

# Data in Action: Data-Driven Decision Making in U.S. Manufacturing

Erik Brynjolfsson<sup>1</sup>  
MIT and NBER

Kristina McElheran<sup>2</sup>  
University of Toronto

This Draft: December 31, 2015

## Abstract

Manufacturing in America has become significantly more data-intensive. We investigate the adoption, performance effects and organizational complementarities of data-driven decision making (DDD) in the U.S. Using data collected by the Census Bureau for 2005 and 2010, we observe the extent to which manufacturing firms track and use data to guide decision making, as well as their investments in information technology (IT) and the use of other structured management practices. Examining a representative sample of over 18,000 plants, we find that adoption of DDD is earlier and more prevalent among larger, older plants belonging to multi-unit firms. Smaller single-establishment firms adopt later but have a higher correlation with performance than similar non-adopters. Using a fixed-effects estimator, we find the average value-added for later DDD adopters to be 3% greater than non-adopters, controlling for other inputs to production. This effect is distinct from that associated with IT and other structured management practices and is concentrated among single-unit firms. Performance improves after plants adopt DDD, but not before – consistent with a causal relationship. However, DDD-related performance differentials decrease over time for early and late adopters, consistent with firm learning and development of organizational complementarities. Formal complementarity tests suggest that DDD and high levels of IT capital reinforce each other, as do DDD and skilled workers. For some industries, the benefits of DDD adoption appear to be greater for plants that delegate some decision making to frontline workers.

Disclaimer and Acknowledgements: Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. All errors are our own. We thank the MIT Initiative on the Digital Economy for generous funding for this research.

---

<sup>1</sup> MIT Sloan School, 100 Main St, room E62-414, Cambridge, MA 02142. [erikb@mit.edu](mailto:erikb@mit.edu)

<sup>2</sup> University of Toronto, 105 St. George Street, Toronto, ON M5S 3E6, [k.mcelheran@utoronto.ca](mailto:k.mcelheran@utoronto.ca)

## 1. INTRODUCTION

Why do seemingly similar firms exhibit large and persistent differences in performance?

Significant evidence points to information technology (IT) as a potential differentiator (e.g., Brynjolfsson and Hitt 2000, Dunne et al. 2004, Syverson 2011). Yet the same basic technologies are available to firms across the economy. Thus differences in performance must depend on differences in how IT is deployed across particular organizational contexts and processes (Bresnahan and Greenstein 1996; Bresnahan, Brynjolfsson and Hitt 2002; Brynjolfsson, Hitt, and Yang 2002; Aral and Weill 2007, McElheran 2015).

Recently, dramatic increases in data storage and processing capabilities have brought renewed interest to the question of how IT may differentially affect firms' performance. Are firms responding by changing how they collect information? In turn, is the ready availability of data changing how managers use it? And, fundamentally, does this translate into changes in firm performance?

Several prior studies have focused on heterogeneous adoption of management practices as an explanation for productivity differences among firms (e.g., MacDuffie 1995; Ichniowski, Shaw, and Prennushi 1997; Bloom et al. 2013, Blader et al. 2015; Hong, Keung and Yang 2015). There is some evidence that data-related management practices may be correlated with superior overall performance in a modest sample of large public firms (Brynjolfsson, Hitt, and Kim 2011). Yet important questions remain about how widespread these practices may be in the broader firm population, to what extent they may be linked to productivity improvements, and whether specific clusters of complementary organizational practices may be required.

We recently worked with the U.S. Census Bureau to design and field a survey aimed at examining these phenomena in more detail. The survey went to a large and representative sample of firms, providing unusual visibility into related activities within firms. It extends our understanding of how firms take advantage of data in several ways. To begin, it sheds light on how these practices are diffusing among establishments of different types and highlights the importance of size, multi-unit status, capital intensity, and complementary investments in information technology and skilled labor for the presence of DDD at the plant. Next, we offer the first evidence that management practices focused on collecting and

using data are correlated with better performance in a wide range of operational settings— even controlling for confounding factors such as: generic IT investment, the adoption of other structured management practices, and perhaps most importantly, unobserved heterogeneity among production units. Specifically, using plant-level fixed effects, we find that substantially more data-intensive decision making is associated with a statistically significant increase of 3% in productivity, on average. For the typical plant in our sample, the output increase is comparable to investing an additional \$5 million in IT capital, or \$60,000 per employee over the five-year period – without actually spending any more on technology. Moreover, a timing falsification exercise reveals that performance increases appear only *after* plants adopt DDD and not before. This is consistent with a causal relationship. We find similar results when using a continuous index of data-related practices, with a positive “dose-response” relationship whereby incremental advances in DDD practices are associated with small but statistically significant differences in productivity.

Interestingly, this productivity differential is not uniform across firms: establishments that belong to multi-unit firms, which also tend to be larger, older, and more capital-intensive, show a lower and noisier correlation between DDD adoption and productivity during our sample period. This is consistent with a diffusion model whereby firms learn about clusters of complementary production practices over time, with more optimal sorting into decision making regimes by more-mature organizations. Reinforcing this learning-based model of diffusion, we see that performance differentials between early and late adopters decrease over time.

What then, makes DDD productive inside certain firms? Our findings highlight a few key organizational characteristics that we observe in the data. The first is the sophistication of the IT infrastructure. Unsurprisingly, later DDD adopters show a disproportionately higher productivity effect when IT capital stock was already high in 2005. Correlation tests are also consistent with complementarities between IT and DDD. We furthermore find that DDD adoption is significantly correlated with the presence of better-educated workers. For the latter, interactions in the performance regressions are not significant, which is consistent with more widespread diffusion of the relevant

practices among these groups.

Finally, an important determinant of DDD effectiveness is the structure of decision making at the plant. Operating environments characterized by joint decision making between managers and frontline workers do not benefit from DDD adoption – unless they are in certain industries. A nuanced examination of the data finds significantly higher productivity with a three-way combination of: 1) increases from 2005 to 2010 in distributed decision making authority over daily tasks (versus manager-centric authority), 2) adoption of data-driven decision making, and 3) being in an industry that is relatively more prone to capital-intensive, “continuous flow” operations.

Consistent with popular accounts of the growing importance of “big data”, the overall pattern and timing of our results suggest that a new, more data-driven approach to management is indeed diffusing across U.S. manufacturing and is bringing economically significant productivity enhancements along with it for certain firms. These findings contribute not only to the rich literature on IT and firm performance (e.g., Brynjolfsson, 1993; Brynjolfsson and Hitt 2003 and Mithas et al. 2012) but to a smaller burgeoning literature in economics and information systems suggesting that way in which firms deploy IT may be more important than sheer levels of investment (e.g., Deveraj and Kohli 2003 and Bloom, Sadun, and Van Reenen 2012). Our paper builds on a growing stream of empirical research on the importance of management quite generally (e.g., Bloom, Eifert et al. 2011; Bloom, Brynjolfsson et al. 2013; Atalay, Hortaçsu, and Syverson 2014; Yang, Kueng, and Hong. 2015), and complementarities in management practices, more specifically (e.g. Novak and Stern 2009 and Hong, Keung and Yang 2015). Most narrowly, we build on prior work in both economics and management emphasizing complementarities specifically between IT and other practices in the firm (e.g., Milgrom and Roberts 1990; Bresnahan, Brynjolfsson, and Hitt 2002; Bartel, Ichniowski, and Shaw 2007; Aral, Brynjolfsson, and Wu 2012; Forman and McElheran 2012; and Tambe, Hitt, and Brynjolfsson 2012)<sup>3</sup> and the use of objective data in managerial decision making (Hoffman, Kahn, and Li 2015). This paper also relates to the rich stream of work in economics and strategic management highlighting the persistent performance differences among

---

<sup>3</sup> See also the review in Brynjolfsson and Milgrom (2013).

firms and emphasizing clusters of management practices as a driver of these differences (e.g., Chew, Bresnahan and Clark, 1990; Baily, Hulten, and Campbell 1992; Jensen and McGuckin 1996; Rivkin 2000; Porter and Sigglekow 2008; Syverson 2011; Gibbons and Henderson 2012). Finally, our results highlighting interactions between improved data collection and changes in decision rights contribute to a small but growing body of empirical work relating IT to delegation within firms (e.g., Baker and Hubbard 2004, Colombo and Delmastro 2004, Bloom, et al. 2014, and McElheran 2014)

## **2. CONCEPTUAL MOTIVATION & PRIOR LITERATURE**

Because organizations can be thought of as “information processors” (Galbraith 1974), it would be surprising if large changes in the costs of data collection, storage and analysis made possible by ever more powerful and affordable computing power did not lead to significant changes in a range of processes and systems in organizations. However, changing human systems, many of which are not well-understood, is costly, time-consuming, and often dependent on other systems within firms (Milgrom and Roberts 1990; Brynjolfsson, Renshaw and Van Alstyne 1997; Rivkin 2000). Thus, declines in the cost of data collection and processing can be expected to have uneven effects on firms as the complementary changes are invented and executed.

In particular, the act of collecting data serves to codify information, which makes it more explicit and less tacit. Lower processing and storage costs make it less expensive to analyze, integrate, and communicate information. Yet to take advantage of these changes, firms must make significant investments on many margins.

For firms with relatively new or highly variable production processes, early steps involve moving from “art” to “science”, whereby managerial efforts focus on greater standardization, mechanization, and instrumentation of the process (Bohn 2005). Even in fairly standardized environments, firms need to invest in the managerial decision making process required to select which data to collect (and take a stand on why it is important). To do this, firms typically go through a lengthy process of “learning what they know” by consulting with employees from throughout the firm and discovering what sorts of knowledge

reside in scattered locations throughout the firm. Importantly, this process is useful for capturing tacit knowledge that employees gain through less formalized channels, and codifies and centralizes it in accessible databases. If firms, like people, “can know more than they can tell” (Polanyi 1969), the steps necessary to implement data-driven decision making may help increase access to more objective information.

The economic value of this information is zero unless it leads to a change in action. Yet, it is not always easy to act on newly-codified and accessible information. In complex modern enterprises, this typically requires not only changes in decision rights and processes, but also complementary changes on multiple dimensions within the organization (performance pay, promotion, hiring, measurement, etc.) and even culture. This is inevitably costly and time-consuming, and often subject to mistakes. On the other hand, once implemented, these changes can be difficult for competitors or quickly imitate, providing a margin of competitive advantage, and measurable differences in performance – at least in the short term.

A key objective of this paper is to shed more light on how enhanced technological possibilities may reinforce or even spark changes in organizational practices. Given the rush to invest in new technologies and all things “data” (Gartner 2014), it is a critical time to better understand how these forces play out in actual firms and begin to unpack the conditions under which this is more or less likely to be productivity-enhancing. To that end, we explore in detail how data-based management practices are diffusing, whether they seem to have any correlation at all with firm performance, and if so, where those benefits tend to cluster and in what type of operating and organizational environments.

### **3. DATA**

Disentangling important strands of this complex phenomenon poses a non-trivial data challenge. This study is one of the first to take advantage of new data on management practices collected by the U.S. Census Bureau in 2011 and released for non-internal use in 2014 (for more details see Bloom et al. 2013). This Management and Organizational Practices Survey (MOPS) was included as a supplement to the Annual Survey of Manufactures (ASM), which targets roughly 50,000 of the over 300,000 establishments

in the U.S. manufacturing sector. It is generated from five-year rotating sample frames that end in years ending with “8” or “3” and are constructed with the intention of generating representative annual coverage of the manufacturing sector.<sup>4</sup>

The survey was sent by mail and electronically to the ASM respondent for each establishment, which was typically the accounting, human-resource, or general plant manager. Non-respondents were given up to three follow-up telephone calls if no response had been received within three months. The ultimate response rate was 78% (Bloom et al. 2013), which is extremely high for surveys of this nature.

The survey comprises 36 multiple-choice questions split into three sections. The first section, labeled “management practices”, is based on work by Bloom and Van Reenen (2007) and has 16 questions focused primarily on monitoring, communication, and incentives at the plant. Examples include the collection, review, and communication of key performance indicators (e.g., production targets, costs, quality, inventory, absenteeism, and on-time deliveries), the speed with which under-performing employees are reassigned or dismissed, and the basis for promotion and performance bonuses.

The second section, labeled “organization,” has 13 questions focused on decision making within the firm. Inspired by work such as Bresnahan, Brynjolfsson and Hitt (2002) and Bloom, Sadun and Van Reenen (2012), it contains a question on delegation *within the plant*, which asks about the prioritization and allocation of daily production tasks (more on this question below). Based on prior work by Brynjolfsson, Hitt, and Kim (2011), there are also two questions on the availability and use of data to support decision making at the plant.

The final section of the survey captures background characteristics of the respondent and establishment, primarily the respondent’s position and tenure at the plant as well as the number and education of the plant’s employees. The full questionnaire is available at <http://bhs.econ.census.gov/bhs/mops/form.html>. Bloom et al. (2013) explores the distribution and

---

<sup>4</sup> As a result, it somewhat over-samples large plants. Establishments with over 1,000 employees are sampled with certainty; the likelihood of sampling is lower but increasing with size for all plants below this threshold. The certainty sample accounted for approximately 67% of the total value shipped in the entire U.S. manufacturing sector in 2007 (see [http://www.census.gov/manufacturing/asm/how\\_the\\_data\\_are\\_collected/index.html](http://www.census.gov/manufacturing/asm/how_the_data_are_collected/index.html)).

performance implications of the practices covered in the first section. While we take advantage of many different aspects of the survey, our point of departure is the data-related questions from the second section, as well as two from the first that are directly related to data collection and use (see below).

