# The effect of Data Analytics on audit efficiency

**Abstract**

Data Analytics (DA) is the application of certain software tools which analyse data from and about the client to gain insights into the clients operations and financial numbers. The application of DA in the audit is relatively new. There exists little empirical research on the effects of DA on the audit. Therefore this paper investigates the effect of Data Analytics (DA) on audit efficiency. This study obtains data from a public accounting firm about their audit engagements and the application of DA during their audit engagements. Audit efficiency is measured by audit hours, audit costs and billed costs. Results show that DA does not lead to an increase in audit efficiency for all efficiency measures. This implicates that audit profession needs to become more experienced with DA to overcome the existing challenges before DA can improve efficiency.

Author: T.H.H. Bender

Supervisor: J. Pierk

Date: 15-08-2017

Student number: 357752

# Table of Contents

# Table of tables

## Table of figures

# 1. Introduction

The profession of auditing is slowly changing. The application of Data Analytics (DA) in accounting creates a new way of auditing which many believe to be the audit of the future. DA in auditing is used to analyse, identify and extract useful information from data of the client for planning or performing the audit (Byrnes, et al., 2014). More commonly, DA is used to gain additional and more thorough insights into the audited company. With the help of specially developed tools, auditors can use data from the client's system to investigate whether the financial statements are truthful and free of material misstatements. Nonetheless, there exists little research on the effects of DA on the audit. Therefore this paper will investigate:

*Does Data Analytics improve audit efficiency?*

Currently the true effect of DA on the audit has not looked into yet. The most likely reason is probably because financial researchers have no access to data which can help tackle the question. For external stakeholders it is not possible to identify whether a public accounting firm applies DA during the audit. There exist no rules or regulations which state that an accountant has to provide such information (Alles, 2015). The only way it is possible to know whether DA is applied, is by acquiring such data from the public accounting firms themselves. Because DA is relatively new and accountants themselves do not fully know the effects of DA, they are not keen on providing such data to the public. Accountants have been under a lot of pressure lately and disclosing such data could lead to more criticism of the profession.

Although there exists little research about the use of DA in the audit, it is important to know how DA affects the audit. DA is a growing market (Gershkoff, 2015). Other professions, like consulting, have already widely adopted the use of DA for their current daily business (Cao, et al., 2015). Current literature such as Liddy (2015) describes how DA can improve insights and risk assessment. Byrnes, et al. (2014) expect audits to become more efficient with the help of DA. Also Earley (2015) believes auditor benefit by allowing more transactions to be tested. This explains why a couple of years ago public accounting firms started to invest and implement DA in their audits (NBA, 2013). Furthermore, clients of public accounting firms will require their financial information to be audited on a larger scale due to the large increase in data generation, which is not possible with the current way of auditing (Alles, 2015). Because DA in auditing will grow in the coming years it is important to have reliable answers on the effects of DA in the audit instead of theories which are based on expectations.

As discussed, data about the utilization of DA during the audit is difficult to obtain. In this paper, data is acquired from a North-West European public accounting firm which regularly applies DA in its audits. An overview is obtained of the firm's audit planning from the fiscal years 2013-2017. This dataset contains all audit hours, audit costs and billed costs per audit engagement. The planning of the DA team is also acquired to determine the hours spent on audit assignments by the DA team members. By measuring the number of hours of the DA department the paper determines whether DA is applied during the audit or not.

The research provides evidence that DA does not affect efficiency during the audit. All three dependent variables, audit hours, audit costs and billed costs, do not change when DA is implemented. There are several reasons why audit efficiency is unaffected by DA. First of all, the implementation of DA costs a lot of time. The auditor has to constantly consult the client and DA team in order to make application of DA possible. Additionally, the DA team develops the DA tool used during the audit. As the DA team has no experience with auditing, the tool might not function to the accountant's needs. Moreover, accountants have to get used to the "new way" of auditing. The application of DA in audits requires a different skillset and a new approach to the audit (Earley, 2015). Besides the required skills and knowledge, auditors might be afraid to trust upon the outcome of DA. This is because there is no legislation on the usage of DA during the audit. Therefore auditors, after using DA in their audit, might still want to perform work which is stated as sufficient according to regulation. Lastly, although DA can process a lot of data, it costs a lot of time to prepare the data in such a way it can be processed. Data comes in different forms and therefore should be evaluated carefully before being used (Brown-Liburd, et al., 2015). Using the wrong data can lead to wrong conclusions which might have a big impact on the assurance given by the auditor.

This study contributes to the current literature because it provides empirical insights in the use of DA during the audit. Currently DA is often described as the more efficient and new way of auditing. This research shows that the current application of DA is not more efficient than normal audits. Public accounting firms should carefully evaluate the challenges of DA and ensure these challenges are overcome before investing in DA.

Although this study shows DA does not improve audit efficiency, public accounting firms should keep investing in DA. The application of DA in auditing is still in its infancy. Accounting firms are investigating by trial and error to learn how to integrate DA in the audit. Once the barrier of entry for DA has been surpassed, the efficiency of audits might increase. Furthermore, besides efficiency, public accounting firms have to take into account the effectiveness of their audit which can be considered to be of higher importance than efficiency. This might be a reason why currently firms are not improving audit efficiency but instead they might be focussing on the effectiveness of DA.

# 2. Theoretical Background

## 2.1 Data analytics defined

Big Data is large and unstructured data produced by people, transactions, and machines. Big Data is commonly labelled by four V's: volume, velocity, variety, and veracity (Kessel, 2014). Volume describes the amount of the data being created. Velocity relates to the speed at which data is created and analysed. Because new data is generated so quickly, information becomes irrelevant faster (Coyne, et al., 2017). Variety illustrates the different forms of data which are produced. And veracity expresses the concerns about the reliability of data. The datasets which Big Data consist of are too complex to inspect with standard tools (Nasser & Tariq, 2015). Therefore specific tools are created which can analyse Big Data and produce understandable information.

According to Cao et al. (2015) "Data analytics can be defined as the process of inspecting, cleaning transforming and modelling Big Data to discover and communicate useful information and patterns, suggesting conclusions, and support decision making". With computerized tools it is possible to identify patterns and anomalies in large and unstructured data sets which can help uncover hidden information. DA is already widely adopted in many business areas. By analysing consumer trends and other patterns consultancy firms are able to determine what the best course of action is in order to reduce costs or increase revenues for a certain company.

However, in auditing the use of DA is not very common yet. A problem for the use of DA in audits is that auditors are no longer allowed to audit a firm and give advice about the business activities of a company according to European laws (PWC, 2015). DA adds value in auditing sector by trying to improve the efficiency and effectiveness of the audits rather than providing clients with innovative and competitive insights (Earley, 2015). Therefore DA has to be applied differently in the auditing practice.

To provide a definition of DA more related to auditing this paper follows Byrnes' et al. (2014) statement that "Data Analytics in auditing is the science and art of discovering and analysing patterns, identifying anomalies, and extracting other useful information in data underlying or related to the subject matter of an audit through analysis, modelling, and visualization for the purpose of planning or performing the audit". Moreover Byrnes' et al. (2014) argue there exist two different modes of DA. Exploratory DA is inductive and mostly applied in the planning phase of the audit. It is used to get an understanding of the firm, identify and assess risks and to design additional audit procedures. Confirmatory DA is deductive and is used in the last two phases. In order to provide assurance, the auditor has to perform substantive procedures to ensure the items in the financial statements are correct. Confirmatory DA determines whether the financial statement is free of material misstatements.

## 2.2 The four audit phases

In general an audit is performed in four stages. The first one is planning and risk identification. During this stage the auditor must understand the company, identify the risk and decide what has to be done to provide reasonable assurance. The second step is strategy and risk assessment. The auditor has to determine what the strategy of the company is and assess how big the risks are at the company. In the third stage, execution, the auditor performs substantive procedures. These are tests to check whether information presented in the financial statement is truthful and free of material misstatements. The final phase is called conclusion and reporting. The public accountant has to perform its lasts procedures and make an official statement regarding its conclusion about the financial statement (Hayes, et al., 2014).

DA can be applied in several stages. Especially in the first two stages DA is expected to be beneficial (Cao, et al., 2015). Big data sets contain large unstructured data. Such data is considered messy and less reliable. This makes the outcome of the DA also less reliable. Therefore it is easier to focus on causation instead of correlation. In the first two stages patterns and trends can be identified, which depend more on correlation than causation. Going back to the terms used by Byrnes et al. (2014) DA applied in the first two stages can be regarded as exploratory DA. When data becomes less reliable it becomes more difficult to implement DA in the third and fourth stage (Cao, et al., 2015). Substantive or analytical procedures are sensitive to data, thus unreliable data gives unreliable outcomes. Considering these stages are used to conclude whether the financial statement is free of material misstatements, it is better to apply traditional audit procedures when Big Data cannot be trusted upon.

## 2.3 Challenges

But why is the auditing business behind on the application of DA? Past experiences show that auditors consistently fail to advance in the current technological developments (Manson, et al., 2007). With regard to DA one of the reasons auditors are lagging behind is because they do not have the required skills to apply DA (Earley, 2015). Accountants learn how debits and credits effect the balance sheet or income statement. They understand the consequences of overstating or understating accounts and how risks related to the financial statement can be minimalized. When applying DA, a different skillset needs to be used. DA identifies patterns and correlations which have to be analysed by the auditor (Brown-Liburd, et al., 2015). This requires a different approach to using the data to come to conclusions. Auditors need to become used and familiar with this new way of analysing financial and non-financial data in order to provide a more efficient and effective audit.

In addition, there is also a need for DA specialists (Business.com, 2017). Developing DA tools is not a requirement for an accountant. At this moment with the current increase in Big Data and DA the new and rising demand for such specialists is growing faster than the amount of people who are schooled to become such specialists (Gershkoff, 2015). This implicates that public accounting firms will have problems finding staff which is able to create DA tools. Moreover, a low demand will most likely lead to a high price for DA specialists as well. Besides the fact that DA specialists are scarce, DA specialist are often not familiar with an audit (Alles, 2015). Because they develop the tool, it might be hard for them to develop a tool which can be used effectively and efficiently during an audit.

Moreover the four V's which describe Big Data also bring forth challenges. First of all the high volumes of data do not only create necessities for storage but also demand analytical programs to have high processing rates of data (Nasser & Tariq, 2015). If programs are processing data slowly, efficiency will more likely decrease instead of increase. Variety of data addresses the difficulty of analysing different forms of data. Numbers are different from texts or graphs but they all can contain information which can affect the conclusion an audit.

DA does not only extract and process a lot of data, it also produces much data at a fast pace (Coyne, et al., 2017). The large amount of data output could create an information overload for the auditor (Brown-Liburd, et al., 2015). When interpreting the data output of DA an auditor should be able to evaluate the useful data and ignore the data which is irrelevant. If the auditor fails to do this, it will lead to improper decisions and a lower quality of the audit.

Another concern is the data availability and data integrity. In order to discard independence issues the auditor should be given access to all data without interference of the client. Besides the fact that auditors often do not have full access, it is also difficult for auditors to capture the data in a way which is useful to the auditor (Adrian, 2013). In order to make sure the data which is analyzed is trustworthy and correct, an IT specialist is needed to sort and process the data. Next to these challenges the data depends on internal and external sources. Therefore the auditor needs to secure that the data is complete, unbiased and has not been altered if he wants to use the data for the audit. According to the PCAOB this is not possible if the auditor relies on externally produced data (Whitehouse, 2014).

Ambiguity of data is another challenge which has to be taken into account. If auditors are not comfortable to make thorough decisions when information is vague or unclear, they tend to neglect other information once the first solution has presented itself (Brown-Liburd, et al., 2015). Precisely because DA produces so much information, an auditor should incorporate all relevant information when making a decision.

Currently, after having spoken with various members of a public accounting firm, accountants consider extraction of data the biggest challenge for applying DA. Extraction of data can be performed by either the client or by the auditor themselves. An advantage of extracting the data by the auditor increases the integrity of the data because the data is directly produced by the accounting system. This lowers the likelihood of data being tampered with and increases its reliability. However extracting data by the auditor themselves requires more time and therefore is a more costly process. Besides that, the accounting system differs client to client. This means that the auditor needs to have different extraction tools at his disposal to extract data. These tools are developed by data specialists which increases audit costs. Besides the requirement of different extraction tools, extracting data is not the auditor's area of expertise. Therefore a special team might be required to extract the data from the client's system in a correct manner.

After the data is extracted, it is not ready for use yet. The data needs to be inserted in the analytics system of the auditor in order to perform DA. The auditor has to specify which data needs to be put into the analytics system to DA. This is another process which requires a team of specialists. The whole process of extracting the data up until the moment it is ready for use can take a few weeks. Because the auditor has to spent more time on obtaining the correct data this can counter the increase of efficiency during the actual audit.

