MASTER THESIS

The value effect of big data in the financial services field

Name:	Tom Luimes
Number:	S2417456
E-mail:	T.luimes@student.utwente.nl

Faculty:	Behavioural, Management and Social science
Master:	Business Administration
Track:	Financial Management

1st supervisor:	Prof. Dr. R. Gutsche
2nd supervisor:	J.R.O. Osterrieder

Date: 29-06-2021

UNIVERSITY OF TWENTE.

Acknowledgements

This thesis is written to finalize my master study Business administration with a specialization in financial management at the University of Twente. I would like to take this opportunity to thank a few people that supported me during this period. First of all, I would like to thank my first supervisor Prof. Dr. R. Gutsche of the department of finance and accounting at the University of Twente. His guidance and useful feedback during my master thesis has really helped me forward in the last couple of months. The feedback also contributed to the improvement of my writing- and statistical skills as well as my presentation skills. Second, I would like to thank my fellow students, together we were able to complete various projects and courses and learned a lot from each other. Last but not least, I would like to thank my family and friends for their unconditional support, not only the last few months but throughout my entire career as a student.

Tom Luimes

June 29, 2021

Abstract

This study investigates the value effect of big data in the financial services field for publicly listed firms operating in the United States. Big data is a relatively new technology that has become more and more popular in the last 10 years. In this study it is investigated how big data can add value, this is done by testing the impact on financial firm performance and market value. There is a lot of literature about firm performance and big data but few studies follow the same methodology and data collection method that is used in this study, this study therefore tries to contribute to the current literature. The practical contribution of this study is that it tries to provide new insights for firms that are thinking about implementing big data but are not convinced of the potential benefits yet. Additionally, this study provides new future research directions as well.

It is expected that early adopters of big data experience higher firm performance and market value because they have had more time to implement and exploit the use of big data, leading to more benefits compared to later adopters. However, the results in this study show that there is no significant difference in the so-called early and late adopters. On the contrary, the study finds significant results for firms that implement big data between 2010 and 2016, if average figures over 2019-2020 were taken whereas figures over 2017-2018 show less significant results. It seems that firms need more time to exploit big data benefits and increase firm performance. Overall, the results implicate that the usage of big data does not have a significant impact on financial firm performance. Additionally, out of the four financial metrics related to firm value, two show significant results in all the regression models. It seems that investors are willing to pay more for a certain stock if this firms announced that they work with big data. However, this higher firm value could be an effect of the firms position towards new innovations and implementing them. Therefore, future research is needed to validate the results of this study and investigate other factors that influence the effect of big data on firm performance.

Key Words*: Big data implementation (BDI), Decision-making, Value effect (VE), Firm performance, First movers advantage, US-listed firms, Financial services field.*

Table of contents

1 Introduction

"The businesses that do not understand or willing to react for these changes will fail miserably"

The quote stated above was addressed in the article of Kumaresan and Liberona (2018) and refers to the importance of changes in, among others, economic and technological changes and how companies react. Big data is a hot topic these days and the application is becoming more visible in all kinds of areas. More recent literature is emerging about the use of big data in the financial field; however, researchers do acknowledge that additional research needs to be done. As stated in the recent article by Huang, Wang, and Huang (2020), it remains unclear which benefits are experienced by firms when adopting big data. Also, the literature about the financial benefits of big data implementation is very limited. According to the survey that was conducted in the study of Huang et al. (2020), 43% of the respondents that plan to invest in big data do not know what the expected return will be. Researchers share the same opinion that big data brings benefits towards the company, but they are not sure how much and which benefits exactly can be achieved. According to Wamba et al. (2017), big data can be considered as a game changer that enables different improvements within the organization, because it has a high strategic and operational potential as well. It is also stated that successful companies are the ones who have developed a big data environment which leads to better decision making. This statement is to some extent confirmed by the study of Shamim, Zeng, Khan, and Zia (2020), they mention that big data driven decision making is not that easy because of the fact that data could be unstructured and not give the correct insights. In the report published by the McKinsey Global Institute[1] it is stated that the potential of big data remains uncaptured by firms. Challenges that firms face are categorized in three different factors, the conclusion that these factors state is that firms are unsure how to use big data and are cautious when it comes to investing in new technologies. Other firms find big data too complicated to start with and rather wait for a few years, according to Suoniemi, Meyer-Waarden, Munzel, Zablah and Straub (2020).

This all brings us to the main focus of this thesis; the investigation of the value effect of big data. In what ways can firms benefit from using big data and does big data actually have a significant impact on firm performance. Therefore, the goal of this thesis is to examine a hot current topic and add to the existing literature by using a different methodology compared to most of the studies about big data and firm performance. As mentioned earlier, researchers do acknowledge that it is not yet clear which benefits big data can bring to the financial services field and additionally in what time frame

[1] The age of analytics: Competing in a data driven world McKinsey (2016): https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.pdf

possible effects can be noticed. In this introduction, the background and context of the stated research question are given.

1.1 Background information

At first, the definition of big data should be clear to fully understand the stated research question. Big data is defined as follows: “Big data is the designation of structured and unstructured data of huge volumes" and “big data is said to be a socio-economic phenomenon associated with the emergence of technological capabilities to analyze huge amounts of data” (Bataev, 2018, p569). However, according to Elgendy and Elragal (2014), big data in itself is not yet valuable. Big data analytics, further BDA, translates data into useful insights, making big data valuable. Earlier studies, for example the research of Elgendy and Elragal (2014), defines big data based on three founding dimensions, namely Volume, Variety, and Velocity. Among others the researchers Rubin and Lukoianova (2014), and Hasan, Popp, and Oláh (2020) add a fourth dimension Veracity to the dimensions that explain big data. Recent articles published about big data even talk about five dimensions, the fifth dimension Value is added in the research of Vitari and Ragueseo (2019) and Shamim et al. (2020). These changes in different dimensions in a relatively short period of time show how constantly evolving this area still is and that researchers have not yet the exact same understanding.

Now that the definition of big data is defined, the following step would be how the underlying process should look like. Big data might not sound very complex, and it does not have to be if the right steps are followed and the process is correctly implemented in the whole organization. This process can be explained by means of the big data chain, the study of Janssen, Van der Voort, and Wahyudi (2017), points out that when big data is used a certain approach has to be followed. Additionally, Shamim et al. (2020) argue that it is not just about having access to big data and decision making based on this data, big data driven decision making follows a chain of activities. To make the implementation of big data within a firm a success, the whole process should be aligned. The study of Janssen et al. (2017) defined four main steps, which were also acknowledged in the study of Shamim et al. (2020). The four steps are the following: data collection, data preparing, data analyzing, and decision making. As displayed in the figure below, without the proper data collection methods, the decision making would rely on false data which is something that a company definitely needs to avoid. The first two steps are as, if not more, important than the last two steps. The implementation of big data in a company should go in good cooperation between IT and Finance to achieve the best results. This thesis focuses mainly on the last part, the decision-making process. Based on the current literature, it is expected that using big data leads to better decision making and eventually will result in higher firm performance.

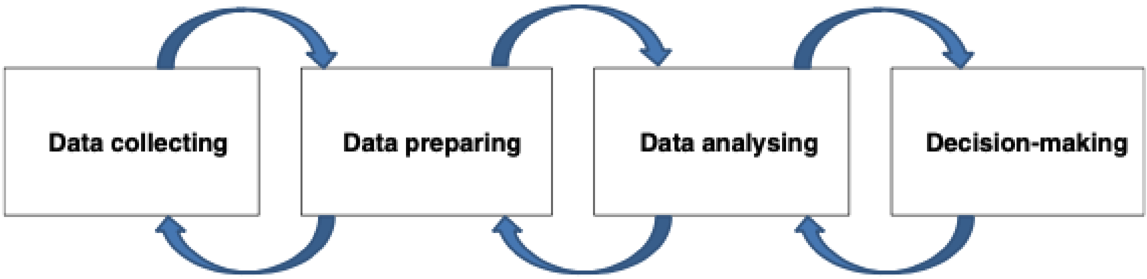

Figure 1: Steps and transfer points in the big data chain (Janssen et al., 2017)

1.2 Research objective and contribution

In the past years more researchers gave attention to the topic of big data and finance, this can be seen in the number of articles published in recent years about this specific topic, compared to earlier years. The study of Huang et al. (2020) states the conclusion that big data positively affects market value and firm performance, however, this study focuses on firms in all branches except IT. As mentioned by Suoniemi et al. (2020), the current available academic research is silent regarding to what extent big data investments have an impact on firm performance.

The research in this thesis is focused on the value effect of big data in the financial services field. The reason why this specific branch is chosen is because Bataev (2018) argues that financial institutions receive a huge amount of information and are therefore leaders in the field of big data. Additionally, according to Sun et al. (2020), big data has revolutionized the financial services industry, firms are moving more towards digitization while using big data because it strengthens the level of firm performance. It is expected that firms that do work with data, and firms that do not work with data, show a significant difference in firm performance, a better decision-making process, and higher market value. However, as said in the introduction of this thesis, the survey conducted in the study of Huang et al. (2020) shows that 43% of the firms do not know what expected return will be. This thesis therefore tries to show practical relevance by testing the hypothesis about significant differences in firm performance and show firms that are thinking about implementing big data what value it can add to the company.

The research questions and hypotheses formulated in this thesis are to some extent based on the research of Huang et al. (2020), their research focuses on the announcement of big data implementation and firm performance. However, in that particular research, the focus is on all branches except IT, it is admitted that their sample is limited and might not have covered all the aspects. In this thesis the focus is more pointed towards one specific branch, the financial services sector, because this field is not yet extensively investigated. Also, the current literature about big data and firm performance related to the financial services industry uses other methodologies and data collection, by using another approach this study adds to the current literature. Additionally, with more recent literature and data available, this thesis tries to contribute new insights to the current existing literature by identifying the value the implementation and use of big data can add to firms operating in the financial services field. This all leads to the following research question that is investigated in this study:

What is the effect of big data implementation on firm performance and firm value of firms operating in the financial services field listed in the United States?

1.3 Outline of the study

The structure of this paper is as follows; in the second chapter the literature review is presented, which consists of the examination of current literature, what firm performance is in this context and the formulation of the hypotheses. The third chapter examines the research methodology, variables and sample size, robustness checks and the sample size and data. The fourth chapter provides the results of the conducted assumptions regression, OLS regression analysis, robustness checks, and the hypotheses testing is presented. The fifth and last chapter five provides the conclusion, limitations and recommendations for future research.

2 Literature review

In this section, a comprehensive literature review is written on the relationship between big data and firm performance. This thesis will mainly focus on the impact of big data on financial firm performance and market value in the financial services field. First of all, the recent literature about big data and finance is examined to see if there are overarching topics or significant differences. Secondly, the role of big data in the financial markets is examined and where possible applied to this study. Thirdly, the value of information related to big data is presented, what effect does big data have on the value of the information within the organization. Furthermore, firm performance is examined based on the current literature, followed by the issues and disadvantages that should be taken into account if a firm chooses for big data implementation. Eventually, the hypothesis development is presented in the last section of this chapter.

2.1 Big data and the financial services field

There has been some interesting literature being published about the advantages of big data. During recent years more and more literature about the effect of big data on all kinds of firm's operations have been published, due to the fact that more firms are taking on big data and acknowledge the importance. In the research of Raguseo and Vitari (2018), it was already stated that big data could be listed as one of the top strategic technology trends with a huge impact for the next five years. In more recent literature, researchers do admit that big data can bring huge benefits to firms. Bataev (2018) and Yadegaridekordi et al. (2020) both state that customer service, operational efficiency, risk management, and legal requirements can be improved if big data is implemented. In addition, according to Hasan et al. (2020), big data technologies provide higher levels of automation which results in lower costs and increased productivity, which ultimately increases profit. In the study of Yadegaridkordi et al. (2020) it is even argued that big data adoption will lead to higher firm performance, under the condition that enough IT expertise is available within the organization to facilitate the big data adoption. Additionally, Huang et al. (2020) argue that big data analytics are positively related to business growth and that useful insights can be created while using big data analytics tools. There is a lot of literature being published about the effect of big data on finance, customer intelligence, operations and many other topics. A more upcoming topic in recent years is the effect of big data on the decision-making process, the study of Sun et al. (2020) tries to contribute to the existing literature and provide new evidence as well to investigate the relationship between the use of big data and improved decision making. Based on the studies outlined in the paragraph above it can be concluded that researchers acknowledge the importance of big data within firms.

The implementation and use of big data can play a huge role in the finance area. According to Wang (2020), big data in finance has become a hot research topic among researchers. Big data, cloud computing and other internet technologies were brought into the financial industry by some internet companies after 2010, from there on big data became more and more important (Wang, 2021). The article of Sun et al. (2020), states that big data can be seen as a key in the development of the financial sector and financial services. According to Hasan et al. (2020), external and alternative data is used by financial analysts to make better investment decisions these days. However, the study of Hasan et al. (2020) also outlines that the extensive view of big data in the financial services field is not done before with proper explanation of the opportunities and influence of big data on finance. Additionally, the research of Bataev (2018) points out that the implementation of big data technology in the financial sector would increase heavily over the upcoming years. The increase in big data use is also mentioned in the study of Sun, Shi, and Zhang (2019), it is stated that big data in finance is becoming one of the most promising areas in the financial sector. This would implicate that firms do acknowledge that the use of big data can be beneficial for their firms. Therefore, the study of Kumaresan and Liberona (2018) tries to understand if a data-driven business model will give financial firms an advantage compared to their competitors and Sun et al. (2019) add that big data can significantly change business models in financial services companies.

The above referenced literature aligns with the goal of this thesis, namely, to investigate what value big data can bring to firms that start working with big data. This is done via the hypothesis testing, to determine if there is a significant difference in firm performance. A more recent paper of Sun et al. (2020), argues that big data is relevant in many research fields but that it is particularly important in the finance area. They also add that finance professionals these days should possess IT skills to some extent to work with big data and other related topics because of the modern business that is constantly changing. Finance professionals themselves do also acknowledge that big data analysis is one of the most important aspects in the analysis of services and financial products (Sun et al. 2020). Finance based on big data has a lot of advantages according to Wang (2020), among others, more transparency, higher participation, lower intermediate costs, and better collaboration is achieved.

Another research of Wang (2021) states that traditional banks, which belong to the financial services area, also should make use of big data and try to keep up with the latest innovations. Another specific type of company within the financial services field are the accounting firms. According to Sun et al. (2020), finance and accounting big data go hand in hand and can be seen as a promising innovation in the accounting area, for example detection of fraud is easier to identify while using big data. Both banks and accounting firms are often categorized in the category of large firms, the study of Sun et al. (2020) addresses that it is interesting to see that companies, mainly in finance, consider big data analysis more and more important, and they acknowledge that it has to be developed. The study of Sun et al. (2020) also argues that the financial services and related sectors are transformed by the upcoming big data usage. Concluded, current literature about big data in the financial services field does acknowledge the importance and potential of this new innovation. However, it remains unclear what benefits can be exploited exactly and which time frames should be taken into account. This thesis tries to prove the benefits of big data while using a different methodology, especially the data collection method, compared to current literature.