**Sample Boundaries.** Our analysis requires that we restrict our attention to establishments that have positive value added, positive employment, and positive imputed capital in the ASM.<sup>5</sup> In order to keep our sample stable across specifications,<sup>6</sup> we further restrict our analysis to records with complete responses to the data-driven decision making questions, headquarters status, and a critical mass of the management practices questions.<sup>7</sup>

**Panel Data Sample.** Although the MOPS only took place once in conjunction with the 2010 ASM, all of the questions asked respondents to report on the state of practices in 2005. The ASM itself provides data on various plant characteristics for both 2005 and 2010. Using differences between what is reported for 2005 and for 2010, we employ a fixed-effects research design to control for time-invariant unobserved heterogeneity at the plant.<sup>8</sup>

Linking the 2010 MOPS sample to the 2005 ASM reduces the size of our analysis sample from roughly 34,000 observations in the 2010 complete-information cross section to a balanced panel of

---

<sup>5</sup> This is to make the standard productivity calculations possible and to exclude low-quality records that may introduce systematic biases to the estimation. To meet these requirements automatically requires a successful match between the MOPS and the ASM and that the establishment be flagged as worthy of tabulation in the national statistics. Another technical condition for the panel analysis (and to get controls such as age) requires the establishment to have a valid linkage to the Longitudinal Business Database (LBD).

<sup>6</sup> A stable analysis sample is essential to avoiding inadvertent disclosure of proprietary firm information and is encouraged by the U.S. Census Bureau when using the non-public data.

<sup>7</sup> Specifically, we require that firms answer at least 5 of questions 1,2,5,6,8,9,11,13,14,15,&16. We explore but do not require responses to questions 3 & 4 on the frequency of with which key performance indicators are reviewed. We exclude these questions from our core analysis because they are highly correlated with our other measures, have a relatively higher percentage of missing data, and offer no additional insights while problematically restricting the sample size.

<sup>8</sup> Of course, the validity of this approach depends in part on the quality of recall. As one gauge of quality, we explored similarities between the 2005 recall questions regarding plant-level employment and actual IRS records for that year and find the differences to be negligible (*pending disclosure*). In cross-sectional explorations (see Appendix Table A5), we included a measure of the discrepancy between the 2005 MOPS and 2005 ASM employment numbers as a “noise” control in many specifications. We also controlled for the tenure of the respondent at the plant, although the impact of these additional controls is trivial.



roughly 18,000 observations.<sup>9</sup> The loss of observations comes primarily from the change in ASM sample frame in 2008 and from the requirement that plants be at least five years old; a small number also lack complete information for 2005 and get dropped as a result. Due to these restrictions, our balanced sample is slightly biased towards plants that are larger and more productive compared to the entire ASM mail-out sample. Comparisons between the 2010 balanced sample, the 2010 complete-information cross section, and the 2010 ASM for key variables are provided in the Appendix (see Tables A1 and A3). Where appropriate, we explored using the ASM sampling weights to generate statistics that are relatively more representative of the entire population of manufacturing establishments; however, the survival condition and inclusion of a large proportion of certainty-sample plants overwhelms the ability of weighting to provide truly population-relevant inference. Thus, while our sample covers the majority of economic activity in this sector, we interpret our results as being principally informative about the population of larger manufacturing establishments in the U.S.

Subject to these limitations, examining these practices in a panel setting is a significant advance, as it expands the scope for addressing unobserved heterogeneity among establishments that could bias estimates of the return to adopting DDD.<sup>10</sup> Finally, our sample is representative of the diversity of activities comprising U.S. manufacturing, covering all industries from food to furniture manufacturing, and everything in-between (86 industries at the 4-digit NAICS level of aggregation).

**Data-Driven Decision Making (DDD).** The key questions about how firms use data to support managerial decision making are questions 27 and 28 of the MOPS. Respondents were asked to choose a value on a 5-point Likert scale according to “what best describes the **availability of data** to support decision making at this establishment”, and “what best describes the **use of data** to support decision making at this establishment.” Empirically, they are highly correlated and for our core results we combine

---

<sup>9</sup> Exact records counts are suppressed in the interest of disclosure avoidance.

<sup>10</sup> We do not take a stand on the direction of the bias, since while unobserved factors could simultaneously promote productivity and DDD adoption, we also have measurement error that could be biasing the effect towards zero (and which will be worse in a fixed-effects model).

information from both to reduce measurement error.

A central concern for measurement is that the *perceived* availability and use of data in U.S. manufacturing was quite widespread by 2010. Over one-half of the establishments responding to the MOPS report being relatively intensive both in collection and use of data (see Appendix Table A1 for details). To address this concern, we take advantage of question 2 of the MOPS, which asks about the number of key performance indicators (KPIs) tracked at the establishment. Respondents are primed with the following example: “Metrics concerning production, cost, waste, quality, inventory, energy, absenteeism and deliveries on time.” Our prior, which has been corroborated by qualitative interviews with plant managers<sup>11</sup> and Census’ own field testing of the survey instrument, was that the number of identified and tracked performance measures is an essential measure of the breadth and/or intensity of data gathering at the establishment. Thus, we combine the more-subjective DDD questions with this objective measure of whether plants collect data on 10 or more KPIs. Adding this restriction reduces the number of firms reporting intensive use of data to 32% in 2010 (see Table 1). Empirically, these measures are correlated, and a polychoric factor analysis supports our use of the combination of measures (see Table 2 and discussion below).

Conceptually, having appropriate targets against which to compare real-time or historical data plays an important role in decision making. Targets help inform managers about whether the production system is performing appropriately (i.e., is the data conveying “good news” or “bad news?”), identifying the locus and magnitude of the problem, and formulating appropriate actions. Again, this interpretation of the role of targets in DDD has been qualitatively corroborated in independent interviews with plant managers. Question 6 of the MOPS asks about the presence and time-frame of production targets (short-term, long-term, or combined). We take the combined approach to as a more advanced engagement with the dimensions of performance that must be monitored and controlled.

Thus, we create an indicator for being at the frontier of data availability (question 27) and use of

---

<sup>11</sup> These were conducted independently of the Census Bureau data collection and survey validation process and do not represent responses from actual firms in the ASM sample frame.

data in decision making (question 28), extensive use of key performance indicators (question 2), and employing a combined approach to target-setting (question 6). We call this cluster of practices “data-driven decision making” (DDD) throughout the rest of this paper.

Relying on this combination of practices to identify DDD is empirically justified by a formal polychoric principal factor analysis (see Table 2). Applying this technique – appropriate for factor analysis of discrete variables – to these four dimensions of practice reports a single factor with an eigenvalue of 2.28 accounting for 57% of the variance in the balanced sample in 2010. An oblique promax rotation confirms a single factor; similar results also obtain for principal-component factor analysis (available upon request). Nevertheless, we further explore the robustness of our findings to a range of different definitions of DDD (see Table A2 in the Appendix).

Finally, while our main focus is on *frontier* use of data-related practices, we investigate less-discrete shifts in DDD by normalizing the responses to questions 2, 6, 27 and 28 (which are on different scales) and summing the scores. We scale this “DDD Index” to lie on the [0, 10] interval to facilitate interpretation and comparison with the binary measure. It is useful to know that a one-standard deviation increase in this index in 2010 would be equivalent to moving to a higher category for two to three questions or making a substantial leap in intensity for one to two questions. For example, this difference would capture the distance between plants that report being “moderate” for both data availability and use and those that have both a “great deal of data” available for decision making (question 27) and “rely heavily on data” in decision making (question 28). The index value and composition for plants that clear our DDD threshold varies considerably, but the mean value is close to the top of the range (*precise values pending disclosure review*).

The first two columns of Table 3a show how adoption of DDD according to these two measures changed over time in the balanced panel.<sup>12</sup> A rapid growth in DDD over our sample frame is apparent in the data. Only 11% of plants reported intensive DDD according to our definition in 2005. By 2010, 30%

---

<sup>12</sup> While we focus on the balanced panel for econometric reasons, adoption rates for the 2010 complete-information cross section -- which has many more small and young establishments and slightly less DDD adoption -- are reported in the Appendix (see Table A1).

of plants in our analysis sample achieve the DDD cut-off. Table 3b splits this out by multi-unit status, a useful distinction to which we return later. Multi-unit establishments go from 11% adoption in 2005 to 34% adoption in 2010; single units go from 5% to 15% on average over the same time frame. Change during our sample period is relatively incremental, with multi-unit establishments increasing their DDD Index by roughly 17% and single-units by a slightly lower 15%. We explore changes in DDD practices over time in more depth in Section 4.

**Management Practices.** A key concern for identifying the relationship between DDD and higher productivity is the possibility that DDD may merely proxy for other “structured” management practices at the firm. Bloom et al. (2013) show robust positive correlations between the adoption of structured management practices and performance measures like to the ones we study here. If DDD and structured management are correlated, then measures of DDD adoption in isolation could simply pick up the effect of these other practice changes. To address this concern, we construct indices of structured management that are similar to those used by Bloom et al. (2013) but that *omit the data-related measures discussed above*. In a departure from prior work, we further eliminate measures that appear to be particularly confounded with performance, such as the likelihood of reassigning or dismissing underperforming workers quickly (which may tend to happen more when the firm, in general, is performing badly) and whether performance bonuses were paid in the prior year (again, which might depend on whether performance was sufficiently good to warrant bonus pay in the first place). We create a Z-score indexing the remaining non-DDD, non-performance questions from the first section of the survey as our main proxy for structured management.<sup>13</sup>

---

<sup>13</sup> This consists of normalizing the response of all of the following questions to the 0-1 interval and summing them to create the composite management score: 1, 5, 8, 9, 11, 13, & 14. These questions cover how the firm reacted to an exception in its production process, whether and where display boards showed output and other key performance indicators, who was aware of production targets at the plant, what the basis (if any) was for performance bonuses for managers and non-managers, and the basis for promotion of managers and non-managers (performance and ability versus other factors such as tenure or family connections). Based on explorations of the broader survey instrument using discrete principal factor analysis, we separately construct and include an index for how quickly an underperforming non-manager or manager is dismissed or re-assigned (questions 15 and 16).

**Investment in Information Technology.** Advances in information technology have had significant impacts on the productivity of U.S. firms (e.g., Brynjolfsson and Hitt 2003). At the same time, IT has changed what is measurable, analyzable, and communicable within firms. Firms that have invested significantly in information technology in prior periods may have a greater volume of digitized information to draw on when they contemplate changes to their management practices, boosting average adoption and/or the productivity benefits of DDD. Conversely, firms that shift their management approach to be more data-driven might subsequently boost their IT investments to provide inputs to managerial decision making.

To grapple with these issues empirically, we calculate IT capital stock using the ASM and CMF questions on hardware expenditure dating back to 2002 and software expenditure questions dating back to 2006. Specifically, we use the Bureau of Economic Analysis (BEA) deflators and a perpetual inventory approach, combining hardware and software investment, imputing values for years in which they are missing<sup>14</sup>, and depreciating at the rate of 35% per year.<sup>15</sup>

**Worker Education.** Human capital is another factor that could significantly influence the productivity of DDD within firms. Questions 34 and 35 ask about the percentage of managers and non-managers, respectively, with bachelor's degrees. We use a combined measure to a) control for labor quality in our performance regressions and b) explore whether DDD is complementary to having a more-educated workforce.

**Learning.** As discussed in Section 2, we hypothesize that the many margins of adjustment required for effective DDD may be difficult for firms to discover and implement. To explore whether the patterns in the data are consistent with a learning-based diffusion mechanism, we take advantage of

---

<sup>14</sup> For plant-years where IT expenditure information is missing, we impute the missing values using the average of the IT investment from the closest before and after years that have non-missing values. For instance, if IT investment in 2008 is missing, we impute it using the average IT investment for the plant in 2007 and 2009 or using the 2007 and 2010 values if 2009 is missing. Similar logic is applied to missing values from other years. Our core results are robust to excluding observations with missing IT data.

<sup>15</sup> Based on the BEA Consumer Price Index for All Urban Consumers: Personal computers and peripheral equipment.

question 29 of the MOPS, which asks “In 2005 and 2010, did the managers at this establishment learn about management practices from any of the following?” Respondents may choose all that apply from: consultants, competitors, suppliers, customers, trade associations or conferences, new employees, and headquarters. We create an index of the number of learning modalities based on the assumption that more ways of bringing information into the firm will facilitate learning about the benefits of DDD, how best to implement it, and what other complementary changes might be beneficial for the particular operating environment

**“Lean” Practices.** Question 1 asks “what best describes what happened at this establishment when a problem in the production process arose?” One of the options respondents may choose states that the establishment “had a continuous improvement process to anticipate problems like these in advance.” This may matter for understanding DDD in two ways. The first concerns measurement. A “continuous improvement” approach to problem-solving is characteristic of a cluster of practices commonly referred to as “Lean Manufacturing” (an iconic example is the Toyota Production System, described in detail in Roos et al. 1991). Managers steeped in these practices for monitoring and coordinating the production process may have particularly high standards for whether they have “enough” data available to them or “enough” use by key stakeholders; plants with these practices in place may under-report their reliance on DDD compared to other plants.

