Faculteit: Geesteswetenschappen
Liberal Arts & Sciences
Lucas van der Schalk (3785521) Economie
Redmar Legius (3591832) Nieuwe Media & Digitale Cultuur

Datum: 18 juni 2015
Begeleider: Elsbeth Spelt
Vakreferent ECO: Yolanda Grift
Vakreferent NMDC: Thomas Boeschoten

Table of Contents

1. Introduction: big data

Big data is a topic that is accompanied by a lot of expectations. Most of them involve great change that goes as far as calling it a revolution and transformation of society. Mayer-Schönberger and Culier wrote a book with the title: Big Data: A Revolution that will transform the way we live, we think and work" (2014). It is a New York Times best seller and is longlisted for the Financial Times/Goldman Sachs Business Book of the Year Award. University Professor Gary King" also claims there is a big data revolution on hand in an interview with Harvard Magazine. He says it is not the quantity that is new but the notion that now, we can finally do something with all our data (Shaw, 2014).

Big data is often called a buzzword. The hype surrounding this term seems to justify this claim. To understand what we are talking about, it is important to define this term. When searching for a definition of big data, no real consensus about one true explanation of the concept comes forward. Thus, defining big data seems a problematic process. A definition that is used by many is the one coined by Gartner inc. in 2001. They describe big data by three characteristics, all beginning with the letter V. This definition says that big data consists of volume, velocity and variety (Laney, 2001). Many scholars see this as the best definition (EG: McAfee & Brynjolfsson (2012), Kerkimäe (Ed.) (2014), Diebold (2012)).

Big data has had an impact on the financial market and this has resulted in jobs removed by computerization as showed by Gsell & Gomber (2009) and Lin (2012) who describe a fundamental change in the financial industry caused by the use of big data. Big data also changed the way financial market is being operated on. "Wall street is essentially floating on a sea of mathematics and computer power" (Financial Crisis Inquiry Commision, 2011). This kind of trading that uses computers to analyze and execute trading opportunities based on complex mathematical models can be divided in two general related forms, algorithmic trading and high frequency trading. Algorithmic trading which utilizes preset formulas to buy, sell and hold positions in various financial instruments (Robert & Lajoux, 2011, p.229 as in Lin, 2013). High frequency trading on the other hand makes use of a computerized platform to execute large numbers of trades at super speed (Aldridge, 2010, p.1). Next to trading, computers with artificial intelligence are being used in asset management and risk assessment (Gerding, 2009).

In this paper, an interdisciplinary research will be presented that combines an economic view with that of a perspective coming from the discipline of Media Studies. We will present an assessment of the pros and cons of big data in finance, as we will reflect on the impact that it might have on society at whole. We will be focusing on the pros and cons (or: risks and costs) in finance, because we identify the financial industry as a precursor when talking about big data. It is one of the first industries that truly implemented this new technology. Big data methods occur within algorithmic trading. Algorithmic trading is taking over trading processes on the stock market (Choria et al., 2013, p.644). So, an assessment of risks and costs of big data within this industry could result in valuable lessons about big data for other industries. In chapter three, this assessment will be presented.

In chapter four the question: *To what extent does the implementation of 'big data' methods have an influence on society?* will be answered. To answer this question, a theoretical framework presented by media scientist Marshall McLuhan will be used. He presented the idea of *the medium is the message* (McLuhan & Fiore, 1967). When we see big data as a new medium through the eyes of McLuhan, which will further be explained in this fourth chapter, we should focus on the influence of this new technology as a method. So, we should focus on the algorithms, because they are the message. It is not about the data that is processed. It is the way that these data are processed, that should be the main focus of interest, because that will have the most significant impact on society.

The integration of our disciplinary insights will help to answer our main interdisciplinary research question:

What are the risks and benefits of the implementation of big data in finance and how does this new technology influence society?

Impacts of technologies cannot be captured by just one discipline since technologies get implemented in various levels of society and affect our lives in many ways. Because of this a broader scope of different disciplines is needed. This is the case with our research to the trustworthiness of big data since artificial intelligence and algorithms are being used in many different ways. Analyzing big data from more than one discipline gives an important insight on how big data is being used, what consequences it has and if it is reasonable to rely on it. Therefore, using an interdisciplinary approach is justified. (Repko, 2012, p.84-85) Furthermore, to investigate the trustworthiness of big data, since it is being used in various ways, makes it necessary to check the relevant instances where we are relying on it.

While we approach this interdisciplinary research from the perspective of both economics and media studies, other potentially relevant disciplines could be mathematics, informatics and artificial intelligence. These could all give a deeper insight in the exact workings and architecture of algorithms that are used in various ways to implement big data processes in finance and other areas. However, in our research we are not focusing on this technical side of the algorithms, we will focus on the way big data has been used in finance and the way it influences society now, and possible effects it will have later. In this way, we do not need an additional discipline that can give us insights in the underlying structure of the algorithms behind big data.

2. What is big data?

2.1 No consensus: a lot of definitions

To work with the term big data in our research question, we need to provide a definition of what we think this term really means. We stumble across a first problem when searching for a single and universally used definition. As Mayer-Schonberger and Cukier put it in their research on the matter: “There is no rigorous definition of big data” (2013). The term big data is probably originated in lunch-table conversations at Silicon Graphics Inc. (SGI) in the mid 1990s, states professor Francis Diebold in his research towards an understanding of the origins of the term big data. Later, he adds: “An unpublished 2001 research note by Douglas Laney at Gartner enriched the concept significantly.” (Diebold, 2012, p.5.). A few insights we found while searching for a single definition are listed below, purely to give an idea of the different understandings about what big data is.

"Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time." (Gold, 2012).

"The term "Big Data" refers to the massive amounts of digital information companies and governments collect about human beings and our environment." (CSA Big Data Working group, 2013).

“Big Data is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software.” (Snijders, Matzat & Repis, 2012).

"Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it." (Ariely, 2013).

"Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale." (Hashem at al., 2015)

"Big data is the new oil" (Kroes, 2014)

"Advancing trends in technology that open the door to a new approach to understanding the world and making decisions." (Lohr, 2012)

"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." (Laney, 2001).

It becomes clear to us that a myriad of definitions can be found. There are a lot of different understandings about what big data really is. We would like to stress defining “big data” is a problematic process.

The “big” in big data makes the term inseparable with the words immense, (too) large, high volume, gigantic etc. This seems to indicate that it is crystal clear that big data also means a lot of data, but that is arbitrary to say the least. As we can see in the definition by Hashem et al.

what is “big” can (and will) change when time passes. (Hashem et al., 2014) This can be seen as a result of the increase in both computing power and the possibility to store large quantities of data and also the rise of digital and mobile communication, which has made the world more connected, networked and processes more traceable has led to the possibility to gather such large scale data sets. (Rainie & Wellman, 2012).

2.2 The Three V’s

The definition by Doug Laney, an analyst of Gartner Inc., dates back from 2001 (at that time Gartner was called META group research). Laney identifies three aspects that are essential and form the core ingredients of big data. This is commonly described as the three V’s. Volume, Velocity and Variety (Laney, 2001). It is a well used definition of big data and many authors in both popular and academic backgrounds refer to it, because it gives the most complete description of the concept (McAfee & Brynjolfsson (2012), Kerkimäe (Ed.) (2014), Diebold (2012)). With volume Laney refers to the amount of data, how big the dataset is, velocity looks at how fast new data is created and variety captures the different forms in the dataset (McAfee & Brynjolfsson, 2012, p. 62-63). The definition as given by Laney is used in this research based on the fact that it gives a well used and round description of the phenomenon.

2.3 “Big data” in finance

Together with technological advances in computer science, the growth of digital information leads to a form of finance where complex mathematics processed by computers at high speed plays a critical role in the decisions concerning capital allocation and risk assessment (Patterson, 2011). If we look at the definition of big data as given by Laney, the use of big data in finance can be identified because the three V’s seem to present themselves in algorithmic trading. This is an important notion because “big data” is not explicitly used in the bulk of economic literature on the subject. The technique of redefinition as stated by Repko helps us with this problem, because it involves modifying or redefining concepts in different texts and contexts to bring out a common meaning (Repko, 2012, p.336). In finance, when looking at the data the computers use with algorithmic trading, volume and velocity would capture the orders and prices that keep changing every second or less. Variety captures the fact that the algorithmic traders not only take the prices and orders but also news messages in their analyses for what to sell or buy. These trading systems seek to capture fleeting anomalies in market prices, profit from statistical patterns within or across financial markets, optimally execute orders, disguise trader's intentions, or detect and exploit rivals strategies (Chaboud et al., 2009).

3. The costs and benefits of algorithmic trading and high frequency trading

In this chapter we will look at the implementation of big data in the financial market, with algorithmic trading and high frequency trading. The impact is analyzed using a cost and benefit approach. The terms costs and benefits are operationalized in a broad sense as to also include advantages and disadvantages on the market. The chapter will begin with scrutinizing the benefits of algorithmic trading and high frequency trading in section 3.1, after that we look how this changed the industry in section 3.2 and then we look at the costs of algorithmic and high frequency trading in section 3.3. We conclude with a table summarizing the benefits and costs in section 3.4 that is going to be used for the integration.

3.1. Benefits of algorithmic trading

To start it is important to make the distinction between algorithmic trading and high frequency trading. The former is the use of electronic systems to place trading orders, including algorithms to decide some or all details of the orders. The distinguishing characteristic of high frequency trading is that trading decisions are made and implemented more quickly than humans possibly could. So high frequency trading is a subset of the much broader algorithmic trading (Kirilenko & Lo, 2013, p.1, Nuti et al, 2011, p.68). The focus will be on high frequency trading for the most part of this analysis. However, as high frequency trading is a subset from algorithmic trading, some costs and benefits from algorithmic trading are also discussed. Many of the benefits attributed to high frequency trading are in fact benefits of algorithmic trading. One of the aspects of algorithmic trading is that it is automated or partly automated, because of this it has several benefits. Algorithmic trading is part of a much broader trend in which computer-based automation has improved efficiency by lowering costs, reducing human error, and increasing productivity (Kirilenko & Lo, 2013: Hendershott et al, 2011, p.31). The computerization of finance has made investing available to a larger public and also reduced the costs for investors by removing the middleman in a lot of instances (Lin, 2013, p.725).

High frequency trading is profitable, grossing approximately 3 billion dollar a year. In addition, high frequency trading is less expensive than traditional market makers. The profitability is related to the volatility, the trading levels slightly increase with a rise of volatility (Brogaard. 2010, p.23). Volatility is the variation of the returns for an asset over time. Another market measurement influenced by high frequency trading is liquidity. Liquidity is about how long it takes to turn the asset into money. A painting is illiquid, since you need to look around for a buyer willing to pay the amount you want for the painting. In the context of trading liquidity can be seen as the amount of stocks traded on a daily basis. “In contrast to a number of public claims, high frequency traders do not as a rule engage in the provision of liquidity like traditional market makers. In fact, those that do not provide liquidity are the most profitable and their profits increase with the degree of 'aggressive,' liquidity-taking activity.” (Kirilenko & Lo, 2013, p.60, Clements, 2012).

