictive analytics use is associated with reduced managerial (relative to staff) payroll.

In terms of performance incentive design, predictive analytics use is associated with headquarters exercising stronger authority through greater use of performance-based incentives. With predictive analytics, firms increasingly tie bonuses, promotions and firing of plant managers to performance outcomes. In addition, performance targets are more frequently used, become more accurate, and are communicated deeper into the organization.

Overall, our results suggest that predictive analytics use is associated with headquarters exerting greater authority over plant activities. Our results are subject to several important caveats. First, while we include numerous controls, fixed effects, and alternative designs including causal forests, our data do not permit us to draw causal links between predictive analytics and organizational authority. In particular, we recognize that use of predictive analytics likely coevolves with decisions concerning, for example, delegation and performance measurement design. Rather, we view our results as suggesting complementarity between predictive analytics use and centralization of authority in headquarters.

Second, we cannot draw normative conclusions because we are not able to observe the full range of outcomes or the costs of implementing a predictive analytics-driven information structure and organizational authority. Subject to these caveats, predictive analytics, and other big data techniques, are an evolving force in the economy about which we have relatively little existing empirical evidence, and are likely to be increasingly important as technology facilitates greater use of data for decision-making and performance management.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jacceco.2022.101526.

Appendix B. Details of ASM and CMF

In years ending in “2” or “7”, the US Census Bureau conducts a full Census of Manufactures (CMF) and sends surveys to roughly 168,000 establishments, representing all but the smallest manufacturing plants (single-plant companies with fewer than 20 employees). Plants are required to provide operational data, including sales, employees, payroll, inventories, capital expenditures, operating expenses, and additional background information. The forms do not directly ask about accounting earnings.[26]

Outside of the quinquennial censuses, Census surveys a subset of manufacturing establishments on an annual basis—the Annual Survey of Manufactures (ASM). Two years after each economic census, Census selects a stratified sample of roughly 51,000 establishments: 33,000 plants from multi-establishment firms and 18,000 from larger single-establishment firms in the CMF. Once selected, this sample receives the ASM form each year for five years, after which a new sample is selected based on the subsequent CMF. The ASM questions are virtually identical to the main CMF questions.

The ASM's stratified sample is designed to capture the most significant economic activity and have adequate coverage of each industry and geographic area. As a result, roughly 15,400 of the largest plants are selected with certainty, accounting for roughly 67% of the overall economic activity in the manufacturing sector. The remaining sample of about 35,600 plants is selected using stratified sampling probabilities for each combination of geography, industry, and employment with the goal of providing representative data for the population of US manufacturing plants. A large portion of plants persist from one five-year sample to the next and the sample of firms surveyed annually remains constant during each five-year window. For both the CMF and ASM, plant responses are required by law. Additionally, the Census aggressively follows up with establishments that do not respond to minimize selection bias, resulting in a response rate of 70–80%. Because responses are confidential, immune from legal process, and can only be used by the government for statistical purposes, there should be little incentive to misreport.[27]

Appendix C. Variable Definitions

A detailed correspondence between the MOPS survey questions (with screenshots of the survey instrument) and our variables is included in Internet Appendix E.

Appendix C.1. Regressors of interest

$PAScore_{p,t}$ Frequency of predictive analytics usage. 1 if respondent checked MOPS Question 29 (how frequently does the establishment rely on predictive analytics) as “Daily”, 0.75 if “Weekly”, 0.5 if “Monthly”, 0.25 if “Yearly”, 0 otherwise.

[26] Because the Census' aim is to derive aggregate manufacturing activity, many expenses such as Selling, General and Administrative Expense are not required to be reported at the plant level.

[27] Researchers granted access to the underlying Census microdata are not permitted to use it to make policy recommendations, eliminating another potential incentive for respondents to misreport.

$StartPA_p$ Indicator if a plant commences using predictive analytics between 2010 and 2015. Using predictive analytics is measured if respondent checked any box for MOPS question 29 (Predictive analytics use frequency) other than “Never”.

