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

PRELIMINARY EXTENDED ABSTRACT– DO NOT CITE OR DISTRIBUTE

Erik Brynjolfsson
Sloan School of Management
MIT
100 Main Street, E62-414
Cambridge, MA 02142
erikb@mit.edu

Kristina McElheran
University of Toronto
105 St. George Street
#7033
Toronto, ON M5S 3E6
kristina.mcelheran@rotman.utoronto.ca

December 2014

Abstract

We investigate data-driven decision-making (DDD) in U.S. firms enabled by the rise in information technology (IT) investment and data collection over the past decade. Our study relies on a novel survey of managerial practices among U.S. manufacturing firms conducted by the U.S. Census Bureau in 2010, combined with other detailed Census data on IT investment and firm performance. Using a differences-in-differences research design, we find that plants adopting DDD to monitor and guide their operations exhibit an average increase of 3% in value-added. This is on par with the productivity benefit associated with moving into the top quartile of the IT capital distribution, yet it is a distinct relationship. We interpret this as evidence that IT *use* is as important for firm performance as IT *investment*. DDD also displays interesting complementarities with other organizational choices. The variation in the timing and distribution of DDD effects points to complex and costly organizational adjustment on multiple margins as firms adapt to new technological possibilities.

* 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.

1. Introduction

Economists and management scholars have long been interested in uncovering the drivers of persistent performance differences among firms that appear otherwise similar in terms of size, industry focus, and investments in labor, capital, and other inputs (e.g., Jensen and McGuckin 1996, Syverson 2011, Henderson and Gibbons forthcoming). A significant body of prior evidence and educated speculation points to information technology (IT) as a potential differentiator among firms (e.g., Brynjolfsson and Hitt 1995, Dunne et al. 2004, Syverson 2011). Yet more-recent work on the role of IT in firm productivity emphasizes important differences among organizations in how they apply generic IT capital in the production process and the extent to which IT is actually integrated into essential firm activities (Bresnahan and Greenstein 1996, Aral and Weill 2007, McElheran forthcoming). Given the striking rise in IT expenditure and recent widespread attention among practitioners to "big data" (Gartner, 2014; Brynjolfsson, Hitt and Kim 2011), this paper focus on one potentially important application of IT: collecting and using data to guide managerial decision-making in firms.

Recent work has focused on management practices as a potentially critical component of heterogeneous firm performance (Bloom et al. 2013). Data -driven decision-making, in particular, is a practice that has been associated with superior firm performance in a modest sample of large public firms (Brynjolfsson et al. 2011).

Our paper leverages data recently collected by the U.S. Census Bureau to examine these phenomena in more detail, in a representative sample of firms, and with unprecedented visibility to organizational design choices and complementary activities at the firm. It extends our current understanding by providing the first evidence that applying IT to the particular practice of data-driven decision making is correlated with better firm performance, even controlling for both generic IT investment and adoption of more structured management practices. In robust cross-sectional analyses as well as our core differences-in-differences specification, we find that increased levels of data-driven decision-making are associated with increases in value-added of over 3% higher, as well as similarly

increased levels of total factor productivity (TFP) and sales (pending disclosure). The magnitude of this effect is comparable to moving into the top quartile of the IT capital distribution (which is quite skewed), suggesting that the *use* of the IT is just as important as the level of investment. However, this productivity differential is not uniform across firms: establishments that belong to multi-unit firms, in particular are less able to leverage DDD adoption or possibly just require more time to do so. Similarly, establishments that are larger or have greater capital intensity, those that are older, and those with a more complex operating environment show a lower correlation of productivity with DDD. This patterns suggests a complex and costly organizational adjustment on multiple margins as firms adapt to new technological possibilities. Consistent with this interpretation, DDD displays interesting complementarities with other organizational choices such as the level of IT investment, the education of the plant's workforce, and the allocation of decision-making within the firm.

2. Conceptual Framework – *in progress*

Data collection hardens information, among other effects. When firms invest in both the IT infrastructure to collect data and the managerial decision making process required to select what data to collect (and take a stand on why it is important), 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 1964), the steps necessary to implement data-driven decision making help more the boundary towards more objective, actionable information.

