# **Big Data Analytics and Firm Productivity**

Liang Guo, Mingtao Fu NEOMA Business School Mont Saint Aignan, France e-mail: Liang.guo@neoma-bs.fr; fu.mingtao.14@neoma-bs.com

Ruodan Lu Cambridge University, UK e-mail: R1508@cam.ac.uk

Abstract—Increasing productivity is a key task for contemporary firms. Although big data analytics has been generally viewed as an effective advanced information processing tool that enables firms to better cope with business operation, thus holding the potential to boost firm productivity, evidence to support this view is lacking in the literature. Equipped with the theory of organizational information processing and resource-based view, we hypothesize that big data analytics systems (BDAS) can help improve productivity and this contribution is influenced by the firm's BDAS capability. These hypotheses are supported with a sample of 45 Chinese retailers over 2012-2014 using Data Envelopment Analysis and Malmquist Productivity Index. This study extends our understanding of the business value of IT by shedding light on the productivity benefit of big data investment. We also underline that for firms that have already implemented BDAS, a greater effort seems necessary to build predictive and prescriptive analytics capability.

#### Keywords-big data analytics; retailing; China.

### I. INTRODUCTION

Today's fast-moving, complex world of increasingly connected people and connected things that are creating vast new digital footprints require enterprises to constantly make sense of this big and fast-moving data and to gain real-time access to powerful insights and deliver them at the point of action [1]. Big data, defined as a holistic analytical approach to manage, process and analyse 5Vs (i.e., volume, variety, velocity, veracity and value, see [2]), is regarded as a powerful weapon to obtain actionable insights for sustained value delivery, to create business disruptions, and to establish competitive advantages [3].

However, such presumption confronts with scare empirical evidence. Recently, skepticism has emerged about the true business value of big data analytics. Given that the implementation of big data requires huge physical and human capital investment, the stake of embracing big data for a company is quite high. Big data may lead to big disappointment. Prior studies find that IT implementation is inherently risky and post-implementation may also be fail.

Although related strand of studies has focused on the nexus of IT investment and firm performance in general, there is not virtually any research that has quantified the performance impact of big data analytics system (BDAS). By performance, we refer to a firm's productivity, also called operation efficiency (i.e. using less input to achieve the same output or using the same input to achieve more output), one of the most important performance measures. [4] find that there is a significant contribution of IT to total factor productivity for the US economy in 2000-2004. However, there are relatively few studies addressing IT's impact on productivity at firm level [5] and among the existing works, most of them focus on the association between productivity and either firm's IT investment Little attention has been paid to whether the use of specific IT system (say, BDAS) is associated with a firm's productivity.

Therefore, identify the true impact of BDAS on firm productivity, if any, would thus provide strong implications for both academic research and practice. In this study, we intend to fill in the void by examining whether BDAS boosts a retail firm's productivity. Grounded in the theory of organizational information processing [6] and resource-based view [7], we sought to theorize the business impact of BDAS. Productivity is commonly measured with Data Envelopment Analysis (DEA), a non-parametric approach to the empirical estimation of production function and with Malmquist Productivity Index (MPI), a special method of time series analysis in DEA [8]. Using a sample of 45 Chinese publically listed retailers, we intend to empirically answer two research questions: 1) what is the impact of BDAS on firm productivity? And 2) how does the impact of BDAS on firm productivity, if any, vary with the capability of BDAS? Our findings support the claim of BDAS's business value and its contribution increases with the firm's capability of BDAS.

In Section 2 we will review relevant literature and propose two hypotheses. We will test the hypotheses in Section 3 and discuss our findings in Section 4.

#### II. LITERATURE REVIEW AND HYPOTHESES

From the late 2000s, the term business analytics has gained prominence over the previously dominant business intelligence principles, by stressing more on the analytical capabilities it offers. In this sense, business analytics refers to the process of transforming data into actions through analysis and gaining insights in the context of organizational decision making and problem solving. Big data analytics resorts to the use of data warehouse, information technology, statistical analysis, and computational intelligence models to help managers gain improved insights about their business operations and make better, fact-based decisions. The contextual application of analytical methods to analyse data sets that are very large in size and complex in terms of their sources and of the level of unstructuredness has led to the growth of big data analytics. It aids to better data analytics that can decipher and understand patterns from the business situations based on the data collected from diverse sources with the aim to predict future outcomes purely based on analytical frameworks. The unique features of big data analytics include the use of advanced data collection, data visualization technologies, feature extraction, and artificial intelligence technologies, which largely increment the speed of mining and analysis of previously ignored unstructured data that reside in disparate sources and different formats.

