UCL
Université catholique de Louvain

Louvain School of Management (LSM)

LOUVAIN
School of Management

The impact of high-frequency trading on the volatility of the financial markets

Mémoire recherche réalisé par
Jérémie Peti

en vue de l'obtention du titre de
Master en ingénieur de gestion, à finalité spécialisée

Promoteur
Leonardo Iania

Année académique 2017-2018

Chaussée de Binche, 151 - 7000 Mons, Belgique www.uclouvain.be/lsm

I thank my thesis’ director, Leonardo Iania, for the help and wise suggestions he gave me this past year.

I am also thankful to my sister who has read my work and given me her feedback.

Finally, I thank my mother, for without her I would not be graduatin

Abstract

A new phenomenon has appeared on the financial markets these past years: high-frequency trading. This technique, which mainly consists of selling and buying stocks at high speed, has since begun to be discussed by the financial community. Indeed, the question of the threat that high-frequency can have for the stability and more precisely the volatility of the financial markets is still discussed. On the 6th of May 2010, the largest U.S. stock markets (Nasdaq, S&P 500, Dow Jones Industrial Average) suddenly lost up to 9% of their value within two minutes before rebounding. The fact that this event was caused by high-frequency trading is still discussed.

This paper aims at answering this question by analyzing data from several stock markets. In order to do that, we take the price variation of the 10% most and least traded stocks on the Nasdaq, S&P 500 and Dow Jones Industrial Average on the day of the Flash Crash and we look at dissimilarity between the samples. Indeed, high-frequency traders operate on liquid stocks so if there is a significant difference between the most and traded stocks, this would mean that high-frequency trading caused the sudden volatility of the most traded stocks on the 6th of May 2010. Finally, we find that high-frequency does not significantly impact the volatility of the financial markets.

Key words: high-frequency trading, volatility, stock markets.

Table of figures

Figure 1. Algorithmic Trading : Percentage of Market Volume in the U.S. 7
Figure 2. Evolution of the Dow Jones on the 6th of May.. 12
Figure 3: £/$ exchange rate on the 7th of October 2016.. 13
Figure 4: High-frequency trading as a share of US equities daily volume 14
Figure 5: Share of High-Frequency Trading in total equity trading 14
Figure 6: Share of high-frequency trading by geography... 15
Figure 7: Coefficients of variation 23
Figure 8: 2 clusters of the DJIA dataset 23
Figure 9: 3 clusters on the DJIA dataset 23
Figure 10: Hierarchical clustering of the DJIA dataset... 24
Figure 11: Hierarchical clustering of the DJIA dataset (with the names of the stocks)........... 25
Figure 12: Hierarchical clustering of the Nasdaq dataset 26
Figure 13: Hierarchical clustering of the Nasdaq dataset (with the names of the stocks)...... 27
Figure 14: 4 clusters of the Nasdaq dataset... 27
Figure 15: 3 clusters of the Nasdaq dataset... 28
Figure 16: 2 clusters of the Nasdaq dataset... 28
Figure 17: Hierarchical clustering of the S&P 500 30
30
Figure 18: Hierarchical clustering of the S&P 500 dataset (with the names of the stocks) 31
Figure 19: 2 clusters of the S&P dataset 31
Figure 20: 3 clusters of the S&P 500 dataset 33
Figure 21: 4 clusters of the S&P 500 dataset 34

Table of Appendices

Annex 1: Volume of Nasdaq stocks from 1/1/2010 to 4/30/2010 43
Annex 2: Volume of DJIA stocks from 1/1/2010 to 4/30/2010 46
Annex 3: Volume of S&P 500 stocks from 1/1/2010 to 4/30/2010 47
Annex 4: code to use with the ‘pdc’ package on R 59
Annex 5: Hierarchical clustering of the S&P 500 dataset 61

Table of contents

1. Introduction **1**

a) Research Question **1**

b) Context **1**

c) Motivations of the research **2**

2. Review of the literature **3**

a) Theoretical framework **3**

1. Trading on the stock exchanges 3
2. Market Liquidity 4
3. Market Volatility 5
4. Algorithmic trading 6

5. High-frequency trading 8

A. Definition 8

B. History of high-frequency trading 10

C. Strategies in high-frequency trading 16

b) Hypothesis to test **18**

3. Methodology **19**

4. Results **23**

DJIA dataset **23**

Nasdaq **26**

S&P 500 .... **30**

5. Discussion **35**

Coefficient of variation **35**

DJIA dataset **36**

Nasdaq dataset **36**

S&P 500 dataset **37**

6. Conclusion **37**

a) Summary of the study **37**

b) Theoretical implications of the study **38**

c) Managerial implications of the study and the recommendations **38**

d) Limitations and suggestions for the future **39**

7. Bibliographie **40**

8. Appendix **43**

1. Introduction

a) Research Question

Does high-frequency trading have an impact on the volatility of the financial markets?

b) Context

Financial markets have a huge impact on our lives whether we like it or not. The financial crisis of 2008 showed us that we cannot simply ignore what is happening on these markets. On the 6th of May 2010, the major U.S. stock indexes briefly collapsed during half-an-hour before quickly rebounding for apparently no particular reason since no major economic or politic news was announced on that day. This event has later been described as ''the Flash Crash of 2010''. Similar strange behaviors of the stock markets have occurred ever since. Indeed, on the 7th of October 2016 the British pound declined by 6% in 2 minutes against the U.S. dollar and on the 22nd of June 2017 the price of Ethereum, a cryptocurrency, fell from $319 to 10 cents in a few seconds and later picked up to its pre-crash price. Again, all these events occurred even though there was no major economic or political news. These episodes shed light on a new phenomenon that had been growing for the past decade: high-frequency trading (or HFT). A subsequent concern appeared: does high-frequency trading have an impact on the volatility of the stock markets?

In its book, ''Flash Boys: A Wall Street Revolt'' (published in 2014), Michael Lewis brilliantly explains how high-frequency trading now represents a large proportion of the trading orders on the stock markets and how high-frequency traders can use their technological edge over the competition to earn millions of dollars by executing thousands of orders in only a few minutes.

This new phenomenon led the regulators in the U.S. and in Europe to intervene since predatory behaviors have been linked to high-frequency traders. Indeed, the fact that some transactions can take less than the blink of an eye to get done means that it is difficult for institutional investors to see if a high-frequency trader is preying on them (using techniques that will be discussed afterwards). Therefore, in the recent years, more and more fines have been given to traders that had such preying behaviors. Moreover, in 2014, the directive of the European Parliament on markets in financial instruments (MIFID II) aimed to define HFT and to regulate HFT companies. Several rules such as notifying the authorities when doing high-frequency trading or keeping the trading algorithms for five years aimed to regulate this market. The MIFID II came into effect in January 2018.

However, the question of the impact of high-frequency trading on the financial markets and more specifically on their volatility is still discussed…

c) Motivations of the research

The main managerial motivation of the research was to put ourselves in the shoes of the regulators. Since high-frequency trading now represents an important part of the transactions that are done on the financial markets, the regulators need to know if it has an impact on the volatility and thus, the stability of those markets. Indeed, even though the stock markets may seem to be only about numbers, their impact on the economy is very real, as everybody brutally learned in 2008. This paper aims to help regulators understand high-frequency trading and therefore be able to correctly regulate it.

The scientific motivation of the paper was to better understand high-frequency trading and how it could affect financial markets. We did it by analyzing data but the difficulty was to answer a complex question with a simple answer. Indeed, there were many ways to tackle this paper such as interviewing people in the banking sector but we decided to analyze data to reach a conclusion based on facts.

The social motivation of the study was to know whether or not high-frequency was dangerous for the stability of the financial markets. It is our responsibility not only as people working in the financial industry but also as citizen to share what we know about a new phenomenon that arose. Then we will be able to understand what we should do about it.

In this research paper we first define and explain the different concepts around high-frequency trading, then we analyze data with datasets that we created using Bloomberg terminals and finally we try to explain those results.

2. Review of the literature

a) Theoretical framework

Since this paper focuses on high-frequency trading, it is important to explain what trading is and what are the different types of trading activities on the financial markets. The activity of trading is defined as ''the buying and selling of shares and money'' (Cambridge Dictionary, 2018). 5 aspects of trading in finance will be covered in this literature review: trading on the stock exchanges, market liquidity, market volatility, algorithmic trading and high-frequency trading.

1. Trading on the stock exchanges

A stock exchange is a place where people can buy and sell (hence the use of the term ''trade") shares of stocks, bonds or other securities. The primary goal of stock exchanges is to allow companies to raise money, which then helps them investing in new businesses and products to further increase their revenues. An investor decides to invest money on a stock because he believes that the firm will generate revenues in the future. Whether one invests for a short-term or long-term gain doesn't change the idea behind the investment: believing that the share price will increase, thus creating a profit for the investor.

The second role of stock markets is to efficiently allocate capital in the economy. Indeed, if share prices are correctly valued, this means that companies that are really performing well will be rewarded by seeing their share price (thus their value) increase. On the contrary, companies that have poor financial results will be punished by seeing their share price decrease and might eventually cease their activities or be acquired. An efficient market is a place where "security prices at any time ''fully reflect'' all available information." (Fama, 1970, pp. 383-417)

Thereby, if we consider that information is accurate, investors can decide to buy shares of firms that are ''riskier'' (but may lead to higher profits) or buy shares of firms that are less risky, depending on their risk profile. This is why the stock markets efficiently allocate capital: they match buyers and sellers that have the same risk profile. Indeed, '' financial markets have 2 important functions for asset pricing: liquidity and price discovery for incorporating information in prices''. (Brogaard, 2014, pp. 2267-2306)

Another role of the stock exchanges is to provide transparency for investors. By going public (being listed on a stock market) a company must disclose several financial pieces of information that it did not have to disclose when it was private. Indeed, if investors want to correctly value a company they need some information about the finances and the overall business of the firm (market share, how the company operates, …). Going public requires to disclose more information than private equity (Bhattacharya, 1995) so stock markets are the best way to provide the most information about a firm to investors.

2. Market Liquidity

The liquidity of a stock market is the ‘’ease with which financial assets can be bought or sold without having an undue impact on prices’’ (European Systemic Risk Board, 2016). Therefore, a liquid market has shares that can be exchanged quickly with a small impact on the shares’ price whereas an illiquid market has shares that will be exchanged with a high impact on the shares’ price.

A typical measure of the liquidity of a stock or a market is the bid/ask spread. It is ‘’the difference between the price that someone will pay for shares, etc. and the price that that they will sell them at’’ (Cambridge Dictionary, 2018). The spread is earned by the person who is selling the asset. This person is the liquidity provider whereas the counterparty is the liquidity demander. Having sufficiently liquid financial markets is important for three reasons.

First of all, liquid financial markets help hinder the volatility of those markets which can help having more stable financial markets (Elliott, 2015). Indeed, if a market is illiquid, it means that there are few transactions occurring everyday so if someone executes a large order (either selling or buying), it is more likely that it will have a big impact on the prices, therefore, leading to a high volatility. On the other hand, a liquid market can better absorb huge transactions since they are more people willing to buy or sell stocks.

Liquid financial markets are also important for risk management. Indeed, ‘’liquid capital markets also facilitate the distribution of financial risks to market participants most able and willing to bear them, and enable investors to manage and hedge against risks, as well as adjust their portfolios effectively’’ (PwC, 2015, p. 20). If one wants to change its risk profile, it would be much easier to do so in a liquid market to quickly adapt his investing strategy.

Lastly, when markets are liquid, the transaction costs are low so investors are more willing to invest on these markets. Therefore, the role of capital markets being to let companies have access to capital, if capital markets are much more liquid, it will become easier for firms to have access to capital, thus decreasing their cost of capital and eventually their ability to invest and create jobs (Jones, 2013).

3. Market Volatility

The NASDAQ (the second largest stock exchange in the world by market capitalization) defines the volatility as ‘’a measure of risk based on the standard deviation of the asset return’’ (Nasdaq, 2018). Therefore, the market volatility refers to variation of the stock prices on this market during a certain period of time. A market with high volatility contains stocks whose prices vary with a great magnitude whereas a market with low volatility contains stocks whose prices vary a little bit. Market volatility is partly influenced by market liquidity. An illiquid market tends to be more volatile. There are less movements on this market so transactions have a higher impact on the price variations. However, a liquid market can still be volatile since market liquidity only partly impacts market volatility. Indeed, other macroeconomic factors affect the market volatility such as government policies, exchange rates, wars, etc. Several risks are associated with a volatile market due to the high variations of its stock prices.

First of all, high volatility can lead to irrationality from the investors. Indeed, when a market is in a high bullish phase (the stock prices increase quickly on a short timeframe), people tend to be irrational and invest more and more in this market, even if the stocks can seem to be completely overvalued. Unfortunately, as the old saying goes ‘’trees don’t grow to the skies’’ and eventually, prices will decrease, which can lead to huge losses for some (or the majority) of the investors. The ‘’tulip mania’’ that occurred during the 1630s is a typical example of such a case: the prices for tulip bulbs reached extremely high prices. Accounts of bulbs worth a house have even been reported (Thompson, 2006) but in February 1637 when they suddenly dropped, incurring losses for investors. More recently, the bitcoin market has been subject to a steep increase during 2017, the price for a single bitcoin going from 900$ in January to

20,000$ in December. One may even personally know someone who invested in this market during this bullish phase. But again, unfortunately, the price for a bitcoin dropped dramatically in December, losing 45% of its value in just 2 weeks (Martin, 2017).