$PAIncrease_p$ Initiation or increase in frequency of predictive analytics usage. 1 if *$PAScore_p$* is higher in 2015 than 2010, 0 otherwise.

$Breadth_{p,t}$ Number of areas influenced by data analysis where the respondent checked a box other than “Never” for MOPS question 28 (values 0–3).

Appendix C.2. Other variables

$Achieve_{p,t}$ 1 if respondent checked one of the first five responses to question 12 of MOPS (when production targets were met, what percentage of managers received a bonus), 0 otherwise.

$ChooseInfoHQ_{p,t}$ 1 if respondent checked “Managers at headquarters and/or other establishments” for question 26 of the MOPS (who chose what type of data to collect at this establishment), 0 otherwise.

$ChooseInfoLocal_{p,t}$ 1 if respondent checked “Managers at this establishment” for question 26 of the MOPS (who chose what type of data to collect at this establishment), 0 otherwise.

$CM_{p,t}$ Cost of materials at the plant in a given year.

$DDD_{p,t}$ Indicator if all of the following conditions are true: 4th or 5th choice for MOPS question 24 is selected (“a great deal of data to support decision making is available” or “all the data we need to support decision making is available”), 4th or 5th choice for MOPS question 25 is selected (“Decision making relies heavily on data” or “decision making relies entirely on data”), number of key performance indicators is 10 or more (third choice for MOPS question 2), and a combination of short-term and long-term production targets are used (MOPS question 6, choice 3) (Brynjolfsson and McElheran, 2016a). If the dependent variable in the regression pertains to targets, we exclude the last criterion.

$DelegationScore_{p,t}$ Average of questions 18–23 from MOPS (Various dimensions of delegation of decision-making). Responses to each question are scaled evenly so the response with the most autonomy is 1 and the least autonomy is 0.

$FamilyCEO_f$ 1 if the respondent checked Yes to CEO of the firm also being a founder or a member of the founder's family (question 45 of MOPS), 0 otherwise.

$FirmAge_f$ 2015 minus year the firms oldest establishment first appeared in the Census Bureau's Longitudinal Business Database.

$FirmPayroll_{f,t}$ Payroll of all establishments in the firm from the Census Bureau's Longitudinal Business Database.

$GrowthFirmPayroll_f$ (2015 Firm Payroll - 2010 Firm Payroll)/2010 Firm Payroll.

$GrowthPlantPayroll_p$ (2015 Plant Payroll - 2010 Plant Payroll)/2010 Plant Payroll.

$IndGrowth_p$ 5-year growth in *TVS_p* (on a percentage basis) for the 4-digit NAICS; obtained from aggregating ASM data using sampling weights.

$ITCapitalStock_{p,t}$ Hardware and software purchases from the ASM and CMF dating back to 2002 and 2006 respectively (imputed if missing), deflated by the BEA consumer price index for personal computers and perhiperal equipment, and depreciated by 35% per year (Brynjolfsson and McElheran, 2016a).

$MgrFdbckScore_{p,t}$ 1 if respondent checked MOPS Question 27, second row (Frequency of use of formal or informal feedback from managers) as “Daily”, 0.75 if “Weekly”, 0.5 if “Monthly”, 0.25 if “Yearly”, 0 otherwise.

$MgtComp_{p,t}$ Non-production workers salaries and wages scaled by *$TVS_{p,t}$*

$MgtEducation_{p,t}$ 1 if respondent checked more than 80% of managers had a bachelors degree, 0.75 if 61–80%, 0.5 if 41–60%, 0.25 if 21–40% and 0 if 20% or less (question 40 of MOPS).

$MultinationalFirm_f$ 1 if respondent checked the firm has production establishments in other countries (MOPS question 46), 0 otherwise.

$MultiplantFirm_{p,t}$ 1 if more than one establishment exists in the Census Bureau's Longitudinal Business Database for 2015, 0 otherwise.