Lastly, the advance of DA in auditing will depend greatly on the acceptance by the public and legislators (Earley, 2015). First auditors and legislators have to deal with the expectation gap. The expectation gap describes the public's expectation of the auditors' role and responsibilities which exceeds the auditors' performance perceived by the public (Ruhnke & Schmidt, 2014). In the past few years the public and legislators have demanded more effective audits and want auditors to provide more assurance (Liddy, 2015). Auditors might expect the gap to become smaller due to more advanced DA. But they should keep in mind to constantly describe what is possible and what is not possible in order to not increase the gap. Moreover, regulations and standards are not designed for DA in audits (Alles, 2015). Although the current standards do not limit the use of DA in audits, they also do not describe how DA can and should be used. Besides the required knowledge and skills for applying DA by the auditor, the fact there is no legislation may be an even bigger problem. Auditors would rather follow the rules and fulfil all requirements according to the legislation than providing more assurance with DA. Within the public accounting firm this is a concern for further development and acceptance of DA within the auditing industry. Every accountant who is involved in DA has concerns about legislation and the progress of using DA in the audit. Especially since partners are personally held accountable, this creates the incentive to audit according to the standards of the regulators instead of using DA to provide more assurance. Considering it takes time to adjust standards and laws, the willingness to apply DA might be limited and therefore might decrease its growth in the auditing business.

From this it becomes obvious that in order for an auditor to apply DA, the auditor should obtain enough knowledge and skills to understand the data used for DA and the information produced by DA. As DA is relatively new, it can be assumed that many auditors do not have these requirements yet. This may lead to complications in the current use of DA.

## 2.4 Benefits

Nevertheless public accountants keep investing in the application of DA during the audit (NBA, 2013). According to EY it is necessary to keep investing in DA because auditors are able to provide more assurance and a more relevant audit (Ramlukan, 2015).

First of all, DA makes it possible to test 100% of the transactions (Byrnes, et al., 2014). It gives the auditor the opportunity to detect anomalies which helps auditors to assess risks and identify trends. This will improve audits year after year, since auditors learn during the audit on which areas they should put extra focus. In addition, better insights in the client's financial system helps auditors forecast estimates or going concern issues more easily. With the help of specifically developed tools, it becomes easier to evaluate non-financial data. Such data can provide extra insights in business risks and other areas (Earley, 2015). As can be seen in table 1, DA can be used to enhance fraud detection. Due to more available data computerized tools, which use mathematical principles, fraud can more easily detected.

**Table 1. Types of data and their role in the audit**

| Type of Data | Current Practice | Potential Future Practice |
|---|---|---|
| Financial Data | Auditors collect and test a sample of transactions and use judgement on areas which are difficult to test | DA tests 100% of the transactions.<br>DA will identify anomalies and patterns in the provided data. This will assist further test work and possibly uncover misstatements. The auditor has to use judgement in assessing anomalies which are uncovered. |
| Non-Financial Data | Rarely used during audits, unless the auditor has specific knowledge of the client or industry. | DA developed to run models or predictive analyses to aid auditors in identifying business risks and areas of focus during planning, aid in fraud detection, and help evaluate and assess going concern. |

Besides the fact that auditors can improve their audit, another reason for the enthusiasm of public accountants for DA is the expectancy that audits will become more efficient (Alles, 2015). Due to more insights the auditor does not only have a better overview of the client’s financial position, but it can also help identify the auditor which accounts require more or less attention. Statistical programs can lighten the workload of auditors by checking whether transactions are abnormal because they do not fulfil certain standard criteria. If their exist any irregularities, the auditor is able to immediately focus on these inconsistencies. This is more effective than spending time checking a random sample. This is the an important reason for auditors to enhance in DA. Public accounting firms are active in a competitive market. If auditors do not invest in DA, they might fall behind of their competitors who are able to provide better services (Alles, 2015). Firms which provide audits which are less effective and more costly are not likely to survive for a long time. Figure 1 shows the expected effect of DA on the audit hours.

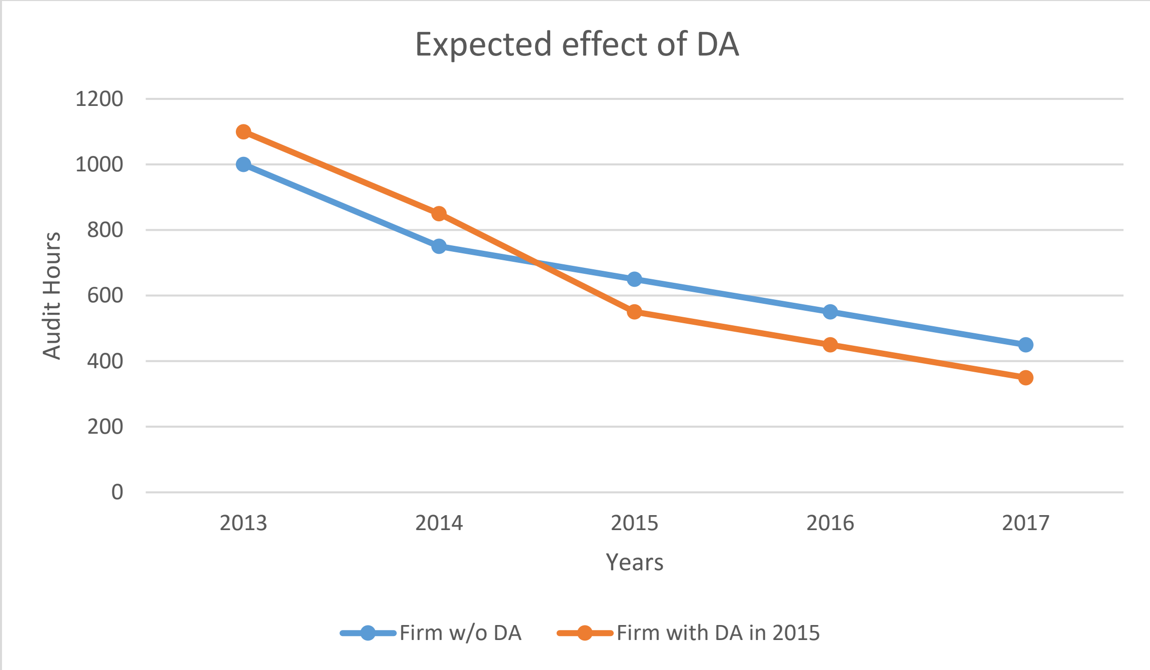



***Figure 1. Expected effect of DA on audit efficiency.***

*The blue line represents the hours required to audit a firm where no DA is applied. The orange line represents the hours required to audit a firm when DA is applied in 2015. The graph shows a steeper decline in audit hours for the year in which DA is applied.*

But public accounting firms' survival do not only depend on the efficiency of their audits. There exists the necessity to invest in DA because Big Data becomes more important for clients (Earley, 2015). Once Big Data becomes a necessity for the strategic business of the entity, clients, regulators and other stakeholders will want auditors to check whether their Big Data is indeed correct and reliable (Alles, 2015). For example, estimates which are based on information produced by Big Data, needs to be checked and confirmed by the auditor once these estimations have a material impact on the financial statement. In order to correctly check whether the information presented by the company is correct, the auditor himself should apply DA to provide reasonable assurance.

Although auditors might not be trained for the usage of DA, the growing market for DA also leads to an increase in developing user-friendly software (Alles, 2015). It is likely that such software will be a facilitating factor in the willingness to adopt DA by the auditors and can explain why we are witnessing an increase in the application of DA in the audit engagements. Although public accountants seem to be to be rather sure that DA proves to be beneficial and adds value to the audit, there does not exist much empirical research on the effects of DA on the audit due to limited data available.

## 2.5 Different forms of Data Analytics during the audit

There exists much literature about the benefits and challenges of DA, there have not been described many real world applications in the literature yet. It is important to provide application examples to visualize DA. Therefore the paper will provide examples how DA is operationalized during an audit. These examples are based on current application of DA tools during real world audits of the same public accounting firm from which the data was obtained.

There are several accounts in the financial statement which, no matter the company or industry, are important to audit. One of these accounts is revenue. Commonly revenue is tested for overstatement because companies prefer to be as profitable as possible. For an investment bank revenue is obtained by customers having to pay for transactions they make. These transactions involve shares or bond investors prefer to buy or sell. Transaction costs are relatively low, but considering there occur a lot of bank transactions like these, a high revenue can be obtained. Because revenue has material value the auditors are required to test this account of the investment bank on management assertions. Management assertions are representations by management about classes of transactions and related accounts in the financial statements (Hayes, et al., 2014). All these transactions are stored on the investment bank's servers. Thus there is data about when these transactions occurred, in which amount, and what the cost for these transactions are. Because all this data is available the auditors of this investment bank wanted to apply DA in order to provide more assurance.

In order to set up a tool which is able analyse all the data the auditors contact the DA department. This team has the skills required to develop a tool which can analyse the data. These skills often involve programming experience because the tools are specially designed programs with their own code. Together with the DA department and the client the auditor determines whether it is possible to create this tool and if they all believe it to be value adding to the audit. After agreeing upon the deal to create the tool, the DA department, IT auditors and financial auditors have to work together to make sure the following things happen. First of all the DA team has to work together with the client IT team, in order to fully understand the system and create a tool which is applicable to the systems data. The financial auditors need to discuss with the DA department what the requirements are for the tool. The investment bank has so many codes and formula's build into their system that it is impossible to write the perfect code. The DA team contacts the auditors when they cannot get the exact same outcome. The auditor has to decide whether the difference can have a material impact on account revenue and when they determined there are no material risks the DA team has to provide a valid reason why the outcomes are not the same. A difference in outcome can be because the codes of the investment bank are too complex or rarely used. When there exists no material risk when the outcomes do not match perfectly and the reasons for not matching them exactly are valid, the DA team will not put in more effort to match the outcomes. Simply because the benefits do not outweigh the costs.

Eventually this process leads to a tool which can, based on the data extracted from the system, calculate what the revenue should be of the year's transactions. This leads to less work for the auditor because the work performed in order to determine whether revenue is correct decreases. For example for every kind of transaction the system should record revenue and bill customers. If for some reason the system does not bill customers for one particular transaction, the audit team can find that the system does not bill customers with that specific tool. Therefore the revenue is overstated because revenue is recorded while it is not received. This leads to a violation of the management assertions and should therefore be identified and solved.

This is an example of customized DA. Customized DA is the application of a custom build in tool for the client. An advantage of such a tool is that it tests the technical connections of a firm's system. Because the tool is developed by a team of specialists in combination with the auditor's expertise of knowing the business, the tool takes away a large proportion of the workload for the auditor for one specific task. This is related to confirmatory DA.

It is not always possible to build a custom tool. Due to the low number of DA specialists, accountants cannot always consult the DA team. For example, the Big Four try to maximize their profits, like most companies. This means that if the DA team members can create an audit tool and earn 90 euros per hour, the firm rather wants them to create a tool for a consulting client if that yields 150 euros per hour.

Another way of applying DA is with the use of more general tools. General tools are not custom designed for the client. The advantage of such tools is the fact that they can be used for all clients without having to consult the DA department. However, just like with custom designed tools, in order to use these general tools, data extraction from the client's system is required. As discussed this brings along challenges, especially because the general tools are not specifically designed for every accounting system. At this point such kind of DA is not yet very common. But in the near future it is very likely there exist common tools, which allow the companies data to be uploaded in a general DA tool. The idea is that this will provide a faster and more focussed audit of the financial statements. Currently there are trials of such programs. An example will be given below.

First of all, one should keep in mind that general DA tools do not test the systems of the client directly, as the custom DA tools can. Therefore they are more closely related to exploratory DA. General DA tools are more applicable to the overall information the data produces and is therefore more suitable for the overall audit process. Therefore the paper will discuss how a general DA tool can be operationalized in the audit.

In the first audit phase, planning and risk identification, the general tool makes it possible to plot client's financial data. The ability to create comparisons over any time period can help the auditor point out which periods are more risky due to a high increase of turnover or other effects.

During risk assessment general DA tools can be used to identify which transactions should be more closely looked at. For instance, if 90% of the purchase orders match with the received goods, the auditors are able to take a closer look to the other 10% and improve their sample. Or the auditor is able to select all journal entries made during free days and see whether entries were made during Sundays or holidays. Another possibility is checking which projects are more risky due to certain criteria, like the time it takes to complete a project.

In the third phase, the auditor can test whether prices, which are made by the management, correspond with market prices. For example, does the value of a piece of land in the system, correspond with the value it would have on the market.

As can be deducted from the above, custom DA tools are more specific and are more likely to induce more direct questions. At this point customized DA tools are more common. Probably because first of all it is easier to create a tool which works on one system than a tool which can be used for multiple systems. Also DA is relatively new in auditing. Thus experimenting with customized tools has lower risks. Although customized tools are currently more common, it might be that in the near future general tools will be more widely used. A reason for this can be because in the end creating multiple tools eventually is more time consuming and costly than building one general DA tool which fits multiple IT platforms.

# 3. Hypothesis development

All the possibilities created by DA to increase audit effectiveness and efficiency audit sound very promising. But as discussed, challenges arise when implementing DA. These challenges might make it difficult for the benefits to become evident. It is therefore important to investigate whether DA is already effective and efficient or whether DA in auditing needs improvement. If the latter is the case, public accounting firms should be cautious about their investment and find possibilities to ensure the challenges are overcome when implementing DA. Although many people believe DA is the future because the opportunities weight out the threats, it is important to investigate where DA in the audit currently stands. Therefore this paper will test whether DA indeed enhances the audit.

According to Byrnes et al. (2014) technological advances in auditing can either be used to increase efficiency or to provide more assurance. Yet most of the time, technology is only used to improve efficiency. In other words, the same level of assurance is given but at a lower cost. This development makes it likely that DA in auditing will mostly be applied to increase the efficiency of the audit. In line with this thought this paper will test the effect of DA on audit efficiency.