2.2 Big data in financial markets

Financial markets are always looking for new innovations in technology that are accepted and have a significant impact on business and therefore will lead to optimism (Sun et al., 2020). The effectiveness of financial markets is determined by the amount of information available and the quality of the data. Several researchers point out that risk management, forecasting, and valuations of overall markets are improved via the use of big data, some examples are the automatization of trading, risk analysis that are easier performed, and investments can be made online without the interference of the banking industry (Sun et al., 2020). Additionally, the study of O'Halloran, Maskey, McAllister, Park, and Chen (2015) points out that big data is used for the regulation of financial markets, by using new data science techniques combined with big data sets it becomes easier to regulate. The second hypothesis is based on the expectation that announcements about new innovations within companies have a positive impact on the market value of a firm. As argued in the study of Begenau (2018), changes in stock prices of a firm are priced based on the available data. When there is more data available about a certain company, usually the risk reduces and therefore the compensation that investors require is less. When a firm implements big data it can also be used to provide more transparency towards potential investors. More insights due to the possibility of extensive data analysis can be shared with investors and other parties to eventually reduce the cost of capital (Hasan et al. 2020). Therefore, it can be stated that the main effect of the use of big data in financial markets is reducing the firm's average cost of capital. Furthermore, big data can provide investors with more transparent information. However, the results in the event study of Huang, Wang, and Tasi (2016) show that the announcement of big data implementation does not directly affects stock prices. Contrary to the event study, this study however focuses on long term performance and investigates if big data implementation has an effect on market value years after the implementation. Also this study takes different measures into account, such as Earnings Per Share and Tobin's Q, to test for significant results. Concluded, big data can lead to more transparency, lower cost of capital, and reducing risk, these topics can lead to investors willing to pay more for the stock of a certain company because higher profitability and lower risk would lead to higher valuation.

2.3 Value of information

One of the aspects that is mentioned by several researchers and eventually can result in higher firm performance is the improvement in the decision-making process. As mentioned in the study of Sun et al. (2020), information gathered from raw data is the basis for the process of decision making. The next step in this process is transforming this raw data into useful visualizations, with all the essential data visualized the management can make the best decisions for the organization. This decision-making process is based on the information gathered from big data, which can also be seen as its value. Additionally, the study of Kościelniak and Puto (2015) points out that big data is not only about the collection of big data, but more about processing and visualization which is essential for obtaining business benefits. They even state that the application of big data will results in real competitive advantage. The study of Elgendy and Elgaral (2014) defines three main areas when it comes to big data: storage and architecture, data analytics processing, and big data analysis. The decision-making process is based on the third area big data analysis, this aligns with the study of Sun et al. (2020). One important implication that was stated in the research of Janssen et al. (2017), is that deciding based on big data is not only about analyzing the big data that you have access to. A chain of activities goes before a big data driven decision making; data needs to be collected, prepared, visualized, and then the analyzed so that big data can help improve the decision making. This study mainly focuses on the last step, the impact of better decision making due to the availability of more important information.

Therefore, to combine all these aspects together, a framework is proposed. The framework brings aspects together that explain how big data can add value and improve firm performance. In the study of Mikalef, Boura, Lekakos, and Krogstie (2019) a framework is developed, which consists of one dependent variable, namely firm performance, and six independent variables that have an impact on firm performance.

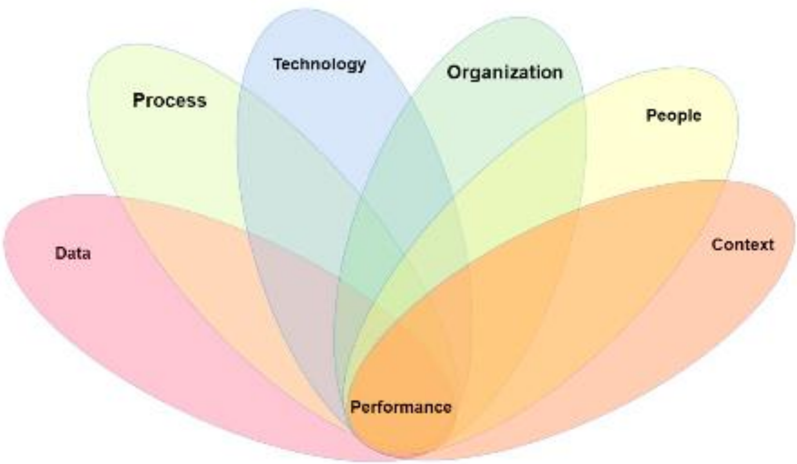

Figure 2: Research framework (Mikalef et al., 2019)

The above framework is the basis for the framework presented in this study as well. There are different aspects that have an impact on firm performance, besides data there are five other aspects as well presented in the framework. This shows that only the access to big data does not instantly result in higher firm performance. The combination of big data availability, the corresponding technology with clear processes, and people that can work and understand big data contribute to the value of big data as well. According to Anfer and Wamba (2019), big data can create new opportunities for firms, and they can improve their business based on new insights and more information that becomes available. An example is that companies use big data analytics to understand customers behavior and improve suggestions.

Additionally, Chen and Lin (2021) argue that converting big data into useful information can increase knowledge about future opportunities and threats and even provide intelligent solutions when it comes to corporate decision making. Another important advantage of using big data is that firms can respond very quickly to new trends. Data is almost instantly available and firms that have faster access to important data can gain competitive advantage (Chen & Lin, 2021). The study of Dong and Yang (2020) add that raw data from for example social media can be used to get more insights about potential customers and how future marketing campaigns can reach their highest potential and lead to more sales and increased revenue. To conclude, based on the current studies it can be stated that several factors that have an influence on firm performance can be improved by using big data. The basis of this study is summarized in a framework and can be visualized as follows:

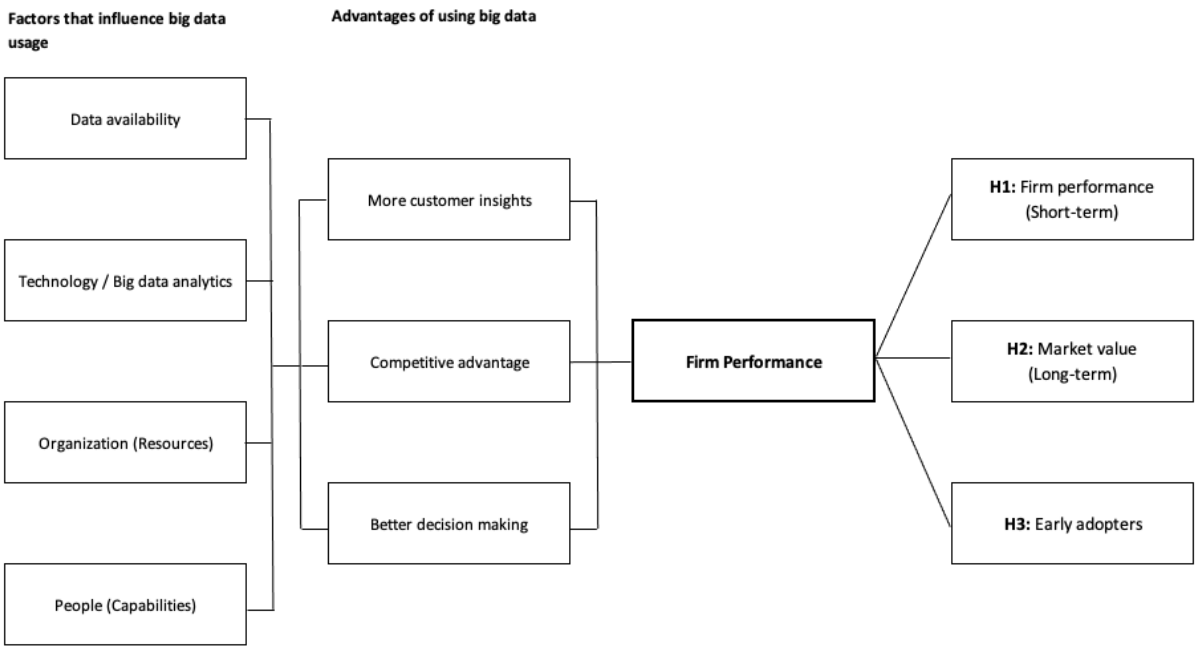

Figure 3: Summarized framework information value big data

2.4 Firm performance

The value effect of big data in this study is measured by testing the impact of this variable on firm performance. The term firm performance should therefore be clarified and described, to make sure that the definition and the eventual effect is understood. According to the study of Santos and Brito (2012), firm performance can be described as follows: “A subset of organizational effectiveness that covers operational and financial outcomes”. Firm performance in this definition is split into two categories, financial and operational performance, in this study the focus is pointed towards the financial performance. Other important aspects that should be considered while using firm performance as a dependent variable, is time frame and the reference point. If new innovations are brought into the company, it takes some time until these effects can be seen in changing firm performance. According to the study of Chakravarthy (1986, as cited in Santos and Brito, 2012), superior financial performance is something that satisfies investors. This financial performance can be divided into market value, financial firm performance. These two categories will be transformed into different financial metrics which will be used during this study in the regression models.

The study of Su et al. (2021) investigates the effect of big data on organizational performance. The results of the study show that big data analytics capabilities have a positive and significant effect on organizational performance. Additionally, the relationship between innovations and organizational performance has become closer than ever (Guo et al. 2017; as cited in Su et al. 2021). Also, innovation is a key factor for firms to obtain a competitive advantage and stay ahead of the competition (Su et al. 2020). The study of Li, Dai, and Cui (2020) mentions that the application of big data and analytics is closely related to enable firms to increase better decision making. Big data can increase efficiency and therefore reduce costs and increase profits. However, the study of Aktar, Wamba, Gunasekaran, Dubey, and Childe (2016) mentions that big data does not pay off for all companies, it appears that only a few companies really benefit from big data advantages and higher firm performance. Additionally, a study conducted in 2014 pointed out that firms that do not adopt big data are expected to experience a decline in market share and momentum[2].

Concluded, current literature agrees that the use of big data can have a significant impact on firm performance. This study focuses on the financial firm performance to measure the impact of big data. In the methodology section more explanation is given how the impact of big data on financial firm performance is measured.

[2] Columbus, L., 2014a. 84% Of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes in the Next Year, Forbes.

2.5 Issues and disadvantages of big data

The paragraphs above have been very positive about the use of big data, however, there are also some limitations when it comes to the use of big data which should be taken into account. According to Sun et al. (2020), the rate of innovation that is brought about by information technology is the main issue in the finance area. This is also mentioned in the study of Hasan et al. (2020), namely, management of big data is the most important factor and the most important issue. This is confirmed by the study of Bataev (2018), they also mention that data protection and confidentiality are important aspects and that qualified personnel is required to fully reach the potential of big data tools. Another aspect that contributes to the issues emerging within companies, is that the transformation in the area of big data is so rapid that financial institutions might not be able to keep up with it (Sun et al., 2020). Additionally, an issue that is also seen a lot, and addressed in the research of Sun et al. (2020), is that when it comes to start working with big data that there is a dissimilarity between the languages of finance and IT, both parties have different interests and do not have the same end goal in mind. As said earlier, huge firms often change faster towards the new innovations such as big data, this is partly because of their available resources and motivation to stay one step ahead of the competition. The downside is that huge firms often are less flexible and that it can take more time to implement new technologies in the organization, whereas smaller companies have the advantage that they can implement faster and change their way of working.

When looking at firm performance in terms of profitability, it should be considered that the implementation of big data usage comes with costs as well. If a company decides to start working with big data, the infrastructure within the organization must be updated in most of the cases. These changes in the organization take time, and a lot of effort from the people working in the firm which can be expensive in some cases. The study of Wang (2020) addresses that many new problems come up when a company decides to actively use big data, for example the processing ability of the current software and the data management. Additionally, Elgendy and Elragal (2014) mention that you want your big data set as big as possible because then you are sure it contains all the information you need. However, the larger the set of data the more difficult it will be to manage, store and secure this data which will come at a higher cost as well. It is therefore not easy for companies to find the perfect balance between the amount of data stored and the management. Companies tend to store as much data as possible to achieve the best representation of reality and to make sure all data is available if needed, however managing this amount of data comes at a cost. Finding the perfect balance is not easy and is often a process of just starting somewhere with data storage and learning from time-to-time which methods fit best.

Concluded, the study of Oussous, Benjelloun, Ait Lahcen, and Belfkih (2018) provides a clear overview of the challenges that can occur while using big data, consisting of: Data capture, searching, storage, analysis, visualization and management. Additionally there are security and privacy issues that might occur in distributed data driven applications, which can be shared internal and external. Most of the challenges related to big data are technical, which is why firms should have enough technical expertise available. Overall, implementing big data can bring benefits to firms, however the disadvantages should be managed correctly, and the risks and challenges should be covered to make sure big data is beneficial and can create competitive advantage.

2.6 Hypothesis development

In this section the hypothesis development is examined. So far, the existing literature regarding big data use in the financial field has been described, this leads to the conclusion that the relationship between the use of big data in the financial services field is not yet extensively investigated. Therefore, three hypotheses are formulated to test the effect of the implementation of big data on firm performance and firm value of companies operating in the financial services field operating in the United States. As mentioned in the study of Ragueso and Vitari (2018), firm performance includes both financial and market performance. Financial performance mainly is about profitability, revenue growth, ROE and other financial figures. Market performance is about the position a firm has compared to their competitors and if this position becomes stronger by using big data. This is measured by variables such as Earnings per share (EPS), Price to book ratio (P/B), Price to earnings ratio (P/E), and Tobin's Q. Overall, based on the study of Brynjolfson, Hitt, and Kim (2011), it is expected that data driven decision making has a positive impact on firm performance, their study shows significant results for return on equity and market value. In the following sections the stated hypotheses are more extensively examined.

2.6.1 Financial firm performance

With the knowledge that large financial firms often have more big data available, the collected data is taken from listed companies operating in the financial services field, which are often relatively large. This is supported by Begenau, Farboodi, and Veldkamp (2018), they state that big firms produce more data because of their extended economic activity and longer firm history. Based on announcements in the newspapers about big data implementation within firms, firm performance is measured to see if a significant difference can be determined. It is expected that firms that work with big data experience higher financial performance. According to Hasan et al. (2020), big data can reduce equity uncertainty and reduce firms' cost of capital. Also, Begenau et al. (2018) stated that "more data processing lowers uncertainty, which reduces risk premia and the cost of capital, making investments more attractive". Additionally, the availability of big data helps firms analyze their risk, Hasan et al. (2020) therefore state that risk management can be improved resulting in higher profitability. Financial figures are therefore expected to be improved because the cost of capital is lower, and more information is available to make the right decisions. Financial firm performance is focused on the short term improvements in for example ROE, ROA and Profit margin. The first hypothesis that is tested to investigate if big data has a significant impact on firm performance is as follows:

H1: *Firms that implement big data are associated with higher short-term financial firm performance than firms that do not implement big data*

2.6.2 Market value

The impact of big data on firm value and the stock market should continue to be explored, according to Hasan et al. (2020). Firm value can be determined in many ways, the way to determine firms' value is in this thesis based on the stock market value. The first hypothesis was about financial firm performance, which is tested by comparing figures of annual reports and therefore more focused on the short term. The second hypothesis is more focused on long term firm performance and value, according to the study of Huang et al. (2020), market value can be seen as a measurement to see a company's long-term performance. Firms that do work with big data analytics enhance the organizations information processing capability, which brings competitive advantages compared to other firms (Chen et al., as cited in Ragueso & Vitari 2018). Their results additionally show that business growth can be increased by using big data analytics. It is expected that firms that announce that they will implement big data have higher market values compared to firms that do not announce they work with big data. Stocks of firms that announce they start implementing big data are expected to be more attractive to investors because the implementation of big data can bring huge benefits to the firms as mentioned before. Additionally, The study of Gunday, Ulusoy, Kilic, and Alpkan (2011) shows that investment in digital systems can have a positive effect on the obtained market share, intensively using big data can thus lead to increased market share which can be the basis for better firm performance and optimistic investors about the long term perspectives resulting in higher market values. All this combined brings us to the second hypothesis that is tested:

H2: *Firms that announce they work with big data are associated with higher long-term firm value*

2.6.3 Early Adopters

The first two stated hypotheses do not consider the potential advantages and risks of early big data adopters. Most of the time, firms that detect and implement new trends first are expected to achieve higher benefits. The study of Huang et al. (2020) takes into account the so called first movers advantage. Applied to this research, it is expected that firms who belong to the group of early adopters of big data in their organization are associated with higher firm performance compared to later adopters. These early adopters already had more time to benefit from data analytics and improve their approach towards data collection, storage and visualization. Also, early adopters might be able to gain superior performance compared to competitors and obtain valuable propositions. According to the study of

Huang et al. (2020) big data started to take off after 2013, early adopters are the companies that started with big data between 2010 and 2013. Therefore, the third hypothesis that is tested in this research is the following:

***H3**: Early adopters of big data are expected to achieve higher firm performance because of first movers advantage*

3 Research method

3.1 Methodology

At first, the different research methods used in other studies related to big data and firm performance or market value are examined. Different research methods are used in the published studies, the most common methods are further examined. The prior study from Huang et al. (2020) that examined a similar research question used the OLS regression model as a research model. The study of Vitari and Raguseo (2020) used the two-stage least-squares regression model, however, they first conducted a confirmatory factor analysis to verify the appropriate properties of the measures used in their study compared to the other studies. There are a lot of studies that make use of the Partial least squares – Structural equation modelling (PLS-SEM) method (Kumaresan & Liberona, 2018; Wamba et al. 2017; Yadegaridehkordi et al., 2020; Shamim et al., 2020). However, the studies that make use of the PLS-SEM method are collecting data via a questionnaire or interviews, which is different compared to this study.