Big data analytics differs from traditional data processing architectures in terms of the speed of decision making, processing complexity, type and volume of data being analysed. The data sources are unstructured in form and rely on newer forms of business analytics like web analytics, click stream analytics and visual analytics that use advanced modelling methods. In general, we argue that big data analytics assist in decision making in the following ways: descriptive, predictive, and prescriptive analytics.

Descriptive analytics is the most common form of business analytics employed by firms to understand the historical business performance and derive useful wisdom as past performance summary. The analytics can be visualized using charts and figures. Its aim is to condense vast chunks of data and decipher useful information from the same. In this sense, it serves as a valuable information reduction technique. Descriptive analytics can be used to drill-down reports to decipher the association between, for instance, an advertising campaign and the product performance through patterns or trends on the basis of historical data. Descriptive analytics, however, cannot be used to test causality.

Predictive analytics analyses past historical performance to forecast future situations by identifying patterns of knowledge based on a variety of statistical tools. For instance, the use of past historical data of a particular advertising campaign to extrapolate the expected response of a new product launch campaign or predict the demand of a seasonal product based on the past years' data. The efficiency of predictive analytics lies in its ability of revealing causality that cannot be undertaken from descriptive analysis. The use of big data comes handy in adding diversity to data sources to reveal more detailed and to elaborate patterns and in modelling complex data originating from different, in-compatible data sources like text, speech, video, and photos.

Prescriptive analytics is a relatively new stream of business analytics that extends beyond descriptive and predictive analytics, by prescribing multiple solutions for future situations and the likely impact of each solution. This form of analytics is also based on a collection of business rules, algorithms, relying on statistical machine learning and simulation-optimization modelling procedures. The difference between predictive and prescriptive analytics lies in the outcome of each option that is offered by the latter method. This is incorporated as a feedback loop to track the impact of each forecasted solution in different what-if scenarios. And finally, based on probabilistic modelling, the most optimal solution is recommended. Some of the leading

applications of prescriptive analytics include optimization models in field of operations, marketing, and finance to identify the best alternatives to minimize or maximize some objective. For instance, its applications in operations management include multiple scenario generation balancing the production planning based on different demand estimates to maximize revenue or natural disaster based recovery options.

The contribution of IT to business performance has been studies from two theoretical perspectives [9]. On the one hand, the theory of organizational information processing [6] suggests that IT systems enable improved information processing and managerial decision making so that a firm can better handle uncertainty, thus achieve better performance. On the other hand, IT is regarded as a kind of key resources for achieving business success and a power tool to help firms gain competitive advantage by altering the competitive forces that collectively determine industry profitability. This study is motivated by these two prior theoretical considerations to examine whether and how BDA influences firm's performance in general and productivity in particular. Answering this question fills a gap in the literature on the possible big business value of BDA.

Prior studies have indicated high level of business risk derived from dynamism and complexity decreases firm productivity. To boost a firm's productivity, managers should lower their business uncertainties by frequently processing a greater deal of information and by making better decisions. TOIP attributes part of firm's risk to the accuracy of information and manager's capability in information processing about uncertainty in the business environment. TOIP proponents argue that IT provides a technological basis for collecting and sharing information from different sources, for integrating business processes and for coordinates decisions makings and actions within and across organizational boundaries. This information processing effect is especially true, when a firm faces high environmental uncertainty and often needs to deal with a large amount of complex information [9]. Prior studies found empirical evidences that support TOIP, indicating improvements in information quality and visibility for decision makers reduce business risk and improve business performance.

We draw upon TOIP to theorize BDA's productivity benefits. Through the automation, optimization and transparentization effects, BDA equips managers with applications that are able to amass and apply information in ways in which upends customer expectations and optimized business operations to unprecedented degrees. In essence, BDA mitigates risks and boosts operational efficiency:

H1: An increase in firm productivity is associated with the implementation of BDA.