The strategies used by the trading systems to gain these profits can be divided into three categories. The first type of strategy is based on trading on news, exploiting a time advantage to act on the news before the market does. This means making use of an information advantage. The second strategy is making use of arbitrage (Fabozzi et al, 2011, p.24). Brogaard found that algorithmic traders and especially high frequency traders use a short-term price reversal strategy (Brogaard, 2012, p.18). Arbitrage is the simultaneous purchase and sale of a financial asset in order to make a profit from the difference in the price of identical or similar assets, on different markets or in different forms. A source working on the North America Exchange reportedly said: “High frequency traders are looking for arbitrage on intra-day mean reversion.” (Fabozzi et al, 2011, p.25). “Arbitrage trading is as old as financial markets, but using algorithms to identify and exploit arbitrage-trading opportunities is a thoroughly modern invention, facilitated by the use of computers, applications of probability and statistics, advances in telecommunications, and the development of electronic markets.”(Kirilenko & Lo, 2013). Next to arbitrage's risk/reward profile, arbitrageurs almost always increase trading activity, so an increase in liquidity and since arbitrageurs exploit temporary mispricing, the price discovery process has become more efficient (Kirilenko & Lo, 2013, p.58). With high frequency trading another way to profit from arbitrage was found, using the time difference between markets and firms that are further away. It takes light approximately 1 millisecond to travel 100 miles (Tabb, 2012, p.4). The third type of strategy is 'front running', trying to make short-term forecasts based on the econometric properties of data (Fabozzi et al, 2011, p.24). If for instance investor A would know that a large order is going to come, investor A would buy what is available, he would then be able to make a profit of it. But algorithmic trading is also used to battle front running from competitors, since it is used to split up large orders into smaller ones and creating fake offers that they cancel the next moment to hide the strategy used by the trader. But the splitting of orders has more benefits.

Algorithmic trading systems place more but smaller orders in comparison to human traders, so they are able to capture the best bid or offer in the order book. These smaller executions also create less market impact than larger executions that could affect more than one price level (Gsell & Gomber, 2009). This also reduces the costs, because almost all stocks have downward-sloping demand curves over a short period of time, so one big order is usually more costly than a sequence of smaller orders (Kirilenko & Lo, 2013).

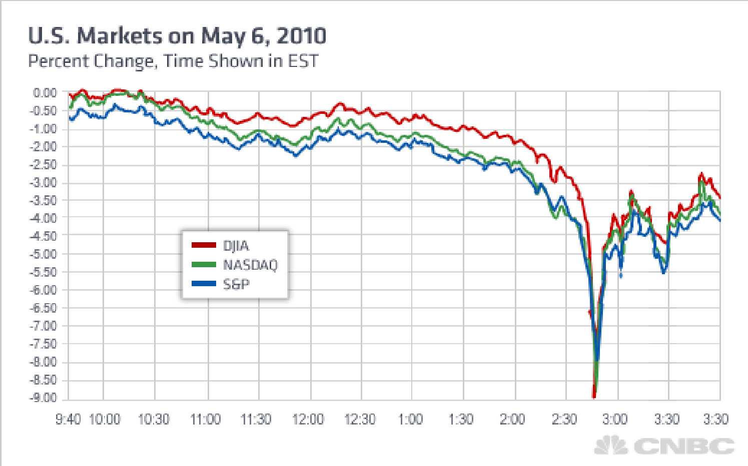
High frequency trading has narrowed the bid-ask spread (Fabozzi et al, 2011, p.24). Earlier volatility and liquidity were mentioned and both are factors that influence the bid-ask spread. The bid-ask spread is the difference between the bid and ask price. A higher volatility increases the bid-ask spread since there is more uncertainty about the return of the asset. If an asset is liquid, there is already a high quantity that is being traded so both the buy and sell price are closer together. Under normal circumstances, high frequency trading can be a positive force in markets, increasing liquidity and decreasing volatility in the short term by enhancing trade volume and execution speeds (Partnoy, 2012, p.43). As a result of that quotes, the price of the last successful trade, have become more informative. High frequency trading helps the price discovery process (Brogaard, 2010, p.36, Hendershott et al, 2011, p.31).

The last benefit of high frequency trading is the data that it uses or requires and the value this data can have for researchers. The fuel for high frequency trading is the High Frequency Data, data taken at intraday frequencies and Ultra High Frequency Data, tick-by-tick data, data relative to each trade. (Ultra)High frequency data, can give us more insight in what goes on on the market. This goes both for what the high frequency traders are doing, but it can also give insight in lower frequency phenomena (Golub et al, 2012). For example using high frequency data it was shown that on a short time scale the influence of stocks is stronger than the influence of the index, the reason for this appears to be the difference in how often it is updated and so was shown how high frequency traders can manipulate the market (Kenett et al, 2013).

To conclude, the benefits and advantages of algorithmic are for a great deal only because of a technological advantage over other market participants. The first benefits are from the automatization of the trading process. These benefits allow the traders the strategies discussed so far, which in turn has had some positive effects on the market. The algorithms can decrease volatility and increase liquidity which in turn can narrow the bid-ask spread and thus help the price discovery process. Another advantage is that algorithms create smaller offers creating less market impact, reducing trading costs. Although some of these benefits and advantages also apply to other market participants, most of the benefits only apply to the traders. An overview of all benefits discussed can be found in table 1 in paragraph 3.3.

3.2. Costs of algorithmic trading

One of the clearest perils of high frequency trading is the flash crash on May 6, 2010 as is graphically represented to the right. On the horizontal axis one sees the time and on the vertical axis the percentage change of the value compared to the previous day closing value. Around 2:32 p.m. a trading algorithm from a mutual fund company started a trade to sell 75,000 of E-Mini S&P futures contracts, worth about $4.1 billion (Bowley, 2010). The program managed to complete the trade in 20 minutes, a sale of this magnitude would take several hours or days to be completed in the days before automated trading (CFTC & SEC, 2010). Other trading programs started to react seconds later by also selling large blocks of S&P futures and this had an effect on the Dow Jones Industrial Average, which caused other stocks to also experience shocks. These shocks were both downward, like with Accenture shares which fell from $40 to $0.01 and also upward like with Sotheby's shares which rose from $34 to $99,999.99 (Haldane, 2011). At around 3:00 p.m. the Dow Jones Industrial Average had recovered most of the decline (Lin, 2012, p.705). The day ended with futures and equities indices closing at losses at around 3 percent from the prior day (CFTC & SEC, 2012, p.1). There has not been a large crash like the one on May 6th, but there have been a lot of small ones which are called mini flash crashes. These mini flash crashes were first identified by Nanex Llc. and there are down crashes as well as up crashes. A study shows that mini flash crashes have a negative impact on market liquidity, which results in a wider bid-ask spread and increased number of locked and crossed National Best Bid and Offer(NBBO) and a decrease in quoted volume (Golub et al, 2012, p.5).



Source: hedgethink.com

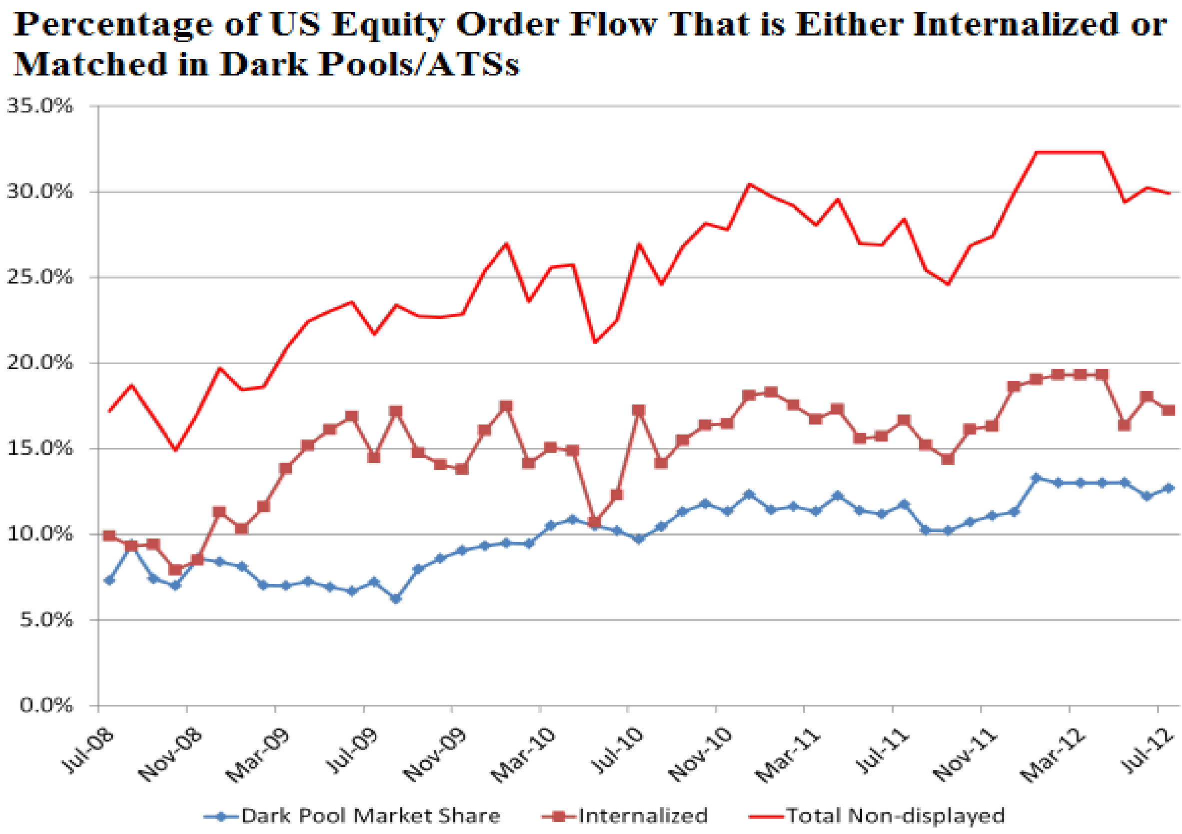
Another threat to the financial markets with the arrival and continuous growth of algorithmic trading is the threat of cybercrime and hacking. “High frequency trading is now estimated to account for 40 to 60 percent of all trading activity, computer and internet integration in finance has opened it up of course against cybercrime and hacking across the universe of financial markets, including stocks, derivatives, and liquid foreign currencies.” (Tabb, 2012, Kirilenko & Lo, 2013, p.59). Of course all industries are vulnerable to this threat, but the modern financial industry is especially vulnerable, because the computer codes and networks are at the very heart of its existence and many of these systems are self-executing with no human control

(Lin, 2012, p.705). A cyberattack causing a crash could cripple the financial system and shatter the confidence (Klein, 2001).