3.1.1 Regression analysis

The regression analysis is the method that is most used in the studies that investigate the effect of big data on firm performance based on quantitative data. The regression analysis is a dependence technique where a single dependent variable is predicted by one of more independent variables. Within the regressions analysis there are some different methods that are used, among others the ordinary least square regression, panel regression, and two-stage least squares regression models. These models will each be further discussed in the following sections.

3.1.1.1 Ordinary least square regression

The ordinary least square (OLS) regression model is a statistical model that estimates between independent variables and a dependent variable. Most of the researchers address that the OLS regression method is one of the best statistical methods that can be used (Souza & Junqueira, 2005). Dismuke and Lindrooth (2005) argue that the OLS regression method is one of the most common techniques used when it comes to multivariate analysis. However, they also mention that it might be the most misused technique in research. OLS is a useful method when parameters are not known and the relationship between independent variables and the dependent variable is hypothesized and needs to be tested. The OLS technique can be used to model a response of the dependent variable based on the independent variables (Craven & Islam, 2011). While using the OLS method there are some assumptions which should be taken into account. The OLS method requires several assumptions related to the model and residuals, such as normality, independency, and homoscedasticity. Additionally, outliers are something that should be considered when conducting the OLS method according to Souza and Junqueira (2005), because this model is very sensitive to the presence of outliers. Therefore, during the data collection process significant outliers that were detected were removed from the sample to avoid this problem.

3.1.1.2 Two-stage least squares regression (2SLS)

The two-stage least squares (2SLS) regression technique is the extension of the OLS method. This method is used in situations with an endogeneity problem. The study of Vitari and Raguseo (2019) uses the 2SLS regression to address the concern of reverse causality and endogeneity related to IT investments. The benefit of using the 2SLS method is that it can assess the relationship between big data implementation and firm performance in both directions. The causality between firm performance and big data implementation can go both ways, it might be that firms that experience superior firm performance are more inclined to implement big data, also because they have the financial resources. While the other way around big data implementation can lead to higher firm performance because of benefits earlier mentioned in this study. The study of Benitez, Henseler, Castillo, and Schuberth (2020) points out that the endogeneity problem can be addressed by using the 2SLS model, they mention that path coefficients used in the OLS regression may suffer from omitted variable bias, the 2SLS can solve this. However, the 2SLS method has some disadvantages as well, compared to the normal OLS regression method the results are more difficult to interpret. Additionally, more skills are needed to conduct this analysis, for example the number of instrument variables should not be too small or huge. This requires a certain skill and is often more time-consuming than conducting a normal OLS regression.

3.1.1.3 Panel regression

The panel regression analysis is used when the same individual data are collected over a time period. It is also known as longitudinal or cross-sectional time-series data. Panel regressions can be divided in two main types, the fixed and random effects. The study of Dong and Yang (2020) uses panel data to avoid repeated measures. The panel regression model allows control for variables that cannot be observed or measured. The use of panel data comes with several benefits. It gives more informative data, more degrees of freedom, more efficiency, less collinearity among variables and more variability. Additionally, panel data controls for individual heterogeneity and is better suited for identifying effects that are not detectable in pure time-series data (Hsia, & Klevmarken, as cited in Baltagi, 2005). There are two types of panel regression that can be used, namely the fixed and random effects. The fixed effects model can be further divided into one-way and two-way models, where one-way is about cross section data or time-series data and two-way is about both. In this study, data is obtained for companies operating in the financial services field over several years after the announcements related to big data. A panel regression could therefore be an option to apply in this study, in the following sections the method used in this study will be further examined.

3.1.2 Method used in this study

Based on previous studies and the stated research question the chosen research method is the ordinary least square (OLS) regression model. The 2SLS method could be used as well, but the OLS method is easier to use and the endogeneity issues that can occur can be avoided by using lagged variables. Additionally, the results of the OLS method are easier to interpret compared to the 2SLS method (Shepherd, 2010). The panel regression can be used as well in this study because data over several years was collected. However, because other studies which use firm performance as dependent variable make use of the OLS regression instead of panel regression. Therefore, the collected data over different years is summed up together and dividend over the years so that the average figures are used in this study. The OLS regression method is in line with the models used in other empirical research regarding the use of big data and firm performance, these studies used lagged variables as well to control for any endogeneity issues. Additionally, the ordinary least square research model has the advantage of being more flexible in the relationship between numerator and denominator and used in the study that follows the same approach (Huang et al., 2020).

3.1.3 Assumptions regression

A few assumptions should be fulfilled regarding the OLS regression method. The first is that there should be no multicollinearity between the independent variables. There are two main methods that can be used to check for multicollinearity (Hair et al. 2010). The first is looking at the correlation coefficients, which can be found in Table 6. Correlations higher than .9 can indicate that multicollinearity might be an issue. This is found for two variables, namely profit margin and Ln_FSIZE, because this only indicates multicollinearity, another method was used to confirm or deny these findings. The second method is looking at the VIF values, whereas VIF values above 10 indicate multicollinearity problems. The VIF values for the independent variables can be found in Appendix D. When the assumptions are fulfilled, an appropriate method to use in this study is the OLS regression. If it might be the case that not all the assumptions are fulfilled then adjustments to the data will be made to make sure that the assumptions are met, such as deleting outliers and not using all variables in the model. Preferably the VIF values should be below 5, in Appendix D it can be seen that all the VIF values for the independent variables are below 5. To conclude, the assumptions regarding the OLS regression method and corresponding independent variables are met and the data can thus be used in the study.

3.1.4 Endogeneity problem

A key issue that could occur in this research is the endogeneity problem, which is about the probability of reversed causality. The endogeneity problem is a common issue in finance studies and therefore should be taken into account, especially for this study because the OLS method is used which does not take into account endogeneity (Wintoki, Link, & Netter 2012). Applied to this research, it could be that firm performance is not affected by the implementation of big data but that firms that perform well, easier implement big data because these firms could afford the relatively high costs. Other factors besides the implementation of big data can have a huge impact on firm performance as well. This endogeneity problem should therefore be taken into account because it could affect and limit the outcomes of this study, according to Schultz, Tan, and Walsh (2010) the presence of endogeneity will produce biased parameter estimates, which should be avoided. Additionally, another endogeneity problem that can occur is that higher firm performance leads to higher market value. So, it does not necessarily have to be that the effects seen in the results are because of big data adoption. To rule out any endogeneity issues the study uses leading variables, this means that the dependent variable firm performance is measured in year t + 1, while the other independent variables are measured as year t.

3.2 Research model

As argued before, the research method aligns with the study conducted by Huang et al. (2020), The findings of that study suggest that big data implementation can affect financial performance and increase market value. However, in this thesis the focus is pointed towards the financial services firms instead of all firms. Huang et al. (2020) also acknowledge that their dataset is limited, by being more specific and adding more recent literature and news articles the dataset used in this thesis should be able to provide new insights.

3.2.1 Firm performance

In order to test the first hypothesis that was stated during the hypothesis development a research model is created to estimate the effect of big data implementation on firm performance. Firm performance is measured by including the dummy variable for big data adoption, the control variable and the standard error term. The following model will be used to test the first hypotheses and eventually the research question:

$$\boxed{\textit{H1: Formula: Firm performance} = \beta_0 + \beta_1(\textit{BDI adoption}) + \beta_2(\textit{Control variables}) + E}$$

Where:

Firm performance = the financial performance of the firm, measured in year t + 1.

β_0 = constant, represents expected firm performance value if all other independent variables are zero.

β_1 *(BDI adoption)* = a dummy variable for the implementation of big data within a company, measured in year t.

β_2 *(Control variables)* = control variables that are expected to have a relationship with financial firm performance, such as leverage (LEV), asset turnover (ATO) and firm size (Ln_FSIZE), measured in year t.

E = random error term (has a mean of zero)

3.2.2 Firm value

The second formula is based on the second hypothesis and tests the effect of big data implementation on market value. The metrics EPS, P/E ratio, P/B ratio, and Tobin's Q are used to express firm value. The other variables are used in the same way as the formula for the first hypothesis.

H2: Formula: Firm market value = β_0 + β_1*(BDI adoption)* + β_2*(Control variables)* + *E*

Where:

Firm market value = the (market) value of the firm, measured in year t + 1.

β_0 = constant, represents expected firm performance value if all other independent variables are zero.

β_1 *(BDI adoption)* = a dummy variable for the implementation of big data within a company, measured in year t.

β_2 *(Control variables)* = control variables that are expected to have a relationship with financial firm performance, such as leverage (LEV), asset turnover (ATO) and firm size (Ln_FSIZE), measured in year t.

E = random error term (has a mean of zero)

3.2.3 Early adopters

The last formula focuses on the third hypothesis, which is about the effect of big data implementation on early adopters. Do firms that implemented big data between 2010 and 2013 experience higher firm performance compared to late adopters. This is tested by including a dummy variable for early adopters in the formula. The other variables are equal to the two previous formulas.

H3: Formula: Firm performance & value = β_0 + β_1*(Early adopter)* + β_2*(Control variables)* + *E*

Where:

Firm performance & value = Firm performance and firm value

β_0 = constant, represents expected firm performance value if all other independent variables are zero.

β_1 *(Early adopter)* = a dummy variable; with the value 1 if company implements big data between 2010 and 2013, 0 if otherwise. Measured in year t.

β2 (Control variables) = control variables that are expected to have a relationship with financial firm performance, such as leverage (LEV), asset turnover (ATO) and firm size (Ln_FSIZE), measured in year t.

E = random error term (has a mean of zero)

3.3 Data and sample size

3.3.1 Data collection

The data used to test the hypotheses related to firm performance is retrieved from ORBIS for publicly listed companies operating in the financial services fields in the United States. Examples of these companies are banks, insurances, financial consultancy, and accounting firms. The total list extracted from ORBIS contained 515 firms operating in the financial services field. After deleting firms with unknown values for the financial metrics used in the regression analysis, the total list came down to 208 firms with useful data. Table 1 provides a more detailed insight in how the data collection was conducted.

Table 1 Data collection ORBIS

Step	Amount	Search description	Extracted -/-
1	3.024.945	All active companies	2.737.734
2	287.211	United states of America	270.405
3	16.806	Publicly listed companies	13.804
4	3.002	Number of employees > 100	2.487
5	515	Finance and insurance, Monetary authories-central bank, Commercial banking, Financial transactions processing, Securities, commodity contracts, and other financial activities, Investment banking, Other financial investment activities,	307
6	208	Manually removing firms that did not have all essential values available.	

Eventually, the list consisting of 208 firms was used to search for announcements related to big data implementation. These announcements about big data implementation are used to test the hypothesis if there is a significant difference between firms that work with big data and firms that do not. Via keywords, news articles or press releases that implicate that firms start working with big data were searched for. The searching method is based on the study of Huang et al. (2020) and was conducted as follows: Key words in combination with the company names retrieved from ORBIS were used to search through Google search, pages were scanned for articles as long as 1 article on the page was about the corresponding firm. If a page did not show any articles or newspapers about the firm anymore the search was stopped, resulting in a value "not found" for that specific company. Also, a filter was set on the google search to only show results for the period 1-1-2010 till 31-12-2016. If announcements or articles are found they are scanned manually to determine if it is about big data before including in the sample. The key words that were used can be found in Appendix A.

After all the 208 firms were searched for, the sample of 208 firms is divided into two groups, one group consists of firms that do announce they work with big data between 2010 and 2016, and one group that does not announce or imply that they work with big data, or at least this is not publicly known. Additionally, the group with firms that work with big data is split into two different sections. The first group is labeled as the early adopters, these are firms that announced or implied they use big data in very early stages. If announcements are found in the years 2010 till 2013, firms are concluded in the first group. The second group consists of firms announcing big data initiatives between 2014 and 2016. Eventually, to test the hypotheses about market value, data related to financial firm performance and market value is retrieved from ORBIS.

3.3.2 Sample size and period

As mentioned in the previous section, the total sample size contains 208 firms. After following the procedure described in the section data collection, the sample distribution can be found in Table 2.

Table 2: Sample distribution

Full detail

Year	Frequency	Percent	Cumulative
2010	4	2%	2%
2011	4	2%	4%
2012	8	4%	8%
2013	14	7%	14%
2014	17	8%	23%
2015	32	15%	38%
2016	21	10%	48%
Not found	108	52%	100%
Total	208	100%	0%

Simplified version

Classification	Frequency	Percent
Non-adopters	108	52%
Adopters	100	48%
(Early adopters)	*(30)*	*(14%)*
(Late adopters)	*(70)*	*(34%)*
Total	208	100%

First of all, Table 2 shows that the total sample size was 208 firms. 108 firms with the value "not found" did not show any article or announcement about big data in the searching period, this group is called the non-adopters. Furthermore, the simplified table on the right in Table 2 shows that the group of so-called early adopters consists of 30 firms and the late adopters group consists of 70 firms, resulting in a total size of 100 for the big data adopters' group. For each of these 100 individual firms a big data implementation announcement in papers or news articles from 2010 till 2016 was found. The period between 2017 and 2020 was not searched for because this allows us to determine if a firm experiences higher performance due to the implementation. Therefore, figures over the years 2017 till 2020 are used to determine firm performance. After some doubting the figures over 2020 were taken into account as well. Even though the Covid-19 crisis had a huge impact on short term performance and value, over the whole year 2020 it does not have a significant impact if looked at financial metrics and stock prices.