$NumProds_{p,t}$ Number of distinct product class codes that the plant ships (https://www.census.gov/programs-surveys/asm/technical-documentation/product-codes-descriptions.html).

$PlantAge_p$ 2015 minus year the establishment first appeared in the Census Bureau's Longitudinal Business Database.

$PlantPayroll_{p,t}$ Payroll of the establishment from the Census Bureau's Longitudinal Business Database.

$ProdPerfIndScore_{p,t}$ 1 if respondent checked MOPS Question 27, first row (Frequency of review of performance indicators from production technology or instruments) as “Daily”, 0.75 if “Weekly”, 0.5 if “Monthly”, 0.25 if “Yearly”, 0 otherwise.

$Promotion_{p,t}$ Response to Question 14 of MOPS (the primary way managers were promoted at the establishment): 1 = “Promotions were based solely on performance and ability”, 0.667 = “Promotions were based partly on performance and ability and partly on other factors (for example, tenure or family connections)”, 0.333 = “Promotions were based mainly on factors other than performance and abillity”, 0 = “Managers are normally not promoted”.

$RemotePlant_p$ 1 if respondent checked that headquarters for the company was not at the same location as this establishment (question 17 of MOPS), 0 otherwise.

$StaffComp_{p,t}$ Production workers salaries and wages scaled by *$TVS_{p,t}$*

$StaffEducation_{p,t}$ 1 if respondent checked more than 20% of non-managers had a bachelors degree, 0.667 if 11–20%, 0.333 if 1–10%, and 0 if 0% (question 41 of MOPS).

$StartBonus_p$ 1 if *$UseBonus_{p,2010}$* = 0 and *$UseBonus_{p,2015}$* = 1, 0 otherwise.

$StartTarget_p$ 1 if *$UseTarget_{p,2010}$* = 0 and *$UseTarget_{p,2015}$* = 1, 0 otherwise.

$TargetAwareness_{p,t}$ Response to Question 8 of MOPS (who was aware of production targets at this establishment): 1 = "All managers and most production workers", 0.667 = "Most managers and most production workers", 0.333 = "Most managers and some production workers", 0 = "Only senior managers".

$TargetEffortExtreme_{p,t}$ 1 if respondent checked "Only possible to achieve with extraordinary effort" for question 7 (target difficulty) of MOPS, 0 otherwise.

$TargetEffortLittle_{p,t}$ 1 if respondent checked "Possible to achieve without much effort" for question 7 (target difficulty) of MOPS, 0 otherwise.

$TargetEffortMoreThanNormal_{p,t}$ 1 if respondent checked "Possible to achieve with more than normal effort" for question 7 (target difficulty) of MOPS, 0 otherwise.

$TargetEffortNormal_{p,t}$ 1 if respondent checked "Possible to achieve with normal effort" for question 7 (target difficulty) of MOPS, 0 otherwise.

$TargetEffortSome_{p,t}$ 1 if respondent checked "Possible to achieve with some effort" for question 7 (target difficulty) of MOPS, 0 otherwise.

$TCE_{p,t}$ Total capital expenditures at the plant in a given year.

$TechInd_p$ 1 if the plant's primary NAICS is in one of the 4-Digit NAICS in Table 1 of Heckler (2005), 0 otherwise.

$Termination_p$ 1 if respondent checked "within 6 months of identifying manager under-performance", 0.5 if "after 6 months of identifying manager under performance", or 0 if "Rarely or never" to question 16 of MOPS (speed of under-performing manager reassignment or dismissal).

$TIE_{p,t}$ Total end of year inventories at the plant in a given year.

$TVS_{p,t}$ Total value of shipments from the plant in a given year.

$UseBonus_{p,t}$ 1 if respondent checked any of the first 4 boxes to question 11 of MOPS (criteria for managerial bonuses).