With the rise of algorithmic trading the market depth has declined. Some large institutions have complained that the decline has hampered investors to trade large volumes without substantial costs. The depth of the book is the ability of an asset to absorb buy and sell orders without the price being moved in either direction. But dark pools have helped investors to trade large orders anonymously (Hendershott et al, 2011, p.31).

With the velocity high frequency trading takes places new perils have come to finance. Billions of dollars move across borders and oceans through cables and spectra in a few milliseconds (Fabozzi et al, 2011, p.8). With this has come the risk know as too fast to save. Nowadays many checks and balances have been sacrificed for velocity and efficiency because of cyborg finances insistence on speed. As a result of this focus on speed, it has become more difficult to stop or prevent bad acts and actors. After the flash crash, regulators implemented circuit breakers to stop the trading of a stock if it the prices falls to fast, but this doesn't fully solve the problem since the trading on less regulated dark pools and electronic markets only grows (Lin, 2013, Kirilenko & Lo, 2013, p.60).

As mentioned before, the trading moves from the public exchanges to the dark pools. Partly because of this regulators can't keep a good eye on how is being traded (Nuti et al, 2011, p.68 : Pasquale, 2015, p.68). In the graph on the next page you can see the increase of internalizing orders and orders being executed on dark pools, it shows that in 4 years, from 2008 to 2012 the number of orders not displayed has doubled from 15% to 30%. Although this isn't a cost of high frequency trading, it is a consequences of it, since a few of the costs and benefits mentioned earlier of it have given the need for dark pools. Another reason why regulators have a problem keeping a good eye on the trading is due to the resource asymmetry (Lin, 2013, p. 723).



Source: Tabb, 2012

The resource asymmetry does not only exist between the regulator and the regulated, but also within the industry there is a resource asymmetry. The computerization of trading has made it more accessible for many, but it has also created barriers to competition. "The increasing dependence on advanced information technology has led to competition for scarce talent and resources that are often captured by the most successful and most moneyed." (Patterson, 2012, p.230). The fact that some players have more resources isn't new, but the differences in resources may not only be of the degree, but also of the kind and this might have an impact on the very function and integrity of the financial system (Lin, 2013, p. 726).

Another risk that has come with algorithmic trading and the change of finance is that everything is now more and more connected to each other. Modern finance exists as an expansive, interconnected network that crosses institutions, industries, states, and products—creating a systemic problem that could be called 'too linked to fail' (Lin, 2013, p.714). This connectedness of finance has enhanced the mobility of capital and reduced some risks (Morgan, 2003). But with this connectedness also has come the risk that if one financial institution falls others are dragged down with it (Kolb 2011, p.xiii).

Another disadvantage of high frequency trading is that it doesn't enhance social welfare or only very little (Pasquale, 2015). Since it operates on the financial market and not on the goods and service market, but as the banking crises has shown, it can also affect the goods and service market. The financial market is very handy for firms to raise capital, but most of the stocks traded involve shares who have already raised capital for the firm and so don't have a direct effect on the firm whose shares are being traded. The price of a share does have some relation with the firm's economic activity, but the trading by high frequency traders just increases the price slightly more quickly. As already mentioned earlier high frequency trading accounts for more and more of the trading done, and also earlier mentioned that this trading is only done by a small group. Only 2% of the 40,000 trading firms were estimated to use high frequency trading in 2010 (Clark, 2010, Nuti et al, 2011, p.68). The numbers that 40%-60% of trading activity is done by such a small group shows that high frequency trading only serves an elite few and because of dark pools and trade secret laws not much is known about these traders. “The number of entities that engage in high frequency trading is reportedly quite small and what is known about them is not particularly illuminating.” (Kirilenko et al, 2013, p.60).

One market fact that has been ascribed to high frequency trading by Professor Voev is the increase in correlation. This makes it very hard for investors to diversify with index tracking or exchange-traded funds. Professor Bauwens noticed this correlation too and said that it might come because all the high frequency traders are guided by econometric models and not fundamentals. This can in turn create artificial price trends (Fabozzi et al, 2011, p.26).

Lastly, as mentioned as one of the benefits, high frequency trading had some positive effects during normal circumstances, sadly the market is not always like that. During periods of high uncertainty, however, high frequency trading can exacerbate volatility and hurt liquidity by removing significant trading positions from the markets at warp speeds (Partnoy, 2012, p.43). Professor Voev has also warned for this, saying that high frequency trading might create volatility spikes (Fabozzi et al, 2011, p.26).

In contrast to the benefits, most costs and disadvantages do not effect only the traders and this has raised concerns about the impact of high frequency trading on the function and integrity of the financial system. The automatization of the trading process to get ever closer to zero has opened it up for cybercrime. To complete trades even faster, safeguards are skipped, since regulators cannot afford the resources the traders use. Trading on dark pools also has increased because the high frequency traders want to keep their intentions hidden. The speed makes high frequency trading to fast to save and the increased connectedness between the institutions but also indexes and funds has also made the financial system to linked to fail. While operating under normal market circumstances algorithmic and high frequency trading can have positive effects, both can also do the opposite when the market is not doing so well, resulting in crashes. All the costs are summed up in the next paragraph in table 3.1.

3.3. Table of cost and benefits of blackbox trading

In the table 3.1 on the next two pages all the benefits and costs of algorithmic trading and high frequency trading are presented. The benefits and costs are summed up in the order they were present in this chapter. This table will be used later on in chapter 5 for the integration of these insights to gain a more comprehensive understanding on the implications of the use of big data.

Benefits §3.1

benefit	***discription***	***reference***
reduced costs	tasks done by machines do not cost as much as done by a worker	Kirilenko & Lo, 2013 : Hendershott et al, 2011, p.31
reduced human error	humans are not good with repetition work	Kirilenko & Lo, 2013 : Hendershott et al, 2011, p.31
increased productivity	computers work faster	Kirilenko & Lo, 2013 : Hendershott et al, 2011, p.31
improved efficiency	with the 3 benefits above due to automatization efficiency has been improved	Kirilenko & Lo, 2013 : Hendershott et al, 2011, p.31
wider available	broader access with online investing	Lin, 2013, p.725
profitable	good money made with algorithmic trading	Brogaard. 2010, p.23
can decrease volatility if market is okay	algoritmic trading can decrease volatility by enhancing trade volume and execution speeds	Partnoy, 2012, p.43
can increase liquidity if market is okay	algorithmic trading can increase liquidity by enhancing trade volume and execution speeds.	Partnoy, 2012, p.43
information advantage(strategy type 1)	acting on news before the rest of the market	Fabozzi et al, 2011, p.24
technological advantage(arbitrage)	simultaneous purchase and sale of identical assets to make a profit from the difference in the price on different markets or in different forms	Fabozzi et al, 2011, p.24 : Brogaard, 2012, p.18 : Kirilenko & Lo, 2013
front running	trying to predict what will happen in the next few seconds on the market	Fabozzi et al, 2011, p.24
creates smaller offers	algorithmic trading executes smaller orders in comparison to human traders.	Gsell & Gomber, 2009
less market impact	smaller offers create less market impact, so less effect on the market price	Gsell & Gomber, 2009
reduced trading costs	since smaller offers have less impact than big offers this also reduce trading costs	Kirilenko & Lo, 2013
narrowed bid-ask spread	reduced the difference between the bid and ask price	Fabozzi et al, 2011, p.24
helps the price discovery process	since the bid and ask price are closer together and that algorithmic trading can increase liquidity and decrease volatility, the price of the last succesful trade has become more informative	Brogaard, 2010, p.36 : Hendershott et al, 2011, p.31
Data can give insights in high frequency trading, but also lower frequency phenomena	The data used by algorithmic trading can give a better understanding in the high frequency phenomena, but also the slower normal market phenomena	Golub et al, 2012

Costs §3.2

cost	*discription*	*reference*
flash crashes	the sudden sharp drop or rise in a financial asset, triggered by an algorithm	Bowley, 2010 : CFTC & SEC, 2010 : Lin, 2013, p.705 : Golub et al, 2012, p.5
Opened up to cybercrime and hacking	with the computerization finance has opened itself to the dangers linked with internet	Lo, 2013, p.59 : Lin, 2013, p.705 : Klein, 2001
less market depth	decrease in the ability of an asset to absorb buy and sell orders without the price being moved in either direction.	Hendershott et al, 2011, p.31
cost up for big orders	because big orders have more market impact	
too fast to save	trading takes places at such a high speed that intervention is next to impossible	Fabozzi et al, 2011, p.8 : Lin, 2013
to gain speed, certain safeguards are skipped	to gain more speed, some safety measures are not take	Lin, 2013, p.725
rise of dark pools	rise of private anonymous exchanges with little to no oversight	Lin, 2013 : Kirilenko & Lo, 2013, p.60
less oversight	regulators can't keep up with those who they should keep in check	Nuti et al, 2011, p.68 : Pasquale, 2015, p.68
resource asymmetry between regulator and regulated	regulators can not make the investment needed to be on the same level technology wise with the algorithmic traders	Lin, 2013, p. 723
resource asymmetry within industry	some market praticipants have better/more resources than others	Patterson, 2012, p.230 as in Lin, 2013
might have a impact on function and integrity of financial system	the worry that some aspects of algorithmic trading and high frequency trading are having such an impact that they might change the financial industry in a bad way	Lin, 2013, p. 726
too linked to fail	because of algorithmic trading everything is now more connected to each other	Lin, 2013, p.714
doesn't enhance social welfare	there is not invested in new business, but making money from trading shares that have already created the capital needed for the business	Pasquale, 2015
serves only elite few	Only 2% of the trading firms is estimated to use high frequency trading(2010)	Clark, 2010 : Nuti et al, 2011, p.68 : Kirilenko et al,2013, p.60
increase in correlation between indexes and funds	Professor Voev noticed that with the increase in correlation, it is becoming harder to diversify for investors	Fabozzi et al, 2011, p.26
artificial price trends	Professor Bauwens ascribes the increase in correlation to the econometric models used by high-frequency traders and that this in turn might create artificial price trends	Fabozzi et al, 2011, p.26
will exacerbate volatility with high uncertainty	algorithmic trading can exacerbate volatility by removing trading positions at warp speeds	Partnoy, 2012, p.43 : Fabozzi et al, 2011, p.26
take liquidity with high uncertainty	algorithmic trading can take liqudity by removing trading positions at warp speeds	Partnoy, 2012, p.43