A second type of risk on a volatile market is the ‘’panic reaction”. This is the fact that a small sudden drop of the prices leads to investors selling all their stocks by fear of seeing the prices keep on decreasing. Logically, as more investors want to sell, the snake bites its own tail: the prices don’t stop declining, investors keep on selling and prices continue to shrink. This risk is linked to the “irrationality risk” since people will tend to sell their stocks quicker if there is a drop of the prices during a bullish phase because they will want to cash in their (potentially) high capital gain.

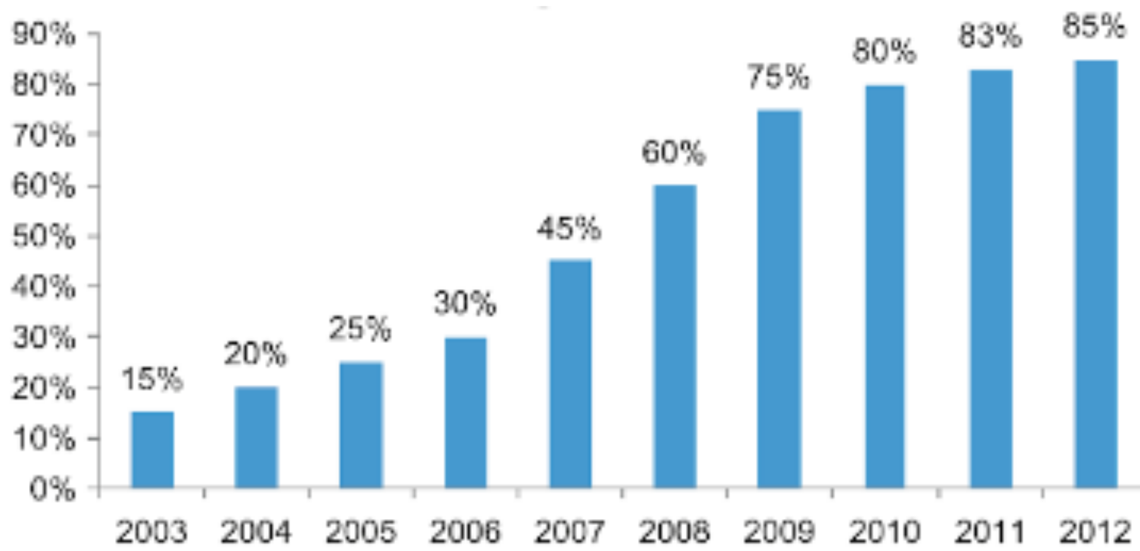
Lastly, a volatile market makes it difficult to know the real value of its stocks. Indeed, when the stock prices of a company soar by 45% in a few months does it really mean that the company’s worth raised by 45%? Did the company actually managed to drastically increase its revenues or is it simply that people believe the company will manage to generate high revenues in the coming years? Or is it because the whole industry of this company is in a bullish phase so everybody wants to buy its stocks? Or is it a combination of all those hypothesis? The fact that all these questions can be raised indicates the difficulties facing investors in a volatile market.

4. Algorithmic trading

Algorithmic trading is ‘’the use of computer algorithms to automatically make trading decisions, submit orders, and manage those orders after submission’’ (Hendershott, 2011, pp. 1-33). The trading decisions can take several factors into account such as the time, the price of the stock and the volume of stock to buy or sell. Moreover, one can decide to create algorithms that predict a certain price for a stock and ask the computer to buy or sell according to a statistical models. The main idea behind algorithmic trading is that if one wants to execute a big order, the markets will directly be affected thus affecting the transaction cost. Therefore, ‘’algorithms were created to break up the block orders into several pieces, which were then executed separately’’ (Hruška, 2016, p. 1912).

The automation of the trading orders began in the 1970s thanks to the apparition of home computers and the growth of their power of computations. In 2012, the share of algorithmic trading as a percentage of market volume had increased to reach approximately 85% as shown on figure 1.

Figure 1. Algorithmic Trading : Percentage of Market Volume in the U.S.



Source : (Glantz, 2014)

Algorithmic trading has mainly three advantages.

First of all, it offers the possibility to backtest. Indeed, when implementing his algorithms, a trader can test them on historical data and take a look at the outputs. He can then improve his model and have a better idea of the expected money he is supposed to earn for every unit of risk he takes.

It is independent of a trader’s feelings since it is automated. Therefore this can prevent the trader to act impulsively in a volatile market by trying to get more and more profit or by deciding to cut his losses. It can force him to ‘’stick to the plan’’ by believing in his algorithms. Moreover, it can decrease the risk of a manual error when executing a trade.

Lastly, thanks to the speed of orders of computers, algorithmic trading can prevent huge losses or generate great profits. Since markets can move pretty quickly, automation can avoid losing time when manually entering an order if a stock goes up or down a certain level. Moreover, according to (Hendershott, 2011) algorithmic trading increases liquidity supply on the markets, and decreases the costs of trading.

However algorithmic trading also has some drawbacks.

Although it is possible to backtest an algorithm, a model can't be 100% accurate when encountering new data. Even if someone expect a certain profit with a 99% confidence interval, this person can nevertheless suffer some losses if the stocks go beyond the confidence interval. This means that there is still a need to constantly monitor the performances of the model. Thus one can't just be blindly confident in his model.

Should a very unlikely event happen (sudden spike or decrease of the price of a stock) even for a fraction of second, the computer might automatically buy or sell the stock while there were no real need to execute that order. The algorithms don't have a common sense and don't understand the concept of ''real value'' of a company.

5. High-frequency trading

A. Definition

High-frequency trading is defined by the Markets in Financial Instruments Directive II (MiFID II) as a trading technique with 3 characteristics:

1. There is no or limited human intervention in the sense that a computer "takes decision at any of the stages of imitating, generating, routing or executing orders or quotes according to pre-determined parameters." (The European Parliament and the Council of the European Union, 2014, p. 385). Indeed, high-frequency traders use algorithms to automate their transactions.
2. The speed of the transactions is very high, more specifically there are "at least 2 messages per second with respect to any single financial instrument traded on a trading venue" or there are "at least 4 messages per second with respect to all financial instruments traded on a trading venue" (The European Parliament and the Council of the European Union, 2014).

3. There is an "infrastructure intended to minimize network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location, proximity hosting or high-speed direct electronic access" (The European Parliament and the Council of the European Union, 2014). The latency is the "time that elapses from the moment a signal is sent to its receipt" (Picardo, 2018).

Since orders are executed in milliseconds or even microseconds, to gain a competitive advantage, high-frequency traders need to execute their orders as quickly as possible. This is why they need to have the lowest latency as possible by using the 3 techniques mentioned in the Mifid II definition:

- Co-location aims at "locating computers (…) in the same premises where an exchange's computers servers are housed" (Picardo, 2018)**.** Therefore, one can access stock prices milliseconds before the rest of the world.
- Proximity hosting is the idea to place a data center (where the algorithms are run) optimally placed between several trading platforms to let traders being as close as possible to multiple trading platforms. Indeed, using co-location for multiple platforms can be expensive so proximity hosting can be a way to optimize the costs of placing the hardware in a strategic place.
- Direct electronic access is the fact that someone is allowed to use any connecting system to transmit orders on a trading venue. This lets the trader avoid having to go through an intermediary to execute his orders.

HFT is a subset of algorithmic trading in the sense that it also uses the automation of orders and monitor the financial markets in real time. However, HFT has additional features compared to algorithmic trading: low latency time, an enormous number of transactions, zero positions at the end of the day for the trader, the use of co-location or proximity hosting and trading of highly liquid assets. Indeed, since high-frequency traders manage a lot of transactions in a short timeframe, they wouldn't be able to act on illiquid markets. Indeed, ''Since the use of high frequency algorithmic trading technique is predominantly common in liquid instruments, only instruments for which there is a liquid market should be included in the calculation of high intraday message rate'' (Commission delegated regulation (EU) of 25.4.2016, 2016, p. 11)

B. History of high-frequency trading

In stock trading, speed of execution has always played an important role in the trading strategies. Indeed, as soon as the first stock market was created in the 16th Century in Amsterdam, people realized that the speed at which one could collect information and execute his orders accordingly could lead to huge differences in the profits made. For instance, it is believed that the Rothschilds used pigeon carriers to be informed of the outcome of the battle of Waterloo in 1815 and exploited this information on the London stock exchange.

After the emergence of the computer in 1970s, the first Bloomberg terminals appeared in 1981. Those computers gave “real-time data on every market, breaking news, in-depth research, powerful analytics and were a communication tools” (Bloomberg Professional Services, 2018). They allowed professional investors to quickly act on the markets and this was a first important step in the computerization of the trading decision-making process. The Bloomberg terminals are still used today by more than 325,000 subscribers worldwide. Bloomberg’s major competitor is Thomson Reuters.

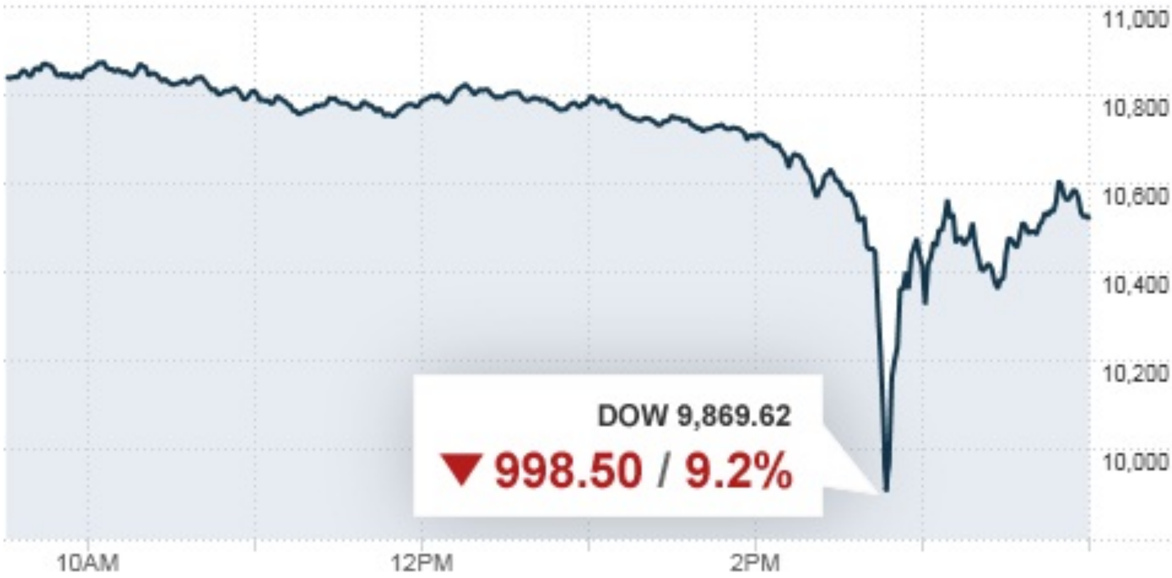
At the end of the 1990s, the U.S Securities and Exchange Commission (SEC) approved the use of electronic communication networks (ECNs**)**. It is a computer system that helps trading financial instruments outside of the classical stock exchanges. This led to the creation of alternative trading systems. The SEC passed this law because it wanted to stop the duopoly of the NASDAQ and the NYSE at that time. This offered the possibility for investors to trade outside of the exchanges and outside the usual trading hours. In 2000, the beginning of the decimalization of the exchanges accelerated the rise of algorithmic trading (Exchange Committe on Decimals, 2000) . Indeed, this new rule made by the SEC required to price the securities in decimal and not in fractions anymore (it used to be in one-sixteenths or 1/16). This meant that the spread of a stock could at least be 1/16 of a dollar whereas under the new regulation the spread could be $0.01 per share. It basically changed the “tick size”, which is the smallest price movement of a financial instrument. This increased the liquidity of the markets and the arbitrage opportunities, which allowed the expansion of high-frequency trading. Indeed, since the spread of a stock went from 1/16 of a dollar to 1/100, the number of possible price movements was multiplied by 6 with this law. As a result, the trading margins decreased by 84% (Kim, 2007).

In 2005, the SEC established a new set of rules: the Regulation National Market System (Regulation NMS). The idea was to strengthen the U.S. exchanges by improving fairness in price execution by ''requiring trades to be automatically executed at the best available price'' (Barker, 2011, p. 49). Moreover, there was an improvement in the access to the quotes (with lower access fees) and the disclosure of market data. It also required trades to be automatically executed These rules helped high-frequency trading to further develop itself. Indeed, high-frequency traders began to have better information on the market data, which allowed to create more efficient algorithms. Moreover, the lower access fees, reduced the execution costs for high-frequency companies. In Europe, the Markets in Financial Instruments Directive (MIFID) implemented in 2007 is considered to be the equivalent to the Regulation NMS.

This is why high-frequency in its modern form (speed of orders of milliseconds) started in the beginning of the 21st Century with the increase amount of computation power the computers could have and the regulation that fostered it.