Yet, the capability to act on this new codified and accessible information is non-trivial. The economic value of information is zero unless it leads to a change in action. In complex modern enterprises, this typically requires not only changes in decision rights and processes, but also complementary changes on multiple dimensions or organization (performance pay, promotion, hiring, measurement, etc) and even

culture. This can be costly and take time. On the other hand, once implemented, they can be difficult for competitors or quickly imitate, providing a margin of competitive advantage.

3. Empirical Approach – *in progress*

Baseline regression:

$$VA_{it} \in \in X_{it} \in \in DDD_{it} \in \in_i \in \in_t \in \in_{it}$$

A test of $\beta<0$ against the null $\beta=0$

X_{it}: Time-varying establishment controls

μ_i: Establishment fixed effect

τ_t: Time fixed effect

OLS: Robust errors clustered by establishment in 2010

Also estimate over two periods: 2005 and 2010

Differences-in differences helps address persistent components of heterogeneity. But we do not have a "natural experiment" per se.

4. Data

This study is one of the first to take advantage of new data on management practices collected by the U.S. Census Bureau in 2010 and released for non-internal use in 2014. The 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 begin in years ending with "4" or "9" and end in years ending with "8" or "3." The ASM sample over-weights large plants with the intention of generating representative annual coverage of the manufacturing sector.[1]

The survey was sent to the ASM respondent for each establishment, which was typically the

[1] Plants with over 2,500 employees are sampled with certainty; the likelihood of sampling is lower but increasing with size for all plants below this threshold. Sampling weights are used throughout to provide representative statistics of the entire population of U.S. manufacturing establishments. The certainty sample accounts for roughly 70% of total value shipped in the entire U.S. manufacturing sector (http://www.census.gov/manufacturing/asm/how_the_data_are_collected/index.html).

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" has 16 questions focused primarily on monitoring, communication, and incentive practices. Examples include the collection 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 the organizational structure of decision-making and the availability and use of data therein. Among others, there are six questions about different types of decisions and whether they are made at the local plant, at headquarters, or both. The final 7 questions capture "background" characteristics of the respondent and establishment, primarily the number and education of employees and their union status. The full questionnaire is available at http://bhs.econ.census.gov/bhs/mops/form.html. Bloom et al. (2013) explores the distribution and performance implications of the practices covered in the first section. We take as our point of departure the data-related questions from the second section, as well as a couple of questions directly related to data collection and use from the first (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.[2] In order to keep our sample stable across specifications,[3] 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

[2] 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 require the establishment to have a valid linkage to the Longitudinal Business Database (LBD).We give more details about sample selection in each step of the sampling process in the Appendix (forthcoming).

[3] A virtue in terms of avoiding inadvertent disclosure of proprietary firm information and requested by the U.S. Census Bureau.

management practices questions.[4]

Panel Data Sample. Although the survey only took place once, in conjunction with the 2010 ASM, all of the questions ask respondents to report what they recall the state of practice and other features of the establishment to have been in 2005. Furthermore, the ASM collects data on various performance metrics for both 2005 and 2010. Thus, we use the differences between what is reported for 2005 and for 2010 to employ a differences-in-differences research design. Of course, the validity of this approach depends in part on the quality of recall, which is impossible to verify directly for the management practices. As one gauge of quality, we explore and control for similarities between the 2005 recall questions regarding plant-level employment levels and actual IRS records for that year and find the differences to be negligible (pending disclosure).

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 roughly 18,000 observations. 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 sample is slightly biased towards plants that are larger and slightly more productive compared to the entire ASM mail-out sample (see Table 1a). Thus, we interpret our core results as being representative for the population of larger manufacturing establishments (and the preponderance of manufacturing-based economic activity) 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 on DDD. Our results suggest that this is a valid concern. That said, we also exploit

[4] Specifically, we require that firms respond to questions 1,2,5,6,8,9,11,13,14,15,&16. Omitted from this required list are questions that we worry have a high potential to be confounded with firm performance, such as how difficult it was to meet performance targets, or what percentage of employees received performance bonuses (which would depend on sufficiently good performance to justify them). We explore but do not require responses to questions 3 & 4 on the frequency of with which KPI's are reviewed, as these have a high percentage of missing data.

the larger and more representative cross section to pursue additional insights into the mechanisms underlying the phenomenon. Finally, our sample is representative of the entire manufacturing sector in the U.S., covering 86 4-digit NAICS industries and using sampling weights to generate statistics that are representative of the entire population of manufacturing establishments.

