Department of Economics and Finance

Chair of Empirical Finance

ORDERFLOW IMBALANCE AND HIGH FREQUENCY TRADING

CANDIDATE
Matteo Pecchiari

SUPERVISOR
Prof. Paolo Santucci De Magistris

CO-SUPERVISOR
Prof. Stefano Grassi

Academic Year 2019/2020

Acknowledgements

I would like to thank my Supervisor Paolo Santucci De Magistris for the countinuous support of my study and research.

I would like to express my sincere gratitude to my mentors, Salvatore Scarano and Fabio Michettoni, for all they have taught me about the industry and about trading.

Finally, I would like to thank my family for their unending support throughout this journey.

Orderflow Imbalance and High Frequency Trading

Matteo Pecchiari

Abstract

This thesis aims to provide a complete description of the relationships between orderflow, trading volume and market returns. Firstly, it was described the trading process and it was done a review of the market microstructure literature in order to explain the differents approaches to the empirical investigations. Thereafter, there are illustrated the properties of the variables and the statistical methods such quantile regression and multiple regression that enable us to highlight many aspects of interconnectedness among these variables. From the outcomes of these empirical studies was possible to build a trading strategy and consequently introduce some improvements.

Contents

Introduction

The technological develpment and the innovations in the financial markets made possible to collect and study the information descending from the market activity in the microstructure.

The information obtained can be extremely useful in order to explain the price movements in the short-term.

The first empirical studies highlight the relevance of trading volume and successively the attention shifted to the orderflow. The orderflow is a more advanced concept respect to the trading volume because it attributes an additional feature to each trade executed making possible to identify where a trade is buyer-initiated or seller-initiated. Consequently, an investor able to get and exploit this information may have an advantage respect the other partecipants who do not, at least in the short-term. The partecipants who, for definition, exploit and trade relying on these signals are the High Frequency Trader (called 'HFT').

They trade at extreme speed against the other investors and aim to gain small profits countless times.

In order to support the thesis that the market activity is a driver of market returns, we investigated with statistical tools the relationships between returns, orderflow and trading volume for four time intervals (30 seconds, 1 minute, 2 minutes, 5 minutes) and we were able to develop a trading strategy based on signals deriving from these variables.

In the first chapter it is illustrated the market framework, the evolution during this last decades and the way how investors trade.

The second chapter describes with the market activity from a more technical view. There are described the methods use to assign the sign to every trade, the first model that riassumes the relations between trading volume, returns with information flow that comes to the market. In this part some studies has been mentioned in order to understand how previous authors deal with these concepts and what they found. Moreover, we describes the most common high frequency trading strategies and analyzed some market inefficiencies.

The third chapter is the main part of this thesis. We studied the distributions of the three variables taken into account, the autocorrelation function of the orderflow and the correlation between returns and orderflow. The first tool used to get a description of these relationships is a multiple linear regression that shows that the effects of the independent variables (orderflow and trading volume) are significant. To examine more deeply the relation between returns and orderflow we availed of quantile regression and found that positive returns are more influenced by orderflow than negative returns.

In the light of these findings a profitable trading strategy has been developed exploiting the orderflow and trading volume statistical features and the relationship with returns. This strategy considers these variable and takes in account the returns of one period ahead.

After having shown that choosing specific thresholds of orderflow and volume can be profitable in the sample considered, another rule is introduced in the strategy deriving from the relation between orderflow and returns. Until the orderflow variable remains positive the trade will not be closed in order to exploiting the buying momentum until it ends. This improves the results of the strategy and shows how the information deriving from orderflow and trading volume can be used in a proper way.

CHAPTER 1

Framework of Financial Markets

1.1 Framework of Financial Markets

Exchanges are the centralized marketplaces for transacting and clearing securities orders. That place may be a physical trading floor, or it may be an electronic system in which traders can easily communicate with each other. The trading rules and the trading system used define the structure.

A regulated market is a system where proposal of buy or sell are put from intermediaries, on behalf of their customers or on their own.