It is important to mention the episode of the Flash Crash that occured on the 6th of May 2010. On that day, the Dow Jones, a stock market index of 30 large public companies, fell over 9% before partially recovering as shown on figure 2. But what was strange wasn't the drop in price but the speed at which the Dow Jones decreased. Indeed, this was the greatest intraday point drop in the history of the Dow Jones. This episode started at 2:32pm and lasted about 36 minutes (Kirilenko, 2017). Other U.S. stock markets such as the S&P 500 and the Nasdaq suffered similar movements on that day.

Figure 2. Evolution of the Dow Jones on the 6th of May



Source: (Nille, 2015)

High-frequency trading became publicly known after that event because the regulators' fear was that high-frequency trading had contributed to this strange behavior on the markets by worsening the price decline. This was later confirmed by a report from the SEC and Commodity Futures Trading Commission (CFTC). This report stated that a firm sold an uncommonly big number of E-Mini S&P contracts (75,000 for a value of $4.1 billion), leading to a shortage of available buyers. Then high-frequency traders started buying the contracts and directly began selling them, contributing to the abrupt decrease in price. Since ''normal'' buyers were still lacking, high-frequency traders started to buy and resell the contracts to each other which the SEC/CFTC report described as ''a ''hot-potato'' volume effect'' (CFTC & SEC, 2010).

However, these conclusions were widely criticized by several people and organizations such as the Chicago Mercantile Exchange, (Leinbewer, 2011) or Nanex, a firm specialized in the analysis of high-frequency trading data.

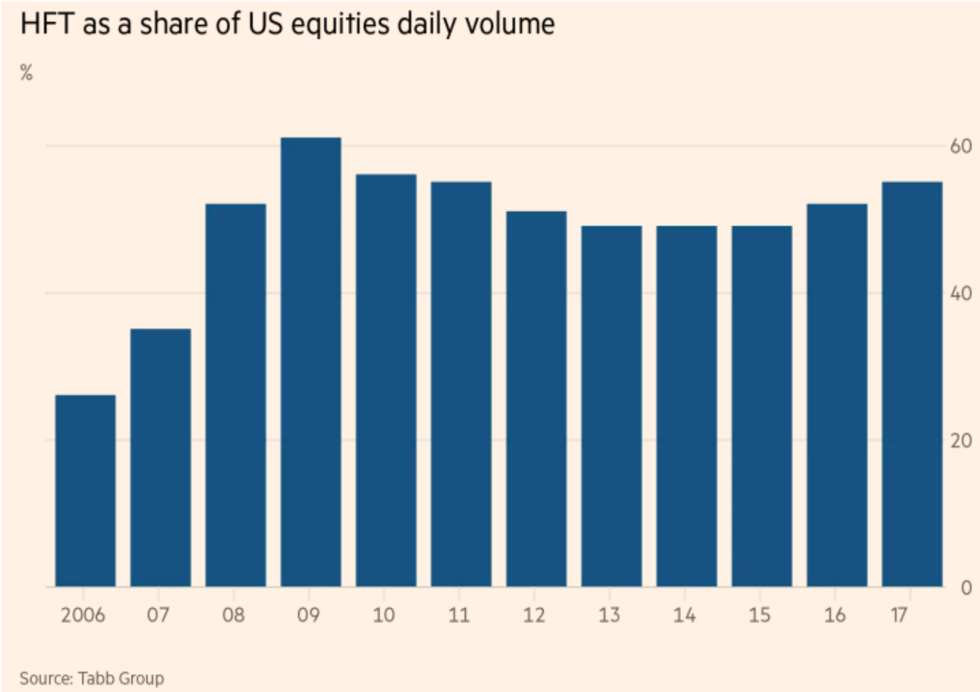
Other ''flash crash'' episodes occurred after 2010. For instance, in April 2013, a false report of explosions at the White House led to $136 billion loss for the Standard & Poor's 500 index in two minutes. On the 7th of October 2016, the value of the pound against the dollar decreased from \$1.26 to \$1.18 (a 6.7% drop) in 2 minutes before rebounding to \$1.24. Figure 3 displays this sudden change in price.

Figure 3: £/\$ exchange rate on the 7th of October 2016

Source: (The Guardian, 2017)

As shown on figure 4, the use of HFT has been rapidly growing in the beginning of the Century to now represent more than half of the volume of U.S. equities transactions. There was a sharp increase of HFT from 2007 to 2009 because of the financial crisis. With constant movements on the stock markets, this crisis offered the possibility for high-frequency traders to apply their strategies (which will be discussed in the next section).

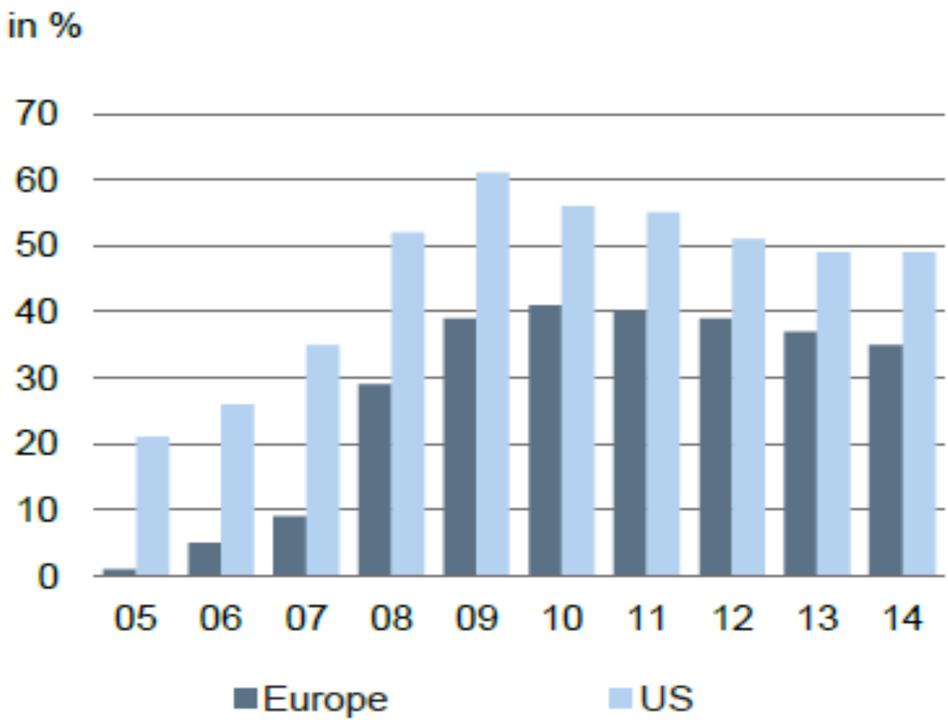
Figure 4: High-frequency trading as a share of US equities daily volume



Source: (Meyer, 2018), Tabb Group

This trend doesn't only concern the U.S. stock exchanges. The European stock markets also saw the rise of HFT, though on a smaller scale than in the U.S, as seen on figure 5.

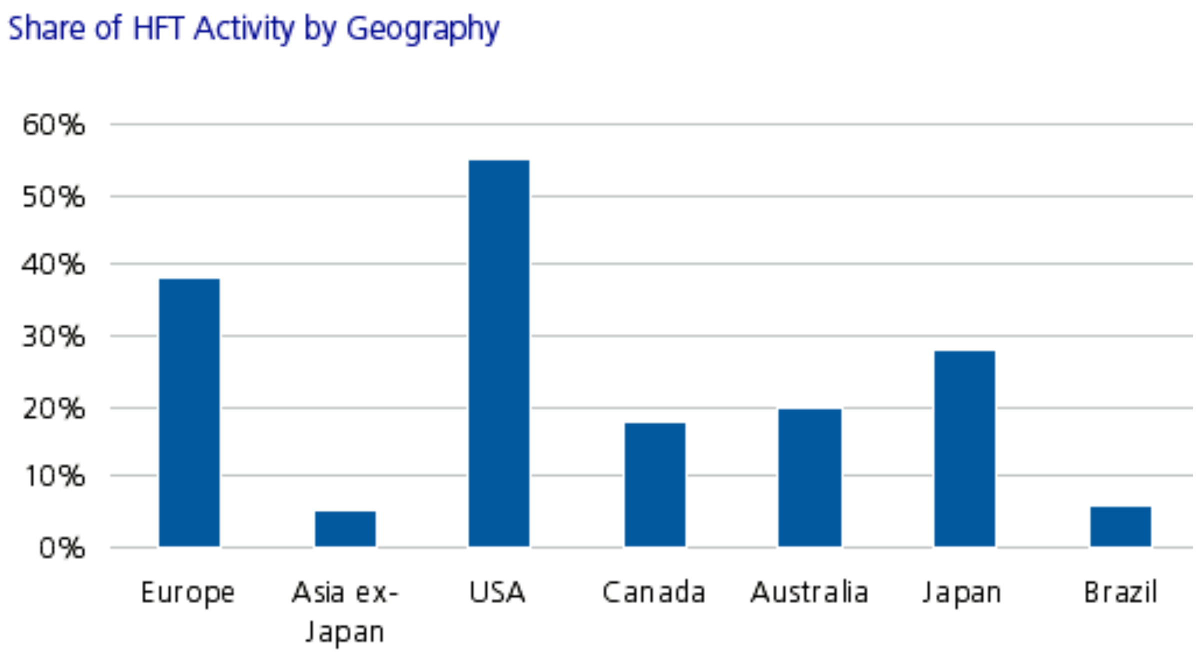
Figure 5: Share of High-Frequency Trading in total equity trading



Source: TABB group, Deutsche Bank Research

Due to the lack of public information available, the data of HFT in APAC and South America only dates back to the beginning of the 2010 decade. However, these data can still give a good idea of the HFT situation in these regions. In 2011, as seen on figure 6, HFT seemed to have marginally penetrated Asia (except for Japan) and Brazil. This is partly due because the markets are less fragmented in those countries, which offers less possibilities of arbitrage to high-frequency traders. Moreover, the regulation in some regions isn't always in favor of HFT. For instance, in China it is forbidden to buy and sell a stock share during the same day. There may also be a lack of cutting edge technology that HFT companies need to operate on those markets. Finally, some stock exchanges such as in Singapore or China have high transactions costs. On the other hand, the high share of HFT in Japan can be explained by several factors: low transaction costs, a market that is fragmented enough for high-frequency traders to act and a tick size that has been lowered in the past year (which had a similar effect to the decimalization in the U.S.) (Kauffman, 2015).

Figure 6: Share of high-frequency trading by geography



Source: (Celent, 2011)

C. Strategies in high-frequency trading

HFT strategies can be divided into two categories: aggressive and passive.

Aggressive

There are two types of aggressive HFT strategies: spoofing and pinging.

With spoofing, a high-frequency trader will generate a large order to quickly change the price of a security on a short timeframe, then cancels his order and at the same time buy or sell his security to profit from the change in price (Haas, 2018). This strategy has two phases:

1. For instance, a high-frequency trader possessing 1,000 shares of firm XYZ at $10 per share will initiate an order to buy 100,000 shares of company XYZ thus leading to an increase in the price of the shares (now quoting at $10.05).
2. He immediately cancels his order and sells his shares, making a profit of $50. The fact that this happened in less than a second is impossible for a human to see what happened. By repeating this strategy, hundreds or thousands of time each day, the cumulative profits can potentially be huge.

Since it is a type of market manipulation, this strategy is forbidden but it is difficult to catch everyone that does it. Indeed, due to the high number of transactions on every stock exchange, it is difficult to analyze all of them. Moreover, these strategies can last for a fraction of second so it is harder to identify them. It can also be hard to know if a transaction has been cancelled to intentionally use spoofing or if it was a bug in the algorithm or in the computer. However, several people have been condemned in the past years with the most famous case being Michael Coscia who was sentenced to pay a $2.8M fine in 2013.

With Pinging, the idea is to find large liquidity providers on the markets. This would be for example an institutional investor that wants to buy a large number of shares of a company. This investor will split his order in several pieces to avoid impacting the price of the stock too much (basically seeing the price increase). Therefore, a high-frequency trader will execute a series of small orders to sell a stock (100 shares for instance). If those orders are immediately filled, this can mean that an institutional investor wants to buy many shares of the stock. Thus the high-frequency-trader will buy shares of this stock and sell them at a higher price to the institutional investor.

Passive

There are two types of passive HFT strategies: market making and arbitrage.

Market making isn't a new concept and the idea behind this is to make a profit on the bid/ask spread of a stock. For instance, a market maker would buy 1000 shares of company XYZ for $100 each rapidly selling these shares when they are traded at $100.05 making a net profit of $50. Of course, the risk for the market maker is that the price of the shares goes below $100 before selling them, thus virtually losing money. Market marking is mainly ''the provision of liquidity in listed instruments which are not liquid'' (Authority for the Financial Markets, 2010, p. 14)

However, with high-frequency trading, this strategy can be enhanced because of three factors:

1. HFT can execute order faster than a human, thus reacting quicker to market changes (a small increase in price for a second for instance).
2. With HFT, it is possible to implement this strategy for multiple financial instruments at the same time and this wouldn't be possible for a human being.
3. HFT being automated, a human just has to monitor the algorithms and the results received but the computers can potentially infinitely be running without taking breaks, which no human is capable of.

Doing market making with HFT means that one will constantly be making a small profit on each transaction (since the algorithms can take advantage of a small bid/ask spread lasting even a second) but since the computers can execute thousands or millions of transactions each day, having a small margin can still lead to huge profits at the end of the day. And at the end of the day, the high-frequency trader, doesn't even hold a position so he doesn't even risk losing money.