In line with TOIP, we suggest that the practice of a broad scope of BDA (i.e. more BDA applications) enhances a firm's information analytics quality and leads to more efficient operations. The practice of a broad scope of BDA allows firms to take the questions that traditional analytics is answering to the next level, moving from a retrospective set of answers to a set of answers focused on predicting performance and prescribing specific actions or recommendations. It is reasonable to believe, as TOIP suggests, a firm's operations will be more efficient thanks to the improvement of information processing capability and quality.

What is more, the analytics expertise and experience accumulated from practicing multiple types of BDA applications form a bundle of strategically important resources ranging from IT hardware assets, BDA human resources, and intangible social-technical resources. RBV theorists assert that firm's key and unique resources are of four properties--value, rarity, inimitability and nonsubstitutability (VRIN) so that a firm's performance depends on management's capability to search for the best usage of these resources. IT-related resources have long be regarded as VRIN, which are firm specific and cannot easily be traded or transferred. Following the IT-RBV perspectives, we argue that a board scope of BDA helps firms to form valuable resources. Like most IT skills, expertise with different analytics applications can only be built through learning by doing practices and by creating the synergies between human and IT-assets. These synergistic intangible resources are value-creating, allowing an organization to capitalize strategic opportunities or diverge from potential threats. BDA professionals who master multiple types of BDA usually possess rare qualities that are much in demand, difficult and expensive to hire and given the very competitive market for their services, difficult to retain. This BDA human capital is rare resources for a firm, as there are not a lot of people with their combination of scientific background and computational and analytical skills.

Likewise, analytics skill shortage makes BDA human capital difficult to be easily imitated by competitors. More importantly, combining big data hardware and skill sets to create an enterprise-specific BDA infrastructure can be inimitable, because creating such synergy requires carefully melding IT and organizational resources to fit firm needs and business priorities. Finally, BDA expertise can be viewed as a social-technical synergy of tacit and explicit operations policies and practices within a firm. It is the result of the longterm, enterprise-wide integration of various BDA applications, business processes, human capital and continuous learning and innovation. BDA expertise is then difficult to substitute, because it is scarce, specialized, and appropriable. And given that data are woven into every sector and function in modern economy, much of advanced business decision making simply could not take place without BDA. There is then virtually no substitutable technology that can bring large pools of data, analyze to discern patterns, and optimize decisions.

Therefore, we argue that combining with organizational expertise gain from the practice of a broad scope of BDA is a source of competitive advantages, enhancing productivity and creating significant value by making better quality of decisions and optimized operations.

H2: An increase in firm productivity is associated with the practice of a broad scope of BDA.

### III. EMPIRICAL STUDY

## A. Sample

We tested these two hypotheses with a sample set of China's publically listed retail firms. The reasons that we chose the retail industry are because this is an industry with a long history of investing and implementing data-intensive IT systems (e.g. POS, ERP, CRM and consumer panel studies) and retailers usually demand a clear justification for the return on IT investment. We chose China's retail sector because of its enormous market size. The total sales of consumer goods in China in 2014 reached USD 4,099.906 billion, representing a growth of 12% [10]. Our final sample included 45 firms. We collected their financial data from their IPO prospectus and annual reports over 2012-2014.

### B. Analyses and Results

Productivity is measured by DEA and MPI [8]. we took number of employee (hereafter A), management & administration expenses (B), marketing expenses (C), cost of sales (D), and inventory (E) as inputs and net revenues (hereafter 1), gross profits (2), net income (3) and market capitalization (4) as outputs. The productivity index was calculated with the averages of these inputs and outputs over 2012-2014. We decomposed the MPI into two componentstechnology change (i.e. technology frontier shift) and technology efficiency change (i.e. a measured of the diffusion of the best-practice technology in the sector). Then, we first calculated MPI, technology change and technical efficiency change with the Model ABCDE123 (i.e. number of employee, management & administration expenses, marketing expenses, cost of sales, and inventory as inputs and net revenues, gross profits, and net income as outputs) for the period 2012-2013 and 2013-2014 respectively. Then, we took the geometric mean of these two periods as the measures.

We measured whether a firm has implemented BDAS based on the answers of our interviews. BDAS Capability is measured with the Big Data and Analytics Maturity Model [11]. We developed a scoreboard of BDAS capability, which contains 16 dichotomous items (1 for have implemented a particular BDAS application in question, 0 for otherwise). The sum of these 16 items was used as the measure of a firm's BDAS capability.