They are executed one against the other one match them with opposite sign. This kind of market are managed by a market management company, authorised from the financial autority.

An alternative to the regulated market are the multilateral trading system (MTF), even these follow the rules imposed by the financial autority but they can be managed by different entities from the market management company autorized to provide investment services. There are less information related to the securities listed.

Systematic Internaliser are intermediary enabled to investment services on their own that in a organized, frequent and systematic way trade financial instruments executing clients orders. It is a bilateral trading system because there is only on intermediary to be the counteparty of every trade. There are no specific rules about information of who issue the securities.

Financial markets can organize the trading sessions in two ways: continuous market sessions and call market sessions.

Continuous market sessions, the most common form, entail in acontinuous trading activity for a lot of hours between traders and permit the price discovery during the day. This allows to determine a significant real price for a given security and consequently to increase the efficiency of the market.

The technology covers a fundamental perspective in the organization of the market with continuous trading session for the purpose of permit the immediate transmission to the market of price movements. It also covers a not less important aspect of collection and trasmission of the trades.

In the continuous market session, a problem that oftenarises is the excess of offer or demand for a security that can produce high volatility in the quotation. The presence of a specialist, or liquidity provider, can resolve thisproblem.

In call market sessions, an auctioneer, in a specified time, invites all traders to trade a security. After all the proposals of buy and sell are made, a price is fixed at a level to satisfyall the offers and demands.

This price is considered teorically the equilibrium price for all operators. The biggest defect of this kind of trading session is that is notable to satisfy the continuous flow from the market and entails the the creation of parallel markets before and after that session. There is no interaction between the market and the traders so the price is notable to react promptly to the arrival of new information making preventing to the market to be efficient and consequently less attractive. After the introduction of automatic quotation this kind of session was abandoned.

It is essential to specify also the way which the market allows to partecipants to interactive among them and how their orders are processed.
The first kind of markets is defined as order driven: in the order driven market can exist limit orders and market orders. The difference between these two orders is that the first can be executed only at the indicated price by the trader, or at a better price (smaller for a buy order, higher for a sell order) and they remain on the order book till the end of the trading session. Contranstly, market orders are executed immediately at the best current price in the other side of the market (best ask for a market buyorder, best bid for a market sell order).
For the same price level, the limit order are sorted with a cronological way, from the oldest to the most recent.
The other kind of market iscalled 'quote driven' is led from the quotation offered by market makers. The market makers exposes continuously two prices: a bid price at which the specialist is willing to buy and an ask price at which it is willing to sell. The difference between these two prices is called bid-ask spread. The market maker forms a fluidifying element for the market because it assumes own positions against other investors. Its profit, deriving from the continuous quotation activity, is the bid-ask spread. The activity represents also an obligation to the other market partecipants in order to permit the trade at the best possible conditions. The order driven is charaterized to be more transparent in terms of liquidity displayed. Contrastly, in the quote driven market ispossible to see only the bid and ask of the market maker.
In this market, the principal advantage is the immediacy in finding a counterpart.

Nasdaq and London stock exchange are examples of quote driven markets.

Recentdevelpments in the financial ecosystem led some markets to assume a hybrid form, combining some aspects of quote driven to other of order driven market.

For example, New York Stock Exchange adopted this form, fundamentally it is an orderdriven market but it requires the presence of a specialist as liquidity provider.

1.2 History and Evolution

In the beginning, the market was based on the Open Outcry auction, a method for communication among traders, and that happened on the pit, or floor, of the market. The orders for buying or selling securities were written down on paper tickets, which needed to be processed by the exchange and each counterparty's clearing firm.

The professional partecipants of the floor trading are the broker-dealers, figures engaged in trading securities for its own account or on behalf of its customers. When executing trade orders on behalf of a customer, the institution is said to be acting as a broker. Brokers help their clients find traders who are willing to trade with them and their profit is the commission charged on the trade.

When executing trades for its own account, the institution is said to be acting as a dealer.

Dealers trade with their clients when their clients want trade. They will buy and sell at bid and ask prices.