In order for audit efficiency to increase, either time spent on the audit or costs of the audit need to decrease. Current literature is positive about DA. However, accountants themselves are more cautious about the challenges DA brings along. Even though public accounting firms are actively investing in DA, they acknowledge that it is a difficult and costly investment. Especially the transition to the new way of auditing is seen as a problem. Accountants need to learn a new way of working and are expected to change the way they look at an audit. Moreover, the tools which they will be using are different than what they are familiar with. Some auditors doubt whether current DA application during the audit improve efficiency and will not be surprised if DA made the audit more costly. However, they do agree upon the fact the audit is more effective with DA and believe effectiveness is more important than efficiency.

On the other hand, as stated by prior literature, DA can measure all transactions and identify anomalies, thus relieving the auditor from looking for anomalies. This will yield a more efficient audit because the auditor has to inspect a smaller sample of observations. Therefore, less time is spent during the planning and identification of risks during the audit. Although accountants are not sure whether efficiency worsen or improves, the paper follows the literature which is positive about the application of DA. This leads to the first hypothesis:

*H1: When data analytics is applied during an audit the amount of time spent to perform the audit decreases.*

Another expectation of DA is the decrease in costs. Reasons for this are the fact time spent on the audit is lower. Thus, considering auditing costs depend on time spent on the audit, the costs for the audit are expected to decrease. However, it is assumed that more experienced auditors are likely to be involved in DA. Therefore, one could state that although time spent during the audit might be less when DA is involved, the average hourly rate per employee might increase. Nevertheless, the paper expects costs to decrease because the number of hours which are saved are likely to be higher than the amount of hours experienced auditors would have to work extra. Moreover, the paper expects that the working hours of experienced auditors will also decrease after the first year DA is applied. Hence the second hypothesis states:

*H2: When data analytics is applied during an audit the costs associated with the audit will decrease.*

It is likely that costs and billed costs go hand in hand. A decrease in costs, will give the public accounting firm a competitive advantage because the firm can lower its audit fee and attract more clients. Moreover, a decrease in working hours of the employees would make it also possible for the employees to work on other clients. Therefore, the third hypothesis is:

*H3: When data analytics is applied during an audit the audit fee for the audit will decrease.*

The hypotheses are stated in alternative form.

# 4. Research Design

In this section the paper will introduce the model and variables on which the analysis is based upon. The paper will investigate what the relation is between the use of DA and audit efficiency. The predictive validity framework (“Libby boxes”) is presented in the appendix A of this paper. The Libby boxes show how the conceptual relation, examined in this thesis, will be operationalized in the research design. The unit of analysis is a year, because the financial statement is audited once per year. Check appendix B for an overview of the variables used in the regressions.

## 4.1 The dependent variable

To determine whether the efficiency of the audit improves when DA is applied efficiency needs to be operationalized. Efficiency can be defined in several different ways. First of all, efficiency describes the amount of work which has to be performed in order to obtain a certain goal and the associated costs for obtaining this goal. Considering auditing a financial statement is a time consuming process, time will be a measurement for efficiency. Material costs for performing an audit is relatively low and is not considered to be a relevant measure. Therefore it can be concluded that costs are mostly related to time spent on the audit by employees of the public accounting firm. Each employee has a job grade, which describes the function a certain employee has. The job grade determines how much an employee costs per hour. Hence, the time spent by the employee auditing a client multiples by a certain amount to determine the cost of that employee. If this is added up for all the employees who performed work related to the audit of the client, total costs related to the audit can be calculated. This provides the second measurement for efficiency, total costs. However, sometimes the client does pay the total costs made. It might be that the client gets a percentage discount per employee or prices are already determined beforehand. Therefore the audit fee can be different from the total costs. The third dependent variable will therefore be total billed costs.

There are three dependent variables to measure efficiency. The first one is time spent on an audit. Time spent will be measured in total hours worked by the audit team since employees of the public accounting firm record their time in hours. It is expected that the amount of hours will decrease because DA should take away some of the workload for the auditors. The second dependent variable is costs of the audit. Audit costs depends on the hours spent on the audit engagement and on the job level of the employee recording hours on that specific engagement. Because the paper expects audit hours to decrease, the paper also expects total costs for the audit to decrease. The last variable to measure efficiency is audit fee. The public accounting firm earns a fee in return for their provided services. Because DA is expected to decrease audit hours, provided services are lower, which would imply that the audit fee decreases as well. All the dependent variables are related to each other. Therefore the study assumes that all the variables will move in the same way. The paper is purely focussing on the costs and potential benefits of DA for the audit and not on the overall costs of applying DA. Therefore hours and costs for the data analytics department which implements DA are excluded from the audit hours and audit engagement costs.

The goal of the application of DA in the audit is to ensure more audit quality and obtain efficiency benefits. It is likely that especially for big audits efficiency benefits can have a major impact on the audit. In order to account for biased results if DA is mainly applied in big audits, the dependent variables have to be deflated. The dependent variables are deflated as follows:

*Deflated Audit Hours = Audit Hours/logarithmic of Total assets*

*Deflated Audit Costs = Audit Costs/logarithmic of Total assets*

*Deflated Billed Costs = Billed Costs/logarithmic of Total assets*

The paper uses the logarithmic of total assets to deflate the dependent variables. Audit hours do not double when assets double, thus if the dependent variable audit hours is divided by total assets, the result might be that big audits become smaller than small audits after this calculation, which would be unrealistic. Therefore to counter the large differences between assets, a more linear distribution of assets, which is created with the logarithmic of total assets, is more suitable to deflate the dependent variables in this case.

### 4.2 The explanatory variable

To determine whether the efficiency during the audit is affected by DA, the effect has to be operationalized. First of all the paper must determine when DA is actually applied. Currently it is difficult to measure the application of general DA tools. Therefore the use of DA will be based on the application of customized DA tools during the audit. The use of DA ("X") will be measured with the help of a dummy variable, which equals 1 for observations in which the auditor makes use of DA during the audit. The dummy variable equals 0 in case the auditor has not applied DA. Based on experiences of the DA team members, it is assumed that DA is applied during an audit when more than 24 hours were spent on an audit assignment by DA team members.

### 4.3 The control variables

Several control variables will be added to the regression in order to ensure there exists a common trend between the control and treatment group. Such control variables will be factors that are likely to be correlated with the paper's measures of efficiency. As discussed we measure efficiency based on hours worked and costs. Although the audit fee is often a fixed amount, the accountant determines the fee based on expected hours and costs. The fixed fee no longer holds when dubious accounts require more investigation or fraud is discovered. Because the research expects the dependent variables to be interrelated, the paper assumes that control variables related to audit fees also apply to hours worked and costs related to the audit. Hay, et al., (2006) performed a meta-analysis of variables which affect audit fees. They evaluate a large amount of audit fee research to test which variables show constant significant results. Important factors are size of the company, company risk and auditor experience. To operationalize these factors we selected the following financial variables. Size of the company often determines more than 70 percent of the audit fee. This element will be defined by total assets, which is the most commonly used account to measure size. To create a more linear distribution of total assets between the companies the paper will use the logarithmic of total assets. This will account for the effect of fixed differences in total assets. It is expected that the dependent variables will increase when total assets increase.

Company risk can be measured by several variables. The first measure will be based on nature of assets and measures inherent risk. Nature of assets evaluates which percentage of total assets consists of receivables and inventory. The value of inventory is based on judgement by management, which increases risk. Likewise, receivables increase risk, because this is revenue which is not yet received. According to Hay, et al., (2006) taking inventory and receivables together is a better proxy then evaluating them separately because it leads to more robust results. The formula for calculating nature of assets is:

*Nature of assets = (Receivables + Inventory)/Total assets*

The next control variable which measures company risk is leverage. This measure describes the company's ability to pay its current obligations. If debt increases compared to equity, the risk of non-payment increases. It is expected that the risk of going-concern increases when the risk of non-payment is higher. A higher going-concern risk requires a more thorough audit which lead to more audit hours and costs (Hay, et al., 2006). Leverage will be measured with the following formula (Hay, et al., 2006):

*Leverage = Debt/(Debt + Equity)*

Another measurement for risk is profitability. The profitability ratio or return on assets, will be used as a variable to measure profitability. If a firm performs badly, there exist a higher going concern risk for the company (Hay, et al., 2006). Therefore the public accounting firm has to perform more work to provide reasonable assurance. According to Hay, et al. (2006) The profitability ratio will be calculated as following:

*Return on assets = Net income/Total assets*

Form of ownership also affects risk. When a company has multiple subsidiaries audits become more complex (Hay, et al., 2006). First of all, intercompany transactions require more attention than regular transactions to ensure there exists no income shifting or other improper accounting. Moreover, if subsidiaries have to be audited as well, audit hours increase. Therefore the amount of subsidiaries a company has will be taken as a control variable to account for the extra hours and costs associated with subsidiaries.

Auditor experience is considered to lower the hours and costs of an audit. When an auditor has to perform the audit of a company a consecutive year the experience gained the previous year, should make the auditor more effective. Therefore it is expected that when an performs multiple audits for a company, the audit hours, total costs and audit fee decrease. The years of experience of the auditor are called auditor tenure.

Normally industry is a control variable which is commonly used in many empirical research of economics. The general consensus is that several industries are harder to audit than other industries which lead to more intensive audits. However, the amount of industry observations in the dataset of this study are too low to use industry as a reliable control variable.

Often auditor size is also regarded as factor which has an effect on the audit fee. For this study auditor size is not applicable because the data used in this study is supplied by one public accounting firm. Thus all the observations are audited by the same company.

## 4.4 The model

This paper uses an empirical research method to test the hypotheses. A difference-in-difference (DID) design will be used to predict the influence of DA on the three dependent variables. DID is a statistical technique which mimics an experimental research using observational data. DID calculates the effect of a treatment on an outcome, by comparing the average change over time for the treatment group with the control group. With the help of Stata, a software which allows for all kinds of statistical analyses, the paper will perform the DID test. The first measure is total hours worked by the auditors during an audit. The regression used to test the first hypothesis takes the following from:

$$\text{SumofHR} = \alpha + \beta 1 * DAt + \beta 2 * \text{POST}t + \beta 3 * DA\text{XPOST}t + \varepsilon t$$

To determine whether the costs differ between audits where DA was applied, the paper uses the same variables as total hours worked. Therefore the regression to test the second hypothesis looks quite similar:

SumofEUR = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \varepsilon t$

The third hypothesis, which measures billed costs, is:

SumofEUR2 = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \varepsilon t$

The variable DA explains the mean difference between treatment group and control group prior to the implementation of DA. The effect of POST is measures by $\beta 2$, which describes changes in the control group after the implementation of DA. The variable of interest in the regressions in DAXPOST. This variable measures the effect of DA after its implementation in the audit. If there exists a difference between the control and treatment group which is applicable to DA it will be captured by DAXPOST.

Hereafter, the model enhances by adding control variables which can affect the outcome of the regression results of the simple models. The other variables are control variables which are used to explain differences which arise due to firm specific circumstances.

SumofHR = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \gamma 1 *$ logcTotalassets$t$ + $\gamma 2 *$ Nature$t$+ $\gamma 3 *$ Leverage$t$ + $\gamma 4 *$ Tenure$t$ + $\gamma 5 *$ ROA$t$ + $\gamma 6 *$ Subsidiaries$t + \varepsilon t$

SumofEUR = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \gamma 1 *$ logcTotalassets$t$ + $\gamma 2 *$ Nature$t$+ $\gamma 3 *$ Leverage$t$ + $\gamma 4 *$ Tenure$t$ + $\gamma 5 *$ ROA$t$ + $\gamma 6 *$ Subsidiaries$t + \varepsilon t$

SumofEUR2 = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \gamma 1 *$ logcTotalassets$t$ + $\gamma 2 *$ Nature$t$+ $\gamma 3 *$ Leverage$t$ + $\gamma 4 *$ Tenure$t$ + $\gamma 5 *$ ROA$t$ + $\gamma 6 *$ Subsidiaries$t + \varepsilon t$

The deflated dependent variables will be tested the exact same way as the original dependent variables. This leads to the following three additional regression models:

DefHR = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \gamma 1 *$ logcTotalassets$t + \gamma 2 *$ Nature$t$+ $\gamma 3 *$ Leverage$t$ + $\gamma 4 *$ Tenure$t$ + $\gamma 5 *$ ROA$t$ + $\gamma 6 *$ Subsidiaries$t + \varepsilon t$

DefEUR = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \gamma 1 *$ logcTotalassets$t$ + $\gamma 2 *$ Nature$t$+ $\gamma 3 *$ Leverage$t$ + $\gamma 4 *$ Tenure$t$ + $\gamma 5 *$ ROA$t$ + $\gamma 6 *$ Subsidiaries$t + \varepsilon t$

DefEUR2 = $\alpha + \beta 1 * DAt + \beta 2 *$ POST$t + \beta 3 * DA$XPOST$t + \gamma 1 *$ logcTotalassets$t$ + $\gamma 2 *$ Nature$t$+ $\gamma 3 *$ Leverage$t$ + $\gamma 4 *$ Tenure$t$ + $\gamma 5 *$ ROA$t$ + $\gamma 6 *$ Subsidiaries$t + \varepsilon t$

The results from the deflated dependent variables will be used to determine whether the original dependent variables are credible or not.