This second way this matters concerns productivity estimation. To begin, a lower subjective interpretation of relatively high DDD practices will assign these more-sophisticated plants to the non-adopter category, potentially attenuating the effect we are trying to measure. Yet the relationship could run the other way, too. Plants that have made significant organizational investments in developing a Lean culture may have workers who are more receptive to standardization and the use of objective measures to guide decision making at the firm. Thus, a Lean orientation at the firm may promote responsiveness to signals coming from data, making the benefits of investing in data collection disproportionately greater to these firms in ways that might otherwise be difficult to observe. Lacking additional data on related

practices, we rely on our econometric approach to control for time-invariant attributes of the establishment and a time-varying indicator of whether or not plants report this Lean-style approach to addressing problems in the production process.<sup>16</sup>

**Delegation/Joint Decision Making.** Finally, we explore whether the relationship of DDD to productivity depends on the organization of decision making within the firm. To do this, we construct an indicator of whether production workers have some authority over the allocation and prioritization of daily tasks. Question 26 asks “In 2005 and 2010, who prioritized or allocated tasks to production workers at this establishment?” We take a response of “managers and production workers jointly”, “mostly production workers”, or “only production workers” as evidence of some delegation to frontline workers. Empirically, relatively few establishments in our sample delegate authority primarily or entirely to frontline workers; thus, our results are informative about the margin between manager-centric authority and joint decision making.<sup>17</sup>

**Performance Measures and Controls.** Our main specifications rely on value added as the dependent variable and plant-level fixed effects to control for time-invariant unobserved heterogeneity. We also include a number of time-varying controls collected from the ASM files. These include depreciated capital stock (calculated following the perpetual inventory method), labor measured in terms of the number of employees, and energy inputs.<sup>18</sup> This approach is useful for measuring a plant’s productivity, because it estimates how much output the plant creates, while controlling for how much it spends on primary inputs. This is similar to a standard total factor productivity (TFP) approach, but does

---

<sup>16</sup> In fact, the relationship between Lean and the collection of data on certain aspects of production such as individual worker performance may be complex and mediated by attributes of firm culture that we do not observe in the MOPS. Gibbons and Henderson (2012) explore this conceptually; Blader et al. (2015) provide empirical evidence from a field experiment.

<sup>17</sup> The MOPS also contains a wealth of information on the allocation of decision making between local plant managers and headquarters. For this study, we use the local measure of delegation to frontline workers at the plant, as it measures delegation within the plant – our unit of analysis – and does not introduce the added complexity needed to model multiple layers of delegation within the parent firm

<sup>18</sup> The energy consumption measure is calculated by combining expenditures on electricity and fuels, logging the value, and then winsorizing it to reduce the impact of outliers and help with disclosure avoidance. We log the capital and labor measures, as well, to address the highly skewed nature of their distributions.

not require the log-linear functional form. As a robustness check, we also conduct a standard TFP analysis, where we estimate a log-linear production function with the aforementioned inputs but with industry-specific factor shares, taking the residual as a measure of plant productivity.

We further explore alternative measures of plant performance that speak to important issues regarding increases in digitization. We use the reported number of production workers at the plant to observe whether DDD and IT may help firms substitute away from labor in their production function – a key concern of the “hollowing out” literature (e.g., Autor and Dorn 2013). We also use the value of e-commerce sales to highlight the tight interconnections between electronically-mediated transactions – which tend to entail a higher volume of digital information – and DDD.

Finally, we use other observable characteristics of the plant to control for important heterogeneity among firms who adopt DDD. An important feature of manufacturing plants is their “multi-unit status”, which is an indicator of whether a plant belongs to a firm with more than one manufacturing establishment. Prior work has shown that larger, more productive plants are more likely to be integrated under the same ownership (Atalay, Hortaçsu, and Syverson 2014), suggesting that this may proxy for differences in underlying firm quality. We also observe and use information on whether or not the plant is a headquarters (specifically, whether it is a headquarters *in addition to being a production unit* – non-production administrative offices are collected in a separate survey and not in the MOPS sample), whether or not the plant exports, and whether or not it accepts orders electronically (e-commerce). Means and standard deviations for all of these variables are reported for the balanced sample in Table 2; pairwise correlations can be found in the Appendix (Table A5).

**Industry Statistics.** In all of our probit analyses, we control for industry using 3-digit NAICS indicators. For understanding the phenomenon, it is separately interesting to observe variation by industry in DDD adoption, a topic which we explore in depth, next.



#### 4. ADOPTION OF DATA-DRIVEN DECISION MAKING

While the non-academic conversation about data has focused primarily on firm performance, it is important to step back and look at the diffusion of DDD over the five-year period we observe. This is not only useful for understanding the phenomenon, but also for illuminating the identification challenges in our setting. In particular, we find that adoption of DDD varies considerably by certain characteristics of industries, plants, and managers, and that firms that report earlier adoption of DDD are different in particular ways from later adopters.

Table 5 provides some more descriptive statistics on DDD adoption, breaking it down by both year and 3-digit NAICS industry for the balanced sample. Top industries for DDD adoption by 2010 include: Transportation (NAICS 336), Beverage and Tobacco (NAICS 312), Food (NAICS 311), Paper (NAICS 322), Chemicals (NAICS 325), and Electrical and Appliance (NAICS 335) manufacturing. DDD prevalence by 2010 in our sample ranges from 34-41% for these industries.

In contrast, laggard industries include Apparel and Leather (NAICS 315 and 316 combined<sup>19</sup>), Furniture (NAICS 337), and Printing (NAICS 323); the average DDD prevalence in these industries by 2010 ranges from 13-19%. The most significant change in the prevalence of DDD practices is in the Beverage and Tobacco manufacturing industries, which vary widely from wineries to cigarette manufacturers, and Transportation, which includes aerospace and automobile manufacturing.

Based on prior work emphasizing the need for greater standardization, monitoring, and responsiveness to exceptions in settings where production tends to be continuous (or near-continuous) and often capital-intensive (see e.g., Hayes and Wheelright 1979), we group together Wood, Paper, Petroleum, Chemicals, Non-Metallic Minerals, and Primary Metal industries into a “continuous flow” category. The average adoption in these industries is also relatively high at 34% (reported in the last row of Table 5), consistent with this story. Unfortunately, we are limited in what we can observe and report

---

<sup>19</sup> We combine these industries in our descriptive statistics to help prevent inadvertent disclosure of firm participation in the survey.

about these production processes outside of relatively broad industry categories. However, we leverage information about their input mix to see if capital intensity predicts higher levels of DDD (see below).

To more systematically explore the potential drivers of DDD adoption, we estimate a standard probit model of adoption (David 1969). Because our performance regressions focus on the late adopters, we first report on the results for adoption between 2005 and 2010 using the sample of firms that has complete information across both years (our 2010 “balanced” sample). Note that this restricts the analysis to firms that report **not** having DDD according to our definition in 2005.

Table 6a reports the average marginal effects calculated at the sample mean of all the covariates, which are lagged to 2005. We control for the industry variation described above with 3-digit NAICS indicators in all specifications. Thus, the results should be interpreted as within-industry relationships.

To begin with the IT-related themes that originally motivated this project, greater investment in IT at the establishment is correlated with the presence of DDD (see columns 1-5, first row). This correlation is consistent with complementarities, between IT and DDD whereby the firms that have more IT reap a greater reward from DDD and vice versa – a relationship we explore further in Section 5. We do not interpret this as causal<sup>20</sup>, although results which we discuss later suggest that the order of adoption associated with the best productivity returns is IT first, DDD second. This effect is largest for the subsample of single-unit firms (not reported).

Because IT adoption could be correlated with a general unobserved push towards standardization and more “structured” management, we also explore the relationship between earlier adoption of structured management practices and DDD. The second coefficient in Column 1 shows that the use of structured management is positively correlated with the adoption of DDD. However, Column 2 suggests that this could largely be a size effect. Large plants are more likely to have both DDD and a more structured approach to management; in specifications that control for the total number of employees, the

---

<sup>20</sup> Moreover, we cannot rule out common cost shocks that could drive correlated adoption between the two without a true complementary relationship (see Athey and Stern 1998).

effect of structured management disappears. To avoid problems of multicollinearity, we omit Structured Management from the subsequent probit specifications.

Column 3 reports that belonging to part of a multi-unit firm is correlated with a significantly higher likelihood of DDD adoption, even controlling for establishment size. Capital intensity, which is also correlated with size, nevertheless has an independent relationship with the likelihood of DDD adoption; note that this is defined as being in the top quintile relative to other plants in the same 4-digit NAICS industry. When we control for capital intensity, the positive association with being in a “continuous flow” industry largely disappears or even becomes negative (primarily for single-unit firms – not reported)<sup>21</sup>, suggesting that this is not necessarily an industry-wide standard practice or norm, but more fundamentally linked to the standardized nature of capital-intensive production (which generates more codified information) and/or high demand for keeping track of the condition and up-time of expensive equipment.

Column 4 explores how “background characteristics” of the plant may be correlated with DDD adoption. Plants with more educated workers, more managers, and more layers of management are more likely to report high levels of data-driven decision making. The significant correlation between DDD adoption and the percentage of workers at the plant with bachelor’s degrees is consistent with complementarities between DDD and skilled labor, something we explore in more detail in Section 5. Having more managers and more layers of management may indicate a greater need for objective measures to facilitate coordination across and between levels of the hierarchy; note that this effect is conditional on size, indicating that hierarchical structures *per se* are more likely to coincide with DDD.

Column 5 explores features of the plant that may substitute for the use of data at the plant and provides some insight into what the alternative to DDD might be. Older plants report a lower likelihood of DDD. While this could be due to the common resistance among older organizations to new technology (and the practices that rely on it), it could also signal a substitution of experience and tacit knowledge for

---

<sup>21</sup> In the interest of space, we do not report the adoption regressions for the multi-unit and single-unit subsamples separately. The pattern of results is very similar, with the exceptions noted above.

objective data. In fact, an alternative hypothesis about the effect of age could be that older organizations have had more time to standardize and instrument their processes, which would lead to a positive relationship between age and DDD. None of these mechanisms are mutually exclusive; in our setting, the net empirical relationship is negative.

Greater tenure of the respondent is also seen in column 5 to be negatively correlated with the adoption of DDD, as is having the CEO fill out the questionnaire. Again, a possible explanation is that these are managers with a high level of experience who rely less on formal data collection and data-driven practices. Or, particularly in the case of the CEO, these individuals may rely less on data to imbue their decisions with sufficient authority (Porter 1996).

Another (not mutually exclusive) possibility is that these respondents are at plants that objectively collect and use a fair amount of data, but that their **subjective** perception of that collection and use is downwardly biased. This could be due to cognitive biases related to how important their informal contribution is to overall decision making at the establishment, or, conversely, due to very high standards for what constitutes a high level of data availability and use. We explore this by splitting out the more objective components of DDD (KPI's and nuanced production targets) from the more subjective ones, and found that the negative relationship between tenure/CEO position was largely confined to the subjective measures, although the coefficients in these regressions were not very precisely measured (available upon request). Because we cannot disentangle the root cause of this regularity in the data, we verify in Section 5 that our core results are robust to excluding CEO and long-tenure respondents.

Much of the above discussion is based on an implicit view of the firm as relatively centralized, where managers are in control of relevant decision making. However, the literature on decision making within firms points to delegation as a powerful tool for facilitating well-adapted action within firms while easing the burden on managers to collect, organize, decipher, and interpret large volumes of information from throughout the firm (e.g., Dessein and Santos 2006; Bloom et al. 2014, McElheran 2014). Question 26 of the MOPS provides information on the involvement of frontline workers in making decisions concerning daily tasks. Empirically, there is little use of very high levels of delegation in our

manufacturing setting. However, there is reasonable – and growing – use of joint decision making between managers and frontline workers that could conceivably substitute for DDD. In the adoption regressions, we find no evidence that this moderate level of delegation affects the adoption of DDD. We explore this dimension of the decision-making environment further in Section 5.

Finally, as we are concerned with the performance implications of DDD, we have to grapple seriously with other factors that might have less to do with the net benefits of DDD at the plant (the model motivating the probit estimation) and more with standard diffusion-model concerns related to knowledge and spillovers (see Karshenas and Stoneman 1993). Even controlling for many plant characteristics, we find that variation among establishments in how they learn about management practices has a significant impact on the likelihood of DDD adoption. Quite simply, firms that learn less about new practices are less likely to adopt DDD.

All of these relationships are statistically significant, most at the 1% level, and for the most part empirically large. Moreover, they are similar across multi-unit and single-unit establishments with the exception of IT investment, which has a much higher magnitude for the smaller stand-alone plants (not reported). These patterns are important to keep in mind when we interpret the relationship between DDD and firm performance in the next section.

Table 6b conducts a similar exercise, but on the entire sample of firms for which we have information in 2005. The covariates here are lagged to 2002. Lacking data on the organizational characteristics, these are sparser specifications and may be missing important organizational details that are simply beyond the reach of our data. For the most part, similar patterns apply to these early adopters, with the exception of IT capital stock: the positive relationship between IT and DDD adoption applies only to the smaller single unit firms (see column 4). The effect of age is also only significant for the single-unit plants.