Who collect and disseminate trade ticket information from trader to trader, or broker, are the "Runners". The runner will usually be responsible for delivering the trade order to the company's broker located in the exchange's trading pit.

Brokers can take advantage of runners to execute trade orders or personally taking, writing and executing them in the exchange'spit, for a compensation represented by the spread charged on the trade.

Broker-dealers provide trading service to different kind of clients: Institutional clients, large corporations, commercial clients and high-net-worth individuals. For this reason, they played a central role in the financial system.

Arrival of internet and the technological progress transformed the financial markets decentralizing the multilayer framework dealers-clients.

The competition between exchanges increased the trading liquidity and consequently decreased the spread between the same security listed in different exchanges.

The first innovation for the exchanges was the introduction of the electronic trading; for the first time the matching engine of orders was managed by a computer network. This happened in 1971, when the National Association of Securities Dealers introduced an automated quotation system that gives the name to the first electronic stock market: NASDAQ.

After this event, in the 1992 the Chicago Mercantile Exchange (CME) launched its first electronic platform, called 'Globex', where was possibile trade with an electronic system for order quotation on the futures listed on CME.

The fragmentation of electronic trading platform caused a strong development in the Electronic Communication Network (ECN), Alternative Trading System (ATS) and Dark Pools and allowed access to trading securities outside the traditional stock exchanges (NYSE or NASDAQ).

The ECN is an electronic system that widely disseminates orders entered by market makers to third parties and permits the order to be executed against in whole or in part. Trading in the ECN eliminates the need for an intermediary to access to the market.

The alternative trading system are automated trading systems alternative to the official exchanges used mainly from the institutional investors and not available for the retail traders. The main feature consists in the opportunity to access to trading activity without the presence of a specialized.

intermediary. The advantages of these trading venues are the speed of execution, reduction of transaction costs, the anonymity of the active players.

Dark Pool are private forum for trading securities in a confidential way. This makes these alternative markets very attractive for institutional investors for buying and selling large block of securities while remaining anonymous. Also the prices and the number of securities traded remain hidden. This aspect makes also appealing the use of this routes because big investors that have the need of buy or sell large blocks of securities are sure that they will not impact the market price. The disadvantage of using these alternative trading venues is that Some studies of FINRA, in the United States there are 43 Dark Pools owned and managed by investment banks and brokers such Credit Suisse, Morgan Stanley, Goldman Sachs.

In the 2012 about 32% of the trades were done in these Alternative Trading System.

1.3 Activity of Market Partecipants

In the financial markets there are different types of operators that trade for different purposes. In this chapter it is highlighted the features and the differences bewteen these and why they trade in that way.

Brokers are agents who arrange trades for their clients charging a commission on every order. They also act as financial advisers for their clients about investments ideas or organize financial plans.

In order driven and quote driven markets, brokers get orders from their customers and match them with orders and quotes displayed by the other partecipants. This activity is related to trades of small or medium size.

In brokered markets, their role is to looking for traders that are willing to buy or sell the securities; in this case the main difference is that they don't act as liquidity provider but they search liquidity.

In the primary market, on the occasion of issue of new securities, they look for traders that want to buy on their behalf.

Lastly, in mergers and acquisition deals, their role is to find a counterparty to conclude the deal.

The dealers are the other big players in the trading industry. Their purpose is to make money trading with their own account, and not for customers differentiating themself from brokers.

Many of them are professional traders who work on the floor or in trading firms and the exchanges often register them

Their activity is to buy and sell to their clients supplying liquidity in the market. This behaviour improves the efficiency and the liquidity presence in the market making easier for the traders that want immediacy in finding a counterparty and executing their trades.