Arbitrage is a similar strategy with the difference that the goal is to profit from price differences of a security between two stock exchanges (Durbin, 2010). Since the same stock share can be exchanged on several platforms, it is possible that at some point in time (even for a fraction of a second), the price of the security on stock exchange A is a bit lower than on stock exchange B. Therefore, being able to see this difference with his algorithms, a high-frequency trader will buy some shares on stock exchange A before quickly selling them on stock exchange B, thus making a small profit. Like the market making strategy, this strategy is based on the principle of having a high volume of transactions with a small margin on each one of them and again, the high-frequency trader doesn't have to hold a position since he knows for sure that he will sell the shares elsewhere.

Due to the dozens of stock exchanges existing, the thousands of stocks available and the millions of transactions executed each day, HFT strategies aren't limited to the ones mentioned above since there can be dozens of different ways to profit from this situation. However, the strategies mentioned in this paper are generally considered as the most used and most known.

In conclusion, we can assert that high-frequency trading is a particular type of trading that now represents a large part of trading transactions. The fact that high-frequency traders invest in cutting-edge technology and manage transactions occurring at high speed means that they have an advantage over the competition and it is tempting to use predatory techniques such as the ones described before. Therefore, the regulators came up with laws to regulate these activities. However, it is still unclear whether or not high-frequency significantly impacts the stability of the financial markets. By stability we mean the volatility of those markets. Is high-frequency trading impacting the volatility of the stock markets?

b) Hypothesis to test

1. The least traded stocks weren't subject to a high volatility on the 6th of May 2010 compared to the most traded stocks (for the Nasdaq, Dow Jones Industrial Average and the S&P 500).
2. There was a significant difference in the price variation between the most traded and least traded stocks on the 6th of May 2010.

These are the two hypothesis that we are testing in this paper by analyzing data and looking at the results.

3. Methodology

In order to compare the volatility of the most and least traded stocks we compute the standard deviation of each stock. The standard deviation of each stock is an absolute value thus we cannot simply compare them because they may have very different values. Therefore, we divide the standard deviation by the mean of each stock. By multiplying this value by 100 we obtain the coefficient of variation (c_v), which is a relative measure. The formula lies hereafter:

$$c_v = \frac{\sigma}{\mu} * 100$$

where σ is the standard deviation of the price of the stock,

μ is the mean of the price of the stock

By comparing the mean of the coefficients of variation of the most and least traded stocks, we are able to know whether or not the least traded stocks were subject to a volatility of the same magnitude compared to the most traded ones.

In order to compare the price variation between the most and least traded stocks we applied the following procedure. Over a period going from the 4th of January 2010 to the 30th of April 2010, we selected the 10% most and 10% least traded stocks by volume, on the S&P 500, the Nasdaq and the Dow Jones Industrial Average[1]. Looking at the volumes of the stocks 4 months before the Flash Crash gave us a good idea of which stocks were traded the most and the least. Choosing a longer timeframe would have meant looking at stock volumes in 2009 but the composition of the Nasdaq and the S&P 500 changes every year so some stocks wouldn't have any volume values for a timeframe longer than 4 months. We took the stock price of each of these stocks during the 6th of May 2010 between 14:00 and 15:59 with intervals of 1 minute. Indeed, the Flash Crash occurred at 14:32 and lasted for about 30 minutes so this two-hour timeframe covers the Flash Crash. We chose 1-minute intervals because we want to see as most variations as possible and 1-minute intervals are the smallest intervals of observations available on the Bloomberg Terminals (from which these data have been gathered). Therefore, we have 120 data points for each stock which will allow us to see the price variation

[1] The numbers representing the volume of the stocks on those markets can be found in the appendice (annex 1-3).

during the Flash Crash. However, data weren't available for some stocks because stocks whose price decreased by more than 60% of their 2:40 p.m. price during the Flash Crash saw their transactions cancelled (Madhavan, 2012). Thus, for the S&P 500 samples, we took 29 of the 50 most and least traded stocks to have samples of equal sizes.

We chose the 6th of May 2010 because on that day, an abnormal high volatility episode occurred on the financial markets. This episode was abnormal because no major politic or economic news appeared on that day. Moreover, deciding to compare stocks on a longer period would mean that there would be a lot of noise in the data. Indeed, stocks' volatility can come from several macroeconomic factors such as changes in monetary policies, wars, etc. Stocks' volatility can also be influenced by microeconomic factors such as speculative investors, unexpected good or bad financial results or unexpected news coming from the company (scandal of corruption, new deals announced, etc.). It would be therefore extremely difficult to choose a timeframe for which we are sure the main driver of the volatility is high-frequency trading. The Flash Crash of the 6th of May 2010 is considered to have been caused by high-frequency traders and this is why it was the best timeframe we could choose.

We took the most traded stocks because they have the highest probability of being targeted by high-frequency traders since their liquidity is the highest on the markets. We selected stocks from the S&P 500, Nasdaq and Dow Jones Industrial Average because those stock markets all suffered a sudden sharp decline and rebound on the 6th of May 2010. Moreover, due to the high volume of transactions on those markets and the fact that high-frequency traders are more likely to operate on those markets, we were more likely to get significant results than with Asian or European stock markets.

If we see a significant difference between the drop of prices between the most traded stocks and the least traded stocks, this would mean that HFT partly impacted the drop in prices. Thus, we are looking for a dissimilarity between the most traded and least traded stocks on the different stock markets.

To measure the dissimilarity we used a package on R (a statistical programming software) called ''pdc''[2]. ''Pdc'' stands for Permutation Distribution Clustering and this package allows to cluster time-series. The goal of this package is to ''formalize similarity as small relative complexity between two time series.'' (Brandmaier, 2015, p. 2) As an approximation of complexity, we use the permutation distribution. The permutation distribution gives a probability to the incidence of some patterns of the ranks of values in a time series (Brandmaier, 2015). In order to do that, the time series is divided into sequences of a certain length, m (referred to as embedding in m-space). For every sequence, the ranks of the values are computed. The permutation distribution is obtained by calculating the relative frequency of the distinct rank patterns (or ''*ordinal patterns''*). Each possible rank pattern can be established by a permutation of the values 0 to m – 1, thus, the name ''Permutation Distribution'' (Brandmaier, 2015).

With this in mind, the dissimilarity between two time series is considered to be the dissimilarity between their respective permutation distributions. This is measured by the squared Hellinger distance[3]. By computing the squared Hellinger distances pairwise between all the permutation distributions we obtain a matrix that can be used for hierarchical clustering. We finally obtain a dendrogram illustrating a tree with branch heights representing the distance between clusters. Then we are able to create a group of 2, 3 or even 4 clusters depending on a chosen height.

The choice of the parameter m impacts the efficiency of the clustering. The bigger the m, the larger the visual potential since more distinct patterns can be detected. Nonetheless, as the m increases, fewer observations of every observation will be found, meaning that we would lose accuracy for the estimation of the frequency.

The Pdc package has properties that make it an interesting candidate for a dissimilarity measure :

- It is invariant to differences in mean and variance in the time series (Montero, 2014, p. 16). Therefore, there is no need to preprocess the data by normalizing them ''which

[2] The code to perform the clustering can be found in the appendice (annex 4).
[3] The squared Hellinger distance is used to compute the similarity between two probability distributions.

often has a serious impact on the clustering results when common metric dissimilarity measures are used, e.g., the Euclidean distance'' (Brandmaier, 2015).

- It has phase invariance, which means it is does not matter when a pattern occur but only that it does occur according to (Brandmaier, 2015).
- The reliance on ranks is a distinct feature of the permutation distribution because the distribution is determined by the observed values relative to each other instead of their absolute values (Brandmaier, 2015).

We finally obtain clusters and should we see that time-series coming from the same sample are clustered together, this would mean that there is a significant dissimilarity between the samples.

4. Results

Figure 7: Coefficients of variation

Sample / Stock market	Most traded stocks	Least traded stocks
DJIA	1.66 %	1.65 %
Nasdaq	1.57 %	1.20 %
S&P 500	1.79 %	1.48 %
Weighted mean	1.73 %	1.42 %

DJIA dataset

Figure 8: 2 clusters of the DJIA dataset[4]

Cluster 1	Cluster 2
Intel	Bank of America
General Electric	
Travelers	
United Technologies	
3M	

Figure 9: 3 clusters on the DJIA dataset

Cluster 1	Cluster 2	Cluster 3
Intel	Bank of America	General Electric
United Technologies		Travelers
		3M

[4] The stocks whose names are in red are the least traded stocks.

Figure 10: Hierarchical clustering of the DJIA dataset

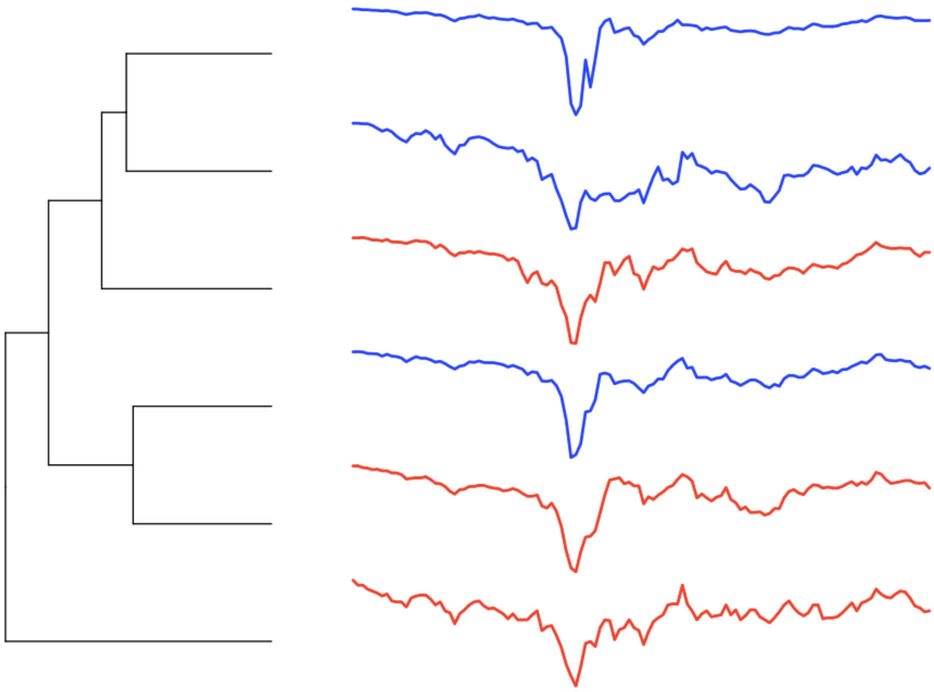
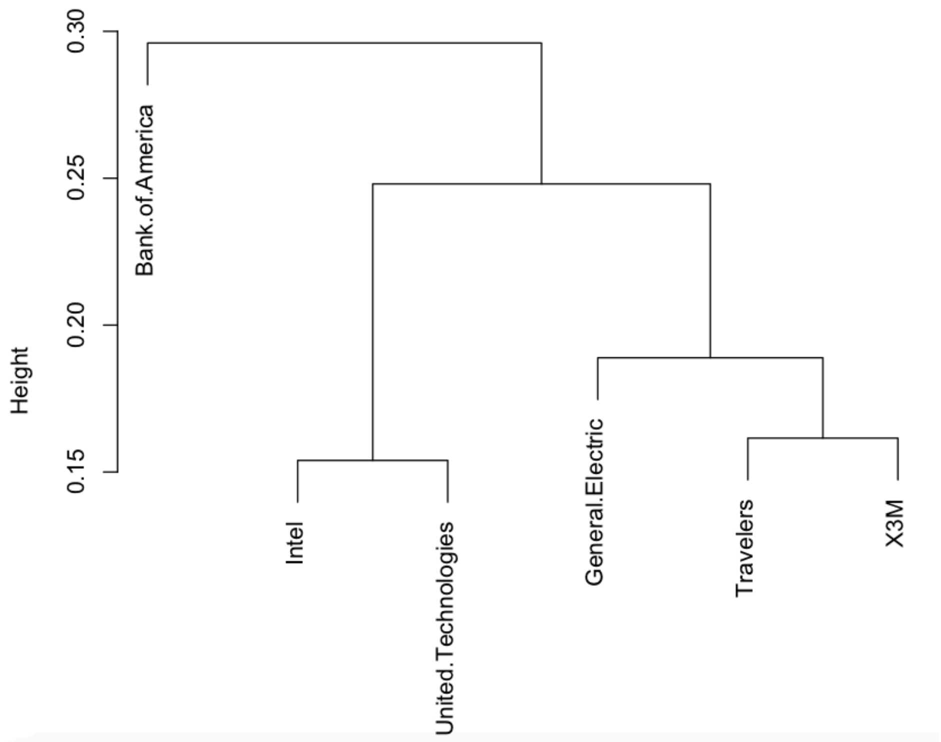


Figure 11: Hierarchical clustering of the DJIA dataset (with the names of the stocks)