4. The algorithm is the message

Big data is hot and happening. It is widely praised and seen as a core ingredient for a new industrial revolution, for future big advances in science and technology and even for solving social problems big data is seen as the solution. The term big data is a bit problematic as it is hard to define. We chose to use a definition coined by Gartner inc. in this research as described in chapter two of this interdisciplinary research. Gartner inc. identifies three main characteristics that define big data, the so called "three V's": Volume, Velocity and Variety. In their report they remark the increasing rate in which data is produced, the increase of a range of methods which is used to deal with this data and an increase in size and quantity of data in general (Laney, 2001). In finance, we see that the implementation of big data has started at an early stage. And lately more industries seem to be interested to incorporate big data in their business plans and cycles (Capgemini & EMC, 2015). The potential benefits and of using big data are indicated by a quote from Neelie Kroes, former EU-commissioner for the Digital Agenda. Neelie Kroes says that "Data is the new oil for the digital age" Kroes later stated that: "There is a completely new economic potential driven by the information society, the use of big data should therefore be enabled in an effective way."(Kroes, 2014). These quotes from a hugely influential person illustrate the potential that is attributed to big data. When talking about big data, often big changes are prospected. The use of big data has an effect on the economy (mostly when talking about efficiency) but also on healthcare, crime prevention and other areas. Further implementation of big data might have a significant impact on society as indicated by the book of professor Mayer-Schönbeger and data-journalist Kenneth Cukier. The title of their book: *Big Data: A Revolution That Will Transform How We Live, Work and Think* speaks for itself (Mayer-Schönberger & Cukier, 2013). Mayer- Schönbeger and Cukier see a lot of change around the corner – often in a positive manner, but this is not always the case - and this makes big data a highly interesting subject for further research.

However, according to the Canadian media theorist and philosopher Marshall McLuhan (1911-1980), data collection is certainly not something new. In the 1960's he wrote: "The effect of script and the ability to make inventories and collect and store data changed many social habits and processes back to early as 3000 B.C." (McLuhan, 2003, p.58). In addition, big data now might contain new elements that will influence these social habits and processes again. When we look at the implementation of big data, the financial sector stands out as one of the industries in which a lot of investments in big data have been made. An industry that was dominated by humans before, has now transformed to an industry that is partly human and partly computer, which has been called “cyborg finance” (Lin, 2013). On the stock market, a large share of the trading is now done by automated processes, powered by supercomputers that rely on complex algorithms that fight to be the best and fastest in a “race to zero” (Gsell & Gomber, 2009). Other industries are also interested and some have been using big data methods already; examples are in retail, healthcare, astronomy and even politics (Murdoch & Detsky, 2013). In a survey conducted by the firms Capgemini and EMC, 56% of the respondents expected to increase their investments in big data (about 1,000 decision makers from 9 industries in 10 countries responded to this online survey). (Capgemini & EMC, 2015). In addition, the effects of big data on the financial sector can be analyzed with theories about new technologies as point of reference. Especially the theory that the Canadian philosopher and scientist Marshall McLuhan provides could be very insightful since he has a broad understanding of what media is. His theory will therefore form the starting point of this paper. His theories can help us understand how the further implementation of big data might affect society. Therefore, the question that will be answered in this chapter is:

To what extent does the implementation of ‘big data’ methods have an influence on society?

4.1. Theory of McLuhan

To answer the question whether the further implementation of big data in the financial sector and other industries has a potential influence on society, the theories that Marshall McLuhan present could be useful. The question is if we could see the algorithms behind high frequency trading in finance as a new medium when looking at the conditions and characteristics that McLuhan offers. To research this insight, a closer look at the way McLuhan sees the medium as a concept is necessary. In addition, the (40-year old) theory will be applied to the concept of big data. As stated above, there are high expectations of the new technology (or *method*) that big data entails. In finance, big data is already implemented and it has truly changed the way the stock exchanges operate (Gsell & Gomber, 2009). This could be a first indication of the fact that we are dealing with a new medium that could have a significant influence on society, and even affect society on an imaginably deeper stage. To understand the processes of affect and influence, different books and interviews of and by Marshall McLuhan will be scrutinized. McLuhan, who is often seen and referred to as the most illustrious exponent of technological determinism, looks at new technologies and the way they influence and form society (Trembley, 2012). Therefore his perspectives offer a strong body of knowledge when examining questions about technologies and their influence on society. Other scholars during his time, such as the Welsh academic Raymond Williams, focused more on the social aspects and the way humans use new technologies that define society. McLuhan, in contrast, argues the other way around. An example is that only looking at how humans extend their sensory abilities and skillset through technology does not reflect strong enough on the possible ways in which technology has an influence on humans. Technology is used by us, but could also "use" us, and has a strong influence on humans (Lister et al, 2009, p. 78). McLuhan is also criticized for being more of a genius with words than with actual science. Philosopher and academic Regis Debray once called McLuhan more a poet than a historian in an interview with *Wired Magazine*, and did not praise him as a systematic analyst. He also said: "McLuhan overemphasizes the technology behind cultural change at the expense of the usage that the messages and codes make of that technology" and: "As he himself said, he was an explorer rather than an explainer" (Joscelyne, 1995). The insights that McLuhan presents help to understand the impact a new medium such as big data can have on society.

4.2. Big data as a new medium

Marshall McLuhan uses a very broad definition of a medium. Any new technology can be defined as a new medium according to McLuhan. One could say that he sees media not as what we would call: *the* media but that any human made technology is a medium. Examples that he presents of media are for instance not only television (programs), radio and newspaper but also trains, light bulbs and even clothes. His main point in his book *The Medium is the Message* is that any technology that has an influence on the social life and environment of humans can be a medium (McLuhan & Fiore, 1967). He also said that "when you put a new medium into play in a given population, all their sensory shifts a lot. This changes their outlook, their attitudes, changes their feelings about studies, about school, about politics" (McLuhan, 2003, p.100). If we follow this theory, a new technology such as big data could also change the view and behavior of the modern day society. Thereby media are all around us in such an imbedded way that media (or: technology) is in no way a bridge between man and nature because they simply are nature (McLuhan, 1969, p. 14). Added to this, the effects that media have change the environment and are in the same way imperceptible for the human eye as water is imperceptible in the eyes of a fish (McLuhan, 1969, p. 22). Some, such as the famous writer and semiotician Umberto Eco for instance, are critical about the broad conception McLuhan uses when talking about a medium. According to these critics the definition of McLuhan would be too broad and simplistic. It lacks precision and would cause confusion by seeing nearly everything as a medium (and thus a message) (Debray, 1996, p.71).

McLuhan stresses that the medium should be the main focus of studies, and not the messages that lie within the content different media present. This is where his most well-known expression derives from: the medium is the message. The fact that the television was brought into our homes and even became the center of the home of many, has for instance a far stronger

influence on history and society than the content that is distributed by this screaming box, according to McLuhan (McLuhan & Fiore, 1967).
In this way a light bulb can be a medium, even though it has not got any content to be analyzed such as there might be on a website, a news article in a paper, or a television program. The social influence is, however, extremely high.

By merely existing, the light bulb creates a new environment. It made working and seeing for a human possible during hours in which that was not possible before, or in the best case very hard. Big data could also create, or better: reshape the environment in which we exist because it is a new method that makes us capable of looking at things in a different way than before. The way in which big data has influenced the financial sector could be seen as practical example in this case: algorithms have become a "game changer" and have taken over the role of the human stock trader (Gsell & Gomber, 2009).

Another important aspect of the theory of McLuhan is the fact that a medium always forms an extension of the human capabilities, or even of the human senses. The wheel extends our legs and thus the capability to go from one place to another. After just more than a century with electricity, McLuhan described the electronic technologies that embraced the world as an extension of the human central nervous system (McLuhan, 1964, p. 19). The phone extended our mouth (or voice) and our ears as well. Extension in these examples also brings a certain upgrade to the human skill set, it improves the capabilities of the human through technology. In the 60s of the last century he described the next step as being an extended version of our consciousness that would combine all media and would collectively extend all human beings (1964, p.19). This prediction of McLuhan is often seen as an all too close predication of the Internet (Kroker, 1995). With regards to big data, specifically the algorithms that are used to compile, analyze and visualize data could be seen as an upgrade or extension of the human brainpower. The idea of technology as an extension of the human body is not entirely new however. It dates back to Aristoteles, who had a similar understanding of technology (Lister et al, 2009, p.90). Through all these theories, McLuhan might come off as a fan of everything that is new, but to understand him better it is necessary to know that he actually said the exact opposite about himself. He stressed that he is resolutely opposed to all innovation and change but that this aversion also forms his main motivation to understand these exact processes (McLuhan, 2003, p.101). When we see big data as a new medium through the eyes of McLuhan, the first answer to the research question "*To what extent does the implementation of 'big data' methods have an influence on society?"* is rather short: the impact of any new medium is vast, and thus the influence of big data on society is too.

4.3. A shift in trust: from human to computer.

The examples of the influence of light bulbs and television learn us what McLuhan really meant with “the medium is the message” and why he felt that focusing on the influence that that medium had was far more significant than the message(s) that were spread by these media. Big data could be thus be seen as a medium when looking at the theory of McLuhan since big data is a new technology that is an innovation which’ promise is a change of view, a new perspective on life through the creation of order in a chaos when data piled on data is organized by algorithms that can handle extreme amounts of information. But in what way does this method (that is big data) extend our senses? McLuhan described a human in the electronic age as follows: “an organism that now wears its brain outside its skull and its nerves outside its hide” (McLuhan, 1964 p. 64).

When taking this in consideration, a “brain outside its skull” could, with some imagination, be seen as the digital computer. Since the software on computers consists of lines of programming that were shaped by humans now help to “compute”. The first “computers” were human (mostly women) beings that, without any disrespect, were used as living calculators as intensively explained in the book “when computers were human” (Grier, 2013). In human nature, we are inclined to sometimes make mistakes, for instance by not doing the same thing exactly in the same way we did it before. Through digital computers human errors were supposed to be eliminated. Computers work, simply put, in sequences that never change from the lines of programming that are thought of by the human brain (unless they are actively told to do so). In a way, the digital computer is a method that comes directly from the human brain and it will always copy this method in the exact same way. The algorithms used in programming (and thus in big data as well) are always a product of the human mind. Data storage on digital computers could be seen as a collection of memories without forgetting. This endless repetition without any deviance in method or process is something that would be hard for a human to perform and thus the actions by the computer can be seen as an extension of the human skillset. Remembering large quantities of memories is for (most) humans also a big task. Big data is all about processing large volumes of data and using algorithms to try to understand what is going on. We use computer power to compensate for human shortcomings, shortcomings of our brain. In that way, if we follow the theory of McLuhan, big data can be seen as a new medium because it is a new way of using computers to help us understand and interpret data. It extends our brain and forms an improvement in speed of execution, remembering and in repetition. The algorithms used by investment companies are an example of how big data is used to operate on the market with brainpower and speed that exceeds the capabilities of any human (Gsell & Gomber, 2009). This shift of trust can be seen as one of the influences that the implementation of big data can have on society.