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 are 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 we combine information from both to reduce measurement error and to help separate out firms that are closer to the frontier of practice. In addition, question 2 asks about the number of key performance indicators (KPIs) tracked at the establishment. We view the number of identified and tracked performance measures as an essential measure of the breadth of data gathering at the establishment.

Logically, having appropriate targets against which to compare real-time or historical data ought to facilitate decision-making on whether the production system is performing appropriately, identifying where it is not (and how badly), and formulating appropriate actions. Question 6 asks about the presence and time frame of production targets (short-term, long-term, or combined). The combined approach to target setting suggests a more nuanced managerial understanding of the dimensions of performance.

Thus, for our core investigation, we create a combined indicator for being at the frontier of data availability, data use in decision making quite generally, extensive use of key performance indicators, and the most nuanced approach to target-setting. We call this "data-driven decision-making" (DDD) throughout the paper.

This combined definition is justified by a polychoric factor analysis. This approach to principal component analysis appropriate for discrete variables reports a single factor with an eigenvalue of 2.28 and accounting for 57% of the variance. Our results are also robust to variations in the measurement of

DDD (see Table 6).

A quick summary of the adoption of these different combinations of data-based managerial practices is provided in Table 1b. The first two columns show how adoption changed over time in the balanced panel. The third shows results for the larger 2010 cross section.

Management Practices. A key concern for identifying the relationship between DDD and higher productivity is the possibility that DDD is merely proxying for "better management" at the establishment. Bloom et al. (2013) show robust positive correlations between the adoption of more "structured" management and performance measures identical 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 a more sophisticated or structured approach to management, in general. 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. Also in contrast to this other work, we further eliminate measures that we believe may be 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, regardless of whether contingent compensation policies are in place). We create a Z-score indexing the remaining non-DDD questions from the first section of the survey as our main proxy for "structured management."[5] Based on discrete principal component analysis – and in an attempt to use as much of the information we have to hand – 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).

[5] 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).

Complements

If firms are adjusting on multiple fronts to take advantage of new technology, this raises the interesting question of what, specifically, they are changing at the same time. An obvious candidate is the level of IT investment. We calculate IT capital stock using the ASM questions on hardware expenditure dating back to 2002 and software expenditure questions dating back to 2006. We calculate IT capital stock using a perpetual inventory approach, combining hardware and software investment and depreciating at the rate of xx% per year. For plant-years where IT expenditure information is missing, we impute values *details of method pending disclosure*.[6] Questions 34 and 35 ask about the percentage of managers and non-managers, respectively, with bachelor's degrees. We use a combined measure to explore whether DDD is complementary to having a more-skilled workforce. Finally, we construct an index of delegation from the 6 questions on where decision-making for different types of activities lies within the firm.

Industry-Level Statistics

Adoption of DDD unsurprisingly varies considerably by industry. Top industries (3-digit NAICS) for DDD adoption include: Chemicals, Paper, and Food and Tobacco manufacturing. Casual empiricism points to similarities among these industries related to the capital intensity of production and the tendency to work on continuous flows of product rather than individual pieces. In contrast, laggard industries include Apparel and Leather, Furniture, and Textile Products. These industries typically involve discrete manufacturing environments and tend to be more labor-intensive. To explore more systematic relationships between the operating environment and DDD, we create a measure of capital intensity – a simple indicator of whether the plant is in the top quartile or not of the capital stock distribution. We also construct a measure of whether the production system will tend to be more of a "continuous flow", creating situations where more standardized processes, greater monitoring and responsiveness to the

[6] Our core results are robust to excluding observations with missing data.

production environment are essential (Hayes and Wheelright 1979) Table 1c shows average rates of adoption in 2010 and the rate of change from 2005 to 2010 by 3-digit NAICS industry.