Nasdaq

Figure 12: Hierarchical clustering of the Nasdaq dataset

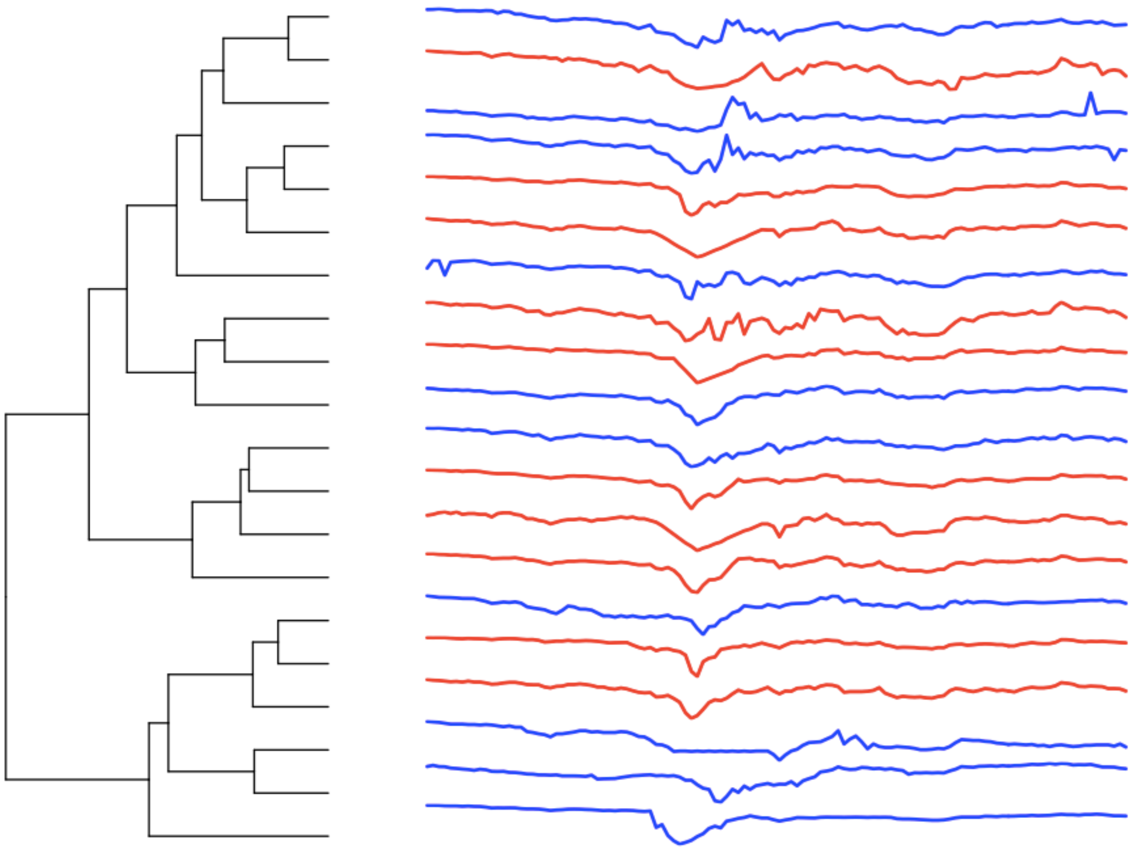


Figure 13: Hierarchical clustering of the Nasdaq dataset (with the names of the stocks)

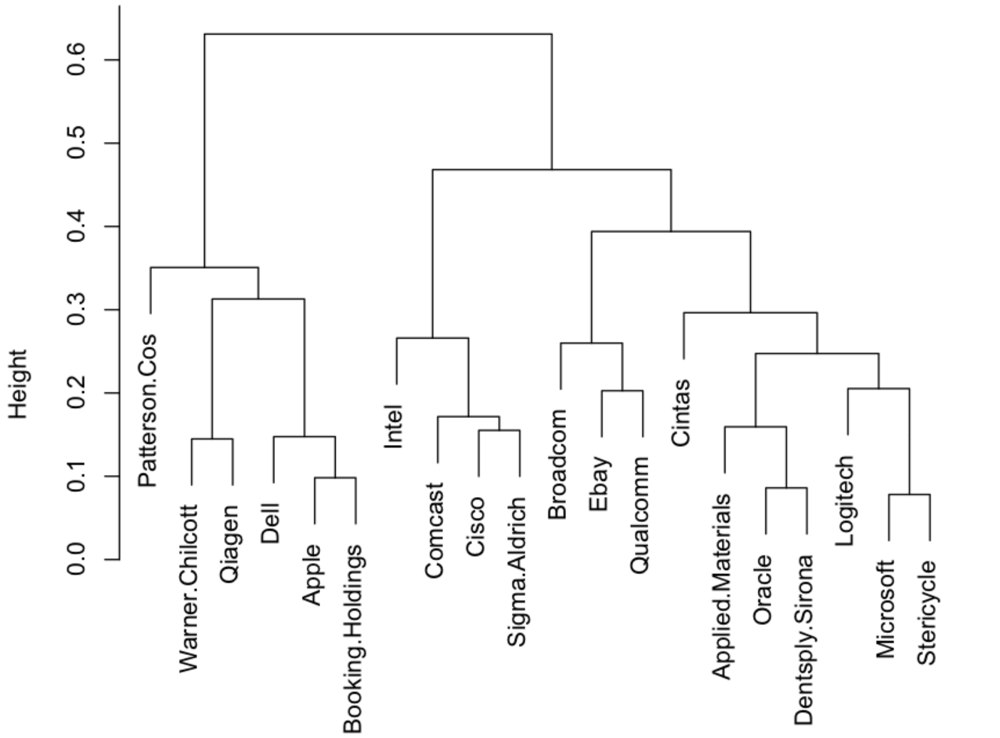


Figure 14: 4 clusters of the Nasdaq dataset

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Apple	Intel	Microsoft	Ebay
Warner-Chilcott	Cisco	Oracle	Qualcomm
Qiagen	Comcast	Applied Materials	Broadcom
Booking Holdings	Sigma-Aldrich	Logitech	
Dell		Cintas	
Patterson Cos		Stericycle	
		Dentsply Sirona	

Figure 15: 3 clusters of the Nasdaq dataset

Cluster 1	Cluster 2	Cluster 3
Apple	Cisco	Qualcomm
Booking Holdings	Comcast	Applied Materials
Dell	Intel	Dentsply Sirona
Patterson Cos	Sigma-Aldrich	Oracle
Qiagen		Ebay
Warner-Chilcott		Cintas
		Logitech
		Microsoft
		Stericycle
		Broadcom

Figure 16: 2 clusters of the Nasdaq dataset

Cluster 1	Cluster 2
Apple	Applied Materials
Booking Holdings	Broadcom
Dell	Cintas
Patterson Cos	Dentsply Sirona
Qiagen	Intel
Warner-Chilcott	Oracle
	Qualcomm
	Stericycle
	Microsoft
	Cisco
	Ebay
	Comcast
	Logitech
	Sigma-Aldrich

S&P 500

Figure 17: Hierarchical clustering of the S&P 500

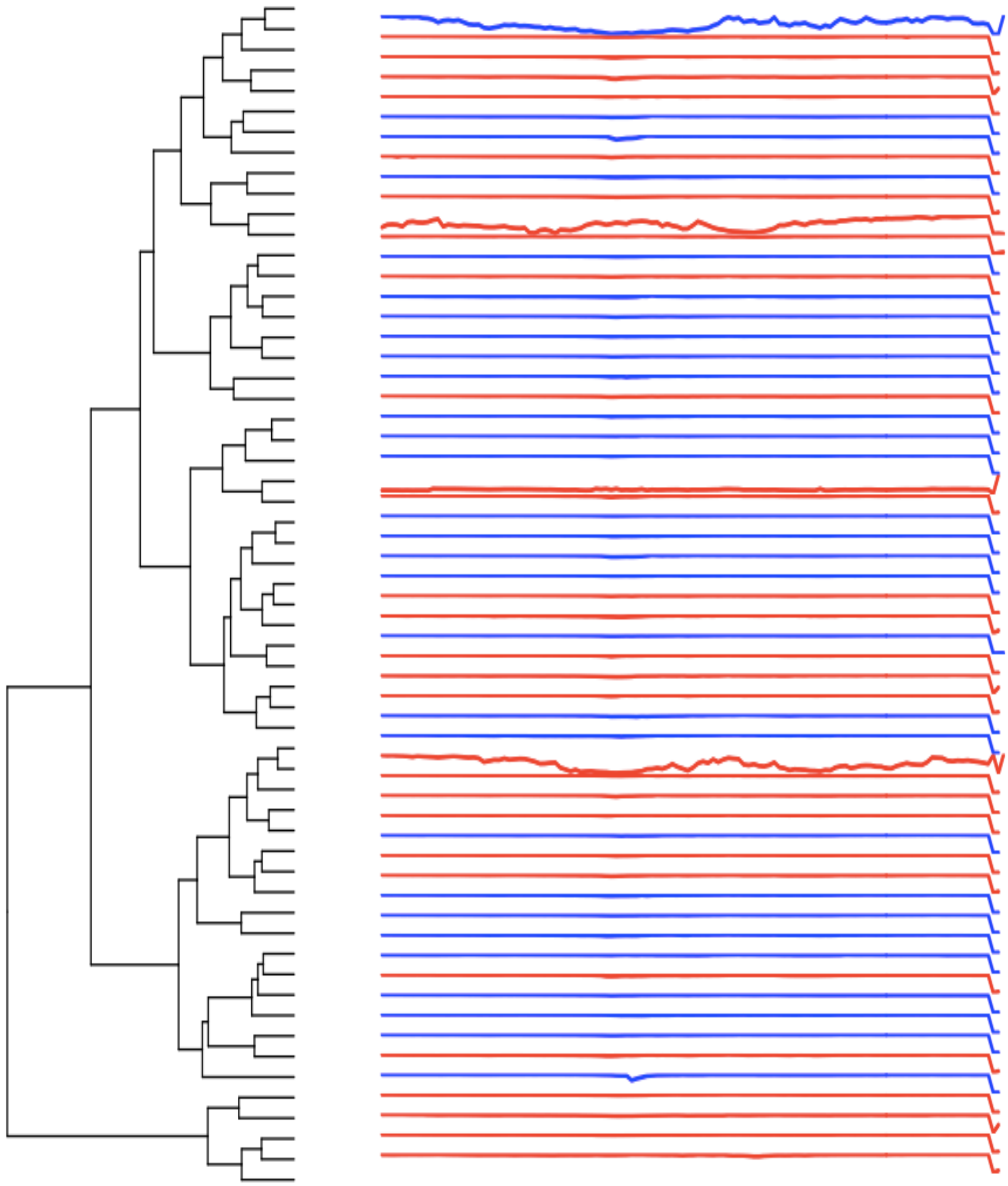


Figure 18: Hierarchical clustering of the S&P 500 dataset (with the names of the stocks)[5]

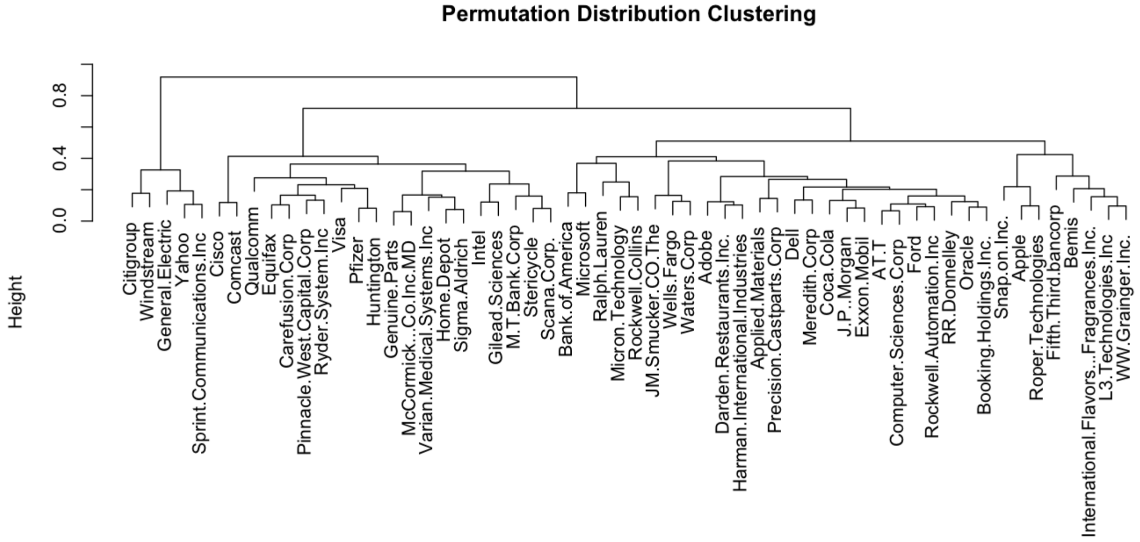


Figure 19: 2 clusters of the S&P dataset

Cluster 1	**Cluster 2**
Adobe	Citigroup
Apple	General Electric
Applied Materials	Sprint Communications Inc.
AT&T	Windstream
Bank of America	Yahoo
Bemis	
Booking Holdings Inc	
Carefusion Corp	
Cisco	
Coca-Cola	
Comcast	
Computer Sciences Corp	
Darden Restaurants Inc.	
Dell	

[5] This figure can be found in annex 5 as a one-page figure.