4.4. Big data hubris; too much trust?

Today, with big data, some might assume that the human aspect could be minimized by using computers. Since computers will execute what they are programmed to do, they imply to remain uninfluenced by emotion. The data in big data seems factual, and it is assumed this data cannot lie. Lying is a human characteristic that computers are incapable of, so seems to be the perception of a lot of people. Of course this might be an observation more than a claim, however, this idea nevertheless exists in society. Computers cannot be fooled, they have no hidden agenda and they seem to be perfectly stoic! It simply does its job. This makes computers trustworthy. Trust in technology is not very different from trust in humans, we rely on quality that is demonstrated in both effectiveness in help, a good looking interface and reliable and steady performance (McKnight, 2005). On the other hand, it is thinkable that too much trust could be a problematic factor. Being too confident in something is a human error not to be overlooked that can corrupt the clear view of even the most critical mind. Big Data hubris is a term Lazer et al. coined when talking about the trust in these data and even the assumption that is easily made: when it is data, when there is a computer (or: algorithm) involved it *has* to be right.

> "Big data hubris" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis. We have asserted that there are enormous scientific possibilities in big data (9–11). However, quantity of data does not mean that one can ignore foundational issues of measurement, construct validity and reliability, and dependencies among data (12). The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis (Lazer et al, 2014, p.1).

This is where the theory of McLuhan really seems to clarify itself. The medium (big data) is transforming how people look at things that possibly were not quantified before. This might affect their view on life itself and could lead to a shift in what is viewed as important in society. According to McLuhan the introduction of the alphabet bombarded sight to be the most important sense, while before literacy ever existed, all senses where more or less equally important (Norden, 1969, p.59). With the contemporary internet age with ever more important quantification and data streams, it might be that the focus of society is now shifting to digital data. The rise in applications and wearable devices that track the steps we take, the calories we burn and the sleep we get, could be a sign that this is actually happening (Swan, 2012). The internet of things and sensor mania are concepts that Swan used to describe the increase in both data collection through sensors all around us, and the datafication of everyday products that are now connected to the internet. The connection to the internet makes them traceable and makes them produce data, where they did not do this before. Everything is becoming "smart". Not only your phone but also your television, your car, your refrigerator, and your washing machine (Swan, 2012). These machines are all creating more and more (big) data. Big data might be viewed as a new way in which we can look at the world and create order, or at least understand the chaos of human life. Some put the method of big data next to the microscope like economist Brynjolffson. The method that is big data gives us another way of looking at the world, similar as the microscope did, only the scale is different (Lohr, 2012).

However, it is important to always keep in mind that in all aspects of big data, the human aspect certainly is present. McLuhan, describes every new medium or technology as an extension of our capabilities that forms the environment we live in. In all extensions of the human capabilities, the human is still the starting point. In big data, this is not different. While it seemingly are only computers that are working with data, in the production of algorithms and collection of data humans always play a role. More so, in analyzing and giving meaning to the data through interpretation. The human aspect should never be overlooked. The dream of a perfect supercomputer without any human "faults" imbedded in the system is a misconception that excludes computers and technology in general from nature entirely. The trust that is put in computers and algorithms can surely be seen as an influence of big data on society, that will be reinforced with success stories of this new technology. However, this trust is not always justifiable.

4.5. Conclusion: Datafication of life; a possible influence?

By implementing big data in finance and other areas, the focus of humans could be shifting to a more datafied life. Data is becoming more important, which is also seen in the fact that every two years, the amount of data doubles, as estimated by research firm IDC (Lohr, 2012). In finance, black boxes are trusted to handle trades in stock, transactions are executed by algorithms. The market that once relied on intuition and where companies and banks trusted on skills and experience of human traders (Lohr, 2012), now is led by computers. The current algorithms that buy and sell stocks could be seen as an extension of those experienced human traders. They now take every single variable in account that throughout history was significant enough to weigh in on a decision. The big traders in stocks are trying to minimize the human aspect in trading that once was arguably the first and foremost condition in which trading goods could actually be executed. Now we let algorithms, sell and buy "things", without any human interaction. This has changed the dynamics of trading fundamentally (Gsell & Gomber, 2009). McLuhan helps us to understand that the changes lie within this new method. The algorithms are the message. It is not about the data that is processed. It is the fact that these data are processed in the way they are that should be the main focus of study, because that will have the most significant impact on society. However, it should not be wise to see big data as a single, revolutionary breakthrough in this possible shift that affects society. Uncountable other factors are at play and it is always a progression that is subtle and slow. The 'moment' algebra was invented could also be seen as the moment that numbers and thus datafication started to have an influence. However, big data could be seen as a strong indicator that indeed, data seems to be more and more in the center of our interest and therefore has an influence on our social existence. It can change society, and Mayer-Schönberger and Cukier even see it as a revolution (Mayer-Schönberger & Cukier, 2013). A strong insight of McLuhan one might add is that it is hard to see what is happening while it is happening and influences that change society often can only be identified after they have taken place "We tend to cling to the rearview mirror of our world", McLuhan said in an interview with Playboy in March, 1969. "In that way we are always one step behind." (Norden, 1969). However, that does not make trying to understand the processes around us that (through media) influence society something we should not try to do.

To answer the research question of this chapter: To what extent do 'big data' methods have an influence on society? We need to look at this new technology as a new medium, as described by Marshall McLuhan. When following the theory of McLuhan, big data is a new technology that is being implemented and thus it has an effect already. Not only can big data presumably help us understanding new things, create order in a seemingly mess of worthless data, it is also changing the way we behave. The seemingly more important role for data in everyday life could be one of them, as is the seemingly trust in data. A deeper understanding of the influence big data has on society can only be found when a scholar during his research focuses on the method and thus algorithms itself, more than on the content the algorithms produce.

5. Integration and conclusion

In this section the integrating process that was used will be explained. The more comprehensive understanding of big data that was a result of this process will be presented as well. The chapter starts with the creation of common ground by connecting the findings of media studies to three topics that were found in both of the disciplinary studies. Then the focus shifts towards each of the topics and the costs and benefits that were found. These findings, including some indirect consequences and relations between the topics, will be presented with numbered lines in figures 1, 2, 3, 4 & 5. Finally, a more comprehensive understanding that can be gained from this research will be described by explaining Figure 6.

A significant step in the integration process is to find and report possible conflicts in use of terminology, definitions, assumptions, or main insights that might create friction between the different disciplines in an interdisciplinary research (Repko, 2012, p.335). In this research, no substantial conflicts where found. Except for one issue that we dealt with in chapter two, where big data was defined by using the definition of Gartner Inc. (Laney, 2001). By using this definition of big data, it was possible to define algorithmic trading as a big data process. The problem we dealt with here, was that even though in the economic literature "big data" was not mentioned literally, big data methods can be identified in finance. By using the technique of redefinition, we were able to identify the slight difference in disciplinary vocabulary, and by "textual integration" (we see algorithmic trading as a process that could be called "big data") we were able to get the both disciplines "on the same page" (Repko, 2012, p.336).

5.1. From disciplinary perspectives to data, automatization and algorithm

To create common ground for the integration of the disciplinary insights, we use the technique of organization as described by Repko in *Interdisciplinary Research, Process and Theory* (2012). This technique requires making a map and clarifying how the concepts are related to each other (Repko, 2012, p.346-347). The costs and benefits of algorithmic and high frequency trading are divided along three topics. To divide the costs and benefits, they were attributed to the following topics: 'data', 'automatization' and 'algorithm', based on which topic they relate to the most. These three topics are a result of the linking of the findings from media studies and the costs and benefits analysis of the economic discipline. In chapter three it was found that big data is a new computerized technology, so the topic automatization captures this aspect. Moreover, big data makes use of algorithms to make sense of the data and to act on the data or to deliver results, this makes algorithm our next topic. Lastly, data is an essential component of this technology, therefore this was added as last topic. With these three topics, the different aspects of big data are captured. The three topics as described are also the general ways in which big data might influence society.

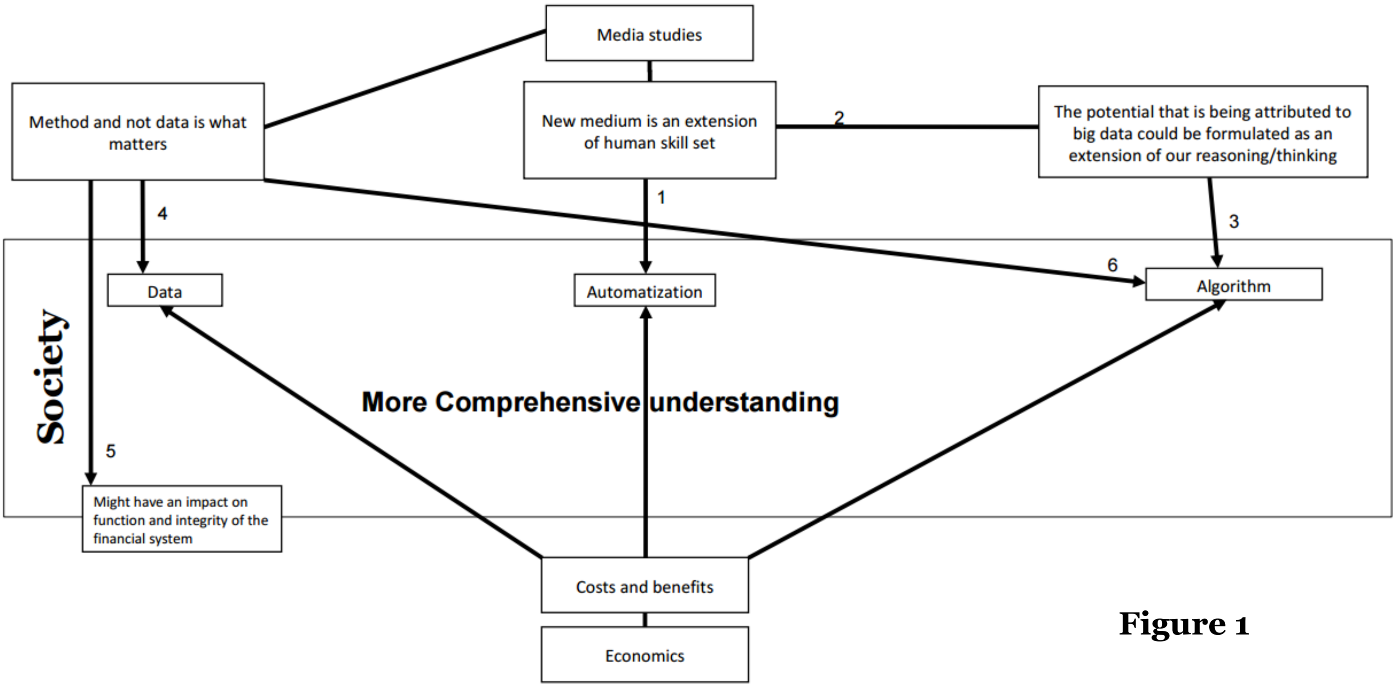


Figure 1

The three topics are linked to the findings of media studies. The first finding was that a medium is an extension of the human skill set, when taking the theory of Marshall McLuhan in mind (McLuhan, 1969). It was concluded that new media is developed to enhance our senses. This finding was scrutinized in a broader context and automatization has always been about machines replacing tasks previously done by humans. Via automatization humans try to enhance their performance (Parasuraman et al., 2000, p.286). (Line 1, Figure 1). Computers do not get "bored" by repetition, as humans might get, for instance. How could we see big data in this light? The potential that is being subscribed to big data (and its application) could be seen as trying to place replace part of a human analysis by a machine analysis. (Line 2, Figure 1). This explains the link to algorithms (Line 3, Figure 1).