CONTROLS

Our main specifications rely on value-added as the dependent variable and control for essential inputs. This include logged depreciated capital stock (calculated following the perpetual inventory method), logged labor measured in terms of the number of employees, and logged energy inputs. The latter are 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. This approach is useful for measuring productivity, because it estimates how much output the plant creates controlling for how much it spends on primary inputs. This is very similar to a standard total factor productivity (TFP) approach, except that it does not impose strong *a priori* assumptions about the contributions of each input. We also conduct a standard TFP analysis, were we calculate the residual of a log-linear production function with the same input measures and industry-specific factor shares.

In order to control for possible heterogeneity in the quality of response to the survey that might vary systematically in ways we can address, we include "noise controls" in many specificaitons. These 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. Table 2 shows pairwise correlations among our primary variables of interest.

5. Results

Table 3 shows results for a probit model of adoption of DDD. We find that plants with more IT, more skilled labor, more structured management, and belonging to multi-unit firms are more likely to adopt DDD. We later explore whether these correlates of adoption appear to be complementary in the production function. *More on complementarities in next draft.*

Table 4 shows results for cross-sectional regressions based on the baseline 2010 sample. Recall that this sample consists of roughly 34,000 plants across the U.S. Manufacturing sector, making it by far the largest and most diverse data set yet available for this type of study. The first column establishes the baseline correlations of value-added with IT investment, controlling for logged capital stock, employment, energy expenditure, and fuel costs. In addition, we control in all specifications for whether or not the establishment belongs to a multi-unit firm. Finally, this and all the columns to follow include industry controls at the 6-digit NAICS level. The coefficient (12%) is large and highly significant.

To separate out the effect of data-driven management practices from generic investment in IT, columns 2 and 3 separately control for IT capital stock. Column 2 uses the logged value of deflated IT capital accumulated at the plant since Census started tracking hardware investment in 2002 and software investment in 2006. This is useful in light of the prior literature emphasizing the relationship between generic IT investment and productivity (e.g., Brynjolfsson and Hitt 1995; Dunne, Foster et al. 2004). Due to the skewed nature of IT investment, Column 3 reports on two IT indicators – one for being in the top quartile of the IT capital stock distribution (High IT) and one for being in the quartile just below it (Medium IT). Interestingly, the coefficient on DDD remains quite robust to this inclusion, suggesting that IT investment and use have separate and important relationships to firm performance.

Column 4 addresses the possibility that firms relying on data in their decision-making may simply have superior management practices in place at the firm. Bloom et al. (2013) shows that structured management is associated with productivity. Column 4 includes a Z-score for all non-data, non-performance, and non-firing questions addressed by the first section of the MOPS survey (details above), as well as a separate index for the dismissal/re-assignment questions (15 & 16) We also control for the skill level at the plant in terms of the percentage of managers and non-managers with bachelor degrees. The coefficient goes down substantially – from .112 to .086 – suggesting that care is required here, but that important differences remain between structured management more generally and the specific practice of DDD.

The coefficient remains largely unchanged in column 5, where we introduce controls for the quality of the survey response such as the day of the week, the seniority of the respondent, and the discrepancy in how many employees are reported on the survey versus IRS tax records. Column 6 reports coefficients for the same model, but restricting the sample to the roughly 18,000 observations in the balanced panel. The point estimate is slightly lower, but the difference is not statistically significant.

Table 5 presents our core differences-in-differences results. These models are run on 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. Column 1 shows the DDD coefficient by itself, which is commensurate with the coefficient from column 6 of Table 3. Column 2 is consistent with the prior table, showing that the effect of DDD appears distinct from investment in generic IT capital. Controlling for structured management practices has a large effect in column 3. This, our preferred specification, reports a coefficient on DDD of 3%. To put that in context, it is the same magnitude as that for moving into the top quartile of the IT capital distribution (High IT). Such a move would entail a roughly 20 times greater level of IT capital accumulation (based on a comparison of the means of these different groups in 2010) and would require an investment of $xx million in the average firm in our sample.

The results in column 3 suggest that unobserved heterogeneity may have a non-negligible impact on the correlation 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 cross-sectional analysis. However, the confidence intervals are very close.