To investigate DA in more detail, the usage of DA will be viewed from various perspectives. As already discussed, the first test will measures the effect of DA which is applied when the DA department performs more than 24 hours on an audit assignment. The second test takes into account implementation effects. As discussed DA is expected to have implementation costs for the auditor, therefore the paper will measure whether DA is more beneficial after the first year of implementation by measuring DA from the second year on. The second test takes on the following form, which is the same as for SumofEUR, SumofEUR2, DefHR, DefEUR and DefEUR2:

$$\text{SumofHR} = \alpha + \beta 1 * DA2t + \beta 2 * \text{POST2}t + \beta 3 * DA\text{XPOST2}t + \gamma 1 * \text{logcTotalassets}t + \gamma 2 * \text{Nature}t + \gamma 3 * \text{Leverage}t + \gamma 4 * \text{Tenure}t + \gamma 5 * \text{ROA}t + \gamma 6 * \text{Subsidiaries}t + \varepsilon t$$

In the third test the study measures whether more advanced DA tools have a stronger effect than simple DA tools. It is expected that more advanced DA are created when more than 80 hours are spent by the DA department on audit engagements. Therefore the paper will perform a regression based on advanced DA. This can provide more insight in the effects of different forms of DA on audit efficiency. The regression for the third test, which is the same as for SumofEUR, SumofEUR2, DefHR, DefEUR and DefEUR2:

$$\text{SumofHR} = \alpha + \beta 1 * DA\text{b}t + \beta 2 * \text{POSTb}t + \beta 3 * DA\text{XPOSTb}t + \gamma 1 * \text{logcTotalassets}t + \gamma 2 * \text{Nature}t + \gamma 3 * \text{Leverage}t + \gamma 4 * \text{Tenure}t + \gamma 5 * \text{ROA}t + \gamma 6 * \text{Subsidiaries}t + \varepsilon t$$

Lastly, DA is implemented in multiple years for different firms. Because the implementation of DA can differ per company per year, the DID design is adjusted in such a way that the effect of DA, no matter the timing is the same. Figure 2 provides a graphical explanation of such a DID design.

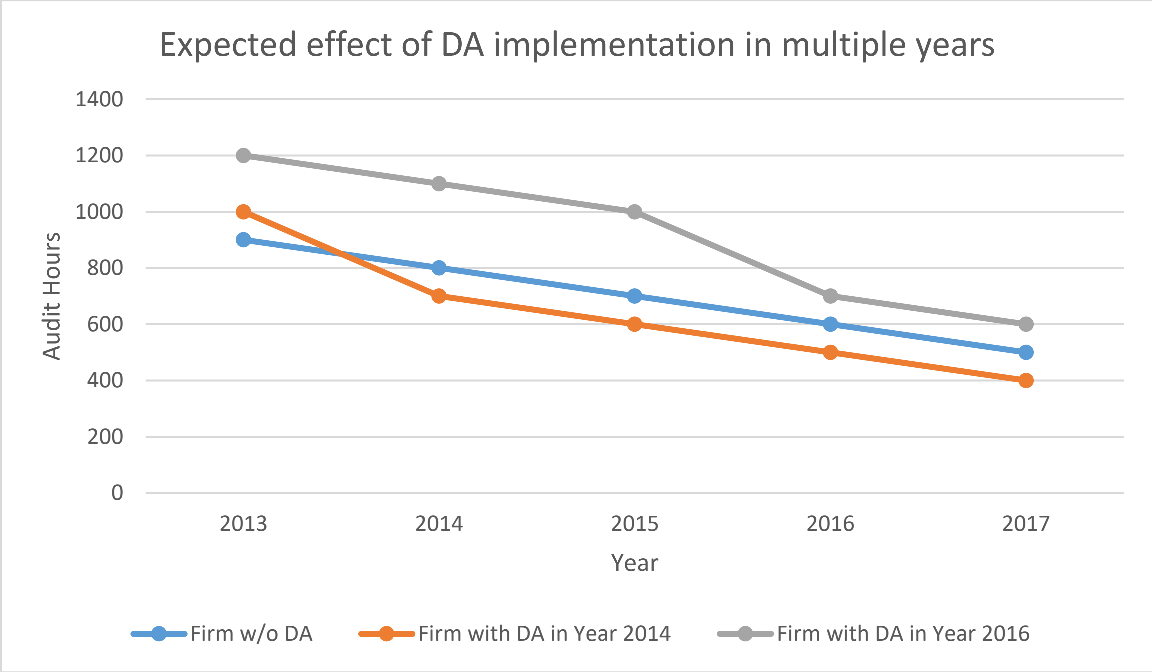



***Figure 2. Graphical explanation of the DID design.***

*The blue line represents the hours required to audit a firm where no DA is applied. The orange line represents the hours required to audit a firm when DA is applied in 2014. The grey line represents the hours required to audit a firm when DA is applied in 2016. When DA is applied there is a steeper decline in audit hours. Although DA is applied in different years, the decline of audit hours is the same for each year.*

## 4.5 Data acquired

To conduct the research, confidential information was received from a North-Western European public accounting firm. The data is not available for the public considering the data contains employee information and other audit information related to clients. First of all, to determine whether DA is used during an audit, data is obtained by consulting the accounting firm's planning. Next all audit information, necessary for determining the efficiency measures, were obtained between the fiscal years of 2013-2017. The reason for choosing this timeframe is due to the fact that the planning of the DA team is not available before the fiscal year of 2013. The public accounting firm's fiscal year ends close after busy season of 2013 which relates to audits of financial statements from 2012. Therefore the fiscal year of 2013 is linked to financial statements of 2012 and so on.

After determining the required control variables, WRDS and Orbis provide the financial data. Through WRDS access is obtained for Bureau van Dijk. With the help of Bureau van

Dijk financial numbers are acquired for the financial years of 2012-2016. Considering that the treatment group is rather small, missing data for companies which are in the treatment group are added. With the help of a domestic website, which gathers and provides the most recently available financial statements, blanks for several companies are filled in.

Unfortunately, out of the originally provided data it is not possible to find the information required for all companies. There are a few reasons behind this; foremost Bureau van Dijk and Orbis do not have the data available. Besides missing data in the databases, the domestic website does not have all the necessary information at its disposal. Due to this, the dataset is narrowed down every step of the way as some information is not available for certain companies and as such cannot be included in the sample. In the next section (4.6 data sampling) the exact steps used to narrow down the data can be found.

### 4.6 Data sample

With the help of the statistical program Stata the paper prepares the dataset and performs the regressions. An overview can be found in Appendix C. After obtaining all the data, the paper first appends the audit engagements of 2013-2017. Audit engagements which have less than 80 hours are dropped from the sample because it is unlikely that a complete audit can be fulfilled in less than 80 hours. Next the planning of the DA team and the financial data from WRDS is merged with the dataset. Duplicates are dropped to ensure there are no duplicate observations and that Stata does not randomly drop certain observations. Data from Orbis and additional financial data from the domestic website is subsequently merged with the initial dataset. Observations which miss values for the control variables are dropped from the sample.

Now there exists a complete dataset for which all variables of the observations have values. However to create a more reliable dataset several actions have to be taken. First of all, first year audits are claimed to cost more time than audits which are performed after the first year . There can be several reasons for more audit hours after the first year audit, like a merger or when a company is in financial distress. Nonetheless, audit engagements which have significantly lower audit hours than the following year are not considered to be a full audit and are taken out of the sample. Lastly, the sample is made more reliable by first taking out companies which have been audited for only one year and second companies which miss audit engagements between two audit years. For instance, if a company has been audited in 2014 and 2016, but has no data for 2015, the whole company is dropped from the sample. From this sample process the eventual amount of observations is obtained. From the 4,423 observations which are used for the regression, 152 of them have a dummy for DA. Because the panel is unbalanced and DA is implemented in multiple years, the distribution of the pre and post control group is determined by and equal to the distribution of the pre and post treatment group.

## 5. Empirical results

In this section the paper will discuss the results of the regression. The regression measures the effect of DA on audit efficiency. Several control variables are used which are described in the previous section.

The descriptive statistics can be found in table 3. The results of the regression for the first three models can be found in table 4. First the regression measures the effect of time before and after the application of DA on the audit for treatment firms. In panel A of table 4 the variable POST, which is a dummy for the time DA was implemented, has an insignificant value. This results describes that for the treatment group there is no significant effect in time which changed the amount of audit hours, audit costs or billed cost before and after the implementation of DA for the treatment group. The outcome of the second regression, recorded in panel B of table 4, provides more insight on the effect of DA on audit efficiency. The treatment variable DA has a positive significant value. As previously discussed in 4.3, the treatment variable describes the mean difference between the treatment group and the control group before DA is applied. Table 4 suggests that on average treated firms have 1014 audit hours more than control firms. In addition, the audit costs are 151780 euro higher and billed costs are 131561 euro more. This implicates that audits which use DA, are more extensive audits. The second variable, POST, captures the difference between the pre and post DA implementation periods for the control group. According to the model this variable is insignificant and thus captures no effect which impacts the control group over time. The last variable in this regression, DAXPOST, captures the effect of DA on the audit efficiency. As can been seen in table 4, variable DAXPOST has no significant value. Therefore the results suggest that DA has no effect on audit efficiency.

Although the simple model already shows that DA has no effect on audit efficiency, the paper uses a third model to measure audit efficiency more accurately. The control variables which are added in this regression model are based on prior literature (Hay, et al., 2006). In the new model, DA stays significant and the variable, although still high, becomes smaller. Also POST and DAXPOST remain insignificant. The first control variable logcTotalassets is a

**Table 3**

***Descriptive statistics***

***Panel A: Distributions of variables***

| Variables | N | Mean | Std. Dev. | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| SumofHR | 4423 | 444.600 | 416.095 | 91.000 | 162.300 | 295.250 | 535.050 | 1618.750 |
| SumofEUR | 4423 | 65871.960 | 65073.460 | 12015.000 | 23202.500 | 41205.750 | 78449.020 | 252344 |
| SumofEUR2 | 4423 | 61315.640 | 58108.430 | 11507.150 | 22251.300 | 39206.700 | 74360.850 | 224341 |
| DefHR | 4423 | 42.135 | 34.297 | 9.659 | 17.748 | 30.510 | 52.816 | 136.368 |
| DefEUR | 4423 | 6148.960 | 5249.935 | 1327.401 | 2540.965 | 4264.054 | 7596.394 | 20761.610 |
| DefEUR2 | 4423 | 5723.879 | 4644.048 | 1261.045 | 2412.635 | 4097.686 | 7199.500 | 18226.72 |
| DA | 4423 | 0.042 | 0.202 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DA2 | 4423 | 0.028 | 0.167 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DAb | 4423 | 0.013 | 0.115 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| POST | 4423 | 0.724 | 0.446 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| POST2 | 4423 | 0.452 | 0.497 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| POSTb | 4423 | 0.529 | 0.499 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| DAXPOST | 4423 | 0.034 | 0.182 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DAXPOST2 | 4423 | 0.016 | 0.126 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DAXPOSTb | 4423 | 0.010 | 0.103 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| logcTotalassets | 4423 | 10.162 | 2.180 | 2.570 | 8.567 | 9.793 | 11.540 | 19.833 |
| Leverage | 4423 | 0.584 | 0.287 | 0.034 | 0.374 | 0.604 | 0.800 | 1.080 |
| Nature | 4423 | 0.401 | 3.310 | 0.000 | 0.100 | 0.372 | 0.656 | 1.000 |
| tenure | 4423 | 1.016 | 0.912 | 0.000 | 0.000 | 1.000 | 2.000 | 4.000 |
| ROA | 4423 | 0.027 | 0.059 | -0.072 | 0.000 | 0. 001 | 0.049 | 0.183 |
| CountofSubsidiaryName | 4423 | 2.290 | 3.631 | 0.000 | 0.000 | 1.000 | 3.000 | 13.000 |

**Table 4**

***Regression results Data Analytics***

*Panel A – Tests POST effect for treatment first*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 1422.238*** | 0.000 | 212733.100*** | 0.000 | 188772.300*** | 0.000 |
| POST | -102.761 | 0.218 | -9669.677 | 0.471 | -6778.122 | 0.562 |
| Adjusted $R^2$ | 0.0028 | | 0.0000 | | 0.0000 | |
| Observations | 190 | | 190 | | 190 | |

*Panel B – Tests without control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 408.122*** | 0.000 | 60952.200*** | 0.000 | 57210.630*** | 0.000 |
| DA | 1014.116 *** | 0.000 | 151780.900*** | 0.000 | 131561.700*** | 0.000 |
| POST | -8.293 | 0.511 | -2316.701 | 0.241 | -2341.136 | 0.187 |
| DAXPOST | -94.467 | 0.164 | -7352.976 | 0.489 | -4436.986 | 0.642 |
| Adjusted $R^2$ | 0.2112 | | 0.2083 | | 0.2011 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -442.528*** | 0.000 | -81199.960*** | 0.000 | -70976.910*** | 0.000 |
| DA | 727.022*** | 0.000 | 104063.000*** | 0.000 | 88713.140*** | 0.000 |
| POST | 7.521 | 0.497 | 443.651 | 0.793 | 234.771 | 0.877 |
| DAXPOST | -58.381 | 0.314 | -1894.808 | 0.831 | 399.787 | 0.960 |
| logcTotalassets | 69.250*** | 0.000 | 12097.550*** | 0.000 | 10903.780*** | 0.000 |
| Leverage | 72.487*** | 0.000 | 8061.627*** | 0.002 | 7464.058*** | 0.001 |
| Nature | 147.473*** | 0.000 | 21460.800*** | 0.000 | 19590.600*** | 0.000 |
| tenure | -14.439*** | 0.007 | -2834.054*** | 0.001 | -2741.115*** | 0.000 |
| ROA | 101.927 | 0.226 | -3975.699 | 0.757 | -823.232 | 0.943 |
| CountofSubsidiaryName | 24.991*** | 0.000 | 3839.028*** | 0.000 | 3414.462*** | 0.000 |
| Adjusted $R^2$ | 0.4261 | | 0.4522 | | 0.4482 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 4 presents the result from an OLS regression of the application of DA. The regression is performed with SumofHR, SumofEUR and SumofEUR2 as dependent variable. Variables definitions are given in Appendix B.

logarithmic of total assets. As can be seen in panel C of table 4, logcTotalassets is positively significant. According to the model for every 1 increase in logcTotalassets, the amount of hours spent on an audit increases by 69 hours. The audit costs increase by 12098 euro for every 1 increase in logcTotalassets, billed costs increase by 10904 euro.