3.4 Measurement of variables

3.4.1 Dependent variables

There are several dependent variables in this study, related to firm performance or market value. In Table 3 financial metrics are defined and based on these metrics the dependent variables are included in the regression model. Firm performance can be measured in different ways, the two most occurring methods are accounting based valuation and market-based valuation. The accounting method focuses on firm figures such as return on assets (ROA) and return on equity (ROE), the market-based valuation is more focused on earnings per share (EPS) and price to book value (P/B). The first hypothesis that was stated is based on the accounting method and therefore more focused on the short-term results. The second hypothesis focuses on market value and is based on the market-based valuation, this method is more focused on the long term because market valuation represents future expectations according to investors.

The dependent variables related to financial firm performance, are Profit margin, ROA, ROE, Debt/equity and Asset turnover (Huang et al. 2020). For the second hypothesis other independent variables are added to test the increase in firm value, the variables are; Price to earnings ratio (P/E), Earnings per share (EPS), Price to book value (P/B) and Tobin's Q. Tobin's Q is calculated by dividing the firm's market capitalization by its total assets[3]. By focusing on the difference in measures related to stock prices after the announcement of big data implementation, it is examined if the announcements have an effect of long-term firm value. The third hypothesis follows in the basis the same formula as the first two hypotheses combined; a divergent variable is the dummy variable for early big data adopters; if a company is considered an early adopter it gets the value 1, if not then 0.

Appendix C provides a more detailed overview of all the variables used in the study. Additionally, based on the study of Huang et al. (2020), Table 3 presents an overview of the variables that are linked to financial performance and market value. However, instead of focusing on operations, metrics for market value are added to the model. The dimensions Profitability and Capital structure are related to firm performance, while the dimensions Market value and Market based are related to firm

[3] Calculate Tobin's Q: https://corporatefinanceinstitute.com/resources/knowledge/valuation/q-ratio/

value. These key financial metrics will be used while testing the hypotheses related to firm performance and market value.

Table 3: Financial metrics

Dimensions	Ratio
Profitability	
	Profit margin
	ROE
	ROA
Capital structure	
	Debt/equity
	Asset turnover
Market value	
	EPS
	P/E ratio
	P/B ratio
Market based	
	Tobin's Q

3.4.2 Independent variables

In this study, two measures will be used, resulting in two types of independent variables, namely the Big data implementation variable (BDI) and the Early big data adopters (EBDI). This dummy variable has the value of 1 if there is a sign of big data implementation in the newspaper for a specific company, otherwise the value will be 0. The OLS regression models are conducted two times, the first time with the independent variables included, and the second time without, to see if there is a significant difference in the adjusted R squared.

3.4.3 Control variables

Control variables were used in the study, because according to Nielsen and Nielsen (2013), some common variables can influence firm performance. These variables could affect the results in the model and should therefore be controlled for. In contrast to the research of Huang et al. (2020), this study does not need to control for the industry effect because the whole sample is operating in the same industry. However, a control variable for firm size was included in the model, because according to the literature, the size of the firm can influence firm performance and market value. Current literature finds that firms with a larger size, on average are associated with higher firm performance (Lee, 2009). To control for firm size, the control variable Ln_FSIZE is created as natural logarithm of the total assets at the end of the year. To minimize the effects of outliers the natural logarithm is used.

Additionally, the different models also control for years to rule out any time specific results, as earlier mentioned, this study uses the average figures over a time span of three years to make sure that the results are representable. Instead of using a control variable for each year the average figures over a time period of four years is taken for both dependent and independent variables. Other control variables that are added to the regression are Asset turnover (ATO) and Leverage (LEV). According to the literature, leverage has an significant impact on firm performance and should therefore be taken into the regression as control variable (Ibhagui & Olokoyo, 2018; Bae, Kim, & Oh, 2017). Furthermore, Asset turnover is used as control variable, because the study of Nurlaela, Mursito, Kustiyah, Istiqomah, and Hartano (2019) finds that asset turnover has a significant impact on financial firm performance. To conclude, in Table 4 an overview is given with all the variables and their definitions used in this study.

Table 4: Variable definitions

Variables	Code	Description
Dependent variables		
Profit margin	PM	Pretax income / revenue
ROE	ROE	Net income / Total shareholders equity
ROA	ROA	Net income / Book value assets
Debt/equity	D/E	Long term debt / Book value common equity
Asset turnover	ATO	Total revenue / Total assets
EPS	EPS	Earnings per share
P/E ratio	P/E	Price to earnings ratio year end
P/B ratio	P/B	Price to book ratio year end
Tobin's Q	TQ	(Book value debt + Market value common equity) / Book value total assets
Independent variables		
BDI	BDI	Dummy variable for big data implementation
EBDA	EBDA	Dummy variable for early big data adopters
Control variables		
Profit margin	PM	Pretax income / revenue
Asset turnover	ATO	Total revenue / Total assets
Leverage	LEV	Leverage measured by dividing long term debt by book value of total assets
Firm size	Ln_FSIZE	Natural logarithm of total assets at the end of the year

3.5 Robustness checks

Robustness checks are conducted to analyze the uncertainty of the models used in the analysis. Additionally, it tests whether effects are sensitive to changes in the model specifications, to make sure that the conclusions that are made hold under different assumptions. a few robustness checks are conducted to make sure that the conclusions that were made hold under different assumptions.

3.5.1 Split sample

The first robustness test tries to examine whether the outcomes are the same if the sample is split in two different groups. Applied to this study, the initial sample of average figures over 2017 - 2020 is divided into two groups, namely 2017-2018 and 2019-2020. This is to make sure that the results hold if the sample is extracted over a different time period. Appendix E shows the results for this robustness test, overall, the results of the conducted split sample test are in line with the general results, in chapter 4 the results are further examined and discussed. By testing the results with a split sample, the possibility of 1 outlier year is avoided, it could be that 1 year, for example 2020 has a huge impact on the average figures. Also, with the split sample robustness test, it can be investigated if the effects of big data are only seen a few years after the implementation, it might be that a longer time period is needed to see significant changes in firm performance.

3.5.2 Alternative measures

The second robustness test that is conducted is the alternative measures method. Firm performance can be measured via different variables. By explaining firm performance by more than one variable, the results can be validated. In Table 3 it can be seen that firm performance is measured with different variables. Profit margin, ROE, and ROA are used to determine firm profitability. Using alternative measures such as ROE instead of ROA is in line with the study of Raithatha and Komera (2016). If only one of the variables shows significant results for the relationship between big data implementation and firm performance, it is not considered a significant relationship. However, if four out of the five variables show significant results, we can conclude that there is a significant relationship. The results for these analyses can be found in Table 7 and Table 8 in the results section.

3.5.3 Lagged variables

As earlier mentioned, endogeneity problems might occur when using the OLS regression method. This study makes use of leading variables to avoid any endogeneity issues. The dependent variable is measured in year t +1 and the other independent and control variables are measured as year t. Applied to the regression results in Table 7,8 and 9, the independent variables are the average over the years 2017-2019, and the dependent variables are the average figures over 2018-2020. Additionally, the robustness checks related to the split sample also make use of the lagged variables. The independent variables in split sample 1 are calculated over 2017 whereas the dependent variables uses variables over 2018. The same applies to the second robustness check, where the independent variables are taken over 2019 and the dependent variables over 2020.

4 Results

4.1 Descriptive statistics

First of all, the descriptive statistics are given, Table 5 presents the descriptive statistics of the variables. The values of the variables are the average of the years 2017 until 2020. Table 5 is split into three different panels, where Panel A provides the descriptive statistics of the whole sample. Panel B focuses on the difference between big data adopters and non-adopters. Panel C is the same as B, but focuses on the difference between early big data adopters and later big data adopters.

A few conclusions can be made based on the descriptive statistics displayed in Table 5. For almost all the firms concluded in the sample the variables are known, only 1 value for the debt to equity ratio was excluded because this was an extreme outlier. Panel A in Table 5 shows that the mean number of employees is 21908 while the median is 3596, this means that the distribution of the number of employees is skewed to the left. The first thing that stands out when looking at Panel B in Table 5, is that the difference in number of employees is relatively high between the different groups. The group of big data adopters has a mean of 21908 employees while the non-adopters group has a mean of 6129 employees. This is also seen in the control variable that is used in the regression model related to firm size, the Ln_FSIZE variable reports a mean value of 8,68 for the adopters group, and a mean value of 5,78 for the non-adopters group.

Table 5 Descriptive statistics

Panel A - Full sample

	N	Mean	Median	Minimum	Maximum	Std. Dev
Big data adoption	208	0,48	0	0	1	0,50
Number employees	208	13715	3596	323	258911	34698
Operating revenue	208	6684743	1258841	184049	111371250	15635748
Net income	208	1192235	236924	-151763	30624000	3241811
Profit margin	208	25,77	28,18	-27,21	91,17	17,11
Total Assets	208	112998588	14367488	358241	2807395000	373070380
Total Equity	208	11566256	2367005	-247875	267569000	33976331
Asset turnover	208	0,26	0,05	0,01	3,87	0,41
Total Liabilities	208	101433083	12038493	128489	2543987000	342643958
Debt/equity	207	5,75	6,31	-55,67	30,61	6,26
ROE	208	18,26	12,72	-23,84	287,57	27,02
ROA	208	2,81	1,22	-4,44	23,57	4,17
Tobin's Q	208	0,63	0,17	0,00	13,93	1,38
EPS	208	4,35	2,86	-10,77	52,53	6,30
P/E ratio	208	19,92	14,60	0,00	226,76	24,14
P/B ratio	208	1,86	1,29	-69,51	42,25	6,21
Ln_FSIZE	208	8,33	8,19	5,78	12,46	1,39

Panel B - Adopter vs Non-Adopter

Variables	*Adopters*			*Non-adopters*		
	N	**Mean**	**Std. Dev**	**N**	**Mean**	**Std. Dev**
Year article	100	2014,16	1,62	-	-	-
Number employees	100	21908	47946	108	6129	9019
Operating revenue	100	10434948	21102219	108	3212331	5969524
Net income	100	1987139	4495024	108	456212	706576
Profit margin	100	25,36	18,05	108	26,15	16,28
Total Assets	100	192568770	518070656	108	39322494	97629168
Total Equity	100	18850698	47262215	108	4821402	8428598
Asset turnover	100	0,27	0,35	108	0,26	0,46
Total Liabilities	100	173719633	475954700	108	34501092	90155268
Debt/equity	99	5,16	7,53	108	6,29	4,78
ROE	100	20,86	35,29	108	15,85	15,74
ROA	100	3,02	4,63	108	2,63	3,70
Tobin's Q	100	0,83	1,86	108	0,45	0,63
EPS	100	5,06	8,00	108	3,69	4,08
P/E ratio	100	21,51	28,35	108	18,45	19,47
P/B ratio	100	2,03	8,87	108	1,70	1,30
Ln_FSIZE	100	8,68	1,55	108	5,78	1,15

Panel C - Early adopters vs late adopters

Variables	*Early adopters (2010-2013)*			*Late adopters (2014-2016)*		
	N	**Mean**	**Std. Dev**	**N**	**Mean**	**Std. Dev**
Number employees	30	30570	59921	70	18196	41745
Operating revenue	30	15669107	27547138	70	8191737	17405826
Net income	30	2871960	5272037	70	1607930	4101482
Profit margin	30	26,48	14,49	70	24,88	19,45
Total Assets	30	254497838	577197140	70	166027742	492632936
Total Equity	30	27454981	59834756	70	15163148	40671781
Asset turnover	30	0,28	0,37	70	0,26	0,35
Total Liabilities	30	227042857	517738662	70	150866822	458910964
Debt/equity	30	5,40	5,60	69	5,05	8,26
ROE	30	21,33	24,29	70	20,66	39,23
ROA	30	3,51	4,67	70	2,80	4,63
Tobin's Q	30	0,82	1,75	70	0,84	1,92
EPS	30	4,61	4,16	70	5,25	9,19
P/E ratio	30	16,44	9,50	70	23,69	33,16
P/B ratio	30	3,06	7,72	70	1,59	9,33
Ln_FSIZE	30	9,01	1,64	70	8,54	1,50

Additionally, when looking at Panel C it seems that bigger firms are more likely to implement big data compared to smaller firms, the first row in Panel C shows a relatively high difference between the mean values for number of employees. Other variables, such as operating revenue and net income, show a higher value for early adopters of big data. The variable operating revenue is expected to be correlated to the number of employees, so it is not surprising that firms with on average more employees have a higher operating revenue. It can be seen that there is a difference in profit margin between the two groups, namely 26,48% and 24,88%, however this difference is too small to say something about the difference in profitability between early- and late adopters.

The “year article” variable stands for the year the article about big data implementation was found, with a mean of 2014,16 within the timeframe of 2010 - 2016 we can conclude that there are relatively less early adopters in the sample. The asset turnover is relatively low with a mean value of 0.27, compared to the study of Huang et al. (2020) which shows a mean asset turnover of 1.523. Additionally, the mean debt to equity ratio is 5.75 whereas the study of Huang et al. (2020) reports a mean debt to equity ratio of 2.57. However, the sample of this study only consists of firms operating in the financial services field, it might be that this sector works with higher leverage, resulting in a higher debt to equity ratio.

Furthermore, Panel B shows that the Adopter group has a higher P/E ratio and P/B ratio compared to the non-adopter group. However, Panel C shows that the P/E ratio is lower for the early adopter group compared to the late adopters. The P/E ratio indicates how much investors are prepared to pay for a stock, it is also based on future expectations and profits. It is interesting to see that early adopters of big data have on average lower P/E ratios, which globally means that investors have less confidence in these firms and are not prepared to pay as much for the stock as for late adopters. However, the P/B ratio is almost twice as high for the early adopter group which was expected to be correlated to the P/E ratio. The P/B ratio says something about the market value compared to the book value, normally a P/B ratio under 1 means that a stock is undervalued. An average P/B ratio of 3.06 indicates that the shares of early big data adopters are rather overvalued compared to late adopters. Concluded, the descriptive statistics already show some interesting differences between big data adopters and early adopters, the regression results in the next sections will provide more insights in these relationships.

4.2 Pearson’s correlation matrix

The Pearson's correlation matrix is used for the bivariate analysis. The most outstanding correlations are discussed to determine if multicollinearity might be an issue. The correlations between the dependent, independent, and control variables are examined. The Pearson’s R coefficient measures the direction and strength of a linear relationship between two variables (De Veaux et al. 2016). Table 6 provides an overview of Pearson's correlations, coefficients between -1 and 1 are given, where -1 means a perfectly negative correlation and +1 means a perfectly positive correlation. 0 implies that there is no correlation at all.

The first column in Table 6, shows that a few variables have a significant correlation with the big data implementation variable. Number of employees, operating revenue, net income, total assets, total equity, total liabilities, and Ln_FSIZE are significantly correlated. This seems logical because these variables say something about firm size and profitability of the firm, if a firm is more profitable and bigger it would easier invest in new technologies such as big data. Additionally, the Tobin's Q variable is also positive and significantly related, which tells something about the market value of the firm. There are also some coefficients close to zero which means no correlation is found. This mostly applies to market value variables and firm performance values, for example: the EPS, and P/B variables have a coefficient close to zero with the total equity and asset turnover variables. Additionally, it can be seen that the big data implementation variable is not correlated with market value variables such as ROE, ROA, EPS, P/B and the P/E ratio.

Another noticeable coefficient is the correlation between profit margin and asset turnover, which is significantly negative (-.400). This means that if the profit margin increases the asset turnover decreases. This is interesting to see because the asset turnover ratio says something about how efficient firms use their assets, if assets are efficient in use it is expected that the profit margin would be positively affected as well. Furthermore, some logical correlations are observed, the total liabilities are significantly positively correlated to the total equity. Also, the correlation between ROE and Asset turnover is positive and significant (.264).