They are considered as passive traders because they trade when other want to trade.
In addition to offering liquidity as brokers do, they speculate on the price changes and consequently they trade as position traders, buying low and selling high.
Dealers quote bid and ask prices, two side market and that usually happens on equity markets. On other hand they can quote only one side, usually in the bond market. Theyaim to capture the bid-ask spread and their profit is called 'realized spread', calculated after that all their positions on the order book are filled with traders' orders. The quotes displayed by dealers have a time expiration, after that they are cancelled to put other orders at better condition.
The best situation for a dealer is when the orderflow on the market is strong so they continuously get their orders filled and update newone.
Consequently, they construct positions on their portfolios, called 'inventories', and they try to keep them in balance to mitigate the market risk.
One of the biggest risk of this partecipants is the adverse selection risk. They may trade against informed traders causing an inventory imbalance after the future price movements.
The way they reduce risk on their portfolios is a continuous hedging on correlated instrument or rising the ask side or decreasing the bid size.
Sometimes, some investors, called block initiators, need to trade block trades and the sizes are too large to be filled using the liquidity present on the market. The demand for such liquidity is found thanks to presence of block dealers, block brokers or large buy-side traders, alsocalled 'block liquidity suppliers'. The consequence of these trades are significant for the volumes and the prices of the security required.

The New York Stock Exchange defines a block trade as 10.000 shares or more and most of them usually exceeded a quarter of a day's average trading volume.

There are some problems related to block trades: latent demand problem that makes it hard to find necessary liquidity, the order exposure problem that makes reluctant the initiators of the trade to show the size for fear to scar the market for a better informed trader and the price discrimination problem that makes the liquidity suppliers reluctant to trade because they that more size will follow after that.

Block Dealers, or Block Positioner, or Block Facilitators take charge these trades from their clients and break them into smaller parts and distribute in the market or they look for other traders interested to do the counterparty of them. They make transaction cost analyses to ensure they will get a profit from these trades.

Block Brokers don't take charge these big positions and consequently they don't assume any risk for the operation but they look for traders that will fill these orders. Usually it becomes easier when the block broker assemble many traders to fill that and that gives the name to them as Block assemblers. Their profit is given by the commission charged for this service.

The biggest typology of investors are the value traders. They buy or sell instruments that they believe are undervalued or overvalued using all the available information on the security that they want to trade. They also act as liquidity providers even if they don't considered themselves in this way. They get into a trade if the price of a security differs from its fundamental value and this can happen in response of the arrival of new information or the movement caused by uninformed traders. The situation when the price moves from its fair value is given by the fact that the dealers do not recognize that the orderflow on the market thatpushes the security price is given by uninformed traders and consequently the adjust their inventory in order to converge to their target levels. In a situation when a large number of uninformed traders demand liquidity selling stocks the the dealers will offer them a bid to satisfy that demand. This will cause a decrease of the bid prices and so the carrying value of the dealers will be lower than the price before the massive sell. Therefore they accumulate long position in their inventories and they will adjust the prices accordingly. At that point the value traders will take action buying the stocks to the dealers that are ready to liquidate their position and get some profits. The price of the security will return to the fundamental value.

The main problem of a value trader is the adverse selection: infact a value trade can be a counterparty of a better-informed traders that possess information not yet disclosed to the market.

A particular category of investors there are the arbitrageurs, speculators that trade on information about relative values. They follow strategies in order to exploit differences in correlated instruments. The correlation is given by the fact that their values depend on common factors and so they tend to behave in the same way. The divergence from this relation gives birth to arbitrage opportunity and it is defined arbitrage spread. In a pratical view, an arbitrageur buy the instruments which is relatively cheaper and sell the instrument which is relatively expensive. The arbitrageurs construct so called 'hedge portfolios', constitute by different legs, one or more long and one or more short in order to reduce the total risk of portfolio. The size of one leg is identify as the arbitrage numerator and the relation between one leg and the other opposite is called 'hedge ratios' of portfolio. There are different strategies around this concept: pure arbitrages for instruments that their value of hedge portfolio is mean reverting. Risk arbitrages or speculative arbitrages are strategies involving a non-stationary component of hedge portfolio. A famous and very used arbitrage strategy is the statistical arbitrage: if there are two securities with a non-stationary feature and their difference is stationary (in the econometric terminology it's called 'cointegrated time-series') for the variance, it means that the spread between the two securities is bounded. The arbitrageur will buy and sell the two instruments when the difference arrives at maximum level observed exploiting the mean reverting feature of the spread.