The last finding is that the method and not the data analyzed/distributed matters with the impact of a new medium. This was linked to data, since it tells us that the focus should not be on the data as a source of influence (Line 4, Figure 1). Line 5 in Figure 1 captures the fact that in finance there has also been found that this new method/medium to trade has resulted in concerns about the impact it has on the industry. Line 6 in Figure 1 shows that since algorithms are the new method/medium, it is this where should be focused on.

5.2. Data

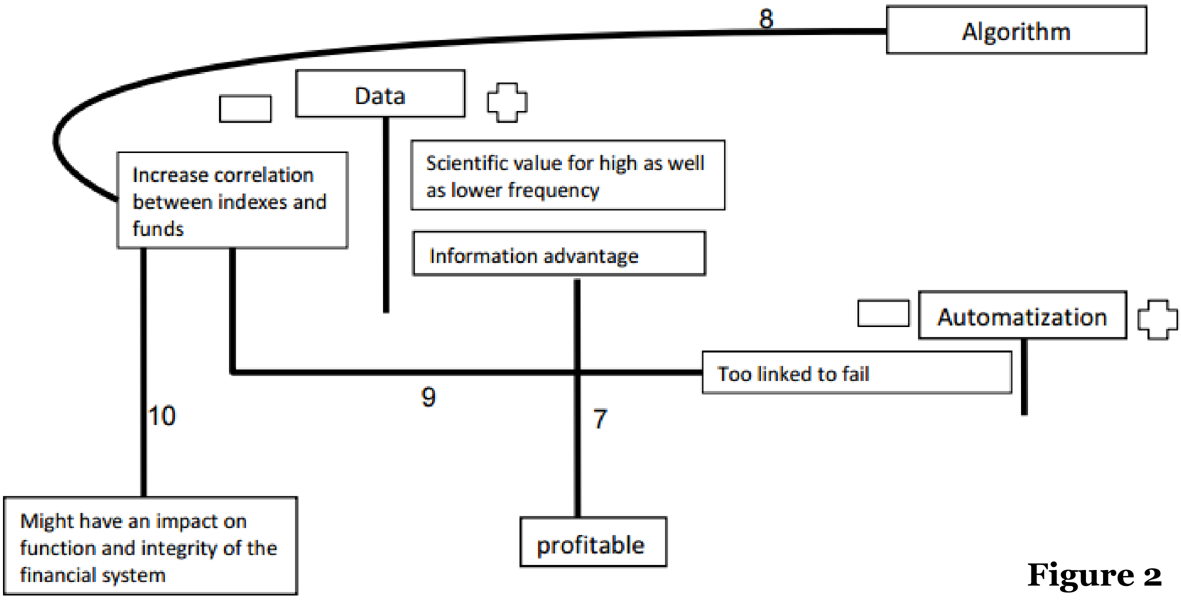


Figure 2

Algorithmic traders basically see the entire financial market as data. This data has to be interpreted, where after action based on this analysis takes place. The data can be used by econometrics and finance to gain new disciplinary insights for low and high frequency phenomena, gaining insights on the events that take place in time windows humans notice, but also the phenomena that take place at speed of (milli-)seconds. Another advantage the algorithmic traders have in this regard is an information advantage, with respect to the data and the algorithms interpreting the data (Line 7, Figure 2). The speed and volume at which assets are bought and sold across different markets by algorithms has created a correlation between indexes and funds (Line 8, Figure 2), because of this everything is now "too linked to fail" (Line 9, Figure 2). There are also concerns about the impact this has on the function and integrity of the system (Line 10, Figure 2).

5.3. Automatization

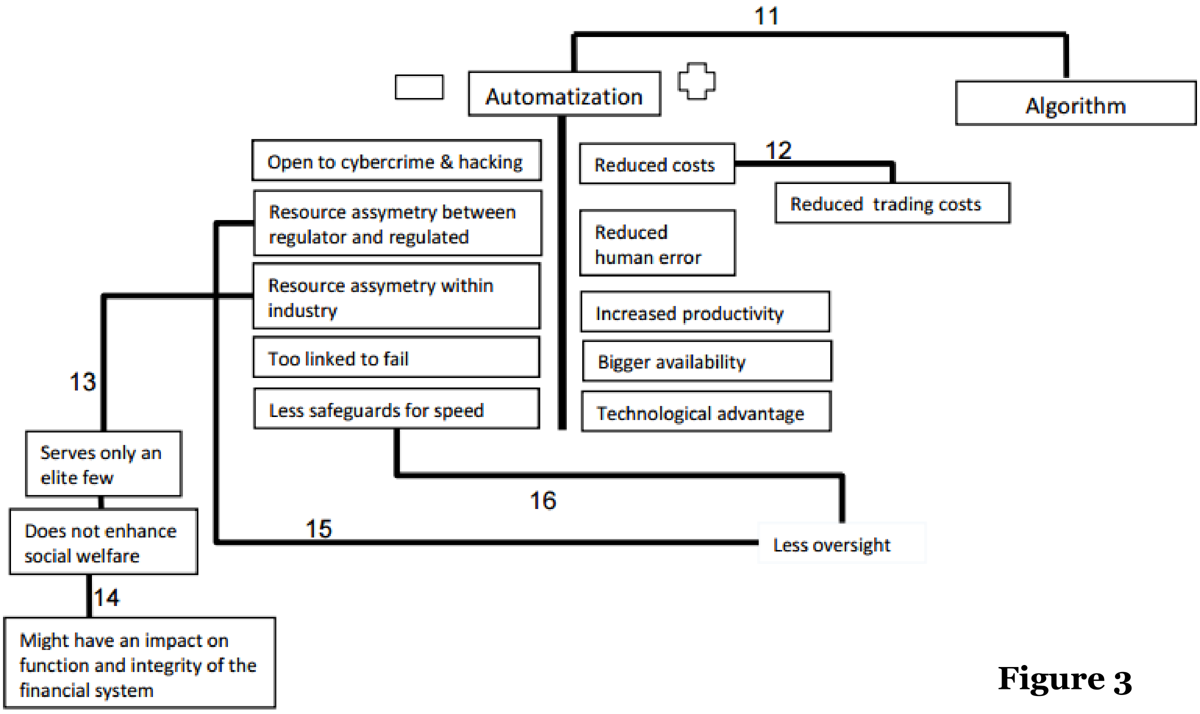


Figure 3

Most of the benefits of high frequency trading are already captured by algorithmic trading, but those benefits are mostly related to the automatization of trading (Line 11, Figure 3).
First of all, automatization reduces personnel costs, human error and increases productivity, by replacing humans with machines. Although part of the cost reduction is also due to trading cost reduction which will be further explained later on (Line 12, Figure 3). With the automatization also came broader availability of trading, making it more accessible to more people, people can now invest with the help of computer programs. However, the automatization of finance also had certain negative consequences. First, it opened the industry to cybercrime and hacking. Second, as mentioned earlier, the industry also became too linked to fail. Another consequence is the resource asymmetry. The consequence of resource asymmetry can be seen as slightly positive when only looking at the industry itself, since competition makes a market more efficient. However, automatization on this scale creates inequality, when the threshold to enter the market is so high that it only serves a select wealthy group, making them richer while the rest stays behind (Line 13, Figure 3), which also adds to the concern about the impact on the function and integrity of the system (Line 14, Figure 3). But the resource asymmetry between the traders and supervisors who control the system is imbalanced, due to this there is currently less oversight (Line 15, Figure 3). One could argue that the automatization of finance was a result of the race to zero, the struggle to make the time it takes to complete a trade/deal as fast as possible. With automatization, some safeguards were sacrificed to gain with respect to other competitors (Line 16, Figure 3). One could also say

however, that because traders currently have more advanced technology than the authorities, they are cutting some corners that might go unnoticed.

Profitability

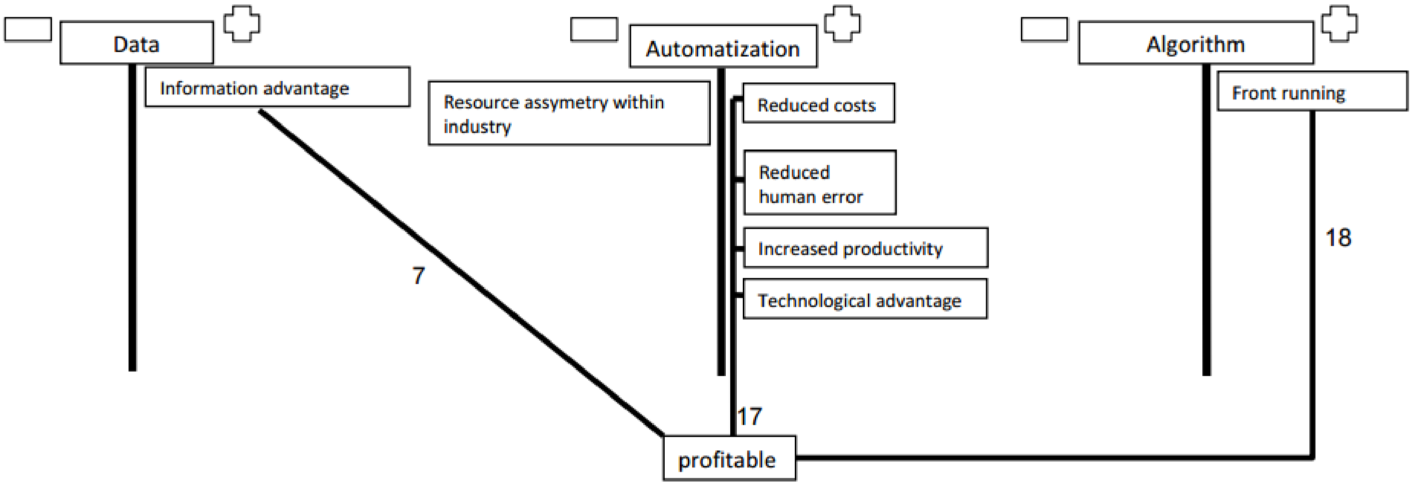


Figure 4

As already stated, most of the increased benefits can be attributed to automatization. In addition, it increases the profitability of algorithmic trading (Line 17, Figure 4). The information advantage (Line 7, Figure 4) that is created using big data and the technological advantage that traders have in comparison to other competitors are also reasons why algorithmic trading is so profitable. The last reason is front running (Line 18, Figure 4), trying to predicting the near future (milliseconds).