(a) Descriptive Analytics Applications:

1. Big data analytics infrastructure platform such as Hadoop, No SQL Database and/or Massively Parallel Processing Databases

2. Data visualization applications or any business intelligence platforms for integrating and visualizing data from multiple sources, taking the raw data and presenting it in complex, multi-dimensional visual formats (e.g. reports and dashboards) to illuminate the information

(b) Predictive Analytics Applications:

3. Basic unsupervised machine learning clustering applications, such as k-means and A Priori association analysis

4. Basic supervised machine learning classification applications, such as, k-Nearest Neighbours and decision trees

5. Regression-based supervised machine learning applications, such as, linear regression, logistic regression, generalized linear models and the exponential family

6. Bayes statistics and kernel-based supervised machine learning applications, such as na we Bayes, graphical models, support vector machines and Gaussian process

7. Latent variable-based machine learning applications, such as mixture models, the EM algorithm, principal component analysis, singular value decomposition

8. Advanced unsupervised machine learning applications, such as ANN, auto-encoder and deep learning

9. Reinforcement machine learning applications, such as Markov decision and MCMC

10. Time series and forecasting applications

11. Natural language processing applications

12. Computer vision applications

(c) Prescriptive Analytics Applications:

13. Linear optimization and sensitivity analysis applications

- 14. Network analysis applications
- 15. Nonlinear optimization applications
- 16. Simulation-based decision making applications

25 out of 45 retailers have scored 100 in full Model ABCDE1234. Among these 25, there are 14 retails that have implemented BDAS. Retailers that have implemented BDAS clearly achieved better productivity than their counterparts that have not in most DEA models, supporting H1.

We then followed [8] to conduct PCA on these 30 DEA models in order to theorize the impact of BDAS on productivity in a parsimonious way and to reveal the similarities and differences that exist between firms in terms of the 30 DEA models in a two-dimensional plot. The first principal component accounts for 44.37% of the total variance while the second for 34.62% (i.e. in total 78.99%). Most models are loaded on Component 1, which can be associated with an overall measure of efficiency. We labelled the second component as "Management & Administration, Inventory and Marketing Efficiency" (see Table 1).

We then calculated the weighted average of two components for each retailer by multiplying the factor loading of each DEA model on its productivity score. We plotted each firm on a two-dimension figure with the first component as the X-axis and the second component as the Y-axis (See Fig. 1). It seems that most BDAS retailers achieved high overall efficiency and scattered along the X-axis while those non-BDAS retailers achieved relatively high efficiency on management and administrative expense, marketing expense and inventory management and scattered along the Y-axis. We conducted an ANOVA analysis by comparing two components between two groups. The results suggest that the overall productivity (i.e. Component 1) of the BDAS retailers outperforms that of those without BDAS (F-score=9.228, p < 0.01). However, there is no significant difference on Component 2 (F-score=0.118, p=0.67). Finally, we aggregated the scores of 30 DEA models into one factor and the result of the ANOVA analysis of this aggregated factor indicates that retailers with BDAS are at an advantage in achieving higher efficiency score (F-score=9.823, p<0.01).



Figure 1: Visualization of DEA Results on Two Principal Components

The MPI analysis shows that 23 out of 40 retailers have experienced productivity progress during 2012-2014. Among these 23. 11 are retails that have implemented BDAS. Company 29, who has implemented BDAS, has registered the highest improvement in MPI (1.442). It can be seen that the progress of MPI for this company was contributed by a significant increase in technology change (1.345) and in technical efficiency change (1.072). Company 39, who also has equipped BDAS, has experienced the largest increase on technology change (1.17). But its decrease in technical efficiency change (0.894) implies that this retailer has failed to conduct proper investment in organizational factors in accordance with its operation. The results of ANOVA analyses showed that on average, retailers with BDAS achieved better MPI (F-score=5.427,p<0.05) and technology change (F-score=4.780, p<0.05) than those without BDAS. But there is no significant difference in technical efficiency change. However, technical efficiency change is a measure of the deviations from the best practice frontier with the sample and it is not directly related to BDAS. So we can safely conclude that H1 is also supported by the MPI analyses.