Looking at the control variable for firm size (Ln_FSIZE), some interesting correlations are found. For example, the correlation between number of employees and profit margin is close to zero and not significant (-.11), this means that almost no correlation is found between those two variables whereas it is expected that bigger firms often have a higher profit margin. Variables that are positively and significantly correlated to the firm size variable are operating revenue, net income, total assets, and total equity. This is very explainable because larger firms often have more assets leading to higher book values and can therefore generate more income. However, a bivariate analysis does not take into account the aspect of independent and dependent variables. So, the correlation coefficient could implicate a relationship both ways. This issue is addressed in the methodology section and therefore accounted for in the models used in this research.

Table 6: Pearson's correlation matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Big data adoption	1																
2	Number employees	,228**	1															
3	Operating revenue	,231**	,897**	1														
4	Net income	,237**	,900**	,880**	1													
5	Profit margin	-0,02	-0,02	-0,07	0,11	1												
6	Total Assets	,206**	,840**	,797**	,915**	0,11	1											
7	Total Equity	,207**	,943**	,870**	,946**	0,05	,905**	1										
8	Asset turnover	0,01	-0,05	0,07	-0,11	-,400**	-,148*	-0,13	1									
9	Total Liabilities	,203**	,821**	,781**	,902**	0,12	,999**	,886**	-,148*	1								
10	Debt/equity	-0,09	0,07	0,09	0,10	,150*	,154*	0,12	-,226**	,158*	1							
11	ROE	0,09	-0,03	0,00	0,05	,218**	0,06	-0,06	,264**	0,07	-,196**	1						
12	ROA	0,05	-0,04	-0,04	-0,03	,144*	-0,13	-0,10	,439**	-0,13	-,296**	,531**	1					
13	Tobin's Q	,139*	-0,04	-0,04	-0,02	0,03	-0,11	-0,09	,349**	-0,11	-,253**	,326**	,757**	1				
14	EPS	0,11	0,12	,218**	,181**	,143*	,138*	0,13	-0,06	,137*	0,04	0,06	0,06	-0,01	1			
15	P/E ratio	0,06	0,02	-0,04	-0,07	-,224**	-0,09	-0,04	,176*	-0,10	-0,13	-0,07	0,07	,338**	-0,07	1		
16	P/B ratio	0,03	0,00	0,01	0,04	0,08	-0,04	-0,02	0,10	-0,04	,462**	0,07	,309**	,465**	0,03	,211**	1	
17	Ln_FSIZE	,240**	,691**	,682**	,617**	-0,11	,534**	,606**	-0,02	,521**	0,03	-0,05	0,00	-0,01	,278**	0,05	-0,01	1

**. Correlation is significant at the 0.01 level (2-tailed)   *. Correlation is significant at the 0.05 level (2-tailed)

4.3 Regression analysis

4.3.1 BDI and Financial firm performance

Several regression analyses are conducted to test the hypothesis. Table 7 reports the results for the OLS regression between the financial metrics stated in Table 3 and the dummy variable created for big data implementation. Also, a control variable for firm size is included in the model. Furthermore, in Table 7 and 8 two models are included in the regression tables, in model 1 the BDI dummy variable is included and in model 2 it is excluded. Differences in R squared can show how much this BDI variable contributes to the explanation of the model. It can be seen that the results reported in the first row, show no significant results, this indicates that big data implementation does not have a significant impact on financial firm performance. Also, the differences between the reported R squared between model 1 and 2 are very small which indicate that the BDI variable does not contribute much to the model.

4.3.2 BDI and Firm value

In Table 8 the regression results for the effect of big data implementation on firm value are given. It can be seen that significant results were reported for the BDI variable, in combination with the Tobin's Q and P/B ratio as dependent variable. It seems that big data implementation has a positive effect on the Tobin's Q value of the firm. As earlier described, Tobin's Q is calculated by dividing the firm's market capitalization by its total assets. It seems that firms that implement big data have a higher market capitalization or less assets, nevertheless, this ratio is higher compared to non-adopters. Also, the R squared increases from .144 to .161 if the BDI variable is included in the model. Furthermore, as said before the P/B ratio reports a positive and significant outcome for the BDI variable as well, this means that firms that implement big data are associated with a higher price-to-book ratio. A higher P/B ratio means that the market value is increasing compared to the book value of the company, it seems that investors are prepared to pay more for a company that implements big data.

4.3.3 Early adopters

Table 9 also reports the results for the OLS regression between the financial metrics, but then in combination with the dummy variable for early adopters of big data. First of all, it can be seen that there are no significant outcomes in the first row, which shows the results for the early big data adopter variable. This means that it does not have a significant impact on firm performance or firm value if a firm belongs to the early adopter group. There are some similarities between the results in Table 7 and 8, and Table 9, overall, the results are in line with each other. The only difference is that Table 8 shows a significant outcome for Tobin's Q when it comes to big data implementation.

If the results of these regression models are compared to the study of Huang et al. (2020), differences arise. While this study shows a very limited amount of significant results, the study of Huang et al. (2020) shows significant outcomes for variables such as return on equity (ROE), return on assets (ROA), and profit margin. However, there are also some similarities between the results presented in both studies. Tobin's Q shows only significant results in Table 8 when it comes to big data adoption, in Table 9 no significant result are reported when it comes to early adopters, the same is found in the study of Huang et al. (2020). Furthermore, in both studies the financial metric asset turnover, measured by revenue and assets, shows a non-significant result for the big data implementation variable. The main difference between both studies is that this study focuses only on firms operating in the financial services field whereas the study of Huang et al. (2020) has a sample with firms operating in different sectors.

It can be concluded that most of the financial metrics that beforehand were appointed to have an influence on big data adoption and performance are not significant in the regression results. Firms seem to experience small benefits from implementing big data when it comes to market value.

Table 7: OLS regression results Financial firm performance (H1)

OLS regression results Hypothesis 1										
Dependent variable	**Profit margin**		**ROE**		**ROA**		**Debt/equity**		**Asset turnover**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Big data implementation (BDI)	.012		.088		.046		.064		.017	
	(.047)		(1.402)		(.769)		(.901)		(.249)	
Control variables:										
Profit margin (PM)			.340***	.340***	.408***	.408***	.199***	.200***	-.412***	-.412***
			(5.001)	(4.989)	(6.316)	(6.322)	(2.632)	(2.643)	(-6.264)	(-6.281)
Asset turnover (ATO)	-.394***	-.394***	.411***	.413***	.597***	.598***	.026	.027		
	(-6.264)	(-6.281)	(6.186)	(6.197)	(9.457)	(9.482)	(.345)	(.362)		
Leverage (LEV)	.166***	.166***	.227***	.233***	-.103*	-.100*			.023	.024
	(2.632)	(2.643)	(3.666)	(3.745)	(-1.747)	(-1.704)			(.345)	(.362)
Ln_Size	-.123*	-.123*	-.030	-.009	.044	.055	.034	.050	-.068	-.065
	(-1.905)	(-1.968)	(-.468)	(-.140)	(.727)	(.937)	(.482)	(.719)	(-1.029)	(-1.001)
Adjusted R2	.186	.190	.236	.233	.312	.313	.021	.022	.148	.152
Observations	208	208	208	208	208	208	208	208	208	208

*Notes: The independent variables are average figures over 2017-2019, the dependent variables are average figures over 2018-2020. This table reports the standardized coefficients. t-statistics are in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ show significance at the 1%, 5%, and 10% level.*

Table 8 OLS regression results Firm value (H2)

OLS regression results Hypothesis 2								
Dependent variable	**EPS**		**P/E ratio**		**P/B ratio**		**Tobin's Q**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Big data implementation (BDI)	.041		.055		.174**		.148**	
	(.600)		(.788)		(2.560)		(2.254)	
Control variables:								
Profit margin (PM)	.176**	.176**	-.172**	-.172**	.147**	.147*	.219***	.219***
	(2.371)	(2.375)	(-2.263)	(-2.265)	(1.992)	(1.964)	(3.076)	(3.044)
Asset turnover (ATO)	.018	.019	.105	.106	.430***	.324***	.431	.433***
	(.248)	(.259)	(1.415)	(1.430)	(4.439)	(4.424)	(6.182)	(6.161)
Leverage (LEV)	.037	.039	-.030	-.027	-.071	-.060	-.071	-.062
	(.545)	(.585)	(-.436)	(-.388)	(-1.052)	(-.881)	(-1.091)	(-.942)
Ln_Size	.287***	.297***	.021	.034	.028	.069	-.016	.019
	(2.987)	(4.433)	(.295)	(.496)	(.406)	(1.028)	(-.247)	(.292)
Adjusted R2	.089	.092	.040	.042	.099	.074	.161	.144
Observations	208	208	208	208	208	208	208	208

*Notes: The independent variables are average figures over 2017-2019, the dependent variables are average figures over 2018-2020. This table reports the standardized coefficients. t-statistics are in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1 show significance at the 1%, 5%, and 10% level.*

Table 9 OLS regression results Early adopters (H3)

OLS regression results Hypothesis 3																		
Dependent variable	**Profit margin**		**ROE**		**ROA**		**Debt/equity**		**Asset turnover**		**EPS**		**P/E ratio**		**P/B ratio**		**Tobin's Q**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Early big data adopter (EBDA)	.077		-.004		.021		-.081		.060		-.079		-.134		.064		-.036	
	(.838)		(-.041)		(.274)		(-.811)		(.637)		(-.800)		(-1.341)		(.620)		(-.393)	
Control variables:																		
Profit margin (PM)			.432***	.431***	.536***	.538***	.264**	.258**	-.399***	-.395***	.158	.151	-.086	-.099	.190*	.196*	.316***	.313***
			(4.560)	(4.598)	(6.468)	(6.548)	(2.442)	(2.397)	(-4.119)	(-4.098)	(1.439)	(1.378)	(-.777)	(-.892)	(1.678)	(1.743)	(3.154)	(3.146)
Asset turnover (ATO)	-.379***	-.377***	.470***	.469***	.704***	.705***	.020	.014			-.007	-.012	.205*	.195*	.176	.181	.530***	.527***
	(-4.119)	(-4.098)	(5.090)	(5.126)	(8.713)	(8.792)	(.185)	(.134)			(-.061)	(-.114)	(1.894)	(1.803)	(1.595)	(1.644)	(5.423)	(5.434)
Leverage (LEV)	.224**	.219**	.288**	.228**	-.140*	-.141*			.018	.013	.032	.039	-.050	-.038	-.041	-.046	-.088	-.085
	(2.442)	(2.397)	(2.608)	(2.634)	(-1.826)	(-1.865)			(.185)	(.134)	(.316)	(.385)	(-.487)	(-.375)	(-.388)	(-.442)	(-.953)	(-.928)
Ln_Size	-.090	-.079	-.017	-.017	.011	.014	.036	.024	-.128	-.119	.276***	.264***	.084	.063	-.013	-.003	-.014	-.019
	(-.975)	(-.864)	(-.195)	(-.204)	(.140)	(.185)	(.361)	(.243)	(-1.350)	(-1.275)	(2.771)	(2.686)	(.830)	(.631)	(-.127)	(-.033)	(-.152)	(-.215)
Adjusted R2	.210	.179	.299	.306	.463	.489	.031	.035	.133	.138	.055	.059	.038	.030	.047	.003	.241	.221
Observations	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

*Notes: The independent variables are average figures over 2017-2019, the dependent variables are average figures over 2018-2020. This table reports the standardized coefficients. t-statistics are in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ show significance at the 1%, 5%, and 10% level.*

4.4 Robustness checks

Several robustness checks are conducted to make sure that the results hold under different circumstances.

4.4.1 Robustness check: Split sample

In Appendix E, the robustness checks for the OLS regression methods are given. Regressions are conducted for the years 2017 till 2018, and 2019 till 2020 to see if the results hold under different assumptions. Even though the current results are based on the average figures over 2017 till 2020, the robustness checks are confirmation that the results are not based on a specific time period. It can be seen that the first split sample over the years 2017 - 2018 does not report any significant results for the big data implementation, expect for the P/B ratio and Tobin's Q. This is in line with the results in Table 7 and 8, the results thus hold under different circumstances, in this case the split sample.

Moving on to the second split sample with average figures over the years 2019 and 2020, it can be seen that Appendix E-3 reports significant results for ROE, P/B ratio, and Tobin's Q. The difference with the results shown in Table 7 is that ROE shows a significant outcome at the 10% level in the second split sample. However, overall the results of the second split sample are in line with the total sample, for example the T value for ROE in the total sample is 1.402 whereas in the split sample it is 1.682, this is relatively close even though one is significant and the other is not. The results in the split sample show two significant outcomes, out of the four variables for firm value, this implies that big data implementation can have an impact on firm value if enough time is taken into account to measure the effects. The alternative measures robustness check in the next section examines the results further, to see if for example 2 significant outcomes are sufficient to say something about the effect of big data on financial firm performance or firm value.

The conclusion of this conducted split sample robustness test is that the results hold under different circumstances, besides the small differences that were found, the overall results point towards the same direction. Additionally, there is however an interesting pattern to see in the results which might be something for future research to more extensively examine. It can be seen that results for the second split sample test show more significant results for the financial metrics than the first split sample. It seems that firms need more time to exploit the benefits from big data implementation, this means that the effects of big data implementation tend to be higher after more time has passed. This however contradicts with the results regarding the third hypothesis where it can be seen that early big data adopters do not experience significant higher firm performance. In the current literature it remains unclear how big data effects can be measured and what impact it really has. It is therefore interesting to investigate how much time it takes on average for firms to experience benefits from implementing big data, and really see changes in firm performance and firm value.

4.4.2 Robustness check: Alternative measures

The models presented in Table 7,8 and 9 make use of different financial metrics as alternative measures. It can be seen that most of the results regarding the financial firm performance are non-significant. Table 7 shows no significant outcomes for the big data adoption variable. This means that the non-significant results hold under alternative measures related to firm performance. In Table 8 significant results are reported for the Tobin's Q, and P/B ratio dependent variable, this means that 2 out of 4 variables related to firm value are significant. The relationship between big data adoption and increased firm value is therefore not very strong, at least a majority of the variables, so at least 3 out of 4 should be significant to really state that there is a significant impact. Since this is not the case the robustness check for alternative measures does not hold. The results reported in Table 9 are all non-significant, which means that all the variables related to the same measurement, show the same outcomes and hold under different measures. Overall, the results of the alternative measures are comparable to the results of the other measures in the same overarching category. The results are all non-significant and therefore the hypothesis which states that early big data adopters experience higher financial firm performance and firm value is rejected based on this regression analyses.

4.4.3 Robustness check: Lagged variables

As mentioned before, in the research model lagged variables are included to control for any endogeneity problems. The regression results in Table 7,8 and 9 use average figures over 2017-2019 and 2018-2020, which on average results in 1 year lagged variables. Additionally, in Appendix E, the results for the split sample robustness check are given, however, these regression results also make use of lagged variables. The split samples measure the independent variables as year t, and the dependent variables as year t + 1. In general, the results show similar outcomes compared to the results presented in Table 7, 8, and 9. The results in Appendix E-3 and E-6 still show no significant outcomes, in line with the results in Table 9. It can be concluded that the results in the appendix are in line with the results in the other regression models used in this study. While using lagged variables the endogeneity issues is addressed and taken into account.