Equifax	
Exxon Mobil	
Fifth Third Bancorp	
Ford	
Genuine Parts	
Gilead Sciences	
Harman International Industries	
Home Depot	
Huntington	
Intel	
International Flavors Fragrances Inc.	
JM Smucker & Co/The	
J.P. Morgan	
L3 Technologies Inc	
McCormick & Co Inc.	
Meredith Corp	
M&T Bank Corp	
Micron Technology	
Microsoft	
Oracle	
Pfizer	
Pinnacle West Capital Corp	
Precision Castparts Corp	
Qualcomm	
Ralph Lauren	
Rockwell Automation Inc.	
Rockwell Collins	
Roper Technologies	
RR Donnelley	
Ryder System Inc.	
Scana Corp	

Sigma-Aldrich	
Snap-on Inc.	
Stericycle	
Varian Medical Systems Inc.	
Visa	
Waters Corp	
Wells Fargo	
WW Grainger Inc.	

Figure 20: 3 clusters of the S&P 500 dataset

Cluster 1	**Cluster 2**	**Cluster 3**
Adobe	Citigroup	Carefusion Corp
Apple	General Electric	Cisco
Applied Materials	Sprint Communications Inc.	Comcast
AT&T	Yahoo	Computer Sciences Corp
Bank of America	Windstream	Equifax
Bemis		Genuine Parts
Booking Holdings Inc.		Gilead Sciences
Coca-Cola		Home Depot
Darden Restaurants Inc.		Huntington
Dell		Intel
Exxon Mobil		M&T Bank Corp
Fifth Third Bancorp		McCormick & Co. Inc. MD
Ford		Pfizer
Harman International Industries		Pinnacle West Corp
International Flavors Fragrances Inc.		Qualcomm
JM Smucker & Co/The		Ryder System Inc.

J.P. Morgan		Scana Corp
L3 Technologies Inc.		Stericycle
Meredith Corp		Varian Medical Syste
Micron Technology		Visa
Microsoft		
Oracle		
Precision Castparts Corp		
Ralph Lauren		
Rockwell Automation Inc.		
Rockwell Collins		
RR DOnnelley		
Sigma-Aldrich		
Snap-on Inc.		
Roper Technologies		
Waters Corp		
Wells Fargo		
WW Grainger		

Figure 21: 4 clusters of the S&P 500 dataset

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Apple	Adobe	Citigroup	Carefusion Corp
Bemis	Applied Materials	General Electric	Cisco
Fifth Third Bancorp	AT&T	Sprint Communications Inc.	Comcast
International Flavors Fragrances Inc.	Bank of America	Windstream	Equifax
L3 Technologies Inc.	Booking Holdings Inc.	Yahoo	Genuine Parts
Roper Technologies	Coca-Cola		Gilead Sciences
Snap-on Inc.	Computer Sciences Corp		Home Depot
WW Grainger Inc.	Darden Restaurants		Huntington
	Dell		Intel

	Exxon Mobil		McCormick & Co Inc.
	Ford		M&T Bancorp
	Harman International Industries		Pfizer
	JM Smucker & Co/The		Pinnacle West Capital Corp
	J.P. Morgan		Qualcomm
	Micron technology		Ryder System Inc.
	Microsoft		Scana Corp
	Oracle		Sigma-Aldrich
	Meredith Corp		Stericycle
	Precision Castparts Corp		Varian Medical Systems Inc.
	Ralph Lauren		Visa
	RR Donnelley		
	Rockwell Automation		
	Rockwell Collins		
	Waters Corp		
	Wells Fargo		

5. Discussion

Coefficient of variation

The coefficient of variation is higher for the most traded samples meaning that these stocks had a higher volatility that the least traded ones. However, when computing the weighted mean of the coefficient of variation for the different stock markets, we observe that the magnitude of the variation of the least traded stocks is comparable as to that of the most traded ones. Therefore, the least traded stocks were subject to the same magnitude of volatility than the most traded shares.

DJIA dataset

When we consider 2 clusters, Bank of America is in one cluster and all the other stocks are put together. There is no significant dissimilarities between the stocks. When we consider 3 clusters we see that it is difficult to spot a difference between the clustering of the two samples. The lack of dissimilarity between the samples of the Dow Jones Industrial Average can come from the fact that the sample size is small (only 6 observations in total). Moreover, even the lowest traded stocks of the DJIA still have a huge volume of transactions. For example, the stock of 3M had about 114 millions volume of transactions during the 4 months preceding the Flash Crash. This means that high-frequency traders would be willing to target these stocks as well. Therefore, a dissimilarity between the most and least traded stocks would not exist as seen in our results.

Nasdaq dataset

We manage to have a better clustering of the Nasdaq sample. Indeed, when we look at the hierarchical clustering performed, if we create 4 clusters we see that for cluster 1, 4 out of 6 observations are from the least traded group, for cluster 2, 1 out of 4 observations are from the least traded group, for cluster 3, 1 out of 3 observations are from the least traded group and for cluster 4, 4 out of 7 observations are from the least traded group. Thus, in each cluster more than half of the observations come from a certain sample, which is a better result than if they had been put in those clusters randomly. Indeed, on average, 66% of observations come from a certain sample.

If we create 3 clusters however, cluster 1 contains 66% of observations from the least traded group and cluster 2 contains 75% of its observations from the most traded group but the third cluster has 50% of its observations coming from both samples. This means that we start losing our ability to correctly cluster the stocks.

When we take the 2 most hierarchical clusters, one cluster contains 4 stocks from the least traded group and the other one has 6 stocks from the least traded group and 8 stocks from the most traded group. Thus, there is no particular difference between the two samples.

The clustering is better with the Nasdaq dataset than with the DJIA dataset but it's still difficult to find a clear dissimilarity between the samples.

S&P 500 dataset

When we consider two clusters, one contains 5 stocks with only one time series coming from the least traded sample. All the other observations go into the other cluster which means there isn't a particular dissimilarity between the samples. When we take three clusters, cluster 1 has 16 stocks from the least traded group for a total of 33 observations. Cluster 2 has 1 stock from the least traded group for a total of 5 observations. Lastly, the cluster 3 contains 11 stocks from the least traded group with a total of 20 observations. Therefore, we have a weighted average of 53% of observations in each sample coming from a particular sample which is not really better than if those stocks had been randomly put into the clusters. If we take 4 clusters, the weighted average is 60%.

In conclusion we find that there was not a clear dissimilarity between the most and least traded stocks when the Flash Crash happened This means that High-Frequency trading doesn't significantly impact the volatility of the stock markets since even the stocks which are less affected by high-frequency traders saw a sharp decline followed by a rapid increase in their stock share.

6. Conclusion

a) Summary of the study

We created a framework to find out whether or not high-frequency trading has an impact on the volatility of the financial markets. This was done by clustering the time series of the most and least traded stocks of several stock markets during the Flash Crash of the 6 May 2010 and comparing their coefficient of variation. Indeed, during this day between 2 pm and 4 pm, a sharp decrease of the stock markets followed by a quick rise occurred even though no major political or economic news occurred at that time. Moreover, this was done based on the hypothesis that high-frequency traders mainly act on the most liquid stocks. Therefore, should a dissimilarity between the two groups appear, this would have meant that the most traded stocks were more volatile than the least traded ones. We found that there was no particular difference between the two groups. Thus, high-frequency trading does not significantly impact the volatility of the financial markets. Which is consistent to the result about in other studies such as the study from (Groth, 2011).

b) Theoretical implications of the study

The results of the study imply that since high-frequency traders operate on liquid stocks, their impact is ‘’absorbed’’ since many transactions occur for those stocks anyway. Although there was a price variation it was quickly corrected and the prices of the stocks came back to their previous level. Episodes of high volatility on a short timeframe have happened before on the stocks markets (krachs of 1929 and 1987) and high-frequency trading had nothing to do with it. Those “Flash Crash” episodes do not pose a threat to the stability of the financial markets. Moreover, the use of circuit breakers on the stock markets (which prevent crashes to appear by stopping trading for a small period of time when a price movement is too high during a short timeframe) protect investors and prevent them from taking irrational decisions. This is why some stock trading of the S&P 500 were not in the dataset since their transactions had been cancelled by those circuit breakers.

Therefore, it would seem that the sudden drop in price was caused by algorithmic trading and not high-frequency trading. As explained earlier, HFT is only a subset of algorithmic trading and the automation of buying and selling orders may have caused the sharp decline in prices.

c) Managerial implications of the study and the recommendations

From a Regulator point of view, high-frequency trading should be regulated to prevent predatory behaviors and some rules stated in MIFID II such as the obligation to keep the records of all the (cancelled) orders or allowing to create higher fees for orders that are cancelled afterwards go in this direction. However, since HFT does not significantly increase the volatility of the financial markets, there would be no need to prohibit it. The Regulator should seek to regularly assess the technological advancements made in this field and assess the records of high-frequency transactions in the near future to continuously monitor the impact of high-frequency trading on the markets.

d) Limitations and suggestions for the future

There are, however, some limitations to the study. Indeed, the data that were collected from Bloomberg Terminals had 1-minute intervals but high-frequency trading occurs at a much higher speed of execution (fractions of seconds). Therefore, the intraday time series of the stocks do not fully capture the price variation of the stocks. If we had been able to get data with 1-second intervals, the algorithm would have better clustered the data. Unfortunately, we could not have access to the proprietary data of banks since they did not want to share them. Thus, we had to come up with assumptions and be creative just as the Regulator would have been before the introduction of MIFID II this year.

Moreover, the least traded stocks may still be liquid enough to attract high-frequency traders. If it were the case this would be the reason why the algorithm did not cluster the most and least traded stocks distinctively. A study with stocks that have a lower volume than those of the S&P 500 would be interesting.

Finally, even though prices came back to their pre-Flash Crash levels, there was still a sudden price movement. It would seem that on the long and medium term, high-frequency trading does not increase the volatility of the financial markets but is involved in small episodes such as the Flash Crash. Since the prices don't just fall down forever, this type of behavior is not a threat to the stability of the financial markets.

For the future, clustering data from European stock markets if a Flash Crash occurs on one of those markets would be interesting since high-frequency traders now also act on those places. Moreover, trying to come up with a new way to analyze stock data (other than the clustering of time series) would be interesting, the only limit being the availability of data related to high-frequency trading. Indeed, for this paper, we wanted to analyze the evolution of the bid/ask spread for each stock on the day of the Flash Crash (since it is linked to the volatility) but no data were available. Bloomberg keeps those data until 140 days before today so the bid/ask spread of the stocks in 2010 was not available.

7. Bibliographie

Authority for the Financial Markets. (2010). *High frequency trading: The application of advanced trading technology in the European marketplace.* Amsterdam: Authority for the Financial Markets.

Barker, W. P. (2011). *The Growth of High-Frequency Trading: Implications for Financial Stability.* Ottawa: Bank of Canada.

Bhattacharya, S. C. (1995, October). Proprietary Information, Financial Intermediation, and Research Incentives. *4*(4), pp. 328-357.

Bloomberg Professional Services. (2018, July 30). *Bloomberg*. Récupéré sur Bloomberg.com: https://www.bloomberg.com/professional/solution/bloomberg-terminal/

Brandmaier, A. (2015, October). pdc: An R Package for Complexity-Based Clustering of Time Series. *Journal of Statistical Software, 67*(5), p. 23.

Brogaard, J. H. (2014). High-Frequency Trading and Price Discovery. *Review of Financial Studies, 27*(8), pp. 2267-2306.

Cambridge Dictionary. (2018, July 30). *Cambridge Dictionary*. Récupéré sur Dictionary Cambridge: https://dictionary.cambridge.org/fr/dictionnaire/anglais/trading

Cambridge Dictionary. (2018, July 30). *Cambridge Dictionary*. Récupéré sur Dictionary Cambridge: https://dictionary.cambridge.org/fr/dictionnaire/anglais/bid-offer-spread

Celent. (2011). *High Frequency trading: Looking to Asia for Succor.*

CFTC & SEC. (2010). *Findings regarding the market events of May 6, 2010.*
Commission delegated regulation (EU) of 25.4.2016. (2016). *supplementing Directive 2014/65/EU of the European Parliament and of the Council as regards organisational requirements and operating conditions for investment firms and defined terms for the purposes ofthe Directive.*

Durbin, M. (2010). *All About High-Frequency Trading.* New-York: The McGraw-Hill Companies.

Elliott, D. (2015). *Market Liquidity: A Primer.* Washington D.C.: The Brookings Institution.

European Systemic Risk Board. (2016). *Market Liquidity and market-making.* Frankfurt: European Systemic Risk Board.

Exchange Committe on Decimals. (2000, July 24). *sec.* Récupéré sur sec.gov: https://www.sec.gov/rules/other/decimalp.htm#secvii

Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance, 25*, pp. 383-417.

Glantz, M. K. (2014). *Multi-Asset Risk Modeling: techniques for a Global Economy in an Electronic and Algorithmic Trading Era.* Oxford: Academic Press.

Groth, S. S. (2011). Does Algorithmic Trading Increase Volatility? Empirical Evidence from the Fully-Electronic Trading Platform Xetra. *Wirtschftsinformatik Proceedings, 112*, p. 11.

Haas, W. K. (2018, July 30). *Alumneye*. Récupéré sur alumneye.com: https://www.alumneye.fr/les-effets-enjeux-et-reglementations-du-trading-haute-frequence/

Hendershott, T. J. (2011, February 6). Does Algortihmic Trading Improve Liquidity? *The Journal of Finance, 66*(1), pp. 1-33.