The second control variable is Leverage. Leverage measures the risk of non-payment of a company by dividing liabilities by liabilities and equity. When leverage increases, going-concern risks increase. In the model of the paper Leverage positively impacts the audit fee. This means that when the leverage worsens, the ratio increases, the amount of audit hours increases as well. The total amount of audit hours increase by 72 when the company is fully debt-financed, in turn audit costs increase by 8062 euro and billed costs by 7464 euro.

Nature is a control variable which takes into account the amount of assets which are not yet secured, namely inventory and receivables. Because the value of these items are often based on judgement by the management, these accounts require extra attention from the audit team. Thus a company has more risks if assets contain high amounts of inventory and receivables. In line with this expectation, the variable nature is positively significant. According to the model the amount of audit hours, audit costs and billed costs will increase by 147, 21461 and 19590, respectively, if all the assets consist of inventory and receivables.

Tenure is another control variable used in the model. Tenure measures the amount of years an accountant is auditing the same company. The general consensus is that every year the public accounting firm better understands the client's financial situation and therefore needs less time to perform the audit. The results from the models are in line with this thought. The amount of hours spent decrease with 14 for every subsequent year the company is audited. On average the audit costs decrease with 2834 euro for each year a firm is audited by the same auditor and the billed costs 2741 euro.

Return on assets measures risk by dividing net income by total assets. In the model the variable ROA is highly insignificant. This implies that the profitability of a company has no effect on audit hours, audit costs and billed costs.

The last control variable is CountofSubsidiaryName. This variable measures the effect of subsidiaries on the audit. Complexity of the audit increases the audit hours. According to the literature complexity increases when the company has more subsidiaries. In line with literature, the model's measure for subsidiaries shows a significant positive value of 25. Indicating that for every subsidiary, the amount of hours spent on the audit increases with 25 hours. According to the model,  the audit cost for each additional subsidiary increases by 3839 euro and the billed cost with 3414 euro.

The model for audit hours has an adjusted $R^2$ of 0,4261 which means that the independent variables used in the regression explain 45,22% of the variation of the dependent variable. For audit costs and billed costs this was  0.4522 and 0.4482, respectively.

The expectation was that DA should improve the audit efficiency. Yet it appears that DA does not affect audit efficiency. This might be explained by the fact that the implementation of DA costs a lot of time. First of all, the accountants need to discuss the possibilities for DA with the client. After the client agrees for DA, the auditor has to communicate with the DA department and the client what the possibilities are for DA. Following this, once the client, auditor and DA team agree upon implementing the DA tool, the auditor has to act as a mediator between the client and DA team to ensure the DA tool is a well functional and correctly operationalized program.

To check whether the implementation of DA has an impact the paper performs a second regression which can be found in table 5. This regression does not take into account the first year DA was applied, to exclude first year implementation costs which might counter the effect of DA on the audit efficiency. First, the POST2 effect for the treatment group is measured. This time POST2 is significant for audit hours, with a p-value below 0.1. The coefficient is negative which suggests that after the implementation of DA there was an effect which lowered the amount of audit hours for treatment firms. POST2 is insignificant for audit costs and billed costs. In the second test, the treatment dummy DA2 has a strong positive significance. POST2 appears to be insignificant for audit hours and costs, but has a significant negative value for billed costs. However, the variable of interest, DAXPOST, has become significant. The negative coefficient provides evidence that on average DA lowers the overall audit hours by 119 hours a year after its implementation and thus improves audit efficiency. For audit costs and billed costs DAXPOST is insignificant. These results suggest that DA does not improve audit costs and billed costs a year after its implementation.

To provide more robust results, control variables are added to the regression to check whether DAXPOST remains significant for audit hours. Panel C in table 5 shows that the regression results have the same sign and significance levels as the regressions from panel C in table 4. This means that DAXPOST is insignificant and does not affect the dependent variables. POST2 stays significant for billed costs.

**Table 5**

***Regression results Data Analytics after first year implementation***

*Panel A – Tests POST effect for treatment first*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 1438.650*** | 0.000 | 221898.100*** | 0.000 | 198336.000*** | 0.000 |
| POST2 | -134.935* | 0.098 | -19737.900 | 0.123 | -17241.160 | 0.124 |
| | | | | | | |
| Adjusted $R^2$ | 0.0138 | | 0.0110 | | 0.0109 | |
| Observations | 128 | | 128 | | 128 | |

*Panel B – Tests without control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 422.0842*** | 0.000 | 62545.110*** | 0.000 | 57210.630*** | 0.000 |
| DA2 | 1016.566*** | 0.000 | 159353*** | 0.000 | 139724.900*** | 0.000 |
| POST2 | -15.780 | 0.178 | -2972.748 | 0.104 | -3124.491* | 0.057 |
| DAXPOST2 | -119.154* | 0.085 | -16765.160 | 0.120 | -14116.67 | 0.144 |
| | | | | | | |
| Adjusted $R^2$ | 0.1480 | | 0.1508 | | 0.1462 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -478.669*** | 0.000 | -86717.780*** | 0.000 | -75743.570*** | 0.000 |
| DA2 | 692.710*** | 0.000 | 105367.900*** | 0.000 | 91261.600*** | 0.000 |
| POST2 | -12.791 | 0.224 | -2205.778 | 0.168 | -2365.089* | 0.099 |
| DAXPOST2 | -66.158 | 0.260 | -8239.894 | 0.358 | -6424.189 | 0.423 |
| logcTotalassets | 73.606*** | 0.000 | 12724.74*** | 0.000 | 11448.270*** | 0.000 |
| Leverage | 80.194*** | 0.000 | 9241.552*** | 0.001 | 8502.242*** | 0.000 |
| Nature | 163.244*** | 0.000 | 23760.010*** | 0.000 | 21566.160*** | 0.000 |
| tenure | -13.090** | 0.022 | -2657.309*** | 0.002 | -2520.554*** | 0.001 |
| ROA | 79.098 | 0.363 | -7143.547 | 0.590 | -3451.669 | 0.771 |
| CountofSubsidiaryName | 25.784*** | 0.000 | 3948.876*** | 0.000 | 3507.222*** | 0.000 |
| | | | | | | |
| Adjusted $R^2$ | 0.3894 | | 0.4207 | | 0.4188 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 5 presents the result from an OLS regression of the application of DA after the first year DA was used during the audit. The regression is performed with SumofHR, SumofEUR and SumofEUR2 as dependent variable. Variables definitions are given in Appendix B.

These results provide evidence that even after the first year that DA is implemented, DA still has no effect on audit efficiency. A reason why audit efficiency does not improve after the first year of implementation is because auditors still have to get used to using DA in their audit. Because DA is a new way of performing work and analysing the results, it might require the auditor more than one year to effectively handle DA. Besides that, it might be that DA requires multiple years of implementation. Especially for big DA projects, tools are carefully tested and edited until the DA team, client and auditor are sure it works properly.

Next the paper takes a closer look at advanced DA. In table 6 the results of this test can be observed. Again panel A starts with the regression for pre and post treatment group effects. The effect is insignificant. Moreover, the $R^2$ is zero, which means that POSTb cannot explain the dependent variable. The simple regression model in panel B shows a negative variable POSTb for audit hours and billed costs. DAXPOST is insignificant and thus has no influence on the dependent variables. Panel C, which provides the results from the complex model, shares the same significant values and signs as panel C of table 4. The only difference is that POSTb has a significant sign. An explanation why DAXPOST is insignificant for advanced DA can be because DA is new for the auditor and the more complex, the longer it takes for the auditor to get used to working with DA. This in turn would lead to audits, which experience a higher complexity of DA, to require more time.

It is odd that the variable POSTb in the advanced DA regression has such a high significance level when POST and POST2, which should measure the same, were insignificant. A reason why POSTb might have such significant negative coefficients can be due to the firms which had implemented DA below 80 hours. After the implementation of DA, firms which have simple DA also apply DA. However these firms are taken as control firms because they do not fulfil the advanced DA requirement and it is expected that simple DA does not improve audit efficiency. The interaction of these firms with the control group could lead to a significant POSTb. Therefore this might explain why POST in panel C of table 4 is not significant and POSTb in panel C in table 6 is. However, if this is the case it might be that simple DA does increase audit efficiency. To check whether this could be an explanation two small tests are performed. The first one drops all firms where simple DA was applied. The second drops all firms where advanced DA is applied. Both run a simple model to measure audit hours, audit costs and billed costs. In appendix E in table 10 results provide evidence that POSTb in table 6 is indeed significant due to the inclusion of firms which applied simple DA. This is due to

**Table 6**

***Regression results advanced Data Analytics***

*Panel A – Tests POST effect for treatment first*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 1441.838*** | 0.000 | 221334.600*** | 0.000 | 199919.000*** | 0.000 |
| POSTb | -39.030 | 0.778 | -3349.606 | 0.880 | -6472.648 | 0.740 |
| | | | | | | |
| Adjusted $R^2$ | 0.0000 | | 0.0000 | | 0.0000 | |
| Observations | 60 | | 60 | | 60 | |

*Panel B – Tests without control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 452.721*** | 0.000 | 67391.410*** | 0.000 | 62695.840*** | 0.000 |
| DAb | 989.116*** | 0.000 | 153943.200*** | 0.000 | 137223.200*** | 0.000 |
| POSTb | -44.914*** | 0.000 | -7529.365 | 0.104 | -6702.932*** | 0.000 |
| DAXPOSTb | 5.883 | 0.964 | 4179.760 | 0.835 | 230.284 | 0.990 |
| | | | | | | |
| Adjusted $R^2$ | 0.0776 | | 0.0797 | | 0.0764 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -517.621*** | 0.000 | -86717.780*** | 0.000 | -81010.950*** | 0.000 |
| DAb | 803.383*** | 0.000 | 122729.900*** | 0.000 | 109153.200*** | 0.000 |
| POSTb | -34.250*** | 0.001 | -5690.046*** | 0.000 | -5011.875*** | 0.000 |
| DAXPOSTb | -92.853 | 0.386 | -12644.850 | 0.438 | -14850.37 | 0.309 |
| logcTotalassets | 78.187*** | 0.000 | 13423.090*** | 0.000 | 12066.170*** | 0.000 |
| Leverage | 90.423*** | 0.000 | 10803.920*** | 0.000 | 9878.511*** | 0.000 |
| Nature | 161.908*** | 0.000 | 23537.770*** | 0.000 | 21406.100*** | 0.000 |
| tenure | -9.850* | 0.074 | -2152.152*** | 0.010 | -2143.238*** | 0.004 |
| ROA | 93.796 | 0.291 | -4833.425 | 0.721 | -1604.977 | 0.894 |
| CountofSubsidiaryName | 27.838*** | 0.000 | 4262.056*** | 0.000 | 3776.415*** | 0.000 |
| | | | | | | |
| Adjusted $R^2$ | 0.3641 | | 0.3968 | | 0.3956 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 6 presents the result from an OLS regression of the application of advanced DA. The regression is performed with SumofHR, SumofEUR and SumofEUR2 as dependent variable. Variables definitions are given in Appendix B.

the interaction of the control group and the firms which apply simple DA and not because simple DA improves audit efficiency.

The regression results for the deflated regressions can be observed in table 7, 8 and 9 in appendix E. In table 7 the variables have the same sign and significant as the variables in table 4. The only difference for DefHR is a significant ROA which implies that, according to the deflated model, ROA has a positive effect on audit hours. For DefEUR and DefEUR2, ROA is insignificant, same as in the original model.

The signs and significances for the coefficients of the regressions for the deflated model which measured DA after the first year it was implemented, table 8, are almost similar to table 5. The difference between the tables is that the deflated model captures a significant effect for the treatment and control group for POST2. Furthermore, just like in table 7, the deflated model has a significant positive coefficient of ROA for audit hours.

For table 9, the deflated model for table 6, the differences are a significant POST2 for SumofEUR in panel B of table 9 compared to an insignificant POST for SumofEUR in panel B of table 6. And, as with all the deflated models, variable ROA in table 9 was positively significant compared to the insignificant ROA of table 6. Lastly, the adjusted $R^2$ of the deflated models is lower than for the initial model. This implies that for the deflated models the variables are less able to explain the result in the regressions. However, this does not necessarily mean that the deflated model is worse than the original model.

Considering the outcome of the multiple regressions, hypothesis 1 cannot be rejected. This implies that DA does not decrease the number of audit hours. Moreover, hypothesis 2 cannot be rejected. Therefore audit costs do not decrease when DA is applied during the audit. Lastly, hypothesis 3 is not rejected. Thus DA does not decrease the billed costs of the auditor.