## 5. DATA-DRIVEN DECISION MAKING AND FIRM PERFORMANCE

In order to investigate the relationship between DDD and performance, we take a conventional approach to modeling the plant production function. Assume that the establishment production function is as given in equation (1):

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} E_{it}^{\gamma} IT_{it}^{\lambda} e^{\mu SM_{it}} e^{\eta X_{it}} e^{\delta DDD_{it}} \quad (1)$$

where  $Y_{it}$  is real value added (output - materials),  $A_{it}$  is productivity,  $K_{it}$  denotes the establishment's capital stock at the beginning of the period,  $L_{it}$  is the labor force,  $E_{it}$  is the establishments consumption of energy inputs,  $IT_{it}$  is the establishment's IT capital stock (hardware and software) at the beginning of the period,  $SM_{it}$  is a measure of structured management at the establishment,  $X_{it}$  is a vector of additional factors like industry and employee education, and  $DDD_{it}$  is our measure of data-driven decision making.<sup>22</sup>

Taking logs provides a tractable form to take to the data:

$$\text{Log}(Y_{it}) = \alpha \log(K_{it}) + \beta \log(L_{it}) + \gamma \log(E_{it}) + \lambda \log(IT_{it}) + \mu SM_{it} + \eta X_{it} + \delta(DDD_{it}) + p_i + \varepsilon_{it} \quad (2)$$

where the productivity term has been decomposed into a set of plant fixed effects  $p_i$  and an added stochastic term,  $\varepsilon_{it}$ .

We estimate equation (2) using the balanced panel of roughly 18,000 plants that report data on the relevant questions for both years and can be linked up to the appropriate ASM surveys for the performance measures and controls. Column 1 of Table 7 shows the DDD coefficient controlling for changes in IT capital stock but not changes in structured management. This estimate of roughly 5% higher value added is robust to excluding the IT control (available upon request). Controlling for changes in structured management practices has a large effect in column 2, justifying our concern about confounding different types of management practices at the plant. In addition to controlling for inputs to the production function, other time-varying plant-level controls include changes in skill mix at the plant (using the percentage of employees with bachelors' degrees), a year-fixed effect, and changes in firing

---

<sup>22</sup>We put the management score and  $x_{it}$  controls to the exponential for the convenience of including them in levels rather than logs.

practices based on questions 15 and 16 (not reported). This, our preferred specification, reports a correlation between value-added and DDD of just over 3%. To put these findings in context: to achieve the same advantage in value-added from IT investment as from DDD adoption would, at the mean of the IT distribution entail an additional \$5 million in IT capital accumulation over the five-year period – without spending any more on technology.<sup>23</sup>

Another insight from column 2 is that unobserved heterogeneity may have a non-trivial impact on the observed relationship between DDD and firm performance. Controlling for time-invariant plant characteristics, the magnitude of the DDD coefficient is roughly half of what it is in the comparable cross-sectional analysis (see Table A5 in the Appendix)

Columns 3 and 4 reveal that the magnitude of this relationship varies a great deal by whether the plant is a single-establishment firm or belongs to a larger, multi-plant firm. Most of the benefit is concentrated among single-unit firms that adopt DDD, as opposed to plants belonging to larger firms. The coefficient for single-unit plants is a striking .144, and significant at the 1% level; that for multi-unit plants is only .011 and noisy. This is one place where the precise definition of DDD might affect interpretation: if we exclude the short-term and long-term targets criterion, the coefficient on DDD is .029 and significant at the 1% level (see Table A2 in the Appendix).

It is useful to have the adoption results in mind when drawing conclusions from these estimates. As described in Table 4a, multi-unit establishments are disproportionately more likely to have cleared the threshold for intensive DDD by 2010. In fact, multi-unit plants have an incidence of DDD that is significantly higher than that for single-unit firms – Table 3b shows that by 2010 they report a mean adoption of DDD of roughly 34%, while single-unit plants report an estimated mean adoption of 15%. This pattern is seen in 2005, as well: 13% of multi-unit plants used data intensively, while only 5% of single-unit firms did.

---

<sup>23</sup> However, as noted previously, the IT capital distribution is quite skewed. Another way to interpret this magnitude is to know that the DDD coefficient is nearly identical to that for the indicator of being in the top quartile of the IT capital distribution (available upon request). In rough numbers, the difference in IT capital stock between an average plant in the top quartile of the IT distribution versus an average plant in the bottom half is approximately \$1 million.

This information is handy for interpretation in two ways. First, we note that the loss of precision for the multi-unit sub-sample is not due to the reduction in sample size – the number of plants that change from non-DDD status to DDD status is much larger for multi-unit plants than single-unit plants. Second, we point out that a much smaller effect is consistent with optimal sorting into decision making regimes by rational managers. As these plants are older, larger, and generally “higher quality” along many dimensions (e.g., Atalay, Hortaçsu, and Syverson 2014), it would be unsurprising that this population of firms is quickest to discover whether and to what extent DDD might be effective in their operating environments -- and make choices to adopt DDD (or not) that are optimal for them. Under these circumstances, we should not see large systematic differences in productivity in equilibrium between adopters and non-adopters. Moreover, the frontier of practice may be moving. What was competitively important for these firms in 2005 may no longer be as much of a differentiator in 2010.

Columns 5-7 perform a similar exercise, but instead rely on a continuous measure of DDD. The pattern of results is quite similar, and shows that it is not necessary to clear our DDD threshold to have a positive correlation with productivity. Moreover, in our sample of firms and at this point in time, more DDD is always better. While one could imagine diminishing returns to DDD and IT investment, it does not appear that the inflection point was reached in our sample as of 2010.

Returning to our hypothesis about optimal sorting - for this to be true, we would also expect that multi-unit plants have learned the most about what investments and practices might be complementary to DDD (Brynjolfsson and Milgrom 2013). Thus, any correlational tests for complementarity should be stronger for this group of firms, which is what we find when we split the formal correlation tests by multi-unit status (*pending disclosure review*).

Table 8 explores variation in these effects by observable plant characteristics. Column 1 reports the results of including an interaction between an indicator of whether an establishment had more than 100 employees in 2005 (roughly the mean of the logged employment distribution for that year). Note that we keep the size indicator fixed at 2005 levels so as not to confound changes in size as well as changes in DDD status. As a result, the direct term is differenced out of the regression (and this approach holds for



the other specifications in this table, as well). In this model, *all* of the performance differential associated with DDD appears among to the smaller establishments, with an interaction effect on size of -7.4% and significant at the 1% level. Column 2 shows that the productivity differential is also only apparent among younger firms, where management may be less experienced. Again, the interaction term of -9% is significant at the 1% level and eliminates the direct effect of DDD. Similar patterns play out for multi-unit status (column 3) and capital intensity (column 4). Column 5 clarifies that these are not all separate effects – there is a great deal of correlation between being multi-unit, large, older, and capital-intensive (see also Table A4 in the Appendix). However, there is an identifiable population of plants that adopt DDD but do not display higher productivity than plants in the same population that refrain from doing so.

If multi-unit plants are doing a reasonably good job of sorting into decision making regimes, then why are the smaller single-unit firms not also showing a similar pattern? We attribute this difference to the difficulty of learning about complementarities among management practices and supporting investments and the fact that these challenges are likely to be higher in younger firms with less-experienced managers and less-developed production technologies. The positive effect on learning in the adoption regression (table 4a) is consistent with this story. In addition, in the absence of differences in complementary investments between early and late adopters, we might expect the performance effects for these two groups to converge – a hypothesis that we can test in the data.

### **Timing Falsification Exercise**

We exploit the rich Census data to gain insights into the timing of these effects. Although we only observe DDD and management practices for 2005 and 2010, we can observe performance and non-MOPS-based controls for every year preceding our sample and for two years afterwards. Figure 1 graphs the coefficient on DDD interacted with year fixed effects for a panel of data from 2002 to 2013. The specification is identical to equation (2) except that it omits the organizational controls (which are unobserved in the pre-period). The top line shows the coefficient on early adoption, the bottom line shows late adoption. By the last year of data, there is no statistical difference between the two, consistent with

our hypotheses about firm learning, complementary investments, and sorting into optimal decision-making regimes. Moreover, we cannot attribute the interaction effects in Table 8 to higher “adjustment costs” for the larger multi-unit establishments, as these plants are well-represented amongst the early adopters, who show an early productivity response to DDD.

This figure is also informative about whether the relationship between DDD and performance can be interpreted as a causal one. Since we lack an exogenous shock to DDD adoption, we hesitate to claim unequivocally that DDD *causes* better firm performance. The pattern in Figure 1, however, shows that the positive correlation between DDD and performance only shows up *after* adoption for the relevant population of firms. For 2005, we do not observe when they adopt, but it is reasonable to assume that practices were improving over the few years prior to reaching our threshold for DDD by 2005. Similarly, we observe whether plants that did not have DDD in 2005 did so by 2010. We observe relatively incremental changes in DDD-related practices over time (see Table 1b), suggesting that to go from very little in the way of availability and use of data to frontier DDD would take some time. The upward trends for DDD start a few years before 2005 for the early adopters and a few before 2010 for later adopters, consistent with a gradual but causal relationship between DDD and performance.

Moreover, If better performance were preceding DDD adoption, a regression of DDD adoption on value-added in our pre-period should show a positive relationship. However, a probit analysis of DDD adoption in 2005 (Table 9) shows that value-added growth from 2002 – 2005 does not predict the presence of DDD in 2005. A similar regression for those plants that did not have DDD in 2005, using growth in the 2005 – 2010 period as the key explanatory variable, again shows that growth in value added does not predict adoption of DDD in the window between 2005 and 2010

Of course, this timing falsification test is not completely dispositive. The classic problem with ascribing positive performance results to any type of technology adoption is that adoption is typically voluntary, and those adopting are more likely to be those who expect to benefit from adoption (David 1969 and Rogers 2010). Good instruments for adoption are also typically difficult to find in large, heterogeneous populations. Our study is subject to these critiques as well, with one important caveat:

DDD is an intangible practice that requires investments in managerial attention and time, but it is not a standard investment in tangibles such as various types of capital that typically requires free cash flow -- and hence it is less likely that only firms with prior good performance will be able to invest in DDD.

### **Complementarities**

Table 10 further examines the complementarities that were hinted at in the adoption regressions. We find evidence consistent with complementarities between IT adoption and DDD. In Column 1, we again fix the indicator for having IT capital stock in the top quartile of the 2005 distribution, so the direct effect is differenced out. The interaction effect has a coefficient of .075 and is significant at the 1% level. The coefficient on DDD by itself is not significantly different from zero, suggesting that almost all of the difference in performance occurs in establishments that already have significant IT investments and subsequently adopt DDD. This is unsurprising to the extent that IT infrastructure is needed to collect, track, and analyze the data inputs to decision making at the firm, a theme that occurred often in our qualitative interviews.

Likewise, more-skilled labor also appears to be complementary with DDD in column 2 of Table 11, with the effect driven entirely by establishments that had greater than the median percentage of managerial and non-managerial employees with bachelor's degrees in 2005 (again, the degree term by itself is differenced out). Column 3 shows that these effects stand up to including both interactions simultaneously (*pending disclosure review*).

Perhaps one of our most tantalizing findings is evidence that DDD is complementary to the way in which decision making is organized within the plant. Using an indicator for whether frontline workers have joint discretion over daily tasks in 2005, and interacting it with DDD adoption, we find in column 4 that the interaction effect has a large and significant negative coefficient, suggesting that DDD is less beneficial in decentralized organizations. In column 5, we allow the decision-making regime to change as well, and a similar negative coefficient obtains (though it is noisier) – this, despite the fact that increased delegation is, on average, associated with higher productivity. The lack of precision is consistent with

important underlying heterogeneity among plants; indeed, we are able to parse this further interacting it with an indicator of being in one of the “continuous flow” industries. Here, we find important variation in the effect: DDD is complementary to joint decision making only in these operational contexts and not in others. The linear combination of DDD, Joint Decision Making, and Continuous Flow indicators is .082 and significant at the 10% level. While some of this variation is likely due to things we cannot observe about the production process, a very similar pattern arises when we substitute our indicator of being in the top quartile of the industry’s (4-digit NAICS) capital distribution (*pending disclosure review*). This is consistent with our hypotheses about the “fit” of DDD in operating environments that are more standardized and instrumented – hallmarks of capital-intensive production environments— as well as the disproportionate benefits of keeping expensive capital equipment operating efficiently and continuously. DDD combined with joint decision-making may be particularly effective in helping workers in these environments respond to exceptions very quickly.<sup>24</sup>

Outside of these environments, centralized coordination – rather than delegation – is apparently important for DDD to be effective in improving productivity. These may be settings where manager expertise is high relative to frontline workers or where coordination is more important relative to responsiveness. Future work is required to disentangle possible explanations. At the least, these empirical regularities help delineate the boundaries of this emerging DDD phenomenon and highlight directions for future research.<sup>25</sup>

Table 11 conducts formal complementarities tests for DDD and IT investment. This relationship passes both the performance and correlation tests, though the evidence is weak that early adopters of DDD were able to benefit from layering on more IT investment between 2005 and 2010, suggesting a certain ordering to the related organizational investments, with IT first and DDD second

---

<sup>24</sup> Three-way interactions between DDD, skilled labor, and delegation were not precisely estimated; thus, we cannot draw clear inferences about whether the relationship between DDD and delegation depends on the human capital at the plant.