Columns 4 and 5 show that the incidence of this relationship varies a great deal by whether the plant is a single-establishment firm or belongs to a larger, multi-unit firm. In fact, when the sample is split according to single-unit status, most of the effect is associated with single-unit firms that enjoy a tremendous benefit (over 12% higher value-added) from adopting DDD; the comparable coefficient for multi-unit establishments is not significantly different from zero. However, column 6 shows that the effect for multi-unit firms may simply be taking more time to show up. When we use 2011 as the end

date, instead, we see that the multi-unit establishments have a coefficient on DDD of 3.8% that is significant at the 5% level. We interpret this as being informative about the adjustment costs entailed in adopting DDD, a point we return to later.

Table 6 shows that these findings are robust to the alternative definitions of DDD. The rows of the table represent the coefficients for: 1) using only question 28 on the use of data in decision making, 2) using question 28 together with question 27 on the availability of data, and 3) combining these two with question 2 about the number of KPIs. The columns show the results by single-unit versus multi-unit status in the same specification as that for column 3 of Table 4. Changing the precise measure of DDD changes the point estimate and precision somewhat, but the results are all statistically indistinguishable from each other.

Table 7 explores variation in these results that depend on characteristics of the firm and help illuminate some of the mechanisms at work. Column 1 shows that the benefits of DDD accrue, at first, to smaller establishments. The interaction effect between DDD and being in the top quartile of the size distribution is negative and has a magnitude that overwhelms all of the benefits of adopting DDD, at least within the five year period we examine. The converse specification, with an indicator for the first quartile in terms of size (available upon request) confirms that smaller plants have a larger association between DDD and productivity. However, lagging the end year to 2011 eliminates this difference, suggesting that adjustment costs may simply be slowing down the impact of DDD at larger establishments.

A similar story plays out in the following columns. Capital-intensive establishments are examined in column 2. Interacting DDD with being in the top quartile of the capital stock distribution shows that the most capital-intensive plants fail to benefit from DDD within five years. Splitting the sample by multi-unit status (available upon request) shows that this primarily takes place among multi-unit plants; moreover, the difference again disappears by 2011. Column 3 of Table 6 shows that younger plants enjoy a bigger relationship between DDD and productivity within the five year period. Columns 4 and 5 explore the implications of having a more complex operating environment in terms of producing a greater number

of products. Plants with more products enjoy a stronger relationship between DDD and productivity, but not until 2011. Again, we interpret this pattern of results as being indicative of costly adjustment to DDD.

Table 8 begins to explore complementarities. We find evidence consistent with complementarities between IT adoption and DDD. This is consistent with IT providing the additional data need to be successful with DDD. Likewise, more-skilled labor also appears to be complementary with DDD. We also find evidence that DDD is complementary with reductions in delegated decision-making, perhaps because central decision-makers can access more of the data needed to make decisions when information is explicit rather than tacit. *More on this in the next draft…*

Robustness checks. *Coming in Table 9.…*We explored other interactions, but found nothing to be statistically significant when explored: continuous improvement practices (Q01), firing practices (Q15 and Q16), nonunion status, or vertical integration measured as the value of within-firm transfers between plants. Also non-significant results with communicating about production targets and frequency of review. Splits we tried but found noisy results: number of products, multi-unit firm, size, exporter, vertically integrated commodity flows.

6. Conclusion

A great deal of anecdotal evidence, speculation, and assertions on the part of industry observers suggest that productivity gains will 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 some of the first evidence that putting data "into action" in these ways is associated with significantly higher productivity. This relationship holds across a variety of industries in the U.S. manufacturing sector, increases over time (particularly for the larger and more complex organizations) and stands up to the inclusion of difficult-to-observe controls for the level of IT investment and other management practices at the organization. Moreover, our findings provide the first

evidence that this correlation is robust to unobserved, time-invariant, plant heterogeneity that might otherwise threaten the validity of these findings.

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.

Also, at the current state of our analysis, we want to stop short of claiming clean causal evidence. Firms in our sample adopt DDD as a choice, and therefore there is no sense in which firms are exogenously "treated with DDD" and enjoy superior performance as a result. However, our findings can be taken as preliminary evidence about the magnitude of the relationship. We hope the sceptics will be spurred to further research into the causal relationship between data-driven decision making and firm performance in manufacturing and other sectors of the economy.