Of the six control variables used to improve the explanation of the model, five are consistently significant. The independent variables show that total assets, leverage, nature of assets and the number of subsidiaries are positively correlated with the dependent variables. Tenure is negatively correlated with the dependent variables. The effect of last control variable, return on assets, is debatable. In most regressions the variable is insignificant. Yet, in the deflated models for audit hours, ROA seems to have an effect on the dependent variable, making it difficult to provide a robust conclusion.

**Table 7**

***Regression results Data Analytics for deflated dependent variables***

*Panel A – Tests POST effect for treatment first*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 121.007*** | 0.000 | 17718.68*** | 0.000 | 15500.090*** | 0.000 |
| POST | -8.373 | 0.207 | -727.001 | 0.483 | -421.040 | 0.632 |
| Adjusted $R^2$ | 0.0032 | | -0.0027 | | -0.0041 | |
| Observations | 190 | | 190 | | 190 | |

*Panel B – Tests without control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 38.236*** | 0.000 | 5635.638*** | 0.000 | 5289.625*** | 0.000 |
| DA | 82.770*** | 0.000 | 12083.040*** | 0.000 | 10210.470*** | 0.000 |
| POST | 0.646 | 0.536 | -14.436 | 0.928 | -23.007 | 0.187 |
| DAXPOST | -9.019 | 0.109 | -712.564 | 0.408 | -398.032 | 0.642 |
| Adjusted $R^2$ | 0.2020 | | 0.1997 | | 0.1883 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 21.348*** | 0.000 | -2803.169*** | 0.000 | -2126.374*** | 0.000 |
| DA | 65.264*** | 0.000 | 8715.600*** | 0.000 | 7248.342*** | 0.000 |
| POST | .915 | 0.350 | 2.627 | 0.986 | 5.563 | 0.966 |
| DAXPOST | -4.310 | 0.405 | 94.464 | 0.902 | 320.683 | 0.640 |
| logcTotalassets | .000*** | 0.000 | 658.881*** | 0.000 | 572.919*** | 0.000 |
| Leverage | 8.951*** | 0.000 | 969.529*** | 0.000 | 917.563*** | 0.000 |
| Nature | 10.962*** | 0.000 | 1938.119*** | 0.000 | 1772.021*** | 0.000 |
| tenure | -1.529*** | 0.001 | -288.197*** | 0.000 | -283.056*** | 0.000 |
| ROA | 28.367*** | 0.000 | 617.590 | 0.580 | 944.133 | 0.344 |
| CountofSubsidiaryName | 2.511*** | 0.000 | 342.132*** | 0.000 | 303.092*** | 0.000 |
| Adjusted $R^2$ | 0.3284 | | 0.3662 | | 0.3539 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 7 presents the result from an OLS regression of the application of DA. The regression is performed with DefHR, DefEUR and DefEUR2 as dependent variable. Variables definitions are given in Appendix B.

# 6. Conclusion

This paper investigates the effect of DA on the audit. With a sample of 4423 audit engagements that took place from 2013-2017 the paper aims to answer the following question:

*Does Data Analytics improve audit efficiency?*

The research shows that DA does not improve audit efficiency. These findings do not support the alternative hypotheses 1, 2 and 3, which measure efficiency by audit hours, audit costs and billed costs, respectively. Because the null hypotheses are not rejected, there is found no evidence that the implementation of DA during the audit improves audit efficiency. The dependent variables for audit hours, costs and billed costs are all unaffected when DA is applied.

The study contributes to existing literature by providing empirical research on the use of DA in the audit. In addition, the paper counters the current arguments in literature which state that DA improves efficiency of the audit. Therefore auditors and clients have to consider whether they are willing to put more time and money in DA.

There are several limitations this study encountered. Firstly, the data was obtained from one public accounting firm. Therefore it is difficult to state whether the outcome of this study applies to all public accounting firms. Secondly, the outcome of this research relates to DA tools which were specifically used for one certain audit. General tools, which are becoming more popular because these can be used for multiple audits, have not been taken into account. Thirdly, it was not possible to obtain data for overtime. Therefore the paper cannot control for clients which required a considerable longer audit due to various reasons. Fourthly, considering the number of observations were too low to account for industry, industry was used as a control variable. Having enough data to account for industry would improve this study because it is likely that several industries require more audit resources than others.

Although DA does not improve audit efficiency, this does not implicate DA is bad. The use and application of DA during the audit is relatively new and public accounting firms need time to adjust to the new way of auditing. Moreover this study only focusses on audit efficiency. More important is the effectiveness of the audit. Future research can focus on measuring whether the audit quality enhances. Furthermore future research can measure whether the use of DA becomes more efficient in time. This can show whether application of DA becomes better when public accounting firms and auditors have more experience.

DA has a lot of advantages for auditing. It can be used from the first to the last phase of the audit and provides the auditor with clearer and more insights. Public accounting firms should keep investing in DA and work to a higher audit quality. Also universities can have a positive effect on the use of DA during the audit by providing analytical courses for accounting students whom are being educated by the university.

# 7. References


Adrian, A., 2013. Big Data Challenges. *Database Systems Journal,* 4(3), pp. 31-40.

Alles, M., 2015. Drivers of the Use and Facilitators and Obstacles of the Evolution of Big Data by the Audit Profession. *American Accounting Association,* 29(2), pp. 439-449.

Brown-Liburd, H., Issa, H. & Lombardi, D., 2015. Behavioral Implications of Big Data's Impact on Audit Judgement and Decision Making and Future Research Directions. *Accounting Horizon's,* 29(2), pp. 451-468.

Business.com, 2017. *business.com.* [Online]
Available at: https://www.business.com/articles/big-data-big-problem-coping-with-shortage-of-talent-in-data-analysis/
[Accessed 15 May 2017].

Byrnes, P., Criste, T., Stewart, T. & Vasarhelyi, M., 2014. *Reimagining Auditing in a Wired World.* [Online]
Available at:
http://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/DownloadableDocuments/Whitepaper_Blue_Sky_Scenario-Pinkbook.pdf
[Accessed 2 May 2017].

Cao, M., Chychyla, R. & Stewart, T., 2015. Big Data Analytics in Financial Statement Audits. *Accounting Horizons,* 29(2), pp. 423-429.

Coyne, E., Coyne, J. & Walker, K., 2017. Big Data Information Governance by Accountants. *International Journal of Accounting and Information Management (forthcoming),* Volume Working paper, pp. 1-34.

Crawford, K. & Boyd, D., 2011. Six Provocations for Big Data. *Oxford Internet Institute's,* p. 17.

Earley, C. E., 2015. Data analytics in auditing: Opportunities and Challenges. *Business Horizons,* Volume 58, pp. 493-500.

Gershkoff, A., 2015. *techcrunch.* [Online]
Available at: https://techcrunch.com/2015/12/31/how-to-stem-the-global-shortage-of-data-scientists/
[Accessed 15 May 2017].

Hay, D., Knechel, R. & Wong, N., 2006. Audit Fees: A Meta-analysis of the Effect of Supply and Demand Attributes. *Contemporary Accounting Research,* 23(1), pp. 91-141.

Hayes, R., Gortemaker, H. & Wallage, P., 2014. *Principles of Auditing.* 3e ed. Harlow: Pearson Education Limited.

Kessel, P. v., 2014. *Big Data: Changing the way business operate,* United Kingdom: Ernst & Young.

Liddy, J., 2015. How Data and Analytics Are Enhancing Audit Quality and Value. *The CPA Journal,* 85(5), p. 80.

Manson, S., McCartney, S. & Sherer, M., 2007. Audit automation: Improving quality or keeping up. In: M. Sherer & S. Turley, eds. *Current Issues in Auditing.* Thousand Oaks,: Sage Publications.

Nasser, T. & Tariq, R., 2015. Big Data Challenges. *Journal of Computer Engineering & Information Technology,* 4(3), p. 10.

NBA, 2013. Big four investeren fors in big data. *Accountant*, 13 11, p. 2.

PWC, 2015. *pwc.com.* [Online]
Available at: https://www.pwc.com/gx/en/audit-services/publications/assets/pwc-fact-sheet-3-summary-of-eu-audit-reform-requirements-relating-to-nas-feb-2015.pdf
[Accessed 08 May 2017].

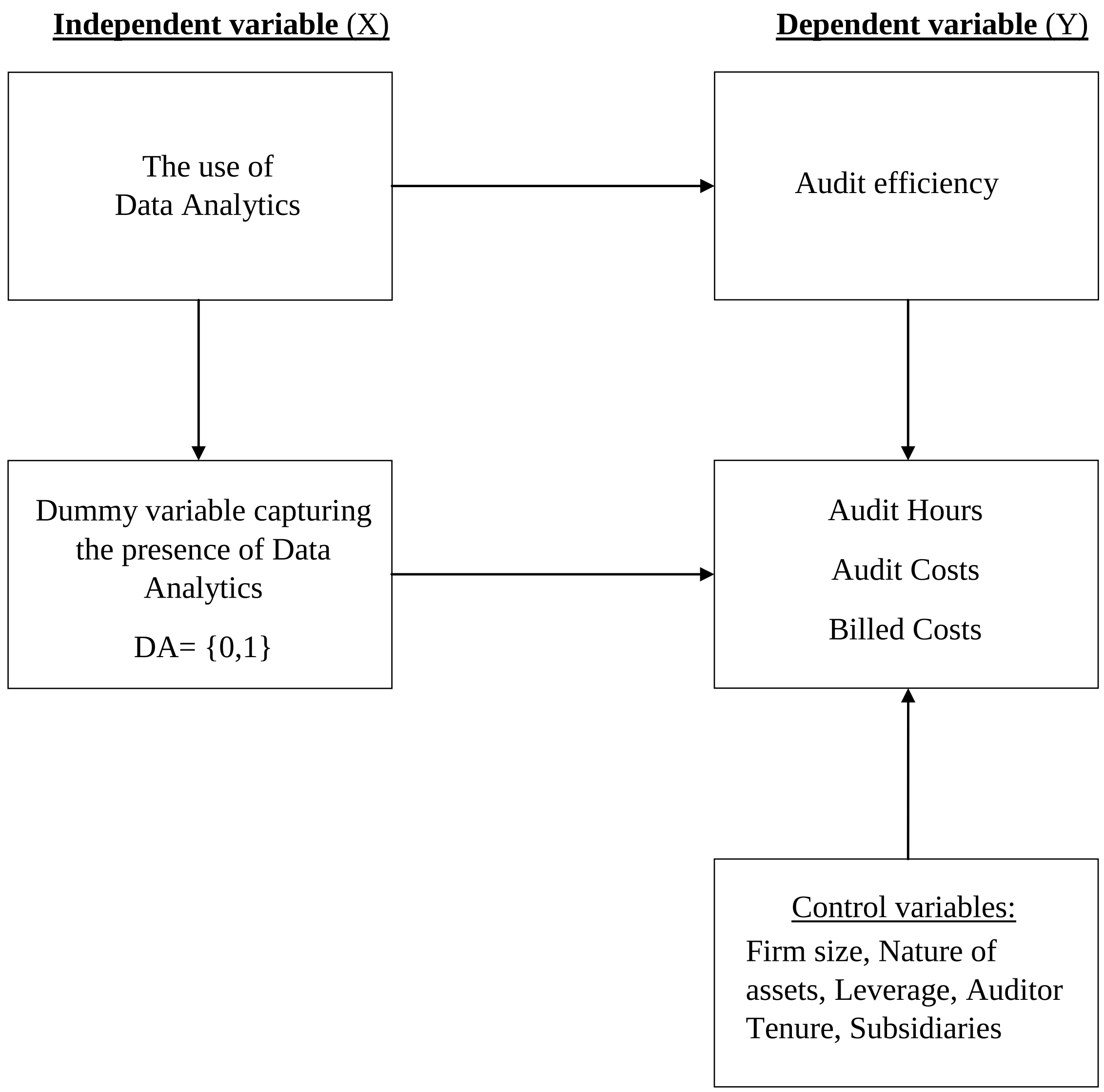
Ramlukan, R., 2015. *ey.com.* [Online]
Available at: http://www.ey.com/gl/en/services/assurance/ey-reporting-issue-9-how-big-data-and-analytics-are-transforming-the-audit#item1
[Accessed 17 05 2017].

Ruhnke, K. & Schmidt, M., 2014. The audit expectation gap: existence, causes,and the impact of changes. *The audit expectation gap: existence, causes,and the impact of changesKLAUS RUHNKEa∗and MARTIN SCHMIDTbaDepartment of Finance, Accounting & Taxation, Freie Universita ̈t Berlin, Berlin, Germany;bDepartment of Financial Reporting and Audit, Chair of Intern,* 5(44), pp. 572-601.

Whitehouse, T., 2014. Auditing in the Era of Big Data. *Compliance Week*, 29 April, p. 2.