5.4. Algorithm

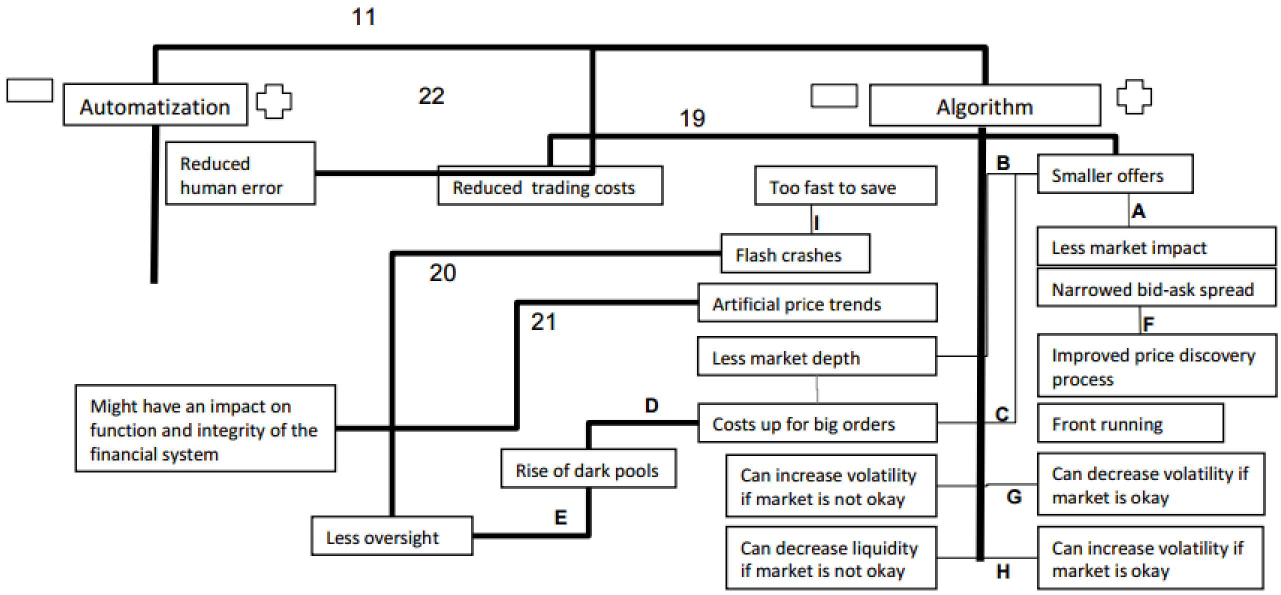


Figure 5

This paragraph will zoom in on algorithms which are at the core of big data. In figure 5 the thin lines represent relations between advantages and disadvantages of algorithmic trading that are two sides of the same coin or a direct result of each other. For instance, the fact that big orders got more expensive, making it logical that more small orders are executed. The other lines are comparable connections to the ones that have been made throughout this chapter. As explained in the chapter three, one of the advantages of algorithmic trading is the fact that algorithms make a larger number of smaller orders instead of executing one big order. Because of that there is less market impact (Thin line A, Figure 5). However, this has caused less market depth since there are less bulk orders waiting to be carried out (Thin line B, Figure 5). In addition, the costs for big orders increased, while the costs for smaller orders decreased (Thin line C, Figure 5). Therefore, trading costs were reduced (Line 19, Figure 5). This was part of the reason for dark pools to come into existence (Thin line D, Figure 5), resulting in less oversight (Thin line E, Figure 5). Two other interrelated benefits of algorithmic trading are: that it improved the price discovery process (1) by narrowing the bid-ask spread (2) (Thin line F, Figure 5). Also, the computation power behind the algorithms allows the processing of the information at such a high velocity that they can front run, or 'predict' a few (milli-)seconds into the future.

If the market is functioning normally and no uncommon events occur, algorithmic trading can decrease volatility (Thin line G, Figure 5) and increase liquidity (Thin line H, Figure 5). However, if the market does not function well and it is not a good moment for traders to operate, algorithmic trading can cause the opposite effect. Possibly the biggest costs of the implementation of big data methods in finance are the flash crash and mini flash crashes that have occurred thus far (Thin line I, Figure 5). This elucidates the point "too fast to save" clearly. The speed at which trading now takes places makes interference impossible when things go wrong. Algorithmic trading occurs with higher speed than humans can possibly keep track of. This makes adequate oversight on the action taken, even if all the safeguards are in place, highly difficult (Line 20, Figure 5). Another concern about algorithms is that algorithmic traders can, either accidentally or purposely, create artificial price trends. This point confirms the concerns about the function and integrity of the financial system (Line 21, Figure 5). Lastly, the connection between reduced human error and algorithm is captured by line 22 (Figure 5). It is often believed that when a computer replaces a human, human errors will be eliminated. However, this is a misconception, there are people creating the algorithms on which all actions are based. In addition, humans analyze the outcomes of big data methods and will use their own interpretation.

5.5. More comprehensive understanding

Insights from new media tell us that it is the method and not the data that has an impact on society, extending our skill set. So even though the name of this new technology (or: new medium) "big data" is not so much about the data itself, but more about the way the data is handled. With the new medium big data, we try to extend our reasoning and thinking. Finance might not present the best example one way, since supercomputers cannot think for themselves yet and the traders can do the same calculations. But in another way, big data in finance is an extension/enhancement of our computing/calculating and could therefore be seen as a good example, since the problem was the time it took for humans.

In finance big data is used to finally do something the industry knew was possible, but never was capable of doing because the medium was not at hand. Industries that are thinking of implementing big data now, think the other way around. They possess tons of data and see big data as an opportunity to make sense of these data and perhaps make a profit of it. The data might present a high number of correlations and show seemingly causal connections when big data methods are applied. However, just a correlation between variables does not mean there is a causal relation in the real world. However, when the correlations are viewed as causal, when this is not the case, the consequences of these conclusions could have significant negative impact. This could lead to false conclusions based on misinterpretation. Furthermore, often the data in datasets were collected for other purposes. These datasets might not have the quality that datasets specifically started for the purpose of analysis have, and might even contain omissions or errors. The increased connectivity between the various data makes the consequences severe since a fault in data will be carried over to other datasets. The success finance achieved should therefore be seen in the right way. Finance found in big data a method that was missing, other sectors that could be inspired by the success of big data will possibly want to use this method as well, while their data is not always suitable for the cause.

In Figure 6 we present an overview of the more comprehensive understanding that resulted from this interdisciplinary research.

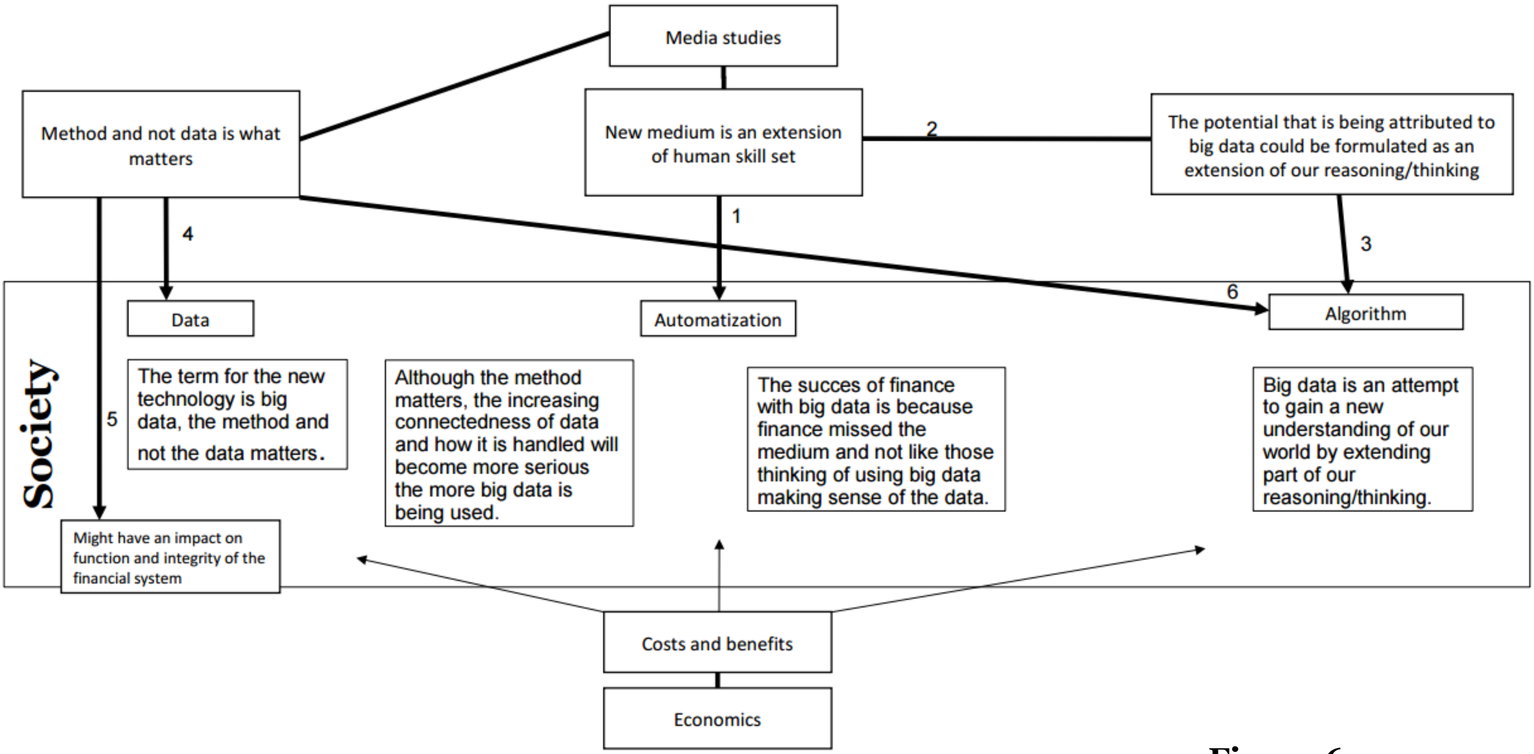


Figure 6

5.6. Conclusion

To answer our main research question:

What are the risks and benefits of the implementation of big data in finance and how does this new technology influence society?