Table 1a. Descriptive Statistics

Variable	Description	Balanced 2010 Sample Mean (estimated)	S.D.	Base 2010 Sample Mean (est)	2010 ASM Mean (est)
Log Value - Added	Log of total value shipped minus cost of goods sold	**9.23**		**8.26**	
Log Sales (TVS)	Log of total value shipped	**10.02**		**8.93**	
Markup	? Look this up	**.31**		**.31**	
Log Exports	Log of total value of goods exported	**3.65**		**2.41**	
Log Employment	Log of the total number of employees at the plant	**4.34**		**3.64**	
Log K stock	Log of capital stock calculated using the perpetual inventory method and BEA capital deflators	**8.91**		**8.26**	
Log Energy Costs	Lot of the total cost of both fuel and electricity consumed by the plant	**5.75**		**4.61**	
Multi-unit Status	=1 if the plant belongs to a multi-unit firm	**.71**		**.52**	
HQ Status	=1 if the plant is a headquarters; equal to 1 for all single-unit firms	**.49**		**.63**	
Exporter status	=1 if the value of exports is >0	**.46**		**.34**	
E-Com Dummy	=1 if the value of e-commerce is >0			**.54??**	
Age	Derived from the year first observed in the LBD. Truncated at 44.	**24.9**		**21.3**	
Log IT Capital Stock	Log of the value of hardware and software stocks calculated using the perpetual inventory method and using BEA deflators (?)	**3.83**		**2.89**	
Structured Management Z-Score	Index created by summing up the normalized values from questions 1,5, ,8,9,11,13 & 14 of the MOPS.	**.607**		**.539**	
% Employees with BA	% of managers and non-managers with bachelor's degrees at the plant.	**.08**		**.07**	

Note: population means estimated using sampling weights.

Table 1b. Adoption of Data-Related Managerial Practices by year

Data-related management practice	**Adoption in 2005 Balanced sample (weighted)**	**Adoption in 2010 Balanced sample (weighted)**	**Adoption in 2010 Cross Section (weighted)**
Top 2 categories for "availability of data" (Q27)	.43	.69	.59
Top 2 categories for "use of data" (Q28)	.41	.66	.55
Top 2 categories for both Q27 & Q28	.32	.58	.47
Top for Q27, Q28, tracking 10 or more KPI's (Q2)	.14	.34	.22
DDD: as above + Targets (Q6)	.09	.25	.16
N	~18,000	~18,000	~34,000

Table 1c. Adoption of DDD by Industry (3-Digit NAICS Code)

3-Digit NAICS Code	Industry	Mean DDD in 2010 (baseline sample)	Mean DDD change from 2005 to 2010 (balanced sample)
311	Food	.617	.142
312	Beverage and Tobacco Products	.664	.163
313	Textile Mills	.655	.102
314	Textile Product Mills	.455	.064
315/316	Apparel and Leather	.414	.023
321	Wood Products	.503	.066
322	Paper	.675	.164
323	Printing and Related Support Activities	.478	.061
324	Petroleum and Coal Products	.569	.077
325	Chemicals	.681	.128
326	Plastics and Rubber	.626	.137
327	Nonmetallic Mineral Products	.504	.069
331	Primary Metals	.614	.139
332	Fabricated Metal Products	.515	.080
333	Machinery	.534	.088
334	Computers and Electronic Products	.624	.108
335	Electrical Equipment, Appliances, and Components	.624	.148
336	Transportation Equipment	.655	.138
337	Furniture	.443	.077
339	Miscellaneous	.491	.074

Table 2. Pairwise Correlations

	logva	DDD4	highIT	medIT	Z_nd	Z_fire	Degree	Exporter	logte	Logkst	Energy	HQ	MU
Logva	1												
DDD4	.294	1											
highIT	.497	.179	1										
medIT	.076	.011	-.338	1									
Z_nd	.447	.327	.254	.049	1								
Z_fire	.150	.127	.091	.030	.292	1							
Degree	.135	.078	.101	.006	.119	.042	1						
Exporter	.344	.118	.205	.054	.205	.055	.108	1					
Logte	.821	.266	.508	.086	.407	.180	.094	.324	1				
Logkst	.725	.272	.462	.040	.393	.125	.110	.261	.675	1			
Energy	.763	.292	.394	.062	.426	.138	.076	.269	.715	.735	1		
HQ	-.231	-.189	-.060	.011	-.301	-.076	-.011	-.037	-.127	-.246	-.311	1	
MU	.337	.211	.177	-.011	.357	.130	.026	.103	.253	.321	.386	-.601	1

MedIT is not significantly correlated with both degree and MU; other than that, all the correlations are significant at least the 5% level.