<sup>25</sup> We explored other interactions, but found little to be statistically significant, including continuous improvement practices (Q01), firing practices (Q15 and Q16), nonunion status, vertical integration measured as the value of within-firm transfers between plants, communicating with the firm about production targets (questions 5 and 8) and frequency of review of KPIs (questions 3 and 4).

We do not report a similar exercise for skilled labor, as the interaction terms are quite noisy and fail the formal test for complementarities in the performance regression. The formal correlation tests for complementarities with skilled labor do pass, as do those for a continuous improvement culture (*pending disclosure*).

Table 12 shows how DDD relates to other performance measures. Column 1 shows a weak relationship to the residual of a standard log-linear production function (total productivity, or TFP). This is largely due to confounding multi-unit and single-unit plants; the coefficient on DDD for single unit plants is both large and statistically significant in column 3. Column 4 provides evidence that DDD is *not* being used to substitute for production workers, and may in fact facilitate retaining employment at the plant during the period we observe, which includes the Great Recession. The coefficient on DDD for non-production workers is .018 and significant at the 10% level. Column 5 shows that the relationship to e-commerce is particularly high, possibly due to the amount of digital information created by electronically-mediated transactions.

Other robustness checks (not reported) shows similar results. We explore whether the results change when including controls for whether or not the plant is an exporter, and whether or not it conducts e-commerce. Additionally, based on qualitative interviews, we checked that the results are robust to restricting the sample to managers who have been at the plant five or more years. We additionally explored whether clustering standard errors by firm affected our results; no differences were found, but the change caused perturbations in the sample due to the need to restrict on plants that did not change ownership over our sample period (and thus are omitted for disclosure -avoidance reasons). Are results also robust to excluding plants where the CEO or a long-tenure respondent filled out the questionnaire.

## 6. CONCLUSION

Theory and case evidence suggest that productivity gains may come from leveraging IT investments to collect and bring data to bear in managerial decision making, tracking performance within the firm, and communicating about the state of the production process. This paper provides statistical evidence that putting data “into action” in these ways is associated with significantly higher productivity – at least in certain firms and in the short run. Importantly, it is not simply the adoption of IT, but the adoption of a specific set of data-driven practices that are correlated with the greatest increases in productivity. This relationship holds across a variety of industries in the U.S. manufacturing sector and stands up to the inclusion of difficult-to-observe controls for the level of IT investment and other management practices at the organization. Our findings provide the first evidence that this correlation is robust to unobserved, time-invariant, plant heterogeneity. Finally, the timing of the adoption decision and performance effects is consistent with a causal relationship between DDD and performance. Moreover, for the period we observe, more DDD is always associated with better performance.

That said, we find ample evidence that firms are able to learn about the set of complementary investments that make DDD productive in their operating environments and that the distinction between adopters and non-adopters may diminish among capable firms that are able to match their decision-making regimes to their particular needs. This sorting mechanism is likely behind the pattern of results for the multi-unit plants, which show no significant differential in productivity due to DDD, and the convergence between early and late adopters (relative to never-adopters) over our sample period. Unfortunately, our study cannot shed light on what the next frontier of practice may be and whether early practitioners of DDD may be better positioned to exploit the next wave of technological progress. This would be a useful direction for future work to explore.

This notion of complementarities is essential, in that it provides a mechanism by which early adopters of technologies can get and stay ahead of competitors who cannot realize performance benefits in time, leading to increasing and persistent performance differences - the puzzle which motivated our study to begin with. We find evidence that prior IT investment, the presence of skilled workers, and

important attributes of firm culture may be important for DDD to be effective. Most interestingly, we find three-way complementarities between DDD, joint decision making between managers and front-line workers, and being in capital-intensive “continuous flow” industries.

Certain limitations of this study are worth noting. For instance, we do not observe directly what types of decision making are being influenced by data. Based on the content of the survey directly preceding the DDD questions, one might infer that respondents have a range of types of decision making in mind: HR decisions such as hiring and pay raises, product strategy decisions such as new product introductions and pricing, marketing decisions such as advertising spend, and financial decisions pertaining to the purchase of new capital assets. Future data collection is aimed at linking the type and amount of data collected with specific decisions to which it may be applied.

Also, we do not know with certainty what respondents have in mind when asked about “data.” We infer from the priming earlier in the survey -- as well as from a number of qualitative interviews -- that respondents at least think about Key Performance Indicators (KPIs) regarding cost, waste, quality, and so on. However, there may be significant differences between firms and industries in whether respondents are thinking about automated data collection and analysis based on significant on-site IT investments, or whether they have in mind some of the relatively “low tech” data management techniques that are nevertheless effective in many lean manufacturing settings. In fact, our dynamic complementarities tests point to different styles of DDD for early and late adopters -- especially among multi-unit plants.

In addition, we do not have objective information about the quality of data or data-driven decision making in our sample. There is an important range of data collection that takes place above the “10 KPI’s or more” threshold, and there is certainly variation in how accurate or fine-grained existing data collection activities may be. To the extent that the frontier of practice is beyond what we are tracking, there may be even bigger short-term benefits to plants that are at the leading edge of this activity. Future data collection efforts are aimed at better identifying variation along these dimensions.

Finally, the potential for bias due to self-selection into DDD persists. The timing of adoption and performance we observe is consistent with a causal explanation. However, there is no sense in which

firms are exogenously “treated with DDD” – our findings can best be taken as informative about the effect of treatment on the treated. They provide evidence about the potential importance of this practice and the usefulness of distinguishing investments in data-related practices from investments in tangible IT capital and other, more structured approaches to management. Our findings suggest that the benefits may be commensurate, while we speculate that the costs may differ significantly in nature and magnitude. In particular, the former may be more in the form of managerial time and attention, and therefore invisible on the balance sheet and difficult to observe and copy – potentially conferring some measure of competitive advantage.

We hope to spur further research into the relationship between data-driven decision making and firm performance in manufacturing and other sectors of the economy – particularly retail and services. Given the large increases we are certain to see in both IT capabilities and the availability of digital data for use in decision making, the effects we identify, and the role of complementary changes in organizations, may grow even more economically important in the coming years.



## REFERENCES

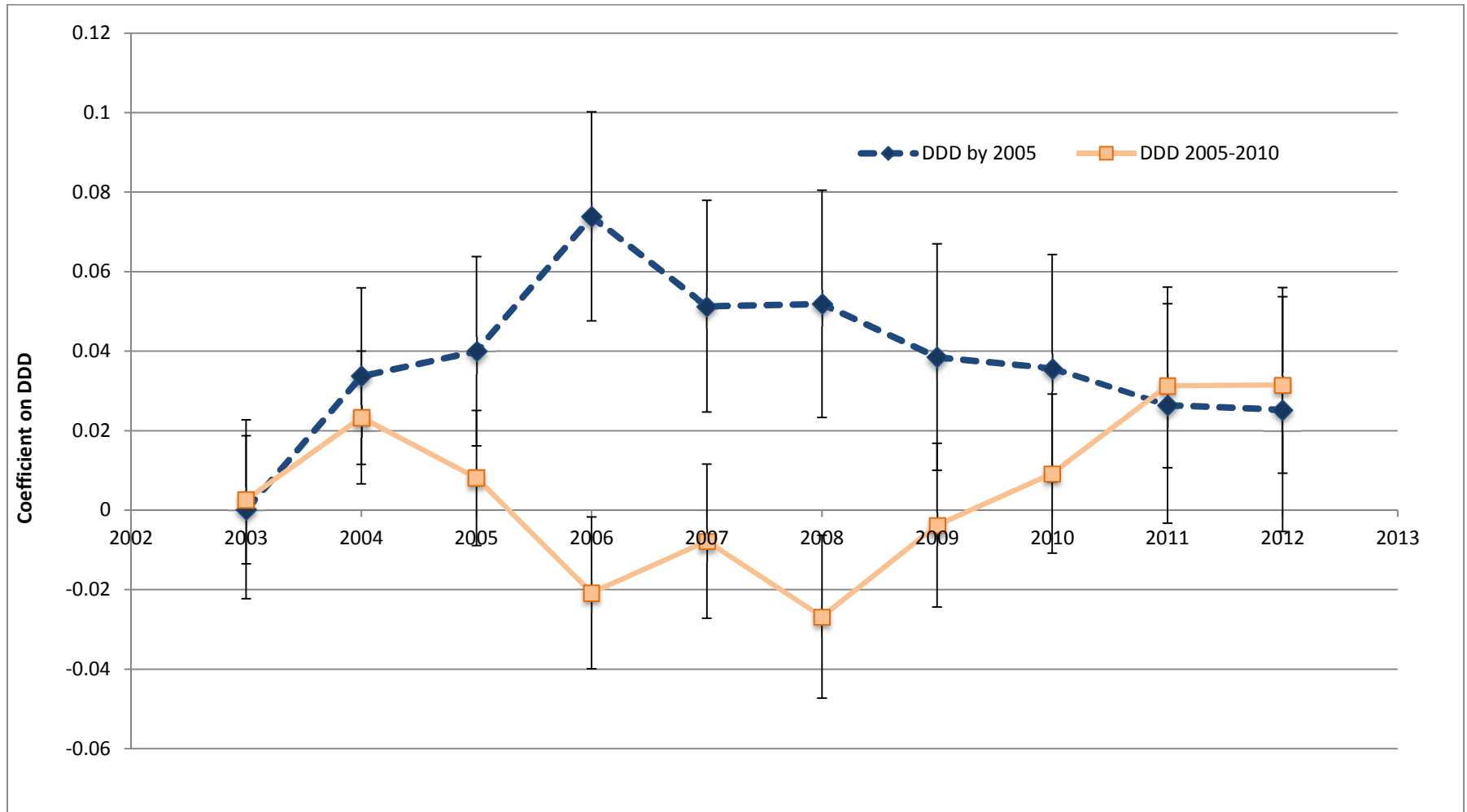
- Aral, Sinan, Erik Brynjolfsson, and Lynn Wu. 2012. "Three-way complementarities: Performance pay, human resource analytics, and information technology." *Management Science* 58, no. 5: 913-931.
- Aral, Sinan, and Peter Weill. 2007. "IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation." *Organization Science* 18, no. 5: 763-780.
- Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2014. "Vertical integration and input flows." *The American Economic Review* 104, no. 4: 1120-1148.
- Athey, Susan, and Scott Stern. 1998. "An Empirical Framework for Testing Theories about Complementarity in Organizational Design." No. 6600. *National Bureau of Economic Research*
- Baily, Martin Neil, Charles Hulten, David Campbell, Timothy Bresnahan, and Richard E. Caves. 1992. "Productivity dynamics in manufacturing plants." *Brookings papers on economic activity. Microeconomics*: 187-267.
- Baker, George P., and Thomas N. Hubbard. 2004. "Contractibility and Asset Ownership: On-Board Computers and Governance in US Trucking." *The Quarterly Journal of Economics* 119, no. 4: 1443-1479.
- Bartel, Ann P., Casey Ichniowski, and Kathryn L. Shaw. 2007. "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Past Trends and Projections in Wages, Work, and Occupations 63 Product Innovation, Process Improvement and Worker Skills." *Quarterly Journal of Economics* 122, no. 4: 1721-1758.
- Blader, Steven, Claudine Gartenberg, Rebecca Henderson, and Andrea Prat. 2015. "The Real Effects of Relational Contracts." *American Economic Review* 105, no. 5: 452-56.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2011. "Does management matter? Evidence from India." No. 16658. *National Bureau of Economic Research*.
- Bloom, N., E. Brynjolfsson, L. Foster, R. S. Jarmin, I. Saporta-Eksten and J. Van Reenen. 2013. "Management in America." *Center for Economic Studies Discussion Papers #CES 13-01*, U.S. Census Bureau.
- Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. 2014. "The distinct effects of information technology and communication technology on firm organization." *Management Science* 60, no. 12: 2859-2885.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "Americans Do IT Better: US Multinationals and the Productivity Miracle." *The American Economic Review* 102, no. 1: 167.
- Bohn, Roger E. 2005. "From Art to Science in Manufacturing: The Evolution of Technological Knowledge." *Foundations and Trends in Technology, Information and Operations Management* 1, no. 2: 1-82.
- Bresnahan, Timothy, Shane Greenstein, David Brownstone, and Ken Flamm. 1996. "Technical progress and co-invention in computing and in the uses of computers." *Brookings Papers on Economic*

- Activity. Microeconomics*: 1-83.
- Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics* 117, no. 1:: 339-376.
- Brynjolfsson, Erik. 1993 "The Productivity Paradox of Information Technology: Review and Assessment" *Communciations of the ACM*, December.
- Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton & Company.
- Brynjolfsson, Erik, and Lorin M. Hitt. 1995. "Information technology as a factor of production: The role of differences among firms." *Economics of Innovation and New technology* 3, no. 3-4: 183-200.
- Brynjolfsson, Erik, and Lorin M. Hitt. 2000 "Beyond computation: Information technology, organizational transformation and business performance." *The Journal of Economic Perspectives*: 23-48.
- Brynjolfsson, Erik, and Lorin M. Hitt. 2003 "Computing productivity: Firm-level evidence." *Review of economics and statistics* 85, no. 4: 793-808.
- Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2011. "Strength in numbers: How does data-driven decisionmaking affect firm performance?." *Available at SSRN 1819486*.
- Brynjolfsson, Erik, Lorin M. Hitt, and Shinkyu Yang. 2002. "Intangible assets: Computers and organizational capital." *Brookings papers on economic activity* 2002, no. 1: 137-198.
- Brynjolfsson, Erik, and Paul Milgrom. 2013. "Complementarity in organizations." *The Handbook of Organizational Economics* edited by Robert Gibbons and John Roberts, 11-55. Princeton: Princeton University Press.
- Brynjolfsson, Erik., Amy Renshaw and Marshal Van Alstyne, 1997 "The Matrix of Change", *Sloan Management Review*, Winter.
- Chew, W. Bruce, Timothy F. Bresnahan, and Kim B. Clark. 1990. "Measurement, Coordination, and Learning in a Multiplant Network." In *Measures for Manufacturing Excellence*, edited by Robert S. Kaplan, 129-62. Boston: Harvard Business School Press.
- Colombo, Massimo G., and Marco Delmastro. 2004. "Delegation of authority in business organizations: an empirical test." *Journal of Industrial Economics*: 53-80.
- David, P. A. 1969. *A Contribution to the Theory of Diffusion*. Research Center in Economic Growth, Stanford University
- David, H., and David Dorn. 2013. "The growth of low-skill service jobs and the polarization of the US labor market." *The American Economic Review* 103, no. 5: 1553-1597.
- Dessein, W. and T. Santos (2006). "Adaptive Organizations." *Journal of Political Economy* **114**(5): 956-995.