4.5 Hypotheses testing

In this section, the three stated hypotheses are examined combined with the regression results shown in Table 7, 8, and 9, and in the appendices. The main findings are also described related to each hypothesis, these will shortly come back in the conclusion of this study.

4.5.1 Hypothesis 1: Firm performance

The first hypothesis that was stated is focused on testing the relationship between implementing big data and increased firm performance. This hypothesis was tested via the OLS regression analysis. If looked at the results presented in Table 7 it can be concluded that financial firm performance does not significantly increase when a firm adopts big data. All the variables show non-significant results in the first row which measures the effect of big data implementation on the dependent variable. Since there at no significant results, no evidence supports the stated hypothesis that firms that implement big data experience higher firm performance.

4.5.2 Hypothesis 2: Firm value

The second hypothesis that was stated is about the effect of big data on firm value. Table 8 shows the results for four metrics related to firm value, namely EPS, P/E ratio, P/B ratio, and Tobin's Q. As can be seen, only significant results are found for the Tobin's Q variable and P/B ratio, when it comes to the BDI independent variable. The Tobin's Q variable expresses the relationship between market value and intrinsic value, it is an estimate which shows if a stock is over or undervalued. The results show that big data adopters are associated with a higher value for Tobin's Q. It seems that investors are prepared to pay more for a stock from firms that work with big data compared to firms that do not. Furthermore, as mentioned in the results section, the P/B ratio reports a positive and significant outcome for the BDI variable as well (2.560), this means that firms that implement big data are associated with an on average higher price-to-book ratio. A higher P/B ratio means that the market value is increasing compared to the book value of the company.

Additionally, the results of the robustness checks show overall the same results. However, there are some small differences that are worth to mention. The robustness checks that use the variables over the years 2019 and 2020 show higher T values for the P/B ratio and Tobin's Q variable. Compared to the robustness check over the years 2017 and 2018, the difference for the P/B ratio is 1.671 in Appendix E-2 and 2.651 in Appendix E-5. The same applies for the Tobin's Q ratio which shows a slightly higher T value in the model using 2019-2020 variables, namely 2.155 over 2.272. Overall, the results in the regression analyses that is conducted over later years, shows higher and more significant outcomes. It seems that big data implementation needs time to be beneficial for firms and that it becomes noticeable in the financial metrics used in this study. Based on these outcomes it can be concluded that if enough time is included in the model, between the implementation and regression data, to experience the benefits of big data, higher firm value can be found. However, it does not necessarily have to be that big data adoption is the only reason for this increase in firm performance, it can also be that the firm's attitude towards new innovations and implementing them results in investors willing to pay more for the stock. To conclude, based on the results we can say that big data adopters are to some extent associated with higher market values, but it should be taken into account that the results might be because of other factors as well, future research could provide more insights on this topic.

4.5.3 Hypothesis 3: Early adopters

The third hypothesis is examined based on Table 9 and the corresponding robustness tests. It can be concluded that, since there are no significant results found for the early adopter variable, it cannot be stated that firms that adopt big data earlier, experience higher firm performance. The same occurs for market value, no significant results were found for the early adopters' variable. The results presented in Table 9 also hold under different circumstances, the conducted robustness checks in Appendix E show the same results. Therefore, it can be concluded that there is no significant difference between early and later adopters within the sample, when it comes to firm performance and market value.

However, some interesting findings occur if the conclusions of the second and third hypothesis are examined. The conclusion of the second hypothesis is that firms need time to exploit the benefits of big data which then can lead to higher market value. The conclusion of the third hypothesis does not support this, because if the conclusion about the second hypothesis is true, it is expected that early adopters of big data experience higher firm performance and firm value because these firms have more time to implement and exploit the benefits of big data. As said before, this is something that could be investigated within future research, preferably with a larger sample size because this study could only find a sample of 30 early big data adopters, by increasing the sample size the results might change or provide more insights.

5 Conclusion

This chapter describes the conclusion and discussion. First, the conclusions based on the conducted regression analysis are given and the stated research question is answered. Second, the limitations of this study are examined, and possible future research directions are recommended.

5.1 Conclusion and Discussion

The amount of literature that is being published about the impact of big data on firm performance is heavily increasing over the last years. Specifically for the financial services industry, additional research is needed to understand the advantages and disadvantages. This study investigates the effect of big data implementation on firm performance and firm value on firms operating in the financial services field, with a different approach compared to current literature. Financial firm performance and firm value are measured by different financial metrics such as ROE, ROA, profit margin, P/E ratio, and Tobin's Q. The research question that was stated in the beginning of this research is:

What is the effect of big data implementation on firm performance and firm value of firms operating in the financial services field listed in the United States?

To answer this research question three hypotheses have been formulated, focusing on financial firm performance, firm value and the difference in early and late adopters. The first hypothesis about the effect of big data implementation on financial firm performance is not supported in this study. The results for the conducted OLS regression models show no significant relationships for the effect of big data on financial firm performance. With no significant outcomes for all the five metrics related to financial firm performance it can be concluded that big data implementation does not have a significant impact.

The second hypothesis about the effect of big data implementation on firm value show more significant results. It can be seen in Table 8 that significant results are found at the 5% level for the P/B ratio and Tobin's Q. However, the robustness tests, in Appendix E and F show even higher significant results when the sample size was extracted over the years 2019 and 2020, whereas the sample extracted over the years 2017 and 2018 reports less significant results. This implies that for big data adopters, it takes time to experience and exploit the benefits of big data, which eventually results in higher firm value.

The third hypothesis about the difference in early- and late big data adopters shows no significant results at all. Based on the conducted OLS regression and sample used in this study, it can be stated that it does not matter if a firm implemented big data between 2010 and 2013, compared to 2014 and 2016. This is interesting to see because it is expected that firms that work longer with big data can exploit more opportunities and experience more benefits because of the increased insights in several processes, which is also the conclusion based on hypothesis 2. Future research might provide more evidence if the sample size is increased and a longer time period is taken into the regressions. Because big data started taking off in 2010, relatively less data is available, researchers in the future might have more data at their disposal, because the amount of data grows exponentially. With more data available, better conclusions about the effect and time period big data needs to improve firm performance, can be made.

The current literature about big data and firm performance states that big data can lead to a competitive advantage and increase in market share. The study of Huang et al. (2020) shows significant outcomes for the effect of big data on firm performance. Additionally, the study of Yadegaridekordi et al. (2020) also finds that big data implementation leads to higher firm performance. On the contrary, this study shows that this is not yet noticeable in the firm performance metrics such as profit margin and return on assets. However, it seems that investors are prepared to pay more for shares of companies that implement big data. This is supported by the significant results that are shown for the P/B ratio and Tobin's Q which indicate that big data adopters have higher ratios for these metrics. Furthermore, the descriptive statistics show relatively high values for big data adopters when it comes to these metrics, this means that these stocks can considered to be more overvalued than undervalued. Nevertheless, these facts should not only be associated with big data implementation but could also be related to firms implementing new innovations in general. Big data is seen as one of the new key innovations and investors are generally more optimistic about firms that are constantly trying to improve themselves and focus on new innovations.

All things considered, the answer to the stated research questions has different elements. First of all, this study shows no evidence of firms experiencing higher financial firm performance if big data is implemented, which contradicts the current literature. Secondly, no evidence was found for the third hypothesis, which was about the first movers advantage. Lastly, based on the results, this study finds that firms that implement big data experience higher firm value if enough time is taken into account to exploit the potential benefits. However, these results vary over time in significance levels, by increasing the sample size, and given the fact that the amount of data grows exponentially, enough opportunities will arise for future research to confirm or deny the stated conclusions in this study.

5.2 Limitations and recommendations for future research

Limitations

As earlier discussed, the results of this study show some evidence that big data implementation has an impact on firm value. However, this study comes with limitations as well. Firstly, despite the fact that the sample selection was very intensively done it is unavoidable that some news articles were missed. On the one hand this is a limitation associated with the methodology of this study, on the other hand it might bring new insights because of the different approach compared to current literature about big data and firm performance. Also, it might be that for bigger firms, more news articles can be found. This can account for the fact that larger firms are more represented in the big data adopters' group.

Secondly, it might be that firms that did not announce they work with big data, still work with big data, this might have an impact on firm performance. However, because only large, listed firms were taken into the sample it is likely that these firms announce new innovations that are implemented in the company, including big data. Also, by using Google search, annual reports of companies were included in the search which provides more insights in new innovations and strategy of the company. Even though the search for big data announcements was done extensively, only 30 firms were found that implemented big data between 2010 and 2013. This is a relatively small sample size and therefore a limitation of the study. No significant outcomes were found related to this sample but an increase in sample size might provide more insights. Overall, this can be applied to the total sample size as well, because of time limitations related to the manual effort that needed to be done the sample size was not that big, with 208 firms. By increasing the sample size the reliability and validity of the results would be increased.

Recommendations

Based on this study, there are some recommendations for future research directions related to big data and firm performance. First of all, more research needs to be done in what time period the benefits of big data implementation can be exploited. The current literature does not point out how much time it takes until the effects of big data implementation can be measured. Can positive effects be measured in for example two, five or even ten years? The short term and long-term effects of big data are not yet sufficiently investigated. More and more data becomes available and easier accessible for firms, especially smaller firms with lower budget will be able to implement big data faster. If the sample becomes bigger and more is known about the effect of big data, then the long-term effects can be better predicted as well.

The second recommendation is about the approach which could be changed in future research to determine te effect of big data. Instead of using a quantitative approach a more qualitative approach can be used. Especially for the data collection, by conducting a questionnaire or interview, it is more certain if a specific firm really implemented big data which results in less bias in the sample. Additionally, a lot of studies in the current literature, including the followed study of Huang et al. (2020), use the OLS regression model. Studies using the OLS method show similar results, and relatively much non-significant results. Different regression models such as the 2SLS or Panel regression could give more significant results because these models account more for the endogeneity problem. Another recommendation related to taking another approach is using a text analysis program, which present the opportunity to search through more platforms besides Google and can also take all the annual reports into account. It is expected that this approach would create a more representable sample with less bias, compared to the manual approach used in this study.

This research has its main focus on firms operating in the United States, this was done because of the availability of information and number of listed firms. However, different results can occur if other regions were taken into account as well, to make sure that the results that were found not only apply to the United States. Additionally, future research could focus more on the decision-making process. In this study it is expected that improved decision making would increase firm performance, and that big data can lead to better decision making. However a qualitative study with in depth interviews with people that work with big data and are responsible for the implementation and interpretation of the data can give new insights. Especially in how and in what ways big data has an impact on the decision making. Possibly outcomes can be that the usage of big data improves the decision making, resulting in more sales, lower costs of capital, increased competitive advantage, and therefore higher firm performance.

6 References

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? International Journal of Production Economics, 182, 113–131. https://doi.org/10.1016/j.ijpe.2016.08.018

Alexey V. Bataev. (2018). Analysis of the application of big data technologies in the financial sphere. *Graduate school of state and financial management.* https://doi.org/10.1109/ITMQIS.2018.8525121

Anfer, O., & Wamba, S. F. (2019). Big data analytics and strategic marketing capabilities: Impact on firm performance: Research in progress. *Advances in Intelligent Systems and Computing, 931*, 633–640. https://doi.org/10.1007/978-3-030-16184-2_60

Bae, J., Kim, S. J., & Oh, H. (2017). Taming polysemous signals: The role of marketing intensity on the relationship between financial leverage and firm performance. *Review of Financial Economics, 33*, 29–40. https://doi.org/10.1016/j.rfe.2016.12.002

Baltagi, B, H. (2005). *Econometric analysis of panel data.* Retrieved from https://himayatullah.weebly.com/uploads/5/3/4/0/53400977/baltagi-econometric-analysis-of-panel-data_himmy.pdf

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, *97*, 71–87. https://doi.org/10.1016/j.jmoneco.2018.05.013

Brynjolfsson, Erik and Hitt, Lorin M. and Kim, Heekyung Hellen, Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? (April 22, 2011). DOI: http://dx.doi.org/10.2139/ssrn.1819486

Chen, Y., & Lin, Z. (2021). Business Intelligence Capabilities and Firm Performance: A Study in China. *International Journal of Information Management, 57.* https://doi.org/10.1016/j.ijinfomgt.2020.102232

De Souza, S. V. C., & Junqueira, R. G. (2005). A procedure to assess linearity by ordinary least squares method. *Analytica Chimica Acta*, *552*(1–2), 25–35. https://doi.org/10.1016/j.aca.2005.07.043

De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2016). *Stats, data and models*. Edinburgh, England: Pearson Education Limited.

Dong, J. Q., & Yang, C. H. (2020). Business value of big data analytics: A systems-theoretic approach and empirical test. *Information and Management*, *57*(1). https://doi.org/10.1016/j.im.2018.11.001

Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8557 LNAI*, 214–227. https://doi.org/10.1007/978-3-319-08976-8_16

Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. (2011). Effects of innovation types on firm performance. *International Journal of Production Economics*, *133*(2), 662–676. https://doi.org/10.1016/j.ijpe.2011.05.014

Hair, J.F., Black, W.C., Babin, B.J. & Anderson, R.E. (2010). Multivariate Data Analysis. 7th edition. New York, Pearson.

Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00291-z

Huang, C. K., Wang, T., & Tasi, Y. T. (2016). Market reactions to big data implementation announcements. In *20th Pacific Asia conference on information systems*. Chiayi: Taiwan.

Huang, C. K., Wang, T., & Huang, T. Y. (2020). Initial Evidence on the Impact of Big Data Implementation on Firm Performance. *Information Systems Frontiers*, *22*(2), 475–487. https://doi.org/10.1007/s10796-018-9872-5

Ibhagui, O. W., & Olokoyo, F. O. (2018). Leverage and firm performance: New evidence on the role of firm size. *North American Journal of Economics and Finance*, *45*, 57–82. https://doi.org/10.1016/j.najef.2018.02.002

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007?

Ji-fan Ren, S., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011–5026. https://doi.org/10.1080/00207543.2016.1154209

Koščielniak, H., & Puto, A. (2015). BIG DATA in Decision Making Processes of Enterprises. *Procedia Computer Science*, *65*, 1052–1058. https://doi.org/10.1016/j.procs.2015.09.053

Kumaresan, A., & Liberona, D. (2018). A case study on challenges and obstacles in transforming to a data-driven business model in a financial organisation. *Communications in Computer and Information Science*, *877*, 263–276. https://doi.org/10.1007/978-3-319-95204-8_23

Lee, J. (2009). Does size matter in firm performance? Evidence from US public firms. *International Journal of the Economics of Business*, *16*(2), 189–203. https://doi.org/10.1080/13571510902917400

Li, Y., Dai, J., & Cui, L. (2020). The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model. *International Journal of Production Economics*, *229*. https://doi.org/10.1016/j.ijpe.2020.107777

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261–276. https://doi.org/10.1016/j.jbusres.2019.01.044

Nurlaela, S., Mursito, B., Kustiyah, E., Istiqomah, I., & Hartono, S. (2019). Asset turnover, Capital structure, and Financial performance consumption industry company in Indonesia stock exchange. *International Journal of Economics and Financial Issues*, *9*(3), 297–301. https://doi.org/10.32479/ijefi.8185

O'Halloran, S., Maskey, S., McAllister, G., Park, D. K., & Chen, K. (2015). Big data and the regulation of financial markets. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*, 1118–1124. https://doi.org/10.1145/2808797.2808841

Oussous, A., Benjelloun, F. Z., Ait Lahcen, A., & Belfkih, S. (2018, October 1). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, Vol. 30, pp. 431–448. https://doi.org/10.1016/j.jksuci.2017.06.001

Raguseo, E., & Vitari, C. (2018). Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research*, *56*(15), 5206–5221. https://doi.org/10.1080/00207543.2018.1427900

Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, *24*, 4–15. https://doi.org/10.7152/acro.v24i1.14671

Santos, J. B., & Brito, L. A. L. (2012, May). Toward a subjective measurement model for firm performance. *BAR - Brazilian Administration Review*, Vol. 9, pp. 95–117. https://doi.org/10.1590/S1807-76922012000500007

Schultz, E. L., Tan, D. T., & Walsh, K. D. (2010). Endogeneity and the corporate governance - performance relation. *Australian Journal of Management*, *35*(2), 145–163. https://doi.org/10.1177/0312896210370079

Shamim, S., Zeng, J., Khan, Z., & Zia, N. U. (2020). Big data analytics capability and decision making performance in emerging market firms: The role of contractual and relational governance mechanisms. *Technological Forecasting and Social Change*, *161*. https://doi.org/10.1016/j.techfore.2020.120315

Shepherd, B. (2010). Introduction dealing with endogeneity examples of IV Gravity Models summary Session 3: Dealing with reverse causality. Retrieved from: https://artnet.unescap.org/tid/artnet/mtg/gravity10_thus3.pdf

Su, X., Zeng, W., Zheng, M., Jiang, X., Lin, W., & Xu, A. (2021). Big data analytics capabilities and organizational performance: the mediating effect of dual innovations.