Table 3. Drivers of DDD Adoption – Probit regressions of adoption, showing average marginal effects at sample means

Dependent Variable:	**DDD**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	IT	**Degree & Structured mgmt. index**	**MU & K-intensity**	**Industry Information**	**Delegation & HQ if MU==1 (pending disclosure)**
High IT	.045*** (.007)	.021*** (.007)	.017*** (.007)	.011 (.007)	
Medium IT	.020*** (.006)	.009 (.006)	.011** (.006)	.010* (.006)	
Degree		.141*** (.020)	.144*** (.019)	.127*** (.019)	
Structured Mgmt index		.662*** (.018)	.616*** (.018)	.629*** (.018)	
Firing Practices		.052*** (.007)	.052*** (.007)	.048*** (.007)	
MU			.070*** (.006)	.084*** (.006)	
Capital Intensity			.023*** (.006)	.044*** (.006)	
Continuous Flow				.013** (.006)	
Delegation					<0 (not sig.)
HQ					<0**
Log Employment	.078*** (.003)	.045*** (.003)	.037*** (.002)	pending	
Observations	~34,000	~34,000	~34,000	~34,000	~34,000
Sample	2010	2010	2010	2010	2010
Fixed effects	NAICS4	NAICS4	NAICS4	omitted	NAICS4

Table 4. Conditional Correlations between Data-Driven Decision Making (DDD) and Firm Performance

Dependent Variable:	**Log (Value Added) Controlling for Logged Capital Stocks, Labor, & Energy**					
	(1) DDD	(2) Add IT Stock	(3) IT Categories	(4) MGMT & Educ.	(5) Add Noise Controls	(6) 2010 Balanced Sample
DDD	**.120*** (.015)**	**.112*** (.015)**	**.112*** (.015)**	**.086*** (.015)**	**.082*** (.015)**	**.067*** (.017)**
Log IT Capital Stock		.048*** (.004)				
High IT			.226*** (.018)	.209*** (.018)	.195*** (.018)	.191*** (.024)
Med IT			.134*** (.013)	.127*** (.013)	.118*** (.013)	.012*** (.020)
Structured management index				.239*** (.042)	.237*** (.051)	.284*** (.062)
% Degree				.309*** (.061)	.297*** (.061)	.408*** (.078)
Observations	~34,000	~34,000	~34,000	~34,000	~34,000	~18,000
Sample	2010	2010	2010	2010	2010	2010 observations with recall data for 2005
Fixed Effects	NAICS6	NAICS 6	NAICS 6	NAICS 6	NAICS 6	NAICS6

Note: Weighted OLS regressions using ASM sampling weights. In all columns the dependent variable is logged nominal value added. The structured management index is the unweighted average of the score for questions 1,5,8,9,11,13 & 14, where each question is first normalized to a 0-1 scale. Unreported controls in all columns include: firing practices (Q15 and A16), whether the establishment belongs to a multi-unit firm, logged capital stock, logged employment, winsorized logged energy expenditures, and whether or not degree information 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. Standard errors are clustered at the firm level. Statistical significance denoted as follows: * 10%, ** 5%, *** 1%.

Table 5. Differences-in-differences estimation of DDD

Dependent Variable:	**Log Value Added**					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	DDD	**DDD + IT**	**Control for Mgmt & Educ.**	**SU Firms**	**MU Estabs**	**2011 End MU Estabs**
DDD	.079*** (.014)	.073*** (.014)	.033** (.016)	.127*** (.038)	.020 (.017)	.038** (.018)
High IT		.046*** (.016)	.031* (.016)	.062** (.028)	.029 (.019)	.058*** (.019)
Med IT		.027*** (.012)	.018 (.012)	.028 (.023)	.016 (.014)	.054*** (.015)
Structured Mgmt Index			.285*** (.048)	.298 (.122)	.275*** (.053)	.419*** (.056)
Degree			-.159 (.115)	.095 (.256)	-.200 (.128)	-.026 (.135)
Log K, Log L and Log Energy	Yes	Yes	Yes	Yes	Yes	Yes
Establishments	~18,000	~18,000	~18,000	~4,000	~14,000	~14,000
Sample	Balanced	Balanced	Balanced	Balanced	Balanced	Balanced