Finally, we then tested H2 by calculating the Spearman correlation coefficients between BDAS capability score with the aggregated DEA efficiency score and the MPI. Both are statically significant (0.689, p<0.01 for DEA and 0.359, p<0.05 for MPI). The relationships between these three measures were visualized by plotting a bubble chat (Figure 2), in which the aggregated DEA is the X-axis, the MPI is the Y-axis and BDAS capability as bubble size. In Table 2, it is clear that retailers with high BDAS capability score concentrate around the upper-right corner, which means they achieve both high DEA and MPI scores. H2 is supported.

#### IV. DISCUSSION AND CONCLUSION

In this study, we theorize how BDAS can help improve firm productivity through the automation and optimization effects. For the first time in the literature, our study empirically examines whether the implementation of BDAS has a significant impact on firm's productivity and whether this contribution is influenced by the firm's BDAS capability. Our findings provide solid evidence that permits an accurate appraisal of the impact of BDAS on firm productivity. There is a general consensus that the spectacular popularity of BDAS in the past few years has opened up new opportunities for firms to re-engineering business processes and to improve their operation efficiencies. Under this context, the analysis of BDAS's effects on firm productivity has been the subject of great attention. Our study complement to the long strand of research that has examined IT's business value in general by revealing the productivity improvement benefits derived from BDAS. Our findings are in line with TOIP and RBV theory, implying that the big data analytics capability of an IT system helps firms develop key and unique resources and then better coordinate and resolve environmental uncertainty.

Our findings are robust as our measure of productivity is calculated with 30 different combinations of multiple/single inputs and outputs DEA models. In addition, our analyses using the MPI approach provides in-depth information on whether BDAS triggers a firm's technology change, which provides managerial implications on what kind of weakness in terms of productivity they should watch out for and remedy. As technology progress is one of the most important factors for future retailing, implementing effectively new technologies like BDAS can help give business an edge in the market place.

Our study made another contribution by developing a measure of BDAS capability. Our findings suggest that only when the implementation of predictive as well as prescriptive big data analytics completed, the benefits from adopting big data start to surface. Descriptive analytics or business intelligence solutions have traditionally required a static, lowdimensional data model for operational reporting. These solutions often are incapable to use streaming data that can provide operational intelligence. However, the emergence of multi-channel selling and fulfilment has increased the need for dynamic, high-dimensional data analytics to process streaming and massive datasets. Predictive and prescriptive analytics businesses can gain new operational insights by taking advantage of the unique capabilities of BDAS and by using the right model to process large volumes of unstructured as well as structured data.

Our study provides strong managerial implications. One of highly productive firm in our sample has successfully deciphered the code of "Big Data Value". With its website and mobile app clickstream data, this retailer can predict the sales of its cash cow products, as well as determine the price range and features that most customers want. Based on the patterns of clicks, it can also determine the popularity of a new product. Another sample firm massively collects social login data (e.g. size of friends on social networks, frequency of posting, number of likes received) of its on-line shoppers and combines with shopping history data. The retailer is able to successfully optimize its retargeting advertising on its website and mobile app while real-time pricing their products according to the utility value that the purchase may generate based on the preferences of the consumers.

Finally, our study has some limitations. Firstly, it does not account for the effect of market power imperative, as the competitive strategy perspective suggests. Future research should extend our study by taking retail industry structure into account. Secondly, although we used market capitalization as an output, our study does not consider a firm's short-term stock market performance. One way to extend our research is to study what is the stock market reaction to a firm's BDAS investment announcement. Third, our measure of BDAS capability is mainly based on self-reported data. We were unable to investigate the quality of each firm's actual usage of the system. In-depth case study or longitudinal field research should be carried out to shed more light on how firms employ BDAS in a greater detail. Last but not least, our research is mainly took a technology-centric view on BDAS. It would be necessary in future works to examine how organization should change to accommodate automated and advanced analytics technology in order to improve firm's performance.