European Journal of Innovation Management, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ejim-10-2020-0431

Subrahmanyam, A. (2019, December 1). Big data in finance: Evidence and challenges. *Borsa Istanbul Review*, Vol. 19, pp. 283–287. https://doi.org/10.1016/j.bir.2019.07.007

Sun, H., Rabbani, M. R., Sial, M. S., Yu, S., Filipe, J. A., & Cherian, J. (2020). Identifying big data's opportunities, challenges, and implications in finance. *Mathematics*, *8*(10), 1–19. https://doi.org/10.3390/math8101738

Sun, Y., Shi, Y., & Zhang, Z. (2019, January 2). Finance Big Data: Management, Analysis, and Applications. *International Journal of Electronic Commerce*, Vol. 23, pp. 9–11. https://doi.org/10.1080/10864415.2018.1512270

Suoniemi, S., Meyer-Waarden, L., Munzel, A., Zablah, A. R., & Straub, D. (2020). Big data and firm performance: The roles of market-directed capabilities and business strategy. *Information and Management*, *57*(7). https://doi.org/10.1016/j.im.2020.103365

Swishchuk, A. (2020). Stochastic Modelling of Big Data in Finance. *Methodology and Computing in Applied Probability*, *22*(4), 1613–1630. https://doi.org/10.1007/s11009-020-09826-6

Vitari, C., & Raguseo, E. (2020). Big data analytics business value and firm performance: linking with environmental context. *International Journal of Production Research*, *58*(18), 5456–5476. https://doi.org/10.1080/00207543.2019.1660822

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. fan, Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wang, Y. (2020). Analysis of financial business model towards big data and its applications. *Journal of Visual Communication and Image Representation*, *71*. https://doi.org/10.1016/j.jvcir.2019.102729

Wang, Z. (2021). Influence of Internet Finance on Commercial Bank Financial Services. *Advances in Intelligent Systems and Computing*, *1233 AISC*, 70–75.

https://doi.org/10.1007/978-3-030-51431-0_11

Wintoki, M. B., Linck, J. S., & Netter, J. M. (2012). Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics, 105*(3), 581–606. https://doi.org/10.1016/j.jfineco.2012.03.005

Yadegaridehkordi, E., Nilashi, M., Shuib, L., Hairul Nizam Bin Md Nasir, M., Asadi, S., Samad, S., & Fatimah Awang, N. (2020). The impact of big data on firm performance in hotel industry. *Electronic Commerce Research and Applications*, *40*. https://doi.org/10.1016/j.elerap.2019.100921

Yang, S. (2020). Research on innovation of financial management model based on cloud computing. *E3S Web of Conferences*, *214*. https://doi.org/10.1051/e3sconf/202021403037

Appendices

Appendix A: Key words

Key words	
Analytics	Could computing services
Big data	Data analytics
Big data analytics	Data center
Big data applications	Data driven
Big data platform	Data processing
Big data technology	Digital analytics
Business analytics	Marketing analytics
Business intelligence	New BI systems
Cloud services	Text mining

Appendix B: Total sample

Company name	Year article	Company name	Year article
JPMORGAN CHASE & CO	2015	FRANKLIN RESOURCES, INC.	Not found
ANTHEM INC.	2012	GENWORTH FINANCIAL INC	Not found
BANK OF AMERICA CORPORATION	2013	HANOVER INSURANCE GROUP INC.	Not found
CITIGROUP INC	2016	WESTERN UNION CO. (THE)	2014
WELLS FARGO & COMPANY	2013	CME GROUP INC	2015
STONEX GROUP INC	Not found	HUNTINGTON BANCSHARES INC	Not found
MORGAN STANLEY	2012	SANTANDER CONSUMER USA HOLDINGS INC	2015
GOLDMAN SACHS GROUP, INC	2014	CNO FINANCIAL GROUP INC.	Not found
METLIFE INC.	2013	PAYCHEX INC	2016
PROGRESSIVE CORP	2014	SVB FINANCIAL GROUP	2016
THE ALLSTATE CORP.	2011	STIFEL FINANCIAL CORP	Not found
AMERICAN EXPRESS COMPANY	2014	MAXIMUS INC	Not found
PRUDENTIAL FINANCIAL INC.	Not found	JEFFERIES FINANCIAL GROUP INCORPORATED	Not found
TRAVELERS COMPANIES INC.	Not found	SELECTIVE INSURANCE GROUP INC.	Not found
CAPITAL ONE FINANCIAL CORPORATION	2015	FIRST HORIZON CORPORATION	2011
US BANCORP	2015	COMERICA INCORPORATED	2010
TRUIST FINANCIAL CORPORATION	Not found	VERISK ANALYTICS, INC.	2012
PAYPAL HOLDINGS, INC.	2016	H&R BLOCK, INC.	2013
MARSH & MCLENNAN COMPANIES INC	2015	EURONET WORLDWIDE INC	Not found
BLACKROCK, INC	2015	FTI CONSULTING INC	2015
PNC FINANCIAL SERVICES GROUP INC	2013	CIT GROUP, INC	2015
FREDDIE MAC	2016	POPULAR, INC	Not found
BANK OF NEW YORK MELLON CORPORATION	2015	FLEETCOR TECHNOLOGIES, INC.	Not found
MASTERCARD	2013	VIRTU FINANCIAL, INC	2015
LINCOLN NATIONAL CORP	Not found	EVERCORE INC.	Not found
CNA FINANCIAL CORP.	2014	HILLTOP HOLDINGS INC	Not found
REINSURANCE GROUP OF AMERICA INC.	2015	PEOPLE'S UNITED FINANCIAL, INC	2010
FIDELITY NATIONAL INFORMATION SERVICES, INC.	2015	TCF FINANCIAL CORPORATION	Not found
CHARLES SCHWAB CORPORATION, THE	2012	SYNOVUS FINANCIAL CORP	2010
STATE STREET CORPORATION	2014	FIRST CITIZENS BANCSHARES	2015
SYNCHRONY FINANCIAL	Not found	FLAGSTAR BANCORP INC	2016
AMERIPRISE FINANCIAL, INC.	Not found	AFFILIATED MANAGERS GROUP, INC.	Not found
DISCOVER FINANCIAL SERVICES	Not found	INTERACTIVE BROKERS GROUP INC	2012
SQUARE, INC.	2014	BOK FINANCIAL CORPORATION	2016
MARKEL CORP.	2016	BGC PARTNERS INC	Not found
W. R. BERKLEY CORP	Not found	COVANTA HOLDING CORPORATION	Not found
RAYMOND JAMES FINANCIAL INC	2016	SLM CORPORATION	Not found
ALLY FINANCIAL INC	2015	SEI INVESTMENTS COMPANY	2014
FIFTH THIRD BANCORP	2016	CREDIT ACCEPTANCE CORPORATION	Not found
GLOBAL PAYMENTS INC	Not found	WINTRUST FINANCIAL CORPORATION	Not found
ALLEGHANY CORP	Not found	FIRSTCASH, INC.	Not found
AMERICAN FINANCIAL GROUP INC	Not found	EAST WEST BANCORP, INC	Not found
ARTHUR J. GALLAGHER & CO.	Not found	SIGNATURE BANK	Not found
FIRST AMERICAN FINANCIAL CORPORATION	Not found	COMPASS DIVERSIFIED HOLDINGS	Not found
CITIZENS FINANCIAL GROUP INC.	Not found	WEX INC.	2016
KEYCORP	2013	UNIVERSAL INSURANCE HOLDINGS INC.	Not found
REGIONS FINANCIAL CORPORATION	Not found	ICF INTERNATIONAL, INC.	2015
NORTHERN TRUST CORPORATION	2015	FEDERATED HERMES, INC.	Not found
INTERCONTINENTAL EXCHANGE, INC.	Not found	CULLEN/FROST BANKERS, INC	2012
T. ROWE PRICE GROUP, INC	Not found	NELNET	2015
LPL FINANCIAL HOLDINGS INC.	Not found	COWEN INC	Not found
M&T BANK CORPORATION	Not found	MORNINGSTAR, INC.	2013
NASDAQ, INC.	2011	COMMERCE BANCSHARES, INC.	2011

VALLEY NATIONAL BANCORP	Not found	RENASANT CORPORATION	Not found
UMPQUA HOLDINGS CORPORATION	2015	WSFS FINANCIAL CORPORATION	2015
UMB FINANCIAL CORPORATION	2014	UNITED COMMUNITY BANKS, INC	2014
HANCOCK WHITNEY CORP	Not found	FIRST INTERSTATE BANCSYSTEM, INC	Not found
ASSOCIATED BANC-CORP.	Not found	AXOS FINANCIAL, INC	Not found
WESTERN ALLIANCE BANCORPORATION	2016	HEARTLAND FINANCIAL USA, INC.	2013
PIPER SANDLER COMPANIES	Not found	WESBANCO, INC	Not found
NAVIENT CORPORATION	Not found	INDEPENDENT BANK GROUP, INC.	Not found
MONEYGRAM INTERNATIONAL, INC.	2013	CATHAY GENERAL BANCORP INC	Not found
FNB CORPORATION	2015	NEW RESIDENTIAL INVESTMENT CORP.	Not found
WEBSTER FINANCIAL CORP	Not found	ARES CAPITAL CORPORATION	2015
PROSPERITY BANCSHARES, INC	Not found	NORTHWEST BANCSHARES INC	2015
NEW YORK COMMUNITY BANCORP, INC	2016	KINSALE CAPITAL GROUP INC.	Not found
PACWEST BANCORP	Not found	OFG BANCORP	2016
OPPENHEIMER HOLDINGS INC	2014	HOPE BANCORP INC	Not found
SOUTH STATE CORP	Not found	META FINANCIAL GROUP, INC	Not found
RLI CORP.	2012	CRA INTERNATIONAL, INC.	2016
PINNACLE FINANCIAL PARTNERS, INC.	Not found	CUSTOMERS BANCORP INC	Not found
RADIAN GROUP INC	Not found	HCI GROUP INC.	2012
UNITED FIRE GROUP INC.	Not found	FIRST MERCHANTS CORPORATION	Not found
AMERIS BANCORP	Not found	ELEVATE CREDIT, INC.	2015
WALKER & DUNLOP INC	2013	SANDY SPRING BANCORP, INC	Not found
PJT PARTNERS INC.	Not found	CVB FINANCIAL CORP	Not found
WADDELL & REED FINANCIAL	Not found	NBT BANCORP, INC.	Not found
UNITED BANKSHARES, INC.	2016	NIC INC.	Not found
CRAWFORD & CO	2010	GREAT WESTERN BANCORP, INC	Not found
STERLING BANCORP INC	Not found	OCEANFIRST FINANCIAL CORP	Not found
TEXAS CAPITAL BANCSHARES, INC	Not found	PROVIDENT FINANCIAL SERVICES, INC.	2015
MOELIS & COMPANY	2015	BERKSHIRE HILLS BANCORP INC	2014
LENDINGTREE, INC.	2015	EAGLE BANCORP, INC.	Not found
B. RILEY FINANCIAL, INC.	Not found	SERVISFIRST BANCSHARES, INC.	2015
SIMMONS FIRST NATIONAL CORPORATION	Not found	FIRST COMMONWEALTH FINANCIAL CORP.	2013
ARTISAN PARTNERS ASSET MANAGEMENT INC.	2015	HOMESTREET INC	2016
OCWEN FINANCIAL CORP	2014	MERCHANTS BANCORP	Not found
HURON CONSULTING GROUP INC.	2015	TRIUMPH BANCORP, INC	2016
FULTON FINANCIAL CORPORATION	2016	S & T BANCORP, INC.	Not found
PROASSURANCE CORP	Not found	REGIONAL MANAGEMENT CORP.	Not found
CURO GROUP HOLDINGS CORP.	Not found	NATIONAL BANK HOLDING CORPORATION	2016
OLD NATIONAL BANCORP	Not found	1ST SOURCE CORPORATION	Not found
EZCORP INC	Not found	REPUBLIC BANCORP INC.	Not found
INVESTORS BANCORP, INC	Not found	LENDINGCLUB CORP	2014
TOWNE BANK	Not found	GREENHILL & CO., INC.	Not found
GLACIER BANCORP, INC	Not found	VERITEX HOLDINGS, INC.	Not found
HALLMARK FINANCIAL SERVICES INC.	Not found	AMERISAFE INC.	Not found
FIRST BANCORP	2014	TOMPKINS FINANCIAL CORP	Not found
RESOURCES CONNECTION INC	2016	WATERSTONE FINANCIAL, INC	2014
FIRST MIDWEST BANCORP, INC	2013	PREMIER FINANCIAL CORP	2015
TRUSTMARK CORPORATION	Not found	BANCORP, INC., THE	Not found
HOME BANCSHARES, INC.	Not found	BROOKLINE BANCORP INC	Not found
MARKETAXESS HOLDINGS INC.	2015	QCR HOLDINGS, INC.	Not found
ENOVA INTERNATIONAL, INC.	2013	UNIVEST FINANCIAL CORPORATION	2014

Appendix C: Variables overview

Code	Description
Assets	Total assets year end
Asset turnover	Total revenue / Total assets
Debt	Total debt year end
Debt/equity	Debt to equity ratio
Equity	Total equity year end
Revenue	Total revenue year end
Net income	Net income year end
Number employees	Number of employees at year end
Profit margin	Profit margin, percentage of net income / revenu
ROA	Return on assets year end
ROE	Return on equity year end
Tobins Q	Firm market capitalization / total assets
EPS	Earnings per share
P/E ratio	Price to earnings ratio year end
P/B ratio	Price to book ratio year end
BDI	Dummy variable for big data adoption (Implementation article found between 2010 and 2016)
EBDA	Dummy variable for big data adoption (Implementation between 2010 and 2013)
Year article	Year the article about big data implementation was found
Ln_FSIZE	Natural logarithm of total assets at the end of the year
Leverage	Total company's debt / shareholders equity - (Debt to equity ratio)