CHAPTER 2

Market Microstructure, Orderflow and High Frequency Trading

2.1 Introduction to Market Microstructure

In the financial markets there are a lot of different securities quoted, but all of themhave a common feature: the way howthey are traded. The area of financial economics that focuses on the trading processis the market microstructure. This branch studies different types of aspectssuchas market structure and design, price formation and price discoveryprocesses, transaction and timing costs, information and disclosure and market maker and investor's behavior.

The market structure and design focuses on the relationship between the process of the prices determination and the trading rules and trading protocols adopted by the exchange. The research addresses to study how the market structure affects the trading costs, the efficiency of the market and the disclosure of the information.

The price formation and price discovery processes focuses on how the fair price is determined in the market studying the behavior of market partecipants. Nowadays, a very important and actual issue aboutthis concept is how the high frequency traders affect the price discovery process because they are able to collect and exploit the arrival of new information faster than everyone else in the market universe.

The transaction and timing costs focuses on how the execution methods impact the returns of an investment. It might be possible to classify the processing costs, adverse selection costs, and inventory holding costs under the heading of transaction costs.

The information and disclosure factor focuses on the availability of market information, the transparency and the impact of the information on the behavior of the market partecipants. The market information widespread are price, breadth, bid-ask spread, reference data, trading volumes, liquidity and securities information.

A fundamental aspect of the market that implies the efficiency is the liquidity. It is considered the immediacy to convert an asset into cash or viceversa and reflects the theability to trade immediately by executingat the best available price.

It has three main features: depth, tightness and resiliency. The depthis the quantity of buy and sell orders of one asset around the current price. A security with a deep orderbook enable to trade large orders without causing large price movements.

Tightness refers to the bid-ask spread and it can be considered the cost of trading. Tight spread means efficiency for the market and represent a small cost to get it or out of a position.

Resiliencyis the ability of the market to recover from an event. If a market is considerable resilient the price discrepancies are less frequent and it can be considered efficient for the price discovery process. The role of liquidity is fundamental to attract investors and to raise the reputation of the market because it leads to have an orderbook full of proposals, it tightens the spread and help to bring back the market price to a fair price after a shock.

The trading ennvironment is defined with rules and protocols to place all the partecipants in the same fair position. These rules cover different aspect of the process such order precedence, requirements for trade sizes, pricing increments, opening/closing procedures and the reactions to shock events.

The order precedence specify how the market orders execute with the limit orders on orderbook. Most markets give the priority to orders with the best price (higher for buy limit orders, lower for sell limit orders) and for the orders put on the same price level, the precedence is given by the time of entry.

The minimum price increments, called 'tick sizes', is the minimum difference between two prices. Bigger the tick size, more profitable is to provide liquidity, or trading with limit orders. In the US equity markets, the stocks have 1 cent of tick (for some penny stocks the SEC imposed the '5 tickspilotprogram', that increases the tick size to 5 cents makes more expensive get in and out of a trade).

2.2 Tauchen and Pitts model

The trading activity raised an interest in the financial research in order to analyze the relation between the trading volume and the returns. Tauchen and Pitts (1983) presented a structural model that concerns on the relationship between the variability of daily price change and trading volume.

The market is composed by *J* active traders that can take long or short position. There are *I* trades. That number of traders is assumed to be fixed during the day The market moves from one equilibrium price to another one reacting to the arrival of new information, called 'Information Flow'.

The desired position $Q_{i,j}$ of the Jth trader is given by the relation

$$Qij = \lambda\ (p^{*}_{i,j} - p_i)$$

with $\lambda > 0$ and p_i is the transaction price.

$p^{*}_{i,j}$ is the Jth trader's reservation price.

A positive value of Qij represent the willing of the Jth trader to take a long position, a negative value means a desired short position. The *J* traders on the market have different reservation prices.