$UseTarget_{p,t}$ 1 if respondent checked any of the first 3 boxes to question 6 of MOPS (timeframe for production targets).

$UsePA_{p,t}$ 1 if respondent checked any box other than "Never" for question 29 of the MOPS (how often does the plant use predictive analytics), 0 otherwise.

Appendix D. Causal Forests

For an excellent primer on regression trees and random forests we refer readers to Chapters 9 and 15 respectively of Hastie et al. (2009).

Regression trees are a non-parametric model for a response variable using successive binary splits on the model's explanatory variables. The greedy algorithm recursively selects covariates and splitting points to partition a training sample into groups with similar response values, creating a binary tree. The tree is grown until the leaves reach a minimum size, at which point the predicted value of the response variable is the average value of training observations in that leaf. The resulting tree is easy to interpret as a series of binary classification steps. Because a variable can be split at different points in the distribution at different branches, the model is inherently non-linear, and because certain variables are used in some branches but not others, the model accommodates interactions.

While conceptually easy to understand, regression trees perform poorly in practice. Building a complicated tree tends to result in a model that fits the training sample well but performs poorly on new data (i.e., is overfit), whereas simple trees do not capture enough variation to sufficiently explain the response variable. Breiman (2001) developed an algorithm that builds an ensemble of trees, each constructed from subsamples, which are averaged together to form a final predictive model. This "forest" results in a smoother function and performs much better than the individual trees in out-of-sample prediction. The key innovation of Breiman (2001) over simple bootstrapping and bagging of trees is incorporating "randomness" in the creation of each tree. At each point in building the trees, the algorithm is restricted to choosing from a random subset of the potential splitting variables, which de-correlates the trees in the forest. As a result, the final average of all the trees captures a large amount of the inherent underlying complexity, but the imposed randomness produces a model that is much more resilient to over-fitting. In effect, the algorithm introduces a small amount of bias when creating trees (by omitting potential important covariates) to reduce the variance of the final model. This bias-variance trade-off is a common feature in the creation of many prediction models (e.g., Ridge regression, LASSO).

While random forests typically perform very well in out-of-sample prediction, they are less useful for inference. As an average of many trees they are hard to visualize and behave somewhat as a black box. Like all unsupervised prediction models, the algorithm keys on the variation that best correlates with differences in the response—not the underlying forces causing the outcomes.

A common approach to causal inference is to employ the unconfoundedness assumption—that after controlling for potential confounds, treatment assignment is independent of the outcome (i.e., it is as good as random) (Rosenbaum and Rubin, 1983). This assumption allows the researcher to link differences in the outcome to differences in treatment, which is the strategy underpinning the use of OLS and propensity score matching for causal inference. Because machine-learning approaches like random forests provide much better models (in terms of prediction) than the ad hoc linear models typically used, there is potential to improve considerably on the reliability of these traditional approaches. Because random forests are non-parametric and resistant to overfitting, using them enables the researcher to control for a broader number of potential confounds and varying functional forms.

Wager and Athey (2018) present an algorithm, further refined in Athey et al. (2019) for generating estimates of the treatment effect employing regression forests. In essence, a regression forest is built to model treatment assignment and a separate forest is built to model the outcome absent differences in treatment. The causal forest combines these two much like propensity score matching combines a selection model and an outcome model. In effect, the treatment assignment forest is used to find “close” observations (likely to receive treatment) and the outcomes forest purges out additional noise in the outcome that can be directly attributed to differences in non-treatment covariates. The key element to obtaining consistency and asymptotic normality of the treatment effect is that the forests are “honest”—distinct samples are used to pick models and estimate effects (concretely, different observations are used to build each tree than the observations used to estimate the predicted value for each leaf). Sample splitting and bootstrapping allow for fulfilling this honesty requirement without considerable sample loss.