- Devaraj, Sarv, and Rajiv Kohli. 2003. "Performance impacts of information technology: Is actual usage the missing link?." *Management science* 49, no. 3: 273-289.
- Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske. 2004. "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment." *Journal of Labor Economics* 22, no. 2: 397-429.
- Forman, Chris, and Kristina Steffenson McElheran. 2012. "Information technology and boundary of the firm: evidence from plant-level data." *Working Paper, Harvard Business School*.
- Galbraith, Jay R. 1974. "Organization design: An information processing view." *Interfaces* 4, no. 3: 28-36.
- Gartner. 2014. "Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014." <http://www.gartner.com/document/2841519>.
- Gibbons, Robert, and Rebecca Henderson. 2012. "Relational contracts and organizational capabilities." *Organization Science* 23, no. 5: 1350-1364.
- Hayes, Robert H., and Steven G. Wheelwright. 1979. "Dynamics of Process-Product Life-Cycles." *Harvard business review* 57, no. 2: 127-136.
- Hoffman, M., L. B. Kahn, and D. Li. 2015. "Discretion in Hiring." No. 21709. *National Bureau of Economic Research*
- Hong, Bryan, Lorenz Kueng, and Mu-Jeung Yang. 2015. "Estimating management practice complementarity between decentralization and performance pay." No. 20845. *National Bureau of Economic Research*.
- Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *American Economic Review* 87, no. 3: 291-313.
- Jensen, J. Bradford, and Robert H. McGuckin. 1997. "Firm performance and evolution: empirical regularities in the US microdata." *Industrial and Corporate Change* 6, no. 1: 25-47.
- Jensen, Michael C., and William H. Meckling. 1976. "Theory of the firm: Managerial behavior, agency costs and ownership structure." *Journal of financial economics* 3, no. 4: 305-360.
- Karshenas, Massoud and Paul L. Stoneman. 1993. "Rank, Stock, Order, and Epidemic Effects in the Diffusion of New Process Technologies: An Empirical Model." *RAND Journal of Economics* 24 no. 4: 503-528.
- MacDuffie, John Paul. 1995. "Human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry." *Industrial & labor relations review* 48, no. 2: 197-221.
- McElheran, Kristina. 2014. "Delegation in Multi-Establishment Firms: Evidence from IT Purchasing." *Journal of Economics & Management Strategy* 23, no. 2: 225-258.
- McElheran, Kristina. 2015. "Do Market Leaders Lead in Business Process Innovation? The Case(s) of E-Business Adoption." *Management Science* 61, no. 6: 1197-1216

- Milgrom, Paul, and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* 80, no. 3: 511-28.
- Mithas, Sunil, Ali R. Tafti, Indranil Bardhan, and Jie Mein Goh. 2012. "Information technology and firm profitability: mechanisms and empirical evidence." *MIS Quarterly* 36, no. 1: 205-224.
- Novak, Sharon, and Scott Stern. 2009. "Complementarity among vertical integration decisions: Evidence from automobile product development." *Management Science* 55, no. 2: 311-332.
- Piore, Michael J., and Charles F. Sabel. 1984. "*The second industrial divide: possibilities for prosperity.*" New York, *Basic Books*.
- Polanyi, Michael, 1969. *Knowing and Being*, Routledge
- Porter, Michael, and Nicolaj Siggelkow. 2008. "Contextuality within activity systems and sustainability of competitive advantage." *The Academy of Management Perspectives* 22, no. 2: 34-56.
- Porter, Theodore M. 1996. *Trust in numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton University Press.
- Rogers, Everett M. 2010. "Diffusion of innovations." *Simon and Schuster*.
- Roos, D., J. P. Womak and D. T. Jones. 1991. *The Machine that Changed the World: The Story of Lean Production*, Harper Perennial.
- Rivkin, Jan W. 2000. "Imitation of complex strategies." *Management science* 46, no. 6: 824-844.
- Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49, no. 2: 326-365.
- Tambe, Prasanna, and Lorin M. Hitt. 2012. "The productivity of information technology investments: New evidence from IT labor data." *Information Systems Research* 23, no. 3-part-1: 599-617.
- Yang, Mu-Jeung, Lorenz Kueng, and Bryan Hong. 2015. "Business Strategy and the Management of Firms." No. 20846. *National Bureau of Economic Research*.

Figure 1. Coefficients on Indicator of Data-Driven Decision Making by Year (2003 – 2013)



**Table 1. Components of Data-Driven Decision Making**

<b>Data-Related Management Practice</b>	<b>Adoption in 2005 Balanced sample</b>	<b>Adoption in 2010 Balanced sample</b>
Top 2 categories for “availability of data” (Q27)	0.45	0.73
Top 2 categories for “use of data” (Q28)	0.44	0.72
Top 2 categories for both availability and use of data (Q27 & Q28)	0.35	0.64
Top for availability and use of data, plus tracking 10 or more KPI’s (Q27, Q28, & Q2)	0.17	0.40
<b>“Data-Driven Decision Making”</b> or <b>DDD</b> indicator: as above plus use of both short-term and long-term targets (Q6)	0.11	0.30
<b>DDD index:</b> sum of the normalized responses to all four DDD-related questions (2, 6, 27, & 28), scaled to lie on the [0, 10] interval	6.1	7.3
N	~18,000	~18,000

**Table 2. Principal Factor Analysis of Data-Related Managerial Practices**

<b>Principal Factor Analysis of 2010 Balanced Sample (~18,000 obs.)</b>					
	<b>Eigenvalue</b>	<b>Proportion of Variance</b>			
Factor 1	2.28	0.570			
Factor 2	0.851	0.213			
Factor 3	0.648	0.162			
<b>Polychoric Correlation Matrix and Factor Loadings</b>					
	Top 2 categories for “availability” of data	Top 2 categories for “use” of data	Track 10 or more KPIs	Use of short-term and long-term targets	<b>Factor 1 Loadings</b>
Top 2 categories for “availability” of data	1				0.870
Top 2 categories of “use” of data	0.778	1			0.854
Track 10 or more KPIs	0.366	0.385	1		0.664
Use of short-term and long-term targets	0.291	0.328	0.343	1	0.595

**Note:** Calculated using the **polychoric** command in Stata13

**Table 3a. Adoption of Data-Driven Decision Making and Potential Complements by Year**

Description		Adoption in 2005 Balanced sample	Adoption in 2010 Balanced sample	2010 Std. Deviation
<b>Data-Driven Decision Making (DDD)</b>				
DDD	Indicator that plant is in top 2 categories for both availability and use of data (Q27 & Q28), uses 10 or more KPI's (Q2) and both short-term and long-term targets (Q6).	0.11	0.30	0.46
DDD Index	Sum of the normalized responses to questions 2, 6, 27 & 28, scaled to lie on the [0, 10] interval.	6.1	7.3	1.7
<b>Potential Complements</b>				
IT capital stock	Value of hardware and software stocks in \$millions at the plant, calculated using the perpetual inventory method and BEA deflators.	0.39	0.46	3.12
Percent employees with BA	Percentage of managerial and non-managerial employees with Bachelor's degrees (Q34 & 35)	7.1	9.6	0.24
Some Delegation	Indicator that the plant reports at least some involvement of production workers in prioritizing and allocating production tasks (Q26)	0.25	0.37	0.48
N		~18,000	~18,000	

**Table 3b. DDD adoption and change by year and multi-unit status**

	Adoption in 2005 Multi-Unit Plants	Adoption in 2010 Multi-Unit Plants	Adoption in 2005 Single-Unit Firms	Adoption in 2010 Single-Unit Firms
<b>DDD</b>	0.13	0.34	0.05	0.15
<b>Percent of plants that change DDD status between 2005 and 2010</b>		22		9.6
<b>DDD Index (ranges from 0 to 10)</b>	6.3	7.6	5.4	6.4
<b>Average percentage change in DDD Index (relative to 2005)</b>		17		15
N	~14,000	~14,000	~4,000	~4,000



**Table 4. Descriptive Statistics for Non-DDD Variables**

<b>Variable</b>	<b>Description</b>	<b>2010 Mean All Balanced Sample (S.D.)</b>	<b>2010 Mean Multi-Unit Plants (S.D.)</b>	<b>2010 Mean Single-Unit Plants (S.D.)</b>
<b>Value Added</b>	Value added at the plant (total value shipped minus total cost of goods sold) in \$millions	61.1 (300.8)	73.4 (337.4)	15.4 (29.3)
<b>Employment</b>	Total number of employees at the plant	227 (477)	257 (529)	113 (133)
<b>Sales</b>	Total value shipped by the plant in \$millions	144 (618)	174 (692)	33.8 (73.9)
<b>Capital stock</b>	Value of non-IT capital stock at the plant in \$millions (calculated using the perpetual inventory method and BEA capital deflators)	41.6 (160)	49.7 (179)	11.3 (31.1)
<b>Energy Costs</b>	Total cost of both fuel and electricity in \$millions consumed by the plant	2.73 (13.2)	3.31 (14.8)	0.57 (1.92)
<b>Multi-Unit Status</b>	=1 if the plant belongs to a multi-unit firm	0.79 (0.41)	N/A	N/A
<b>Computer Hardware Expenditures</b>	Expenditures on computer hardware and other equipment in \$millions	0.13 (1.07)	0.15 (1.19)	0.06 (0.34)
<b>Computer Software Expenditures</b>	Expenditures on purchases of prepackaged, custom coded or vendor customized software in \$millions	0.07 (0.95)	0.08 (1.06)	0.02 (0.15)
<b>Structured Management Z-Score</b>	Index created by summing up the normalized values from questions 1, 5, 8, 9, 11, 13 & 14 of the MOPS	0.64 (.17)	0.67 (.15)	0.53 (.20)
<b>Age</b>	Establishment age	26.1 (9.9)	26.3 (10.0)	25.4 (9.8)
<b>Number of Layers</b>	Number of layers of direct reports in the establishment from the factory floor to the plant manager, inclusive	3.4 (3.2)	3.5 (3.2)	2.9 (2.8)
<b>Number of managers</b>	Number of managers in the establishment	16.3 (37.6)	18.0 (41.6)	10.1 (13.5)
<b>CEO respondents</b>	Indicator of whether the CEO was the respondent for the MOPS	0.18 (0.38)	0.11 (0.31)	0.44 (0.50)
<b>Tenure of respondent</b>	Number of years the respondent worked in the establishment	13.1 (10.5)	12.7 (10.5)	14.6 (10.7)
<b>N</b>		~18,000	~14,000	~4,000

**Table 5. Adoption of Data-Driven Decision Making by Industry (3-Digit NAICS Code)**

<b>3-Digit NAICS Code</b>	<b>Industry</b>	<b>Mean DDD in 2010 Balanced Sample</b>	<b>Mean DDD change 2005 to 2010 Balanced Sample</b>
311	Food	0.39	0.23
312	Beverage and Tobacco Products	0.40	0.28
313	Textile Mills	0.32	0.18
314	Textile Product Mills	0.21	0.15
315/316	Apparel and Leather	0.13	0.06
321	Wood Products	0.22	0.12
322	Paper	0.37	0.24
323	Printing and Related Support Activities	0.19	0.14
324	Petroleum and Coal Products	0.32	0.13
325	Chemicals	0.37	0.21
326	Plastics and Rubber	0.31	0.22
327	Non-metallic Mineral Products	0.24	0.14
331	Primary Metals	0.33	0.22
332	Fabricated Metal Products	0.25	0.17
333	Machinery	0.24	0.17
334	Computers and Electronic Products	0.33	0.21
335	Electrical Equipment, Appliances, and Components	0.34	0.24
336	Transportation Equipment	0.41	0.25
337	Furniture	0.18	0.14
339	Miscellaneous	0.26	0.17
321, 322, 324, 325, 327, 331	“Continuous Flow” industries	0.34	0.19

\* Mean adoption estimated using ASM sampling weights.