We have integrated the findings from both media studies and finance to come to a better understanding of the implications of the implementation of big data in society. We presented an analysis of the costs and benefits of algorithmic trading in chapter 3 and the findings of the media studies disciplines in chapter 4. The integration process in chapter 5 was a result of the common ground that was created, through connecting both disciplines. Three topics were found that were identified in both of the disciplinary studies, namely data, automatization and algorithm. A more comprehensive understanding of the implications of big data was the result of this process, which led us to certain key insights.

First of all, we found the method matters and not the data, so when looking at the influence of big data on society it is not primarily the data that should be the main focus, but the handling of this data. Secondly, we found that when big data is seen as a new medium through the eyes of theorist McLuhan, it is an attempt to extend part of our skillset, specifically extending our thinking or reasoning. Thirdly, we found that the industries thinking of implementing big data that are inspired by the success made in finance should keep in mind that for finance it was a problem of insufficient computational power. Finance missed a solution for a known problem, while other industries now might force a prospected solution upon unknown problems. Lastly, we found that the method should be the focus and not the data when looking at big data.

To answer the main research question, we found that most of the benefits attributed to algorithmic trading are actually due to the automatization of the trading process while most of the risks or "costs" are related to one form of algorithmic trading: high frequency trading. High frequency trading acts faster than humans can keep track off, which makes for the both the positive and negative aspects of this new method of trading. In this light the attempt to extend part of our reasoning/thinking with big data can have some serious consequences for society. When data is more important in society, this could lead to a more datafied life. Moreover, the trust that is put in data is high, but not always justifiable as is the notion of an algorithm that is uncorrupted by any human error. Some other consequences could be that with the increased speed and connectivity humans will not know something went wrong until after it happened, while possible errors that are embedded in a connected chain of datasets could act like a domino effect when they transgress into other datasets.

A limit in this research was the time frame, which was determined at ten weeks. Becoming an expert on any topic is very time consuming. We would, therefore, recommend further research on this topic, this will be explained below. Another possible limit is the fact that there is no consensus on the definition of big data, so we had to choose one definition to work with. The results of this research could possibly be different when we would have used another definition (for example one of the definitions that are stated in paragraph 2.1. Therefore, future research is needed to get a better grasp of the interconnectedness of data and how this might affect the algorithms using the data feeding other algorithms. Another interesting research could be how this interconnectedness might influence artificial intelligence as a form of interaction with other artificial intelligence. Research on how the extension of our reasoning by using this new medium affects our reasoning and or thinking, could also give us a better understanding of the implications of this new medium.

6. Literature

Articles:

Ariely, D. (2013, January 06). Dan Ariely's Facebook page [Facebook post]. Retrieved from: https://www.facebook.com/dan.ariely/posts/904383595868. Last visited on June 10 2015.

Bowley, G. (2010). Lone $4.1 billion sale led to 'flash crash' in May. *The New York Times*, *1*.

Brogaard, J. (2010). High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*.

Capgemini & EMC (2015) *Big & Fast Data: The Rise of Insight-Driven Business*. March, 2015. Retrieved from: https://www.capgemini.com/resources/big-fast-data-the-rise-of-insight-driven-business

CFTC & SEC. (2010). Findings Regarding the Market Events of May 6, 2010, Sept. 2010, Retrieved from: http://www.sec.gov/news/studies/2010/marketevents-report.pdf. Last visited on 4 June 2015.

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, *69*(5), 2045-2084.

Chordia, T., Goyal, A., Lehmann, B. N., & Saar, G. (2013). High-frequency trading. *Journal of Financial Markets*, *16*(4), 637–645. http://doi.org/10.1016/j.finmar.2013.06.004

Clark, C. (2010). Controlling risk in a lightning-speed trading environment.Chicago Fed Letter, 272. Retrieved from: https://chicagofed.org/~/media/publications/chicago-fed-letter/2010/cflmarch2010-272-pdf.pdf. Last visited on 4 June 2015.

Clements, M. T. (2012). Curbing the dangers of high-frequency trading. *Economists'Voice, issue: 1*(9).

CSA. (2013). *Expanded Top Ten Big Data Security and Privacy Challenges*. Retrieved from: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf. Last visited on 10 June 2015.

Debray, R. (1996). Media manifestos. *Trans. Eric Rauth. London and New York: Verso*.

Fabozzi, F., Focardi, S. M., & Jonas, C. (2011). High-frequency trading: methodologies and market impact. *Review of Futures Markets*, *9*, 7-38.

Financial Crisis Inquiry Commission. (2011). *The Financial Crisis Inquiry Report: The Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States Including Dissenting Views*. Cosimo.

Gerding, E. F. (2009). Code, Crash, and Open Source: The Outsourcing of Financial Regulation to Risk Models and the Global Financial Crisis. *Wash. L. Rev.*, *84*.

Gsell, M., & Gomber, P. (2009). Algorithmic trading engines versus human traders-Do they behave different in securities markets? In *ECIS,* 98-109.

Golub, A., Keane, J., & Poon, S. H. (2012). High frequency trading and mini flash crashes. *arXiv preprint arXiv:1211.6667.*

Haldane, A. (2011). *The race to zero*. Speech given at: International Economic Association Sixteenth World Congress, Beijing, China, 8 July 2011.

Hashem, I. et al. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98-115. Advance online publication. Retrieved from: http://www.sciencedirect.com/science/article/pii/S0306437914001288. Last visited on 10 of June, 2015.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, *66*(1), 1-33.

Joscelyne, A. (1995) "Debray on Technology". *Wired Magazine*. Retrieved from: http://www.generation-online.org/p/fpdebray1.html. Last visited on 10th of June 2015.

Kenett, D. Y., Ben-Jacob, E., & Stanley, H. E. (2013). How high frequency trading affects a market index. *Scientific reports*, *3*.

Kirilenko, A. A., & Lo, A. W. (2013). Moore's Law Versus Murphy's Law: Algorithmic trading and its discontents. *The Journal of Economic Perspectives*, 51-72.

Klein, J. (2001). Closework: Why We Couldn't See What Was Right in Front of Us. *The New Yorker*, Oct 1st: 44-49.

Laney. D. (2001) *3d data management: Controlling data volume, velocity and variety*. Gartner. 6 February 2001. Retrieved from: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lazer, D., R. Kennedy, G. King, and A. Vespignani. (2014). "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6176) (March 14): 1203–1205.

Lin, T. C. (2012). New Investor, The. *UCLA L. Rev.*, *60*, 678.

Lin, T. C. (2014). The New Financial Industry. *Alabama Law Review*, *65*(3), 567.

Lohr, S. (2012). The Age of Big Data, *The New York Times*, February 11, 2012. Retrieved from: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, United States of America: Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data. *The management revolution. Harvard Bus Rev*, *90*(10), 61-67.

McKnight, D. H. (2005). "Trust in Information Technology." In G. B. Davis (Ed.), *The Blackwell Encyclopedia of Management. Vol. 7 Management Information Systems*, Malden, MA: Blackwell, pp. 329-331.

Morgan, J. (2003). How Reality Ate Itself: Orthodoxy, Economy and Trust. *Post-Autistic-Economics-Review*, *18*.Retrieved from: http://www.paecon.net/PAEReview/issue18/Morgan18.htm, last visited 3-6-2015.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama, 309*(13), 1351-1352.

Norden, E. (1969). The Playboy Interview: Marshall McLuhan. *Playboy Magazine*, 53-74.

Nuti, G., Mirghaemi, M., Treleaven, P., & Yingsaeree, C. (2011). Algorithmic trading. *Computer*, *44*(11), 61-69.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *30*(3), 286-297.

Shaw, J. (2014). Why "Big Data" is a big deal. *Harvard Magazine*, *3*, 30-35.

Swan, M. (2012). Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Journal of Sensor and Actuator Networks*, *1*(3), 217-253.

Snijders, C., & Matzat, U.,& Reips, U.-D. (2012). "'Big Data': Big gaps of knowledge in the field of Internet". *International Journal of Internet Science, 7*, 1–5.

Tabb, L. (2012). Written testimony to the United States Senate Committee on Banking. *Housing and Urban affairs*. Washington DC, Sept 2012. Retrieved from: http://www.banking.senate.gov/public/index.cfm?FuseAction=Files.View&FileStore_id=268afb1b-4c5e-4ec0-8028-fc840b973843, last visited June 4th, 2015.

Books:

Aldridge, I. (2009). *High-frequency trading: a practical guide to algorithmic strategies and trading systems*. John Wiley and Sons.

Grier, D. A. (2013). *When computers were human*. Princeton University Press.

Kerikmäe, T. (Ed.). (2014). *Regulating E-Technologies in the European Union: Normative Realities and Trends*. Springer.

Kolb, R. W. (Ed.). (2011). *Financial contagion: The viral threat to the wealth of nations* (Vol. 604). John Wiley & Sons.

Lister, M. et al. (2009). *New media: A critical introduction*. Taylor & Francis.

McLuhan, M., & Fiore, Q. (1967). T*he medium is the message*. New York.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.

McLuhan, M. (1969) *Marshall Counterblast*, London: Rapp and Whiting.

McLuhan, M. (2003). *Understanding Me*. eds. S. McLuhan and D. Staines. McLelland & Stewart, Toronto.

Monks, R. A., Reed Lajoux, A., & LeBaron, D. (2011). *Front Matter*. John Wiley & Sons.

Partnoy, F. (2012). *Wait: The art and science of delay*. PublicAffairs.

Pasquale, F. (2015). *The Blackbox Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, Massachusetts.

Patterson, S. (2013). *Dark pools: The rise of the machine traders and the rigging of the US stock market*. Crown Business.

Patterson, S. (2011) *The Quants: How a new Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed It*. Crown Books.

Repko, A. F. (2012). *Interdisciplinary research: Process and theory*. Sage.

Stearn, G. E. (1968) *McLuhan Hot & Cool*, Bear Books.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable fragility* (Vol. 2). Random House.

Video Files

Kroes, N. (2014, October 15). Data: the new currency? Neelie Kroes keynote speech [Video file]. Retrieved from: http://www.europeanvoice.com/event/data-the-new-currenc