References

Abernethy, M.A., Bouwens, J., Van Lent, L., 2004. Determinants of control system design in divisionalized firms. Account. Rev. 79, 545–570.
Athey, S., Imbens, G.W., 2017. The state of applied econometrics: causality and policy evaluation. J. Econ. Perspect. 31, 3–32.
Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. Ann. Stat. 47, 1148–1178.
Athey, S., Wager, S., 2019. Estimating treatment effects with causal forests: an application. Observ. Stud. 5, 36–51.
Baiman, S., Evans, J.H., 1983. Pre-decision information and participative management control systems. J. Account. Res. 371–395.
Bloom, N., Garicano, L., Sadun, R., Van Reenen, J., 2014. The distinct effects of information technology and communication technology on firm organization. Manag. Sci. 60, 2859–2885.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M., 2002. Information technology, workplace organization, and the demand for skilled labor: firm-level evidence. Q. J. Econ. 117, 339–376.
Brynjolfsson, E., McElheran, K., 2016a. Data in Action: Data-Driven Decision Making in Us Manufacturing. US Census Bureau Center for Economic Studies. Paper No. CES-WP-16-06.
Brynjolfsson, E., McElheran, K., 2016b. The rapid adoption of data-driven decision-making. Am. Econ. Rev. 106, 133–139.
Bushman, R.M., Indjejikian, R.J., Penno, M.C., 2000. Private predecision information, performance measure congruity, and the value of delegation. Contemp. Account. Res. 17, 562–587.
Campbell, D., 2008. Nonfinancial performance measures and promotion-based incentives. J. Account. Res. 46, 297–332.
Campbell, D., Datar, S.M., Sandino, T., 2009. Organizational design and control across multiple markets: the case of franchising in the convenience store industry. Account. Rev. 84, 1749–1779.
Demski, J.S., Feltham, G.A., 1976. Cost determination: a conceptual approach. Iowa State Press.
Dilda, V., Hippe, M., Mori, L., Noterdaeme, O., Schmitz, C., Van Niel, J., 2017. Manufacturing: Analytics Unleashes Productivity and Profitability. White paper, McKinsey & Company.
Eyring, H., Narayanan, V., 2018. Performance effects of setting a high reference point for peer-performance comparison. J. Account. Res. 56, 581–615.
Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. J. Soc. Econ. 40, 35–42.
Garicano, L., 2000. Hierarchies and the organization of knowledge in production. J. Polit. Econ. 108, 874–904.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
Heckler, D.E., 2005. High-technology employment: a NAICS-based update. Mon. Labor Rev. 128, 57.
Laaper, S., Yauch, G., Wellener, P., Robinson, R., 2018. Embracing a Digital Future: How Manufactures Can Unlock the Transforamative Benefits of Digital Supply Networks. White Paper, Deloitte.
Matějka, M., 2018. Target setting in multi-divisional organizations. J. Manag. Account. Res. 30, 13–27.
Merchant, K.A., 1998. Modern Management Control Systems: Text and Cases. Prentice-Hall.
Merchant, K.A., Stringer, C., Shantapriyan, P., 2018. Setting financial performance thresholds, targets, and maximums in bonus plans. J. Manag. Account. Res. 30, 55–73.
Moers, F., 2006. Performance measure properties and delegation. Account. Rev. 81, 897–924.
Nagar, V., 2002. Delegation and incentive compensation. Account. Rev. 77, 379–395.
Prendergast, C., 2002. The tenuous trade-off between risk and incentives. J. Polit. Econ. 110, 1071–1102.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Swaminathan, J.M., 2018. Big data analytics for rapid, impactful, sustained, and efficient (rise) humanitarian operations. Prod. Oper. Manag. 27, 1696–1700.
Trentmann, N., 2019. Predictive analytics give a boost to Diageo's cost-savings efforts. Wall St. J.
Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Am. Stat. Assoc. 113, 1228–1242.
Webb, R.A., Williamson, M.G., Zhang, Y., 2013. Productivity-target difficulty, target-based pay, and outside-the-box thinking. Account. Rev. 1433–1457.