**Table 6a. Adoption of Data-Driven Decision Making between 2005 and 2010:  
Marginal effects of probit regression with lagged (2005) plant and firm characteristics**

Dependent Variable:	Indicator of Adopting Data-Driven Decision Making between 2005 and 2010					
	(1)	(2)	(3)	(4)	(5)	(6)
<b>2005 Log IT capital stock</b>	0.014*** (0.002)	0.003* (0.002)	0.003* (0.002)	0.003 (0.002)	0.003* (0.002)	0.003* (0.002)
<b>2005 Structured management</b>	0.064*** (0.021)	0.010 (0.021)				
<b>2005 Log employment</b>		0.059*** (0.003)	0.046*** (0.004)	0.036*** (.005)	0.036*** (0.005)	0.032*** (0.005)
<b>Multi-unit status</b>			0.118*** (0.010)	0.109*** (0.010)	0.089*** (0.010)	0.082*** (0.010)
<b>2005 High capital stock (top 25% of industry)</b>			0.021** (0.009)	0.019** (0.009)	0.014* (0.009)	0.013 (0.009)
<b>2005 Percent workers with BA degrees</b>				0.094*** (0.030)	0.080*** (0.030)	0.054* (0.030)
<b>High number of managers†</b>				0.015* (0.008)	0.017** (0.008)	0.014* (0.008)
<b>High number of layers†</b>				0.066*** (0.011)	0.058*** (0.011)	0.054*** (0.011)
<b>Age</b>					-0.0006* (0.0003)	-0.0006* (0.0003)
<b>Longer tenure* for respondent</b>					-0.052*** (0.007)	-0.051*** (0.007)
<b>CEO respondent</b>					-0.068*** (0.010)	-0.067*** (0.010)
<b>2005 Joint decision making</b>						-0.002 (0.008)
<b>2005 number of learning sources</b>						0.016*** (0.002)
<b># Establishments</b>	~16,000	~16,000	~16,000	~16,000	~16,000	~16,000
<b>Fixed effects</b>	NAICS3	NAICS3	NAICS3	NAICS3	NAICS3	NAICS3

†Greater than median

**Note:** Weighted Maximum likelihood probit estimation of the likelihood that an establishment adopts DDD between 2005 and 2010. Sample is balanced sample conditional on not adopting DDD by 2005. Reporting marginal effects calculated at sample means of the covariates. Columns 1-3 report on the sample of plants with information for both 2005 and 2010 as well as not clearing the threshold for DDD as of 2005, column 4 restricts this sample to plants belonging to multi-unit firms, column 5 consists of the single-unit firms only. All columns include industry controls at the 3-digit NAICS level, as well as establishment age and indicators for missing covariate data. High capital stock is calculated relative to other plants in the same 4-digit NAICS industry. Robust standard errors are clustered at the firm level and reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**Table 6b. Early adoption of Data-Driven Decision Making (by 2005):  
Marginal effects of probit regression with lagged (2002) plant and firm characteristics**

<b>Dependent Variable:</b>	<b>Indicator of Data-Driven Decision Making (DDD) in 2005</b>			
	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
<b>Logged IT capital stock</b>	-0.002 (0.001)	-0.002 (0.001)	-0.002 (0.001)	0.003* (0.002)
<b>Logged employment</b>	0.039*** (0.003)	0.027*** (0.003)	0.027*** (0.004)	0.025*** (0.005)
<b>High capital stock</b> (top quartile for industry)		0.041*** (0.006)	0.045*** (0.007)	Suppressed
<b>Multi-unit status</b>		0.042*** (0.007)		
<b>Age</b>	-0.0004 (0.0003)	-0.0003 (0.0003)	-0.00004 (0.0003)	-0.0009** (0.0003)
<b># Establishments</b>	~18,000	~18,000	~14,000	~4000
<b>Sample</b>	<b>All 2005</b>	<b>All 2005</b>	<b>Multi-unit 2005</b>	<b>Single-unit 2005</b>
<b>Fixed effects</b>	NAICS3	NAICS3	NAICS3	NAICS3

**Note:** Maximum likelihood probit estimation of whether an establishment reports relatively intensive use of data-driven decision making (DDD) in 2005. Reporting average marginal effects calculated at sample means of the covariates. Columns 1-2 report on all plants for which 2005 data is available. Column 3 contains only plants belonging to multi-unit firms. Column 4 contains only single-unit plants; the control for high capital stock (top quartile for industry) is omitted due to small cell concerns, although the results are robust to its inclusion. All columns include industry controls at the 3-digit NAICS level and indicators for missing covariates. IT capital stock, employment, and conventional capital stock values are from 2002. Column 3 includes a control for headquarters status. Robust standard errors are clustered at the firm level and reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**Table 7. Fixed-Effects Estimation of Data-Driven Decision Making and Firm Performance**

Dependent Variable:	Log Value Added						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Control for IT	Control for Mgmt. & Educ.	SU Plants	MU Plants	All	SU Plants	MU Plants
<b>DDD</b>	<b>0.052***</b> (0.015)	<b>0.030*</b> (0.016)	<b>0.144***</b> (0.038)	<b>0.011</b> (0.017)			
<b>DDD index</b>					0.011*** (0.005)	0.030*** (0.009)	0.007 (0.006)
<b>Log IT capital stock</b>	0.006* (0.003)	0.006* (0.003)	0.008 (0.007)	0.006 (0.004)	0.006* (0.003)	0.008 (0.007)	0.006 (0.004)
<b>Structured mgmt. index</b>		0.262*** (0.053)	0.389*** (0.139)	0.213*** (0.057)	0.205*** (0.062)	0.262* (0.145)	0.173** (0.068)
<b>Percent Employees with BA degrees</b>		-0.172 (0.116)	0.205 (0.247)	-0.234* (0.129)	-0.187 (0.116)	0.144 (0.252)	-0.244* (0.129)
<b>Year</b>	0.027*** (0.008)	0.010 (0.009)	-0.038* (0.020)	0.025** (0.011)	0.008 (0.010)	-0.043* (0.020)	0.024** (0.011)
<b>Capital, Labor, and Energy</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b># Establishments</b>	~18,000	~18,000	~4,000	~14,000	~18,000	~4,000	~14,000
<b>Sample</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>

**Note:** Two-period linear regression with establishment-fixed effects. In all columns the dependent variable is logged nominal value added. The structured management index is the unweighted average of the score for each of the first 16 questions, omitting questions 1,5,8,9,11,13 & 14, where each question is first normalized to be on a 0-1 scale. Unreported controls in all columns include: firing practices (questions 15 & 16), whether the establishment belongs to a multi-unit firm, logged capital stock, logged employment, winsorized logged energy expenditures, and whether or not information on employees' education is missing. The sample in columns 1, 2 and 5 is all MOPS observations with complete answers to the data and HQ questions, at least 5 non-missing responses to the non-data management questions for both 2005 and 2010, a successful match to ASM in both years, which were also included in ASM tabulations for both 2010 and 2005, have positive value added, positive employment and positive imputed capital in the ASM for both 2005 and 2010. Columns 3, 4, 6, and 7 split this sample by multi-unit status. Robust standard errors are reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**Table 8. Variation in Performance Effects by Plant Characteristics**

Dependent Variable:	Log Value Added				
	(1)	(2)	(3)	(4)	(5)
<b>DDD</b>	0.073*** (0.020)	0.110*** (0.048)	0.122*** (0.032)	0.053*** (0.018)	0.334** (0.061)
<b>DDD x Large Size in 2005</b>	-0.073*** (0.029)				-0.034 (0.034)
<b>DDD x &gt;5 years in 2005</b>		-0.090* (0.049)			-0.130** (0.053)
<b>DDD x Multi-unit</b>			-0.103*** (0.034)		-0.167*** (0.037)
<b>DDD x High Capital Stock</b> (75% tile for industry)				-0.063** (0.029)	-0.015 (0.036)
<b># Establishments</b>	~18,000	~18,000	~18,000	~18,000	~18,000
<b>Sample</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>

**Note:** Two-period linear regression with establishment-fixed effects. In all columns the dependent variable is logged nominal value added. Large Size is an indicator of whether the plant had more than 100 employees in 2005. “>5 years” is an indicator for whether the establishment was more than five years old in 2005. Multi-unit is an indicator for whether the establishment comes from a parent firm with multiple establishments. High Capital Stock is an indicator for whether the capital stock of an establishment is above the 75 percentile within its 4-digit NAICS industry. Unreported controls include the structured management index, firing practices (questions 15 & 16), whether the establishment belongs to a multi-unit firm, logged capital stock, logged employment, winsorized logged energy expenditures, and whether or not information on employees’ education is missing. The sample in columns 1, 2 and 5 is all MOPS observations with complete answers to the data and HQ questions, at least 5 non-missing responses to the non-data management questions for both 2005 and 2010, a successful match to ASM in both years, which were also included in ASM tabulations for both 2010 and 2005, have positive value added, positive employment and positive imputed capital in the ASM for both 2005 and 2010. Robust standard errors are reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**Table 9. Timing of DDD Adoption is Inconsistent with Reverse Causality (both late and early adoption): Marginal effects from probit regression of DDD adoption in 2005 and 2002**

Dependent Variable:	DDD adoption	
	(1) DDD in 2005	(2) Adopted DDD in 2005-2010 period
Value-added growth 2002 - 2005	-0.00004 (0.00005)	
Value-added growth 2005 - 2010		0.00001 (0.00002)
IT capital stock 2002	-0.001 (0.001)	
IT capital stock in 2005		0.003* (0.002)
Log Employment in 2002	0.035*** (0.003)	
Log Employment in 2005		0.052*** (0.004)
Structured management index		-0.032 (0.021)
% Bachelor's Degree		0.110*** (0.030)
# Establishments	~18,000	~16,000
Sample	Balanced	Balanced & not DDD in 2005

**Note:** Maximum likelihood probit estimation of DDD adoption. Reporting marginal effects calculated at mean values of the covariates. Unreported controls include whether the establishment belongs to a multi-unit firm and its age. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**Table 10. Fixed Effect Estimation of Interaction with Potential Complements**

Dependent Variable:	Log Value Added				
	(1)	(2)	(4)	(5)	(6)
<b>DDD</b>	0.005 (0.019)	-0.004 (0.024)	0.046*** (0.017)	0.046*** (0.020)	0.074*** (0.022)
<b>DDD x High IT in 2005</b>	0.075*** (0.027)				
<b>DDD x High %BA Degree in 2005</b>		0.054** (0.028)			
<b>DDD x Joint Decision Making (fixed) in 2005</b>			-0.061* (0.032)		
<b>Joint Decision Making</b>				0.060*** (0.020)	0.053** (0.021)
<b>DDD x Joint DM</b>				-0.042 (0.028)	-0.081*** (0.033)
<b>DDD X CTS Flow</b>					-0.097** (0.042)
<b>Joint DM X CTS Flow</b>					0.027 (0.043)
<b>DDD X Joint DM X CTS Flow</b>					0.134*** (0.062)
<b>Linear Combination of DDD, Joint DM, and CTS Flow</b>					0.082* (.046)
<b>Establishments</b>	~18,000	~18,000	~18,000	~18,000	~18,000
<b>Sample</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>

**Note:** Two-period linear regression with establishment-fixed effects. In all columns the dependent variable is logged nominal value added. High IT in 2005 is an indicator of whether the plant was in the top quartile of the IT capital distribution in 2005. High Degree in 2005 is an indicator for whether the establishment had greater than the median percentage of employees with Bachelor's degrees in 2005. Joint Decision Making is an indicator of whether the frontline workers involvement index is higher than the sample mean. Note that these values are fixed in 2005 in columns 1-3 and therefore the direct effect is differenced out. CTS Flow is an indicator of the plant being in the following industries: Wood, Paper, Petroleum, Chemicals, Non-Metallic Minerals, and Primary Metal industries (3-digit NAICS definitions). Unreported controls include the structured management index, firing practices (questions 15 & 16), whether the establishment belongs to a multi-unit firm, logged capital stock, logged employment, winsorized logged energy expenditures, and whether or not information on employees' education is missing. The sample in columns 1, 2 and 5 is all MOPS observations with complete answers to the data and HQ questions, at least 5 non-missing responses to the non-data management questions for both 2005 and 2010, a successful match to ASM in both years, which were also included in ASM tabulations for both 2010 and 2005, have positive value added, positive employment and positive imputed capital in the ASM for both 2005 and 2010. Robust standard errors are reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.