The market clearing condition requires that $\sum_{j=1}^{J} Qij = 0$ that implies that the

$$p_i = \frac{1}{J} \sum_{j=1}^{J} p^{*}_{i,j}$$

and defining

$$\Delta p^*_{i,j} = p^*_{i,j} - p^*_{i-1,j}$$
$$\Delta p_i = p_i - p_{i-1}$$

It is possible to derive the trading volume

$$v_i = \frac{\lambda}{2} \mid \Delta p^*_{i,j} - \Delta p_i \mid$$

The assumptions support that the changes in the reservation values are due to the changes in global information and trader-specific information:

$$\Delta p^*_{i,j} = \phi_i + \psi_{i,j}$$

where ϕ_i is the common component and $\psi_{i,j}$ is the individual component.

The moments are:

$$\mathrm{E}[\phi_i] = \mathrm{E}[\psi_i] = 0 \text{ and } \mathrm{Var}[\phi_i] = \sigma^2_\phi \text{ and } \mathrm{Var}[\psi_i] = \sigma^2_\psi$$

It is possible define the return as

$$r_i = \Delta p_i = \phi_i + \frac{1}{J}\sum_{j=1}^{J} \psi_{i,j}$$

If a good news hit arrive to the market, the common component ϕ_i increases causing an increase in r_i.

If a trader has a good private news, the return increases but it is averaged by other traders (from the component $\frac{1}{J}$).

The trading volume becomes

$$v_i = \frac{\lambda}{2} \sum_{j=1}^{J} | \psi_{i,j} - \overline{\psi_i} |$$

The unconditional moments of r and v implied by the Tauchen and Pitts model are:

$$\mu_r = \mathrm{E}\,[r_i]$$
$$\sigma_r^2 = \sigma_\phi^2 + \sigma_\psi^2 / J$$
$$\mu_v = \frac{\lambda\, J}{2} \sqrt{\frac{2}{\pi}}\ \sigma_\phi \left(\frac{J-1}{J}\right)^{\frac{1}{2}}$$
$$\sigma_v^2 = \left(\frac{\lambda}{2}\right)^2 \mathrm{Var}[| \psi_{i,j} - \overline{\psi_i} |]$$

These results lead to the conclusion that as the number of J traders increases, the expected volume increases and the volatility of returns decreases.

The model supports the idea that the information flow that comes into the market in form of private or common information is reflected by the activity of the traders and influences the moments of the returns and the trading volume.

This is one of the first models that give importance to the trading volume and treats it as a dependent variable. Furthermore, the trading volume will be studied as explanatory variable of the returns or the magnitude of the returns.

2.3 Limit Order Book

The different instructions to buy or sell are represented by the different types of orders sent to the market. It is important to highlights these two types because they influence the trading process in different ways: market orders and limit orders.

A market order is an instruction to trade immediately at the best price available. The trader that chooses to buy or sell a security with this order is considered aggressive and he wants immediacy to enter in the market wihout having certainty about the execution price. It is called 'price taker' and from a liquidity perspective, it removes liquidity.

Instead, a limit order is an instruction to buy at a maximum price or to sell at a minimum price. The trader that sends a limit order is considered passive and he waits for the execution have certainty about the price he will get. It is called 'price maker' and he adds liquidity to the market

In the financial ecosystem, the exchanges allow to the listed securities to be traded with the engine that matches buyers to sellers: the Limit Order Book.

It is a centralized database of pending limit orders for a specified price or better and with a given size. It is a transparent system for the market partecipants and the matching algorithm for limit orders with market orders works with a price time priority algorithm: it works with matching priority for price, time and lastly visibility.

A buy limit order with the highest price represents that buyer is willing to pay more than the other traders and he gets the priority of execution when a sell market order will arrive. Consequently, a market order is executed on the best proposal, best bid or best ask (in the US market the first levels are guaranteed by NBBO, National Best Bid Offer). After that, it follows the FIFO principle: First in First out.

The limit order that is submitted earlier will be executed before the later arrival.

Therefore the queue of limit orders follows this chronological rule.

In most markets, there is the possibility to put hidden orders, that stands on the book without being displayed. They are the last limit orders to be executed because the precedence is given to the visible orders.