# 8. Appendix

## Appendix A – Libby Boxes



# Appendix B – Variable definitions list

| **Definition** | **Variable description** |
| --- | --- |
| Audit Hours | SumofHR |
| Audit Costs | SumofEUR |
| Billed Costs | SumofEUR2 |
| Hours DA department | SumofHRDA |
| Deflated Audit Hours (Audit Hours divided by the logarithmic of total assets) | DefHR |
| Deflated Audit Costs (Audit Costs divided by the logarithmic of total assets) | DefEUR |
| Deflated Billed Costs (Billed Costs divided by the logarithmic of total assets) | DefEUR2 |
| Dummy for DA | DA |
| Dummy for the year 2016 and after | POST |
| Measurement of the effect of DA on the audit efficiency. | DAXPOST |
| Total assets | cTotalassets |
| Logarithmic total assets | logcTotalassets |
| Net income | cNetIncome |
| Receivables | cReceivables |
| Inventory | cInventory |
| Liabilities | cLiabilities |
| Equity | cEquity |
| Current audit hours / consecutive year audit hours | perH |
| Auditor tenure | tenure |
| Amount of subsidiaries | CountofSubsidiaryName |
| Profatibility ratio (Net Income divided by total assets) | ROA |
| Nature of assets (Receivables and Inventory divided by total assets) | Nature |
| Leverage (defined by total liabilities divided by total liabilities and equity) | Leverage |

## Appendix C – Sample selection process

**Table 2**

***Sample Selection Process***

| | | |
|---|---|---|
| **Initial Number of obtained Audit engagements (2013-2017)** | | **54,929** |
| Audits engagements which have less than 80 hours | 33,315 | |
| Observations where Total assets are missing | 13,541 | |
| Observations where Equity & Liabilities are missing | 7 | |
| Observations with negative Liabilities | 1 | |
| Observations without Net Income | 3 | |
| Observations where Receivables are missing | 2 | |
| Observations where Inventory are missing | 0 | |
| **Final Sample of observations with financial data** | | **8,060** |
| Drop if percentage previous year audit is less than 80% | 1,581 | |
| Drop if only one year audit occurred | 1,526 | |
| Drop if audit engagements are more than two years apart | 530 | |
| **Final Sample of observations** | | **4,423** |
| **Control group** | | **4,271** |
| **Treatment group** | | **152** |

# Appendix D – Descriptive statistics

**Table 3**

***Descriptive statistics***

***Panel A: Distributions of variables***

| Variables | N | Mean | Std. Dev. | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| SumofHR | 4423 | 444.600 | 416.095 | 91.000 | 162.300 | 295.250 | 535.050 | 1618.750 |
| SumofEUR | 4423 | 65871.960 | 65073.460 | 12015.000 | 23202.500 | 41205.750 | 78449.020 | 252344 |
| SumofEUR2 | 4423 | 61315.640 | 58108.430 | 11507.150 | 22251.300 | 39206.700 | 74360.850 | 224341 |
| DefHR | 4423 | 42.135 | 34.297 | 9.659 | 17.748 | 30.510 | 52.816 | 136.368 |
| DefEUR | 4423 | 6148.960 | 5249.935 | 1327.401 | 2540.965 | 4264.054 | 7596.394 | 20761.610 |
| DefEUR2 | 4423 | 5723.879 | 4644.048 | 1261.045 | 2412.635 | 4097.686 | 7199.500 | 18226.72 |
| DA | 4423 | 0.042 | 0.202 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DA2 | 4423 | 0.028 | 0.167 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DAb | 4423 | 0.013 | 0.115 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| POST | 4423 | 0.724 | 0.446 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| POST2 | 4423 | 0.452 | 0.497 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| POSTb | 4423 | 0.529 | 0.499 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |
| DAXPOST | 4423 | 0.034 | 0.182 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DAXPOST2 | 4423 | 0.016 | 0.126 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| DAXPOSTb | 4423 | 0.010 | 0.103 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| logcTotalassets | 4423 | 10.162 | 2.180 | 2.570 | 8.567 | 9.793 | 11.540 | 19.833 |
| Leverage | 4423 | 0.584 | 0.287 | 0.034 | 0.374 | 0.604 | 0.800 | 1.080 |
| Nature | 4423 | 0.401 | 3.310 | 0.000 | 0.100 | 0.372 | 0.656 | 1.000 |
| tenure | 4423 | 1.016 | 0.912 | 0.000 | 0.000 | 1.000 | 2.000 | 4.000 |
| ROA | 4423 | 0.027 | 0.059 | -0.072 | 0.000 | 0. 001 | 0.049 | 0.183 |
| CountofSubsidiaryName | 4423 | 2.290 | 3.631 | 0.000 | 0.000 | 1.000 | 3.000 | 13.000 |

*Panel B: Correlations audit hours*

| **Variables** | SumofHR | SumofEUR | SumofEUR2 | DefHR | DefEUR | DefEUR2 | DA | DA2 | DAb | POST |
|---|---|---|---|---|---|---|---|---|---|---|
| SumofHR | 1.0000 | | | | | | | | | |
| SumofEUR | 0.9823*** | 1.0000 | | | | | | | | |
| SumofEUR2 | 0.9802*** | 0.9905*** | 1.0000 | | | | | | | |
| DefHR | 0.9795*** | 0.9492*** | 0.9484*** | 1.0000 | | | | | | |
| DefEUR | 0.9735*** | 0.9816*** | 0.9732*** | 0.9772*** | 1.0000 | | | | | |
| DefEUR2 | 0.9678*** | 0.9673*** | 0.9792*** | 0.9756*** | 0.9880*** | 1.0000 | | | | |
| DA | 0.4596*** | 0.4566*** | 0.4486*** | 0.4495*** | 0.4474*** | 0.4345*** | 1.000 | | | |
| DA2 | 0.3840*** | 0.3875*** | 0.3813*** | 0.3722*** | 0.3773*** | 0.3660*** | 0.8148*** | 1.000 | | |
| DAb | 0.2744*** | 0.2775*** | 0.2714*** | 0.2699*** | 0.2736*** | 0.2637*** | 0.5535*** | 0.4811*** | 1.000 | |
| POST | 0.0268 | 0.0184 | 0.0184 | 0.0205 | 0.0127 | 0.0120*** | 0.0358** | 0.0642*** | 0.0198 | 1.000 |
| POST2 | -0.0084 | -0.0118 | -0.0158 | -0.0173 | -0.0195 | -0.0253* | -0.0314** | 0.0381** | -0.0124 | 0.5606*** |
| POSTb | -0.0365** | -0.0399*** | -0.0404*** | -0.0421*** | -0.0451*** | -0.0464*** | -0.1175*** | -0.0805*** | 0.0636*** | 0.6540*** |
| DAXPOST | 0.3999*** | 0.4009*** | 0.3950*** | 0.3910*** | 0.3931*** | 0.3834*** | 0.8904*** | 0.8114*** | 0.4929*** | 0.1163*** |
| DAXPOST2 | 0.2678*** | 0.2716*** | 0.2674*** | 0.2585*** | 0.2627*** | 0.2553*** | 0.6072*** | 0.7452*** | 0.3556*** | 0.0793*** |
| DAXPOSTb | 0.2431*** | 0.2468*** | 0.2401*** | 0.2390*** | 0.2429*** | 0.2330*** | 0.4944*** | 0.4766*** | 0.8932*** | 0.0646*** |
| logcTotalassets | 0.5014*** | 0.5397*** | 0.5406*** | 0.3667*** | 0.4189*** | 0.4107*** | 0.2546*** | 0.2386*** | 0.1569*** | 0.0467*** |
| Leverage | 0.0846*** | 0.0724*** | 0.0733*** | 0.1045*** | 0.0918*** | 0.0947*** | 0.0649*** | 0.0505*** | 0.0348** | -0.0060 |
| Nature | -0.0014 | -0.0254* | -0.0236 | 0.0548*** | 0.0252* | 0.0314** | 0.0075 | -0.0101 | 0.0246 | -0.0188 |
| tenure | -0.0213 | -0.0297** | -0.0331** | -0.0343** | -0.0431*** | -0.0485*** | 0.0131 | 0.0366** | 0.0063 | 0.2144*** |
| ROA | 0.0671*** | 0.0543*** | 0.0573*** | 0.0650*** | 0.0509*** | 0.0551*** | -0.0009 | 0.0004 | -0.0184 | 0.0298** |
| CountofSubsidiaryName | 0.3904*** | 0.4004 | 0.3996*** | 0.3586*** | 0.3753*** | 0.3719*** | 0.1351*** | 0.1157*** | 0.0160 | 0.0008 |
| **Variables** | POST2 | POSTb | DAXPOST | DAXPOST2 | DAXPOSTb | logcTotalassets | Leverage | Nature | Tenure | ROA |
| POST2 | 1.000 | | | | | | | | | |
| POSTb | 0.8135*** | 1.000 | | | | | | | | |
| DAXPOST | 0.0080 | -0.0808*** | 1.000 | | | | | | | |
| DAXPOST2 | 0.1415*** | -0.0506*** | 0.0080 | 1.000 | | | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| DAXPOSTb | 0.0100 | 0.0987*** | 0.5552*** | 0.4004*** | 1.000 | | | | | |
| logcTotalassets | 0.0297** | 0.0071 | 0.2302*** | 0.1828*** | 0.1549*** | 1.000 | | | | |
| Leverage | -0.0201 | -0.0179 | 0.0654*** | 0.0419*** | 0.0411*** | -0.0010 | 1.000 | | | |
| Nature | -0.0105 | -0.0102 | 0.0032 | -0.0013 | 0.0207 | -0.2764*** | 0.1492*** | 1.000 | | |
| tenure | 0.3453*** | 0.1347*** | 0.0372** | 0.1190*** | 0.0219 | 0.0184 | -0.0499*** | 0.0118 | 1.000 | |
| ROA | 0.0380** | 0.0248* | -0.0004 | -0.0035 | -0.0187 | 0.0986*** | -0.1968*** | 0.1440*** | 0.0507*** | 1.000 |
| CountofSubsidiaryName | -0.0019 | -0.0207 | 0.0966*** | 0.0514*** | 0.0073 | 0.3757*** | -0.0079 | -0.1053*** | -0.0060 | 0.0574*** |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

This table presents descriptive statistics for variables used in an OLS regression. Panel A presents distributions; correlations are presented in Panel B. Variables definitions are given in Appendix B.

## Appendix E – Regression Results

**Table 4**

***Regression results Data Analytics***

*Panel A – Tests POST effect for treatment first*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 1422.238*** | 0.000 | 212733.100*** | 0.000 | 188772.300*** | 0.000 |
| POST | -102.761 | 0.218 | -9669.677 | 0.471 | -6778.122 | 0.562 |
| Adjusted $R^2$ | 0.0028 | | 0.0000 | | 0.0000 | |
| Observations | 190 | | 190 | | 190 | |

*Panel B – Tests without control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 408.122*** | 0.000 | 60952.200*** | 0.000 | 57210.630*** | 0.000 |
| DA | 1014.116 *** | 0.000 | 151780.900*** | 0.000 | 131561.700*** | 0.000 |
| POST | -8.293 | 0.511 | -2316.701 | 0.241 | -2341.136 | 0.187 |
| DAXPOST | -94.467 | 0.164 | -7352.976 | 0.489 | -4436.986 | 0.642 |
| Adjusted $R^2$ | 0.2112 | | 0.2083 | | 0.2011 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -442.528*** | 0.000 | -81199.960*** | 0.000 | -70976.910*** | 0.000 |
| DA | 727.022*** | 0.000 | 104063.000*** | 0.000 | 88713.140*** | 0.000 |
| POST | 7.521 | 0.497 | 443.651 | 0.793 | 234.771 | 0.877 |
| DAXPOST | -58.381 | 0.314 | -1894.808 | 0.831 | 399.787 | 0.960 |
| logcTotalassets | 69.250*** | 0.000 | 12097.550*** | 0.000 | 10903.780*** | 0.000 |
| Leverage | 72.487*** | 0.000 | 8061.627*** | 0.002 | 7464.058*** | 0.001 |
| Nature | 147.473*** | 0.000 | 21460.800*** | 0.000 | 19590.600*** | 0.000 |
| tenure | -14.439*** | 0.007 | -2834.054*** | 0.001 | -2741.115*** | 0.000 |
| ROA | 101.927 | 0.226 | -3975.699 | 0.757 | -823.232 | 0.943 |
| CountofSubsidiaryName | 24.991*** | 0.000 | 3839.028*** | 0.000 | 3414.462*** | 0.000 |
| Adjusted $R^2$ | 0.4261 | | 0.4522 | | 0.4482 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 4 presents the result from an OLS regression of the application of DA. The regression is performed with SumofHR, SumofEUR and SumofEUR2 as dependent variable. Variables definitions are given in Appendix B.