Appendix D: Assumptions regression

D-1: Shapiro-Wilk Test

Tests of Normality: Shapiro- Wilk			
Variable	**Statistic**	**df**	**Sig.**
Profit margin	0,984	208	.019
ROE	0,481	208	.000
ROA	0,633	208	.000
Debt/Equity	0,059	208	.000
Asset turnover	0,584	208	.000
EPS	0,589	208	.000
P/E ratio	0,418	208	.000
P/B ratio	0,291	208	.000
Tobin's Q	0,405	208	.000
Ln_FSIZE	0.964	208	.000

D-2: Collinearity statistics

D-2 VIF values ROE

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,887	1,128
Ln_FSIZE	0,106	2,167
Profit margin	0,510	1,654
Asset turnover	0,206	2,008
Debt/equity	0,891	1,123
ROA	0,249	3,410
Tobin's Q	0,285	3,468
EPS	0,211	1,174
P/E ratio	0,708	1,413
P/B ratio	0,716	1,397

a Dependent Variable: ROE

D-2 VIF values ROA

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,896	1,120
Ln_FSIZE	0.671	2,144
Profit margin	0,549	1,613
Asset turnover	0,253	3,947
Debt/equity	0,829	1,206
ROE	0,622	1,607
Tobin's Q	0,474	1,852
EPS	0,213	1,169
P/E ratio	0,740	1,351
P/B ratio	0,705	1,419

a Dependent Variable: ROA

D-2 VIF values Profit margin

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,882	1,134
Asset turnover	0,609	1,642
Tobin's Q	0,286	3,499
EPS	0,865	1,156
P/E ratio	0,693	1,444
P/B ratio	0,738	1,355
Ln_Size	0,477	2,095
ROE	0,58	1,724
ROA	0,26	3,843

a Dependent Variable: Profit margin

D-2 VIF values Debt/Equity

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,884	1,131
Asset turnover	0,615	1,627
Tobin's Q	0,292	3,425
EPS	0,867	1,153
P/E ratio	0,739	1,354
P/B ratio	0,749	1,336
Ln_Size	0,488	2,051
ROE	0,581	1,722
ROA	0,263	3,800
profit margin	0,761	1,314

a Dependent Variable: Debt/equity

D-2 VIF values Asset turnover

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,885	1,131
Tobin's Q	0,292	3,424
EPS	0,883	1,132
P/E ratio	0,761	1,313
P/B ratio	0,751	1,332
Ln_Size	0,492	2,033
ROE	0,582	1,717
ROA	0,286	3,498
profit margin	0,761	1,314

a Dependent Variable: Asset turnover

D-2 VIF values EPS

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,883	1,132
Tobin's Q	0,287	3,484
P/E ratio	0,690	1,450
P/B ratio	0,737	1,356
Ln_Size	0,488	2,051
ROE	0,577	1,733
ROA	0,283	3,528
profit margin	0,756	1,323

a Dependent Variable: EPS

D-2 VIF values P/E ratio

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,881	1,135
Tobin's Q	0,341	2,935
P/B ratio	0,738	1,355
Ln_Size	0,471	2,121
ROE	0,577	1,732
ROA	0,302	3,311
profit margin	0,767	1,304
EPS	0,855	1,169

a Dependent Variable: P/E ratio

D-2 VIF values P/B ratio

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,884	1,131
Tobin's Q	0,318	3,141
Ln_Size	0,473	2,116
ROE	0,582	1,719
ROA	0,283	3,53
profit margin	0,738	1,356
EPS	0,856	1,169
P/E ratio	0,69	1,449

a Dependent Variable: P/B ratio

D-2 VIF values Tobin's Q

Collinearity Statistics		
Variable	**Tolerance**	**VIF**
Big data adoption	0,91	1,098
Ln_Size	0,471	2,125
ROE	0,584	1,712
ROA	0,578	1,73
profit margin	0,736	1,359
EPS	0,858	1,165
P/E ratio	0,821	1,218
P/B ratio	0,82	1,219

a Dependent Variable: Tobin's Q

Appendix E: Robustness checks (Split sample & Lagged variables)

E-1 Robustness check hypothesis 1 (Variables 2017 - 2018)

OLS regression results Hypothesis 1

Dependent variable	**Profit margin**		**ROE**		**ROA**		**Debt/equity**		**Asset turnover**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Big data implementation (BDI)	.008		.074		.054		.063		.034	
	(.123)		(1.422)		(.922)		(.902)		(.535)	
Control variables:										
Profit margin (PM)			.290***	.291***	.401***	.385***	.255***	.266***	-.473***	-.473
			(4.924)	(4.924)	(6.316)	(5.803)	(2.893)	(2.915)	(-7.415)	(-7.433)
Asset turnover (ATO)	-.450***	-.410***	.413***	.416***	.597***	.648***	-.012	-.009		
	(-7.415)	(-7.015)	(7.175)	(7.215)	(9.457)	(10.019)	(-.155)	(-.122)		
Leverage (LEV)	.176***	.176***	.555***	.560***	-.115*	-.101*			-.010	-.008
	(2.893)	(2.793)	(10.644)	(10.728)	(-1.747)	(-1.728)			(-.155)	(-.122)
Ln_Size	-.120*	-.107*	-.029	-.011	.041	.013	.050	.066	-.077	-.069
	(-1.933)	(-1.983)	(-.542)	(-.210)	(.757)	(.224)	(.711)	(.957)	(-1.197)	(-1.104)
Adjusted R2	.248	.251	.469	.467	.328	.328	.040	.041	.210	.213
Observations	208	208	208	208	208	208	208	208	208	208

*Notes: The independent variables are 2017 figures, the dependent variables are 2018 figures. This table reports the standardized coefficients. t-statistics are in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1 show significance at the 1%, 5%, and 10% level.*

E-2 Robustness check hypothesis 2 (Variables 2017 - 2018)

OLS regression results Hypothesis 2								
Dependent variable	**EPS**		**P/E ratio**		**P/B ratio**		**Tobin's Q**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Big data implementation (BDI)	.014		-.048		.118*		.137**	
	(.205)		(-.678)		(1.671)		(2.115)	
Control variables:								
Profit margin (PM)	.279***	.279***	-.193**	-.194**	.161**	.163**	.208***	.210***
	(3.719)	(3.730)	(-2.408)	(-2.418)	(2.021)	(2.026)	(2.862)	(2.856)
Asset turnover (ATO)	.048	.049	-.022	-.024	.166**	.171**	.482***	.488***
	(.657)	(.667)	(-.282)	(-.308)	(2.129)	(2.184)	(6.785)	(6.811)
Leverage (LEV)	.093	.094	.005	.002	.041	.049	-.085	-.076
	(1.407)	(1.426)	(.073)	(.030)	(.585)	(.689)	(-1.319)	(-1.178)
Ln_Size	.294***	.297***	-.069	-.080	.044	.073	.002	.035
	(4.835)	(4.575)	(-.961)	(-1.156)	(.622)	(1.044)	(.032)	(.545)
Adjusted R2	.140	.144	.015	.018	.026	.017	.190	.176
Observations	208	208	208	208	208	208	208	208

*Notes: The independent variables are 2017 figures, the dependent variables are 2018 figures. This table reports the standardized coefficients. t-statistics are in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1 show significance at the 1%, 5%, and 10% level.*

E-3 Robustness checks hypothesis 3 (Variables 2017 - 2018)

OLS regression results Hypothesis 3																		
Dependent variable	**Profit margin**		**ROE**		**ROA**		**Debt/equity**		**Asset turnover**		**EPS**		**P/E ratio**		**P/B ratio**		**Tobin's Q**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Early big data adopter (EBDA)	.063		.076		.011		-.068		.046		-.059		-.013		.112		-.017	
	(.712)		(1.162)		(.148)		(-.688)		(.481)		(-.614)		(-.133)		(1.172)		(-.190)	
<u>Control variables:</u>																		
Profit margin (PM)			.428***	.431***	.515***	.515***	.302***	.300***	-.373***	-.372***	.277***	.274***	-.051	-.052	.303***	.308***	.284***	.283***
			(5.989)	(6.027)	(6.505)	(6.549)	(2.896)	(2.892)	(-3.779)	(-3.785)	(2.645)	(2.632)	(-.472)	(-.480)	(2.899)	(2.941)	(2.897)	(2.907)
Asset turnover (ATO)	-.462***	-.459***	.418***	.422***	.717***	.718***	-.012	-.016			-.003	-.007	.267**	.266**	.384***	.390***	.544***	.543***
	(-5.249)	(-5.236)	(6.044)	(6.105)	(9.366)	(9.437)	(-.113)	(-.150)			(-.034)	(-.068)	(2.530)	(2.540)	(3.795)	(3.857)	(5.733)	(5.760)
Leverage (LEV)	.230***	.226**	.531***	.525***	-.153**	-.154**			.031	.028	.074	.078	-.088	-.087	-.093	-.102	-.091	-.089
	(2.628)	(2.593)	(7.905)	(7.826)	(-2.054)	(-2.082)			(.317)	(.284)	(.751)	(.802)	(-.857)	(-.854)	(-.948)	(-1.035)	(-.983)	(-.977)
Ln_Size	-.093	-.084	-.019	-.008	-.026	-.024	.042	.032	-.126	-.119	.282***	.273***	-.035	-.037	.021	.037	-.005	-.008
	(-1.048)	(-.957)	(-.293)	(-.122)	(-.354)	(-.338)	(.420)	(.324)	(-1.313)	(-1.263)	(2.944)	(2.894)	(-.355)	(-.381)	(.218)	(.395)	(-.056)	(-.086)
Adjusted R2	.249	.253	.592	.591	.499	.504	.061	.066	.111	.119	.125	.131	.052	.062	.128	.124	.232	.240
Observations	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

*Notes: The independent variables are 2017 figures, the dependent variables are 2018 figures. This table reports the standardized coefficients. t-statistics are in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ show significance at the 1%, 5%, and 10% level.*

E-4 Robustness checks Hypotheses 1 (variables 2019 - 2020)

OLS regression results Hypothesis 1

Dependent variable	**Profit margin**		**ROE**		**ROA**		**Debt/equity**		**Asset turnover**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Big data implementation (BDI)	.032		.099*		.084		.078		-.013	
	(.454)		(1.682)		(1.311)		(1.108)		(-.208)	
<u>Control variables:</u>										
Profit margin (PM)			.212***	.212***	.291***	.290***	.208***	.207***	-.463***	-.463***
			(3.063)	(3.044)	(4.045)	(4.031)	(2.639)	(2.631)	(-7.341)	(-7.356)
Asset turnover (ATO)	-.143**	-.143**	.254***	.256***	.511***	.513***	.000	.002		
	(-2.057)	(-2.059)	(3.693)	(3.710)	(7.172)	(7.185)	(.001)	(.022)		
Leverage (LEV)	.061	.064	-.459***	-.459***	-.100	-.101			-.031	-.031
	(.882)	(.919)	(-7.577)	(-7.565)	(-1.599)	(-1.608)			(-.492)	(-.491)
Ln_Size	-.096	-.088	-.005	.018	-.011	.009	.063	.082	-.069	-.072
	(-1.351)	(-1.282)	(-.078)	(.297)	(-.167)	(.149)	(.891)	(1.188)	(-1.068)	(-1.153)
Adjusted R2	.015	.019	.259	.254	.204	.201	.033	.032	.202	.206
Observations	208	208	208	208	208	208	208	208	208	208

*Notes: The independent variables are 2019 figures, the dependent variables are 2020 figures. This table reports the standardized coefficients. t-statistics are in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1 show significance at the 1%, 5%, and 10% level.*

E-5 Robustness checks Hypotheses 2 (variables 2019 - 2020)

OLS regression results Hypothesis 2								
Dependent variable	**EPS**		**P/E ratio**		**P/B ratio**		**Tobin's Q**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Big data implementation (BDI)	.101		.102		.166***		.151**	
	(1.459)		(1.450)		(2.651)		(2.272)	
Control variables:								
Profit margin (PM)	.137*	.137*	-.123	-.124	.216***	.215***	.206***	.205***
	(1.754)	(1.741)	(-1.555)	(-1.559)	(3.069)	(3.010)	(2.752)	(2.711)
Asset turnover (ATO)	.069	.071	.014	.016	.348***	.351***	.407***	.410***
	(.891)	(.916)	(.177)	(.203)	(4.991)	(4.969)	(5.477)	(5.464)
Leverage (LEV)	.058	.057	-.018	-.019	-.353***	-.354***	-.061	-.062
	(.854)	(.839)	(-.259)	(-.272)	(-5.744)	(-5.685)	(-.934)	(-.946)
Ln_Size	.217***	.241***	.114	.138	.034	.074	-.020	.017
	(3.090)	(3.528)	(1.597)	(1.993)	(.545)	(1.193)	(-.294)	(.251)
Adjusted R2	.056	.051	.028	.022	.238	.216	.135	.117
Observations	208	208	208	208	208	208	208	208

*Notes: The independent variables are 2019 figures, the dependent variables are 2020 figures. This table reports the standardized coefficients. t-statistics are in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1 show significance at the 1%, 5%, and 10% level.*

E-6 Robustness check hypothesis 3 (variables 2019 - 2020)

OLS regression results Hypothesis 3

Dependent variable	**Profit margin**		**ROE**		**ROA**		**Debt/equity**		**Asset turnover**		**EPS**		**P/E ratio**		**P/B ratio**		**Tobin's Q**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Early big data adopter (EBDA)	.090		-.067		.040		-.064		.038		-.056		-.115		.039		-.025	
	(.887)		(-.803)		(.452)		(-.641)		(.421)		(-.553)		(-1.141)		(.460)		(-.270)	
<u>Control variables:</u>																		
Profit margin (PM)			.276***	.271***	.318***	.321***	.258**	.253**	-.473***	-.471***	.099	.095	-.166	-.175	.320***	.323***	.297***	.295***
			(2.928)	(2.885)	(3.149)	(3.201)	(2.285)	(2.253)	(-5.227)	(-5.237)	(.865)	(.833)	(-1.456)	(-1.535)	(3.299)	(3.352)	(2.804)	(2.806)
Asset turnover (ATO)	-.133	-.129	.254***	.249***	.557***	.561***	.024	.019			.016	.011	.069	.060	.467***	.470***	.499***	.497***
	(-1.294)	(-1.258)	(2.690)	(2.645)	(5.500)	(5.568)	(.214)	(.169)			(.137)	(.099)	(.600)	(.521)	(4.801)	(4.866)	(4.705)	(4.720)
Leverage (LEV)	.087	.087	-.549***	-.548***	-.100	-.100			-.082	-.083	.055	.056	.011	.012	-.363***	-.364***	-.039	-.038
	(.854)	(.852)	(-6.523)	(-6.525)	(-1.108)	(-1.118)			(-.912)	(-.924)	(.539)	(.547)	(.105)	(.118)	(-4.192)	(-4.215)	(-.410)	(0.409)
Ln_Size	-.070	-.057	.023	.013	-.040	-.033	.072	.062	-.108	-.103	.234**	.225**	.148	.130	.058	.063	-.016	-.019
	(-.688)	(-.567)	(.278)	(.161)	(.441)	(-.379)	(.712)	(.626)	(-1.210)	(-1.167)	(2.299)	(2.250)	(1.458)	(1.300)	(.668)	(.748)	(-.165)	(-.209)
Adjusted R2	.038	.020	.333	.336	.234	.240	.028	.034	.233	.230	.016	.023	.023	.020	.293	.299	.161	.169
Observations	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

*Notes: The independent variables are 2019 figures, the dependent variables are 2020 figures. This table reports the standardized coefficients. t-statistics are in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ show significance at the 1%, 5%, and 10% level.*