The ordebook events can be mainly identified in three types: submission of limit orders, cancellation of limit order, execution of market orders.

The empirical studies in the financial literature focus their attention on studying the price impact of the orderbook events. The outstanding limit orders, called Market Depth, concept related to the market liquidity, is associated with the magnitude of price changes. In a thin orderbook, where the market depth is low given by the small size and the low number of limit orders, it is easier to cause a bigger price change given to the fact that there is less need of market orders, respect a thick orderbook, to buy or sell all the first level proposals and than to move the price.

The price dynamics are also conditioned by the arrival, the size and the direction of the market orders.

There are other kinds of orders in order to address to the different needs of investors. For example, conditional orders allow to traders to put one or more conditions before the order can be submitted to the market. These conditions may activate an order if the price is above or below one price level, or before or after a specified time. A stop order, usually used by traders to close a position at target price or at maximum loss, is a conditional order that activated a market order when the market price reaches a point.

Limit if-touched order is a standing limit order but it remains hidden from the order book until the the price reaches a specified price level.
Investors that need to submit large volume orders may wish to hide the full size of their orders to avoid adverse behaviours from other market partecipants given by the big quantity displayed. To overcome this problem, they can use iceberg orders, that submit a specified portion of the total size. After this part is executed, another part of the total order is submitted until all the quantity is executed.
To get a more precise empirical investigation of the relationship between the market response and the microstructure activity, is not only important to take into account the trading volume, but also to make a buy-sell classification of the ordeflow.
There are different trade classification algorithms: Tick Rule, Lee-Ready Algorithm and Bulk Volume Classification.
The Tick Rule is a level-1 algorithm that classifies a trade as buy if the trade price is higher than the price of the trade before, called 'uptick'. Conversely, if the trade price is lower than the price of the trade before it is considered a seller-initiated trade, called 'downtick'. In the event where the price is equal, it is called 'zero-tick trade' and it needs to look the closest prior trade price to attribute a zero-uptick trade or zero-down trade.
This kind of classification requires only trade data without the need to have the current bid and ask quote.
The Lee and Ready's algorithm [8] is a level-2 algorithm and it is a more complex trade classification. It needs the TAQ Data (Trades and Quotes, trade price, best bid and best ask price) to assign a feature to the trade.
If a trade occurs in the mid-point between bid and ask, the classification works using the tick rule. Instead, it uses a quote rule that assigns the

sign to the trade if it has an execution price above or below the midpoint, classifying buyer-initiated or seller-initiated feature.

The bulk volume classification is a probabilistic method that assigns the sign of trading volume aggregating the trading intervals into bars composed of equal size of volume. For each bar, a fraction of volume is determined to be buyer-initiated, while the seller-initiated amount is given by the difference between the total volume and the buyer-initiated volume. The formula of the BVC algorithm takes in account the standardized price change for the reason that if the last trade price of a volume bar increases respect to the last trade price of prior volume bar the buyer-initiated volume increases.

The feature of limit orderbook that is the presence of bid and ask price raises an important issue for the empirical market microstructure research and for who is involved in working with high frequency data: the bid-ask bounce.

The first problem is to determine the true price of the security and the presence of two prices at the same moment makes it tricky.

In some researches or in the phase of planning of a trading strategy the price used is the mid-quote between ask and bid.

The second problem, illustrated by Roll (1984), is the impact of bid-ask bounce on the time series of returns. He proved that this feature induces negative autocorrelation in observed returns.