**Table 5**

***Regression results Data Analytics after first year implementation***

*Panel A – Tests POST effect for treatment first*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 1438.650*** | 0.000 | 221898.100*** | 0.000 | 198336.000*** | 0.000 |
| POST2 | -134.935* | 0.098 | -19737.900 | 0.123 | -17241.160 | 0.124 |
| | | | | | | |
| Adjusted $R^2$ | 0.0138 | | 0.0110 | | 0.0109 | |
| Observations | 128 | | 128 | | 128 | |

*Panel B – Tests without control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 422.0842*** | 0.000 | 62545.110*** | 0.000 | 57210.630*** | 0.000 |
| DA2 | 1016.566*** | 0.000 | 159353*** | 0.000 | 139724.900*** | 0.000 |
| POST2 | -15.780 | 0.178 | -2972.748 | 0.104 | -3124.491* | 0.057 |
| DAXPOST2 | -119.154* | 0.085 | -16765.160 | 0.120 | -14116.67 | 0.144 |
| | | | | | | |
| Adjusted $R^2$ | 0.1480 | | 0.1508 | | 0.1462 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -478.669*** | 0.000 | -86717.780*** | 0.000 | -75743.570*** | 0.000 |
| DA2 | 692.710*** | 0.000 | 105367.900*** | 0.000 | 91261.600*** | 0.000 |
| POST2 | -12.791 | 0.224 | -2205.778 | 0.168 | -2365.089* | 0.099 |
| DAXPOST2 | -66.158 | 0.260 | -8239.894 | 0.358 | -6424.189 | 0.423 |
| logcTotalassets | 73.606*** | 0.000 | 12724.74*** | 0.000 | 11448.270*** | 0.000 |
| Leverage | 80.194*** | 0.000 | 9241.552*** | 0.001 | 8502.242*** | 0.000 |
| Nature | 163.244*** | 0.000 | 23760.010*** | 0.000 | 21566.160*** | 0.000 |
| tenure | -13.090** | 0.022 | -2657.309*** | 0.002 | -2520.554*** | 0.001 |
| ROA | 79.098 | 0.363 | -7143.547 | 0.590 | -3451.669 | 0.771 |
| CountofSubsidiaryName | 25.784*** | 0.000 | 3948.876*** | 0.000 | 3507.222*** | 0.000 |
| | | | | | | |
| Adjusted $R^2$ | 0.3894 | | 0.4207 | | 0.4188 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 5 presents the result from an OLS regression of the application of DA after the first year DA was used during the audit. The regression is performed with SumofHR, SumofEUR and SumofEUR2 as dependent variable. Variables definitions are given in Appendix B.

**Table 6**

***Regression results advanced Data Analytics***

*Panel A – Tests POST effect for treatment first*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 1441.838*** | 0.000 | 221334.600*** | 0.000 | 199919.000*** | 0.000 |
| POSTb | -39.030 | 0.778 | -3349.606 | 0.880 | -6472.648 | 0.740 |
| | | | | | | |
| Adjusted $R^2$ | 0.0000 | | 0.0000 | | 0.0000 | |
| Observations | 60 | | 60 | | 60 | |

*Panel B – Tests without control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 452.721*** | 0.000 | 67391.410*** | 0.000 | 62695.840*** | 0.000 |
| DAb | 989.116*** | 0.000 | 153943.200*** | 0.000 | 137223.200*** | 0.000 |
| POSTb | -44.914*** | 0.000 | -7529.365 | 0.104 | -6702.932*** | 0.000 |
| DAXPOSTb | 5.883 | 0.964 | 4179.760 | 0.835 | 230.284 | 0.990 |
| | | | | | | |
| Adjusted $R^2$ | 0.0776 | | 0.0797 | | 0.0764 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -517.621*** | 0.000 | -86717.780*** | 0.000 | -81010.950*** | 0.000 |
| DAb | 803.383*** | 0.000 | 122729.900*** | 0.000 | 109153.200*** | 0.000 |
| POSTb | -34.250*** | 0.001 | -5690.046*** | 0.000 | -5011.875*** | 0.000 |
| DAXPOSTb | -92.853 | 0.386 | -12644.850 | 0.438 | -14850.37 | 0.309 |
| logcTotalassets | 78.187*** | 0.000 | 13423.090*** | 0.000 | 12066.170*** | 0.000 |
| Leverage | 90.423*** | 0.000 | 10803.920*** | 0.000 | 9878.511*** | 0.000 |
| Nature | 161.908*** | 0.000 | 23537.770*** | 0.000 | 21406.100*** | 0.000 |
| tenure | -9.850* | 0.074 | -2152.152*** | 0.010 | -2143.238*** | 0.004 |
| ROA | 93.796 | 0.291 | -4833.425 | 0.721 | -1604.977 | 0.894 |
| CountofSubsidiaryName | 27.838*** | 0.000 | 4262.056*** | 0.000 | 3776.415*** | 0.000 |
| | | | | | | |
| Adjusted $R^2$ | 0.3641 | | 0.3968 | | 0.3956 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 6 presents the result from an OLS regression of the application of advanced DA. The regression is performed with SumofHR, SumofEUR and SumofEUR2 as dependent variable. Variables definitions are given in Appendix B.

**Table 7**

***Regression results Data Analytics for deflated dependent variables***

*Panel A – Tests POST effect for treatment first*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 121.007*** | 0.000 | 17718.68*** | 0.000 | 15500.090*** | 0.000 |
| POST | -8.373 | 0.207 | -727.001 | 0.483 | -421.040 | 0.632 |
| Adjusted $R^2$ | 0.0032 | | 0.0000 | | 0.0000 | |
| Observations | 190 | | 190 | | 190 | |

*Panel B – Tests without control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 38.236*** | 0.000 | 5635.638*** | 0.000 | 5289.625*** | 0.000 |
| DA | 82.770*** | 0.000 | 12083.040*** | 0.000 | 10210.470*** | 0.000 |
| POST | 0.646 | 0.536 | -14.436 | 0.928 | -23.007 | 0.187 |
| DAXPOST | -9.019 | 0.109 | -712.564 | 0.408 | -398.032 | 0.642 |
| Adjusted $R^2$ | 0.2020 | | 0.1997 | | 0.1883 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 21.348*** | 0.000 | -2803.169*** | 0.000 | -2126.374*** | 0.000 |
| DA | 65.264*** | 0.000 | 8715.600*** | 0.000 | 7248.342*** | 0.000 |
| POST | .915 | 0.350 | 2.627 | 0.986 | 5.563 | 0.966 |
| DAXPOST | -4.310 | 0.405 | 94.464 | 0.902 | 320.683 | 0.640 |
| logcTotalassets | .000*** | 0.000 | 658.881*** | 0.000 | 572.919*** | 0.000 |
| Leverage | 8.951*** | 0.000 | 969.529*** | 0.000 | 917.563*** | 0.000 |
| Nature | 10.962*** | 0.000 | 1938.119*** | 0.000 | 1772.021*** | 0.000 |
| tenure | -1.529*** | 0.001 | -288.197*** | 0.000 | -283.056*** | 0.000 |
| ROA | 28.367*** | 0.000 | 617.590 | 0.580 | 944.133 | 0.344 |
| CountofSubsidiaryName | 2.511*** | 0.000 | 342.132*** | 0.000 | 303.092*** | 0.000 |
| Adjusted $R^2$ | 0.3284 | | 0.3662 | | 0.3539 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 7 presents the result from an OLS regression of the application of DA. The regression is performed with DefHR, DefEUR and DefEUR2 as dependent variable. Variables definitions are given in Appendix B.

**Table 8**

***Regression results Data Analytics after first year implementation for deflated dependent variables***

*Panel A – Tests POST effect for treatment first*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 121.887*** | 0.000 | 18488.600*** | 0.000 | 16290.440*** | 0.000 |
| POST2 | -11.401* | 0.086 | -1715.372* | 0.089 | -1433.639* | 0.096 |
| Adjusted $R^2$ | 0.0155 | | 0.0151 | | 0.0109 | |
| Observations | 128 | | 128 | | 128 | |

*Panel B – Tests without control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 40.605 *** | 0.000 | 5921.563*** | 0.000 | 5555.727*** | 0.000 |
| DA2 | 81.281*** | 0.000 | 12567.040*** | 0.000 | 10734.710*** | 0.000 |
| POST2 | -1.883* | 0.052 | -315.296** | 0.033 | -332.661** | 0.012 |
| DAXPOST2 | -9.518 * | 0.096 | -1400.076 | 0.109 | -1100.977 | 0.156 |
| Adjusted $R^2$ | 0.1395 | | 0.1434 | | 0.1353 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -13.176*** | 0.000 | -3290.897*** | 0.000 | -2531.555*** | 0.000 |
| DA2 | 60.423*** | 0.000 | 8998.814*** | 0.000 | 7601.310*** | 0.000 |
| POST2 | -1.199 | 0.196 | -186.109 | 0.180 | -206.316* | 0.095 |
| DAXPOST2 | -4.181 | 0.420 | -559.700 | 0.471 | -339.675 | 0.623 |
| logcTotalassets | 3.800*** | 0.000 | 711.868*** | 0.000 | 618.544*** | 0.000 |
| Leverage | 8.843*** | 0.000 | 1071.999*** | 0.000 | 1006.593*** | 0.000 |
| Nature | 14.991*** | 0.000 | 2133.346*** | 0.000 | 1936.551*** | 0.000 |
| tenure | -1.473 *** | 0.003 | -278.585*** | 0.000 | -266.102*** | 0.000 |
| ROA | 14.193* | 0.064 | 353.456 | 0.758 | 731.656 | 0.474 |
| CountofSubsidiaryName | 2.324*** | 0.000 | 351.372*** | 0.000 | 310.762*** | 0.000 |
| Adjusted $R^2$ | 0.3012 | | 0.3314 | | 0.3215 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 8 presents the result from an OLS regression of the application of DA after the first year DA was used during the audit. The regression is performed with DefHR, DefEUR and DefEUR2 as dependent variable. Variables definitions are given in Appendix B.

**Table 9**

***Regression results advanced Data Analytics for deflated dependent variables***

*Panel A – Tests POST effect for treatment first*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 123.177*** | 0.000 | 18575.43*** | 0.000 | 16518.760*** | 0.000 |
| POSTb | -3.426 | 0.746 | -356.614 | 0.829 | -554.598 | 0.695 |
| | | | | | | |
| Adjusted $R^2$ | 0.0000 | | 0.0000 | | 0.0000 | |
| Observations | 60 | | 60 | | 60 | |

*Panel B – Tests without control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 43.010*** | 0.000 | 6298.358*** | 0.000 | 5863.469*** | 0.000 |
| DAb | 80.167*** | 0.000 | 12277.070*** | 0.000 | 10655.290*** | 0.000 |
| POSTb | -4.069*** | 0.000 | -658.028*** | 0.000 | -586.336*** | 0.000 |
| DAXPOSTb | 0.642 | 0.952 | 301.413 | 0.853 | 31.738 | 0.982 |
| | | | | | | |
| Adjusted $R^2$ | 0.0757 | | 0.0781 | | 0.0729 | |
| Observations | 4,423 | | 4,423 | | 4,423 | |

*Panel C – Multivariate tests with control variables*

| | DefHR | | DefEUR | | DefEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | -16.530*** | 0.000 | -3777.455*** | 0.000 | -2962.778*** | 0.000 |
| DAb | 68.230*** | 0.000 | 10216.610*** | 0.000 | 8843.464*** | 0.000 |
| POSTb | -3.112*** | 0.000 | -495.449*** | 0.000 | -436.698*** | 0.000 |
| DAXPOSTb | -3.843 | 0.684 | -543.107 | 0.701 | -688.949 | 0.584 |
| logcTotalassets | 4.198*** | 0.000 | 770.964*** | 0.000 | 669.748*** | 0.000 |
| Leverage | 9.734*** | 0.000 | 1204.659*** | 0.000 | 1120.888*** | 0.000 |
| Nature | 14.858*** | 0.000 | 2111.944*** | 0.000 | 1921.117*** | 0.000 |
| tenure | -1.180** | 0.015 | -233.062*** | 0.001 | -232.955*** | 0.000 |
| ROA | 15.536** | 0.047 | 560.482 | 0.632 | 893.801 | 0.391 |
| CountofSubsidiaryName | 2.506*** | 0.000 | 378.468*** | 0.000 | 333.528*** | 0.000 |
| | | | | | | |
| Adjusted $R^2$ | | 0.2727 | 0.3048 | | 0.2962 | |
| Observations | | 4,423 | 4,423 | | 4,423 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 9 presents the result from an OLS regression of the application of advanced DA. The regression is performed with DefHR, DefEUR and DefEUR2 as dependent variable. Variables definitions are given in Appendix B.

**Table 10**

***Regression results simple and advanced Data Analytics split***

*Panel A – Tests without simple DA firms*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 408.882*** | 0.000 | 60575.310*** | 0.000 | 56661.580*** | 0.000 |
| DAb | 1032.955*** | 0.000 | 160759.300*** | 0.000 | 143257.500*** | 0.000 |
| POSTb | -1.075 | 0.925 | -713.256 | 0.690 | -668.669 | 0.677 |
| DAXPOSTb | -37.954 | 0.753 | -2636.349 | 0.889 | -5803.979 | 0.731 |
| | | | | | | |
| Adjusted $R^2$ | 0.0898 | | 0.0917 | | 0.0875 | |
| Observations | 4,319 | | 4,319 | | 4,319 | |

*Panel B – Tests without advanced DA firms*

| | SumofHR | | SumofEUR | | SumofEUR2 | |
|---|---|---|---|---|---|---|
| **Variable** | **Coefficient** | **p-value** | **Coefficient** | **p-value** | **Coefficient** | **p-value** |
| Intercept | 402.982*** | 0.000 | 60550.100*** | 0.000 | 56732.320*** | 0.000 |
| DAs | 1010.210*** | 0.000 | 148213.100*** | 0.000 | 126895.400*** | 0.000 |
| POSTs | 2.895 | 0.819 | -1128.424 | 0.570 | -1116.420*** | 0.531 |
| DAXPOSTs | -135.070 | 0.101 | -11458.210 | 0.374 | -5802.689 | 0.616 |
| | | | | | | |
| Adjusted $R^2$ | 0.1459 | | 0.1415 | | 0.0729 | |
| Observations | 4,375 | | 4,375 | | 4,375 | |

*,**,*** Indicate significance at 0.10, 0.05 and 0.01 levels, respectively

Table 10 presents the result from an OLS regression of splitting simple and advanced DA. The regression is performed with DefHR, DefEUR and DefEUR2 as dependent variable. Variables definitions are given in Appendix B.