Hruška, J. (2016). Aggressive and defensive high-frequency trading and its impact on liquidity of german stock market. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 64*(6), p. 1912.

Jones, C. (2013, March 21). What do we know about high-frequency trading? *Columbia Business School Research Paper, 13*(11), p. 56.

Kauffman, R. (2015, June 9). Will high-frequency trading practices transform the financial markets in the Asia Pacific Region? *Financial Innovation, 1*(4), p. 27.

Kaya, O. (2016). *High-frequency trading: reaching the limits.* Frankfurt: Jan Schildbach.

Kim, K. (2007). *Electronic and Algorithmic Trading Technology, The Complete Guide.* Burlington, Massachussetts, United States: Academic Press.

Kirilenko, A. K. (2017, February 25). The Flash Crash: High-Frequency Trading in an Electronic Market. *The Journal of Finance, 72*(3).

Leinbewer, D. (2011). Avoiding a Billion Dollar Federal Financial Technology Rat Hole. *The Journal of Portfolio Management, 37*(3), pp. 1-2.

Madhavan, A. (2012, July/August). Exchange-Traded Funds, Market Structure, and the Flash Crash. *Financial Analysts Journal, 68*(4), pp. 20-35.

Martin, W. (2017, December 22). *business insider uk*. Récupéré sur uk.businessinsider.com: http://uk.businessinsider.com/bitcoin-price-drops-plunges-friday-december-22-2017-12?IR=T

Meyer, G. B. (2018, January 1). *financialtimes.com.* Récupéré sur ft.com: https://www.ft.com/content/d81f96ea-d43c-11e7-a303-9060cb1e5f44

Montero, P. V. (2014, November). TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software, 62*(1), p. 16.

Nasdaq. (2018, July 30"). *Nasdaq*. Récupéré sur Nasdaq: https://www.nasdaq.com/g00/investing/glossary/v/volatility?i10c.encReferrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmJlLw%3D%3D&i10c.ua=1&i10c.dv=14

Nille, J. (2015, April 22). *lechoblogs.* Récupéré sur l'echo: http://lechoblogs.typepad.com/fairtrade/2015/04/sarao-le-trader-derri%C3%A8re-le-flash-crash.html

Picardo, E. (2018, August 1). *Investopedia*. Récupéré sur Investopedia: https://www.investopedia.com/articles/active-trading/042414/youd-better-know-your-highfrequency-trading-terminology.asp

PwC. (2015). *Global financial markets liquidity study.* London.

The European Parliament and the Council of the European Union. (2014). *Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014.* The European Parliament.

The Guardian. (2017, January 13). *The Guardian.* Récupéré sur theguardian.com: https://www.theguardian.com/business/2017/jan/13/pound-flash-crash-traders-sterling-dollar

Thompson, E. (2006, June 26). The tulipmania: Fact of artifact? *Public Choice, 130*(1), pp. 99-114.

8. Appendix

Annex 1: Volume of Nasdaq stocks from 1/1/2010 to 4/30/2010

Nasdaq	Volume (in billions) from 1/01 to 04/30
Apple	12,62
Intel	5,474
Microsoft	4,862
Cisco	4,06
Ebay	3,324

Comcast	3,32356
Oracle	2,39
Qualcomm	2,35578
Applied Materials	2,24157
Dell	2,2116
Gilead Sciences	1,92
News Corporation	1,616
Baidu	1,54
Starbucks	1,52
Activision Blizzard	1,39
Nvidia	1,37
Seagate Technology	1,26
Yahoo	1,2
Blackberry (Research in motion)	1,18612
Marvell Technology Group	1,052
Symantec Corporation	0,97
Henry Schein	0,92636
SanDisk	0,852
Directv	0,845
Amazon	0,77232
Ross Stores	0,766
Staples	0,715
Altera	0,69968
Flextronics (FLEX)	0,686
Mylan	0,609
Electronics Arts	0,59757
Adobe Systems	0,59735
Celgene	0,577
Cognizant	0,575
Xilinx	0,572
Amgen	0,57084
Infosys	0,54
NetApp	0,517
Steel Dynamics	0,488
Teva Pharmaceutical Industries	0,469
Fiserv	0,457
Qurate Retail	0,42608
Virgin Media	0,3825
CA Inc	0,367343
Linear Technology	0,366
Hologic	0,35048
Express Scripts	0,34904
KLA-Tencor	0,333

Akamai Technologies	0,32469
Citrix	0,32109
Vodafone	0,31846
Komatsu Mining corp (JOY)	0,308025
Cerner	0,304
Autodesk	0,3
Google	0,2981
Costco	0,296
Verisign	0,296
Urban Outfitters	0,284
Dish Network	0,283
Maxim Integrated Products	0,273
Bed Bath & Beyond	0,27
Genzyme	0,265
Intuit	0,259
Automatic Data Processing	0,25759
Paychex	0,252
Foster Wheeler Ltd. (FWLT)	0,248
Apollo Education Group	0,24715
Microchip Technology	0,245
Paccar	0,244
Fastenal	0,239
Wynn Resorts	0,234
Lam Research	0,23
Biogen Idec	0,225
Expedia	0,211
First Solar	0,21
BMC software	0,19716
NII Holdings	0,183
Liberty Global	0,18059
Expeditors International	0,18
Illumina	0,176
Flir Systems	0,17
Garmin Corp	0,168
Life Technologies (LIFE)	0,167
CheckPoint Software Technologies	0,165
CH Robinson Worldwide	0,16
Cephalon	0,144
Intuitive Surgical	0,142
Vertex Pharmaceuticals	0,1407
JB Hunt	0,13
O'Reilly Automotive	0,13

Sears Holdings	0,12247
Warner Chilcott	0,118
Logitech	0,112
Cintas	0,1
Patterson Cos (PDCO)	0,0961
Dentsply Sirona Inc (XRAY)	0,09337
Qiagen	0,08311
Booking Holdings	0,08175
Sigma-Aldrich	0,071
Broadcom	0,061
Stericycle	0,047
Milicom International Cellular	0,044

Annex 2: Volume of DJIA stocks from 1/1/2010 to 4/30/2010

DJIA	Volume (1/1 to 4/30) in Billions
Bank of America	16,84
General Electric	6,71
Intel	5,48
Pfizer	5,11
Microsoft	4,86
Cisco	4,05963
J.P. Morgan	3,60774
AT&T	2,54085
Exxon Mobil	2,24134
Coca-Cola	1,82386
Verizon	1,66265
Mondelez	1,56011
Alcoa	1,47477
Merck	1,34889
Home Depot	1,34
hewlett-Packard	1,13623
Walmart	1,12
Walt Disney	1,007
P&G	0,952
Johnson & Johnson	0,944
Chevron	0,858
American Express	0,82375
Caterpillar	0,767
Dupont	0,63827
Mcdonald's	0,598
IBM	0,555
Boeing	0,52118

United Technologies Corporation	0,41458
Travelers	0,39798
3M	0,11471

Annex 3: Volume of S&P 500 stocks from 1/1/2010 to 4/30/2010

S&P 500	**Volume (1/1 to 4/30)**
Apple Inc	3,91
Bank of America Corp	2,7
Yahoo! Inc - Set up for Fundamental Purp	2
Intel Corp	1,63
Microsoft Corp	1,48
General Electric Co	1,46

Ford Motor Co	1,32
Cisco Systems Inc	1,26
Comcast Corp	1,16
Pfizer Inc	1,12
Oracle Corp	0,82198
QUALCOMM Inc	0,80379
JPMorgan Chase & Co	0,7675
Wells Fargo & Co	0,75997
Sprint Communications Inc	0,7213
Applied Materials Inc	0,68131
Gilead Sciences Inc	0,68092
Citigroup Inc	0,67765
Micron Technology Inc	0,66941
Dell Inc	0,65571
Exxon Mobil Corp	0,63097
AT&T Inc	0,60884
Adobe Systems Inc	0,59735
Freeport-McMoRan Inc	0,59616
Coca-Cola Co/The	0,57052
Darden Restaurants Inc	0,5454
Starbucks Corp	0,54408
Huntington Bancshares Inc/OH	0,54141
Twenty-First Century Fox Inc	0,52355
Qwest Communications International Inc	0,49405
Fifth Third Bancorp	0,48543
eBay Inc	0,46733
Boston Scientific Corp	0,46474
NVIDIA Corp	0,43043
Regions Financial Corp	0,41511
Verizon Communications Inc	0,4123
EMC Corp	0,39343
Visa Inc	0,39214
Mondelez International Inc	0,37406
Merck & Co Inc	0,36892
Symantec Corp	0,3655
Home Depot Inc/The	0,35781
Berkshire Hathaway Inc	0,35029
Walmart Inc	0,34696
Morgan Stanley	0,33459
Kroger Co/The	0,32233
HP Inc	0,31994
KeyCorp	0,31813
CSX Corp	0,30762

Altria Group Inc	0,30588
Mastercard Inc	0,30127
DIRECTV	0,29739
Johnson & Johnson	0,29725
Ross Stores Inc	0,29633
Bristol-Myers Squibb Co	0,29028
Chevron Corp	0,28642
Abercrombie & Fitch Co	0,285132
Procter & Gamble Co/The	0,28382
Advanced Micro Devices Inc	0,28209
Schlumberger Ltd	0,28117
Goldman Sachs Group Inc/The	0,2794
SanDisk Corp	0,27805
NIKE Inc	0,27735
Corning Inc	0,27734
Walt Disney Co/The	0,27652
UnitedHealth Group Inc	0,27119
Amazon.com Inc	0,27081
US Bancorp	0,27009
Halliburton Co	0,26785
Texas Instruments Inc	0,26638
Broadcom Corp	0,26127
United States Steel Corp	0,25951
Lowe's Cos Inc	0,2587
Altera Corp	0,25362
ConocoPhillips	0,24914
CF Industries Holdings Inc	0,247
Alcoa Corp	0,23883
Chesapeake Energy Corp	0,23645
Valero Energy Corp	0,23381
Amgen Inc	0,23159
Southwest Airlines Co	0,23067
Celgene Corp	0,22684
American Express Co	0,22674
Staples Inc	0,22565
Electronic Arts Inc	0,21975
XTO Energy Inc	0,2181
CVS Health Corp	0,21808
Cognizant Technology Solutions Corp	0,21715
Philip Morris International Inc	0,21658
Zions Bancorporation	0,2134
Dow Chemical Co/The	0,21207
TJX Cos Inc/The	0,20879

Xilinx Inc	0,20429
Time Warner Inc	0,20312
Macy's Inc	0,20213
Genworth Financial Inc	0,20206
AES Corp/VA	0,19982
PepsiCo Inc	0,19864
McDonald's Corp	0,19631
Caterpillar Inc	0,19302
Abbott Laboratories	0,19231
NetApp Inc	0,18986
El Paso LLC	0,18946
LSI Corp	0,18854
Union Pacific Corp	0,18358
Host Hotels & Resorts Inc	0,18244
Aflac Inc	0,17991
Fiserv Inc	0,179447
EI du Pont de Nemours & Co	0,1794
CBS Corp	0,17918
Cabot Oil & Gas Corp	0,17812
Newmont Mining Corp	0,1775
American International Group Inc	0,17737
Western Union Co/The	0,17641
MetLife Inc	0,1763
Hartford Financial Services Group Inc/Th	0,17619
Gap Inc/The	0,17577
Bank of New York Mellon Corp/The	0,17139
International Business Machines Corp	0,17102
SunTrust Banks Inc	0,1703
Marshall & Ilsley Corp	0,16785
Tellabs Inc	0,16754
Micro Focus Software Inc	0,16737
BB&T Corp	0,16706
Whole Foods Market Inc	0,16655
Marathon Oil Corp	0,16547
Smith International Inc	0,16431
Eli Lilly & Co	0,16412
Denbury Resources Inc	0,16289
Allergan Inc/United States	0,16199
Hudson City Bancorp Inc	0,16186
Walgreens Boots Alliance Inc	0,1618
Viavi Solutions Inc	0,16077
salesforce.com Inc	0,16044
Boeing Co/The	0,15857

Monsanto Co	0,15776
Eastman Kodak Co	0,157
Prologis	0,1569
AK Steel Holding Corp	0,15678
PNC Financial Services Group Inc/The	0,15596
Hasbro Inc	0,15549
Mylan NV	0,1544
Coca-Cola European Partners PLC	0,15427
Aetna Inc	0,15206
Williams Cos Inc/The	0,15095
Colgate-Palmolive Co	0,15067
Alpha Appalachia Holdings Inc	0,15065
Best Buy Co Inc	0,14905
Interpublic Group of Cos Inc/The	0,14845
United Parcel Service Inc	0,14808
DR Horton Inc	0,14803
Capital One Financial Corp	0,14756
Linear Technology Corp	0,14676
Travelers Cos Inc/The	0,14655
EOG Resources Inc	0,14548
Baker Hughes a GE Co LLC	0,14351
Target Corp	0,14276
Express Scripts Holding Co	0,14216
PulteGroup Inc	0,14134
International Paper Co	0,13951
Occidental Petroleum Corp	0,13857
CA Inc	0,13785
National Oilwell Varco Inc	0,13694
United Technologies Corp	0,13675
People's United Financial Inc	0,13539
Archer-Daniels-Midland Co	0,13516
Reynolds American Inc	0,13292
Mattel Inc	0,13233
Honeywell International Inc	0,1321
State Street Corp	0,13166
Nabors Industries Ltd	0,13118
Citrix Systems Inc	0,13066
Akamai Technologies Inc	0,13015
Safeway Inc	0,12914
General Mills Inc	0,12897
Medtronic PLC	0,12865
Andeavor	0,12858
CNX Resources Corp	0,12855