#### REFERENCE

- Capgemini, "Big & Fast Data: The Rise of Insight-Driven Business," Capgemini, Paris: Capgemini, 2015.
- [2] S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study," Int. J. Prod. Econ., vol. 165, pp. 234–246, 2015.
- [3] Forbs, "Data Analytics Dominates Enterprises' Spending Plans For 2015," Retrieved: 2016.09.
- [4] D. W. Jorgenson, M. S. Ho, and K. J. Stiroh, "A retrospective look at the US productivity growth resurgence," J. Econ. Perspect., vol. 22, no. 1, pp. 3–24, 2008.
- [5] C. Q. Romero and D. R. Rodríguez, "E-commerce and efficiency at the firm level," Int. J. Prod. Econ., vol. 126, no. 2, pp. 299–305, 2010.
- [6] J. R. Galbraith, "Organization design: An information processing view," Interfaces, vol. 4, no. 3, pp. 28–36, 1974.
- [7] J. Barney, "Firm resources and sustained competitive advantage," J. Manage., vol. 17, no. 1, pp. 99–120, 1991.
- [8] C. Serrano-Cinca, Y. Fuertes-Callén, and Cecilio Mar-Molinero, "Measuring DEA efficiency in Internet companies," Decis. Support Syst., vol. 38, no. 4, pp. 557–573, 2005.
- [9] F. Tian and S. X. Xu, "How Do Enterprise Resource Planning Systems Affect Firm Risk? Post-Implementation Impact," Mis Q., vol. 39, no. 1, pp. 39–60, 2015.
- [10] Deloitte, "China Power of Retailing 2015," Retrieved: 2016.09.
- [11] Oracle, "An Architect's Guide to Big Data," Retrieved: 2016.09.



Figure 2: The Visualization of the Correlation between DEA, MPI and BDAS Capability

# TABLE I: FACTOR LOADING OF 30 DEA MODELS

### TABLE II: MPI OF 45 FIRMS

ON TWO PRINCIPAL COMPONENTS

Model	Component 1	Component 2
ABCDE1234	0.394	
ABCDE1	0.383	
ABCDE2	0.644	
ABCDE3	0.814	
ABCDE4	0.787	
A1234	0.769	
A1	0.336	
A2	0.805	
A3	0.914	
A4	0.821	
B1234		0.679
B1		0.689
B2		0.769
B3	0.644	
B4	0.752	
C1234	0.58	
C1		-0.696
C2	0.836	
C3	0.874	
C4	0.733	
D1234	0.871	
D1	0.868	
D2	0.868	
D3	0.894	
D4	0.763	
E1234	0.52	
E1		0.454
E2		0.609
E3	0.741	
E4	0.685	

Firm ID	RDAS(1=with)	MAINDIITST INDEX	TRCHNICAL CHANGE	REFICIENCY CHANGE
1	0	0.824	1.006	0.819
2	1	0.914	1.012	0.903
3	0	0.841	0.993	0.848
4	0	0.859	0.979	0.877
5	1	1.068	0.964	1.108
6	1	1.031	1.031	1
7	1	1.076	1.022	1.053
8	1	0.717	0.932	0.77
9	1	0.843	0.855	0.986
10	0	0.98	0.983	0.598
11	0	0.866	0.977	0.887
12	0	1.056	0.986	1.071
13	0	1.019	0.987	1.032
14	1	1.041	0.95	1.096
15	0	0.98	0.98	1
16	1	1.002	1.002	1
17	0	0.771	0.771	1
18	0	1.039	0.986	1.054
19	0	1.085	1.02	1.063
20	0	0.576	0.576	1
21	0	1.039	1.021	1.017
22	0	0.998	0.942	1.059
23	0	0.928	0.928	1
24	0	1.03	0.975	1.057
25	0	0.995	1.019	0.977
26	0	1.07	1.002	1.068
27	0	1.071	1.043	1.027
28	0	1.032	1.017	1.015
29	1	1.442	1.072	1.345
30	1	0.698	0.879	0.794
31	1	0.967	0.932	1.037
32	0	1.179	1.081	1.09
33	0	1.028	0.971	1.059
34	0	0.718	1.007	0.713
35	1	1.004	1.028	0.976
36	0	0.951	1.074	0.886
37	1	1.148	1.016	1.13
38	1	1.11	1.028	1.08
39	1	1.043	1.167	0.894
40	1	0.919	1.076	0.855
41	1	1.194	0.904	1.321
42	0	0.987	1.013	0.975
43	0	0.95	1.018	0.933
44	0	1.049	1.059	0.99
45	0	0.961	0.961	1