Note: Weighted linear regressions using ASM sampling weights. 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 (Q15 and A16), whether the establishment belongs to a multi-unit firm, logged capital stock, logged employment, winsorized logged energy expenditures, and whether or not degree information is missing. The sample columns 1-3 is all MOPS observations with at least 11 non-missing responses to management questions for both 2005 and 2010 and 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 4 and 5 split this sample by multi-unit status. Column 6 has as the dependent variable logged value added in 2011. Observations with missing controls for 2011 are imputed MORE DETAILS HERE to strictly maintain the same number of observations.

Table 6. Differing definitions of DDD

Dependent Variable:	**Log Value Added**	
	(1)	**(2)**
	Single-Unit Firms	**Multi-Unit Establishments**
Q28	.063***	.025
	(.013)	(.017)
Q27 & Q28	.090***	.053***
	(.027)	(.017)
Q27, Q28, & KPI's	.103***	.052***
	(.035)	(.017)
Establishments	~18,000	~18,000
Sample	Balanced	Balanced

Table 7. Variation in the Effects

Dependent Variable:	**Log Value Added**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	Establishment Size	**K Intensity**	**Age**	**Number of Products**	**Number of Products End 2011**
DDD	.037* (.020)	.054*** (.017)	.025 (.016)	.017 (.020)	.020 (.022)
DDD x Top Size Quartile	-.078*** (.029)				
DDD x K-intensive		-.062** (.029)			
DDD X Young			.087* (.049)		
DDD X Number of Products				.038 (.027)	.061** (.029)
Establishments	~18,000	~18,000	~18,000	~18,000	~18,000
Sample	Balanced	Balanced	Balanced	Balanced	Balanced

Table 8. Interaction Effects

Dependent Variable:	**Log Value Added**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	HighIT	**High IT, Single Unit**	**Educated Workers**	**Delegation**	**Delegation End 2011**
DDD	.004 (.019)	.131*** (.042)	-.002 (.024)	.042** (.020)	.043** (.022)
DDD x High IT in 05	.087*** (.027)	.002 (.074)			
DDD x High Degree in 05			.055** (.028)		
DDD x High Delegation in 05				-.057* (.030)	-.006 (.033)
Sample	**Balanced**	**Balanced SU establishments**	**Balanced**	**Balanced**	**Balanced**
Establishments	~18,000	~4,000	~18,000	~18,000	~18,000

Note: the term interacted with DDD is fixed at 2005 levels.

Table 9. Robustness checks

REFERENCES

Aral, S. and P. Weill (2007). "IT Assets, Organizational Capabilities, and Firm Performance: How Resource Allocations and Organizational Differences Explain Performance Variation." Organization Science **18**(5): 763-780.

Bloom, N., E. Brynjolfsson, et al. (2013). Management in America. Center for Economic Studies Research Program. Washington D.C., U.S. Census Bureau.

Bresnahan, T. F. and S. Greenstein (1996). "Technical Progress and Co-Invention in Computing and in the Uses of Computers." Brookings Papers on Economic Activity. Microeconomics **1996**: 1-83.

Brynjolfsson, E. and L. Hitt (1995). "Information Technology As A Factor Of Production: The Role Of Differences Among Firms." Economics of Innovation and New Technology **3**(3-4): 183-200.

Brynjolfsson, E., L. M. Hitt, et al. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?, MIT Working Paper.

Dunne, T., L. Foster, et al. (2004). "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment." Journal of Labor Economics **22**(2): 397-429.

Gartner (2014) Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014 http://www.gartner.com/document/2841519.

Jensen, J. B. and R. H. McGuckin (1996). Firm Performance and Evolution: Empirical Regularities in the U.S. Microdata. Center for Economic Studies Working Paper CES 96-10. U.S. Census Bureau.

Hayes and Wheelright (1979).

Henderson and Gibbons, forthcoming

McElheran, K. (forthcoming). "Do Market Leaders Lead in Business Process Innovation? The Case(s) of E-Business Adoption." Management Science.

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