**Table 11. Tests of dynamic complementarity with IT Adoption**

Dynamic complementarity test for DDD and IT				
2010 Adoption				
2005 Adoption	(1) Neither Base group	(2) IT	(3) DDD	(4) Both
(1)Neither		<b>0.053*</b> <b>(0.028)</b>	<b>0.015</b> <b>(0.021)</b>	0.013 (0.044)
(2)IT		0.058*** (0.020)	0.090** (0.038)	<b>0.101***</b> <b>(0.028)</b>
(3)DDD			0.052** (0.027)	-0.050 (0.062)
(4)Both				0.086** (0.039)
Performance Test 1: compare adding DDD only from 2005 - 2010 (row 1 column3) to adding DDD by 2010 to existing high IT in 2005 (row 2 column 4) : PASS (adding DDD to high IT is better than adding it by itself) Coefficient = .086*** SE = .032				
Performance Test 2: compare adding high IT only from 2005 - 2010 (row 1 column 2) to adding high IT by 2010 to existing DDD in 2005 (row 3 column 4) is better (i.e., adding IT to existing DDD is better than adding it by itself) : FAIL (possibly due to small sample problem) Coefficient = -.103 SE= .067				
NOTE: Few establishments fall back from 2005 to 2010 and are not reported for disclosure avoidance reasons.				

**Correlation Test**

Year = 2005		High IT	
		0	1
DDD	0	~12,500	~4,000
	1	~1,000	~500
Pearson Chi2(1) = 112.85 Pr < 0.001; N = ~18,000			

Year = 2010		High IT	
		0	1
DDD	0	~10,000	~2,500
	1	~3,500	~2,000
Pearson Chi2(1) = 319.11 Pr < 0.001; N = ~18,000			

**Table 12. Other Outcome measures**

Dependent Variable:	Log TFP	Log TFP MU only	Log TFP SU only	Log Number of Production Workers	Log Ecommerce
	(1)	(2)	(3)	(4)	(5)
<b>DDD</b>	<b>0.011</b> <b>(0.009)</b>	<b>0.003</b> <b>(0.009)</b>	<b>0.059**</b> <b>(0.023)</b>	<b>0.018*</b> <b>(0.011)</b>	<b>0.462*</b> <b>(0.106)</b>
<b>IT capital stock (logged)</b>	0.003* (0.002)	0.003 (0.002)	0.005 (0.004)	0.010*** (0.002)	0.012 (0.022)
<b>Structured management index</b>	0.126*** (0.032)	0.094*** (0.034)	0.221** (0.086)	0.024 (0.033)	1.610*** (0.328)
<b>% Employees with Bachelor's degrees</b>	-0.126* (0.070)	-0.135* (0.079)	-0.057 (0.145)	-0.219** (0.085)	1.442** (0.663)
<b>Year Trend</b>	0.105*** (0.005)	0.116*** (0.005)	0.073*** (0.009)	-0.160*** (0.006)	4.503*** (0.055)
<b># Establishments</b>	~18,000	~14,000	~4,000	~18,000	~18,000
<b>Sample</b>	<b>Balanced</b>	<b>Balanced MU</b>	<b>Balanced SU</b>	<b>Balanced</b>	<b>Balanced</b>

**Note:** Two-period linear regression with establishment-fixed effects. Unreported controls include firing practices (questions 15 & 16), logged total value of shipment (column 3), logged capital stock (columns 2, 3, and 4), logged employment (columns 2 and 4), winsorized logged energy expenditures (columns 2, 3, and 4), and whether or not information on employees' education is missing (all columns). The sample in columns 1, 4 and 5 is all MOPS observations with complete answers to the data and HQ questions, at least 5 non-missing responses to the non-data management questions for both 2005 and 2010, a successful match to ASM in both years, which were also included in ASM tabulations for both 2010 and 2005, have positive value added, positive employment and positive imputed capital in the ASM for both 2005 and 2010. Columns 2 and 3 split this sample by multi-unit status. Robust standard errors are reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**APPENDIX:**

**Table A1: Adoption Rates of DDD Components by Different Subsamples**

<b>Data-Related Management Practice</b>	<b>Adoption in 2005 Balanced sample</b>	<b>Adoption in 2010 Balanced sample</b>	<b>Adoption in Entire 2010 Cross-section sample</b>	<b>Adoption in 2010 Weighted by ASM sampling weights</b>
Top 2 categories for “availability of data” (Q27)	0.45	0.73	0.67	0.59
Top 2 categories for “use of data” (Q28)	0.44	0.72	0.65	0.55
Top 2 categories for both availability and use of data (Q27 & Q28)	0.35	0.64	0.56	0.47
Top for availability and use of data, plus tracking 10 or more KPI’s (Q27, Q28, & Q2)	0.17	0.40	0.32	0.22
<b>“Data-Driven Decision Making”</b> or <b>DDD</b> indicator: as above plus use of both short-term and long-term targets (Q6)	0.11	0.30	0.23	0.16
<b>DDD index:</b> sum of the normalized responses to all four DDD-related questions (2, 6, 27, & 28), scaled to lie on the [0, 10] interval	6.1	7.3	6.9	6.3
N	~18,000	~18,000	~34,000	~34,000

**Table A2. Robustness to Definitions of DDD**

<b>Coefficients for DDD from Table 7, Column 2 specification</b>			
	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
	<b>Full Sample</b>	<b>Single-Unit Firms</b>	<b>MU Plants</b>
<b>Top 2 categories for Q27</b>	0.039*** (0.014)	0.035 (0.027)	0.038** (0.016)
<b>Top 2 categories for Q28</b>	0.007 (0.013)	0.049* (0.027)	-0.003 (0.015)
<b>Top KPIs</b>	0.030** (0.014)	0.069** (0.032)	0.023 (0.016)
<b>Q27 &amp; Q28</b>	0.030** (0.013)	0.069*** (0.025)	0.021 (0.015)
<b>Q27, Q28, &amp; KPI's</b>	0.040*** (0.014)	0.104*** (0.032)	0.029*** (0.016)
<b>Q27, Q28, ST and LT targets</b>	<i>Pending</i>	<i>Pending</i>	<i>Pending</i>
<b>Q27, Q28, KPI &amp; high frequency of review</b>	0.031* (0.017)	0.074** (0.037)	0.027 (0.018)
<b>Q27, Q28, KPI, high frequency of review, and ST &amp; LT targets</b>	0.017 (0.017)	0.103*** (0.040)	0.009 (0.019)
<b># Establishments</b>	~18,000	~4,000	~14,000
<b>Sample</b>	<b>Balanced</b>	<b>Balanced</b>	<b>Balanced</b>

**Note:** Two-period linear regression using establishment-fixed effects. Replicating the specification in column 3 of Table 5, but with differing definitions of DDD. Columns 1 and 2 split the sample by multi-unit status. Column 3 reports the full sample (equivalent to column 3 of Table 5). Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.

**Table A3. Comparison Sample Descriptive Statistics**

<b>Variable</b>	<b>Description</b>	<b>2010 Balanced Sample Mean (estimated)</b>	<b>Entire 2010 Sample Mean (estimated)</b>	<b>2010 ASM Mean (estimated)</b>
<b>Log Value Added</b>	Log of value added (in \$K) at the plant, which is total value shipped minus total cost of goods sold	9.23	8.25	7.89
<b>Log Sales (TVS)</b>	Log of total value shipped (in \$K) by the plant	10.02	8.93	8.50
<b>Markup</b>	Operating profit per sales at the plant, calculated as value added minus wages and salaries, divided by total value of shipments	0.31	0.31	0.13
<b>Log Exports</b>	Log of total value of goods exported (in \$K) by the plant	3.65	2.41	1.80
<b>Log Employment</b>	Log of total number of employees at the plant	4.34	3.64	3.39
<b>Log K stock</b>	Log of capital stock (in \$K) at the plant, calculated using the perpetual inventory method and BEA capital deflators	8.91	8.26	N/A
<b>Log Energy Costs</b>	Log of total cost of both fuel and electricity consumed (in \$K) by the plant	5.75	4.61	4.28
<b>Multi-Unit Status</b>	=1 if the plant belongs to a multi-unit firm	0.71	0.52	0.52
<b>HQ Status</b>	=1 if the plant is a headquarters; equal to 1 for all single-unit firms	0.49	0.63	N/A
<b>Exporter Status</b>	=1 if the value of exports is >0	0.46	0.34	0.25
<b>E-Com Dummy</b>	=1 if the value of e-commerce at the plant is >0	0.61	0.54	0.44
<b>Age</b>	Derived from the year first observed in the LBD. Truncated at 44 due to LBD starting in 1976	24.9	21.3	N/A
<b>Log IT Capital Stock</b>	Log of value of hardware and software stocks (in \$K) at the plant, calculated using the perpetual inventory method and using BEA deflators.	3.83	2.89	N/A
<b>Structured Management Z-Score</b>	Index created by summing up the normalized values from questions 1, 5, 8, 9, 11, 13 & 14 of the MOPS	0.61	0.54	N/A
<b>% Employees with BA</b>	% of managers and non-managers with bachelor's degrees at the plant	0.08	0.07	N/A
<b>Number of sites</b>	Number of establishments belonging to the parent firm	107.22	60.95	N/A
<b>Number of products</b>	Number of products (at the 7-digit NAICS code level) produced at the plant.	3.55	3.47	3.15

**Table A4. Pairwise Correlations**

	Log VA	DDD	IT Stock	Mgmt	% BA Degree	Exp	Log TE	Log K stock	Energy	MU	# Sites	# Prod	Age
<b>Log Value-Added</b>	1												
<b>DDD</b>	0.294	1											
<b>IT capital stock</b>	0.561	0.183	1										
<b>Structured Management</b>	0.447	0.327	0.276	1									
<b>% BA Degree</b>	0.135	0.078	0.115	.119	1								
<b>Exporter</b>	0.344	0.118	0.236	0.205	.108	1							
<b>Log Total Employment</b>	0.821	0.266	0.581	0.407	0.094	.324	1						
<b>Log K stock</b>	0.725	0.272	0.497	0.393	0.110	0.261	0.675	1					
<b>Energy</b>	0.763	0.292	0.431	0.426	0.076	0.269	0.715	0.735	1				
<b>Multi-Unit Status</b>	0.336	0.211	0.141	0.357	0.026	0.103	0.253	0.321	0.386	1			
<b># Sites</b>	0.134	0.106	0.060	0.131	0.023	-0.011	0.094	0.125	0.138	0.195	1		
<b># Products produced</b>	0.116	0.026	0.095	0.020	0.021	0.103	0.120	0.075	0.072	0.030	0.033	1	
<b>Establishment Age</b>	0.297	0.082	0.191	0.099	0.011	0.163	0.333	0.262	0.300	0.039	0.045	0.109	1

**Note:** All correlations are significant at least the 5% level except the correlation between % with degree and establishment age

**Table A5. Conditional correlations between data-driven decision making (DDD) and firm performance**

Dependent Variable:	Logged Value Added					
	(1) DDD	(2) Add IT Stock	(3) IT Categories	(4) Mgmt. & Education	(5) Noise Controls	(6) 2010 Balanced Sample
<b>DDD</b>	<b>0.120***</b> (0.015)	<b>0.112***</b> (0.015)	<b>0.112***</b> (0.015 )	<b>0.086***</b> (0.015)	<b>0.082***</b> (0.015)	<b>0.067***</b> (0.017)
<b>IT capital stock (logged)</b>		0.048*** (0.004)				
<b>High IT</b>			0.222*** (0.018)	0.209*** (0.018)	0.195*** (0.018)	0.191*** (0.024)
<b>Med IT</b>			0.133*** (0.013)	0.127*** (0.013)	0.118*** (0.013)	0.102*** (0.020)
<b>Structured management index</b>				0.239*** (0.042)	0.237*** (0.043)	0.284*** (0.062)
<b>% Employees with Bachelor's degrees</b>				0.309*** (0.062 )	0.297*** (0.063)	0.408*** (0.078)
<b>Capital, Labor, and Energy inputs (logged)</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b># Establishments</b>	~34,000	~34,000	~34,000	~34,000	~34,000	~18,000
<b>Sample</b>	<b>All 2010</b>	<b>All 2010</b>	<b>All 2010</b>	<b>All 2010</b>	<b>All 2010</b>	<b>Balanced 2010</b>
<b>Fixed Effects</b>	NAICS6	NAICS 6	NAICS 6	NAICS 6	NAICS 6	NAICS6

**Note:** Weighted OLS regressions using ASM sampling weights. The dependent variable is logged nominal value added at the plant. Unreported controls in all columns include: firing practices (questions 15 & 16), whether the establishment belongs to a multi-unit firm, logged capital stock, logged employment, winsorized logged energy expenditures, and whether or not information on employees' education is missing. Noise controls (column 5 onwards) include: (1) measures for the distance between ASM and MOPS reported employment for 2005 and 2010; (2) online filing indicator; (3) date of filing and date; (4) day of week; (5) tenure of the respondent; (6) seniority of the respondent. Industry controls are included at the 6-digit NAICS level. The sample in columns 1-5 is all MOPS observations with complete answers to the data and HQ questions, at least 5 non-missing responses to the non-data management questions for 2010, a successful match to the ASM, which were also included in ASM tabulations for 2010, have positive value added, positive employment and positive imputed capital in the ASM for 2010. Column 6 restricts to the establishments that fit these criteria for 2005 as well. Robust standard errors are reported in parentheses. Statistical significance is denoted as follows: \* 10%, \*\* 5%, \*\*\* 1%.