Southwestern Energy Co	0,12831
Discover Financial Services	0,12766
Anthem Inc	0,12714
NRG Energy Inc	0,12708
SunEdison Inc	0,12682
Johnson Controls Inc	0,12641
Cerner Corp	0,12579
Baxter International Inc	0,12552
Progressive Corp/The	0,12508
Kimco Realty Corp	0,12474
Allstate Corp/The	0,123298
National Semiconductor Corp	0,12267
Weyerhaeuser Co	0,12225
GameStop Corp	0,12208
KLA-Tencor Corp	0,12096
Lennar Corp	0,12092
JC Penney Co Inc	0,12009
CME Group Inc	0,11781
TEGNA Inc	0,11672
Nucor Corp	0,11512
3M Co	0,11471
Western Digital Corp	0,11318
Emerson Electric Co	0,11258
Deere & Co	0,11256
Danaher Corp	0,11248
Bed Bath & Beyond Inc	0,11189
Costco Wholesale Corp	0,11011
Autodesk Inc	0,10965
Anadarko Petroleum Corp	0,10906
Cleveland-Cliffs Inc	0,1084
XL Group Ltd	0,1083
Lorillard LLC	0,10823
Exelon Corp	0,10736
International Game Technology	0,10694
Textron Inc	0,10673
Tapestry Inc	0,10656
SLM Corp	0,1062
Devon Energy Corp	0,10604
Office Depot Inc	0,10585
VeriSign Inc	0,10578
Urban Outfitters Inc	0,10568
Yum! Brands Inc	0,10521
Tyson Foods Inc	0,10508

Juniper Networks Inc	0,10452
Franklin Resources Inc	0,10449
Charles Schwab Corp/The	0,10444
Masco Corp	0,10442
Viacom Inc	0,10428
Southern Co/The	0,10405
Invesco Ltd	0,10331
Teradyne Inc	0,10326
Intuit Inc	0,10322
Avon Products Inc	0,10301
Genzyme Corp	0,10182
Lincoln National Corp	0,10138
Carnival Corp	0,10018
Cigna Corp	0,09935
Automatic Data Processing Inc	0,0983
HCP Inc	0,09825
Biogen Inc	0,09736
Intercontinental Exchange Inc	0,09712
Alphabet Inc	0,09704
Conagra Brands Inc	0,09685
Sysco Corp	0,09627
L Brands Inc	0,0951
Simon Property Group Inc	0,09494
Discovery Inc	0,0948
PACCAR Inc	0,09432
Kohl's Corp	0,09402
H&R Block Inc	0,09329
Paychex Inc	0,09321
Harley-Davidson Inc	0,09294
Spectra Energy Corp	0,09278
Analog Devices Inc	0,09235
Noble Energy Inc	0,09223
Estee Lauder Cos Inc/The	0,0922
Fastenal Co	0,09189
Novellus Systems Inc	0,09175
Goodyear Tire & Rubber Co/The	0,09172
Marriott International Inc/MD	0,09149
Apache Corp	0,09146
Cameron International Corp	0,09125
Cardinal Health Inc	0,09102
Eaton Corp PLC	0,09086
NYSE Euronext	0,09046
Forest Laboratories Inc	0,09006

Hess Corp	0,09006
Newell Brands Inc	0,08971
Illinois Tool Works Inc	0,08938
Principal Financial Group Inc	0,08846
Prudential Financial Inc	0,08828
Apollo Education Group Inc	0,08729
Marsh & McLennan Cos Inc	0,08726
CenterPoint Energy Inc	0,08717
Equity Residential	0,08628
Dominion Energy Inc	0,08502
QLogic LLC	0,08446
Thermo Fisher Scientific Inc	0,08444
Microchip Technology Inc	0,08367
Medco Health Solutions Inc	0,08358
KP Pharmaceuticals LLC	0,08295
VF Corp	0,08278
CBRE Group Inc	0,08206
AmerisourceBergen Corp	0,08198
American Tower Corp	0,0814
Comerica Inc	0,08106
American Electric Power Co Inc	0,08003
NextEra Energy Inc	0,07996
Rowan Cos Plc	0,07983
PPG Industries Inc	0,07974
Wynn Resorts Ltd	0,07971
FirstEnergy Corp	0,07906
FedEx Corp	0,07871
Jabil Inc	0,07851
TechnipFMC PLC	0,07816
Public Service Enterprise Group Inc	0,07751
Chubb Corp/The	0,07738
Nasdaq Inc	0,07712
CenturyLink Inc	0,07649
St Jude Medical Inc	0,07634
First Solar Inc	0,07632
First Horizon National Corp	0,07612
Ball Corp	0,0761
Nordstrom Inc	0,0756
Agilent Technologies Inc	0,07507
Omnicom Group Inc	0,07469
Kimberly-Clark Corp	0,07433
RS Legacy Corp	0,07414
Unum Group	0,07265

Xerox Corp	0,07229
Norfolk Southern Corp	0,07209
Expedia Group Inc	0,07181
Motorola Solutions Inc	0,07158
Allegheny Energy Inc	0,07133
Hormel Foods Corp	0,07102
Fidelity National Information Services I	0,07092
Starwood Hotels & Resorts Worldwide LLC	0,0708
Moody's Corp	0,07068
BMC Software Inc	0,07015
Amphenol Corp	0,06938
Republic Services Inc	0,06907
Compuware Corp	0,06891
Raytheon Co	0,06841
PG&E Corp	0,0684
Kraft Heinz Foods Co	0,06837
Titanium Metals Corp	0,06832
Fluor Corp	0,06769
Loews Corp	0,06753
Waste Management Inc	0,06717
Cummins Inc	0,06694
Kellogg Co	0,06661
CMS Energy Corp	0,06596
Janus Capital Group Inc	0,06561
Lockheed Martin Corp	0,06546
Expeditors International of Washington I	0,06499
Legg Mason Inc	0,06439
Humana Inc	0,06429
Northern Trust Corp	0,06397
Range Resources Corp	0,06397
Time Warner Cable Inc	0,06332
General Dynamics Corp	0,06321
E*TRADE Financial Corp	0,06311
Owens-Illinois Inc	0,06298
New York Times Co/The	0,06266
Campbell Soup Co	0,06264
FLIR Systems Inc	0,06231
Sunoco Inc	0,06206
Quanta Services Inc	0,06195
Ameriprise Financial Inc	0,06181
Family Dollar Stores Inc	0,06166
McKesson Corp	0,0615

CH Robinson Worldwide Inc	0,06093
Helmerich & Payne Inc	0,06074
ONEOK Inc	0,05995
PPL Corp	0,05993
Xcel Energy Inc	0,0592
Monster Worldwide Inc	0,05907
T-Mobile US Inc	0,05898
Tiffany & Co	0,05838
Vornado Realty Trust	0,0583
Praxair Inc	0,05812
T Rowe Price Group Inc	0,05792
Northrop Grumman Corp	0,05786
Life Technologies Corp	0,05783
Ecolab Inc	0,05672
Pepco Holdings LLC	0,0562
McAfee Inc	0,05599
Duke Energy Corp	0,05575
Red Hat Inc	0,05505
NiSource Inc	0,05493
Hershey Co/The	0,05412
Flowserve Corp	0,05407
Wyndham Destinations Inc	0,05404
S&P Global Inc	0,05352
Diamond Offshore Drilling Inc	0,05289
Intuitive Surgical Inc	0,05287
Sempra Energy	0,05263
Dr Pepper Snapple Group Inc	0,05239
Stanley Black & Decker Inc	0,05219
Cephalon Inc	0,05202
Mead Johnson Nutrition Co	0,05171
Lexmark International Inc	0,05148
AutoNation Inc	0,05129
Murphy Oil Corp	0,05123
Edison International	0,05094
Stryker Corp	0,05088
Progress Energy Inc	0,05063
Allegheny Technologies Inc	0,05004
Plum Creek Timber Co Inc	0,04966
Pactiv LLC	0,04932
Air Products & Chemicals Inc	0,04921
Ventas Inc	0,04913
Pioneer Natural Resources Co	0,04905
Pitney Bowes Inc	0,04901

Apartment Investment & Management Co	0,04864
Aon PLC	0,04837
Zimmer Biomet Holdings Inc	0,04737
O'Reilly Automotive Inc	0,04731
Eastman Chemical Co	0,04709
Constellation Energy Group Inc	0,04665
Total System Services Inc	0,04661
Consolidated Edison Inc	0,04529
Welltower Inc	0,04432
Constellation Brands Inc	0,04422
Public Storage	0,0436
Torchmark Corp	0,04321
WEC Energy Group Inc	0,04274
Dover Corp	0,04218
TECO Energy Inc	0,04214
Ameren Corp	0,04186
Boston Properties Inc	0,04153
Molex LLC/US	0,0415
Jacobs Engineering Group Inc	0,04115
Parker-Hannifin Corp	0,04068
Iron Mountain Inc	0,04039
Tenet Healthcare Corp	0,04001
FMC Corp	0,03962
Dominion Energy Questar Corp	0,0394
DaVita Inc	0,03929
Robert Half International Inc	0,03895
Quest Diagnostics Inc	0,03872
EQT Corp	0,03856
Cintas Corp	0,03847
PerkinElmer Inc	0,03806
Leggett & Platt Inc	0,03801
Assurant Inc	0,03786
Dean Foods Co	0,03761
Patterson Cos Inc	0,0373
Becton Dickinson and Co	0,03665
Hillshire Brands Co/The	0,03658
Avery Dennison Corp	0,03637
Coventry Health Care Inc	0,03623
DENTSPLY SIRONA Inc	0,03621
WestRock MWV LLC	0,03599
Entergy Corp	0,03588
Big Lots Inc	0,03579

Whirlpool Corp	0,03473
Sherwin-Williams Co/The	0,03471
Teradata Corp	0,03471
Molson Coors Brewing Co	0,0346
Eversource Energy	0,03458
Brown-Forman Corp	0,03435
Airgas Inc	0,03424
Sealed Air Corp	0,03421
DTE Energy Co	0,03406
Cincinnati Financial Corp	0,034
Federated Investors Inc	0,03352
AvalonBay Communities Inc	0,03339
Goodrich Corp	0,033
Clorox Co/The	0,03253
Sears Holdings Corp	0,03158
Vulcan Materials Co	0,03142
Rockwell Automation Inc	0,03119
EMD Millipore Corp	0,0309
Sigma-Aldrich Corp	0,03048
Equifax Inc	0,03017
Allergan PLC	0,02843
Pinnacle West Capital Corp	0,02805
Varian Medical Systems Inc	0,02805
CareFusion Corp	0,02762
Computer Sciences Corp	0,02759
Booking Holdings Inc	0,0275
M&T Bank Corp	0,02708
Precision Castparts Corp	0,02704
Waters Corp	0,02685
Laboratory Corp of America Holdings	0,02584
Rockwell Collins Inc	0,02575
Leidos Holdings Inc	0,02554
Beam Suntory Inc	0,02538
Jefferies Financial Group Inc	0,0247
Adtalem Global Education Inc	0,02466
ITT Inc	0,02407
Hospira Inc	0,02376
Pall Corp	0,02369
L3 Technologies Inc	0,02366
Ralph Lauren Corp	0,02347
Bemis Co Inc	0,02324
Genuine Parts Co	0,02322
Harris Corp	0,02254

Harman International Industries Inc	0,02205
Stericycle Inc	0,02139
JM Smucker Co/The	0,02137
Ryder System Inc	0,02016
RR Donnelley & Sons Co	0,01997
McCormick & Co Inc/MD	0,01979
CR Bard Inc	0,01936
Integrys Energy Group Inc	0,01823
SCANA Corp	0,01807
Roper Technologies Inc	0,01785
WW Grainger Inc	0,01762
Scripps Networks Interactive Inc	0,01677
International Flavors & Fragrances Inc	0,01444
SUPERVALU Inc	0,01341
AutoZone Inc	0,01215
Nicor Inc	0,01166
Dun & Bradstreet Corp/The	0,01091
Snap-on Inc	0,01065
Peabody Energy Corp	0,00914
Meredith Corp	0,00857
Frontier Communications Corp	0,00518
Windstream Holdings Inc	0,00331
Graham Holdings Co	0,00141

Annex 4: code to use with the ‘pdc’ package on R

```
data <- read.csv2(file="DJIA.csv") #Read the csv file
data = as.matrix(as.data.frame(lapply(data, as.numeric))) #Transforms the dataframe into a matrix containing numeric numbers (necessary to perform the permutation distribution)
clustering <- pdclust(data) #Performs the clustering of the time series
plot(clustering, cols=c(rep("red",29),rep("blue",29))) #Plots the dendrogram with a group in red and the other one in blue
plot(clustering, timeseries.as.labels = FALSE,labels = names(colnames(data))) #Plots the dendrogram with the names of the
```

```
 stocks
cutree(clustering, k = 4) #Returns the cluster membership of each time
 series, given a certain number of clusters
```

Annex 5: Hierarchical clustering of the S&P 500 dataset