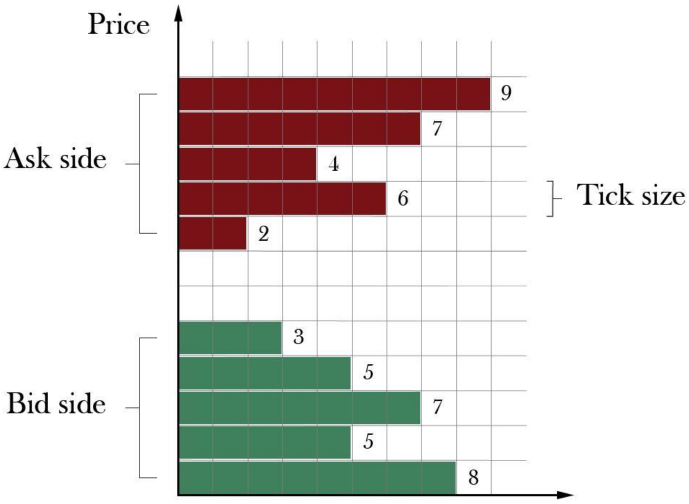


Figure 2.1: Limit Order Book

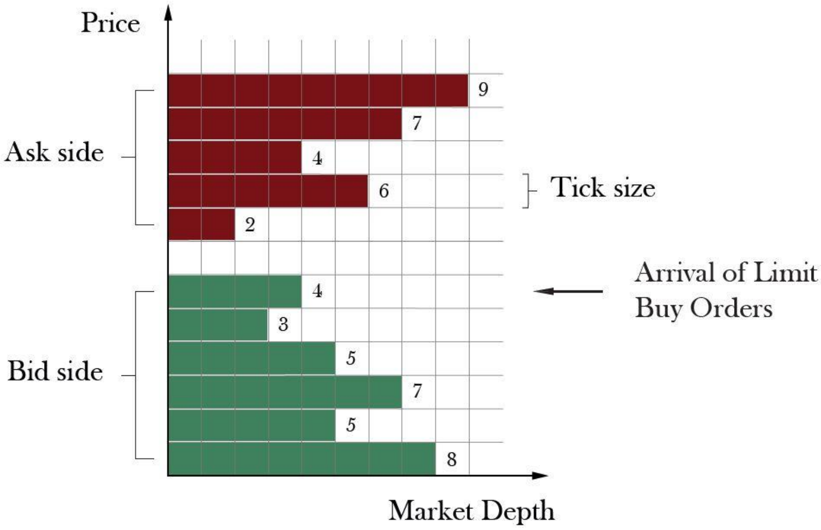


Figure 2.2: Arrival of Limit Buy Order on bid side

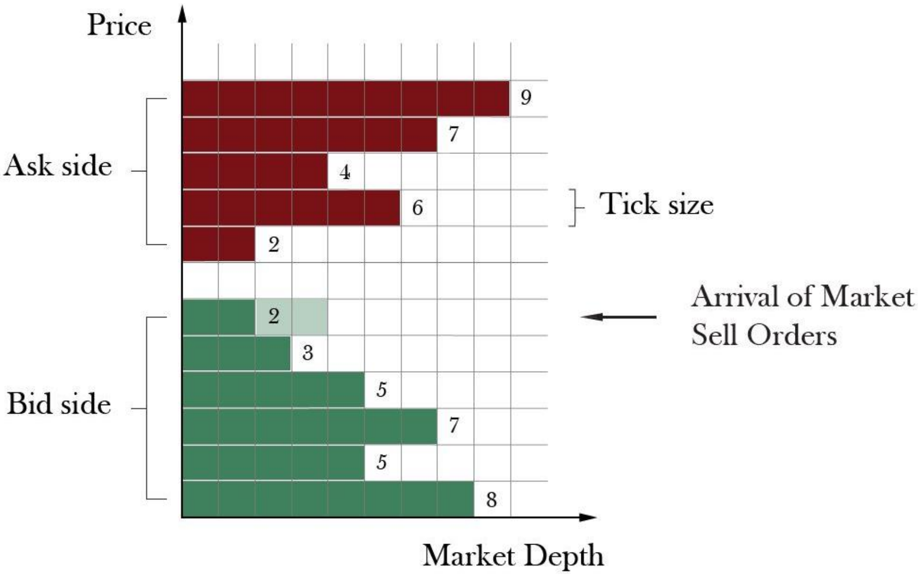


Figure 2.3: Arrival of Market Sell Orders executed on the best bid

2.4 Orderflow Imbalance

The limit order book contains all the information available to investigate how the events of submission or cancellation of limit orders and arrival of market orders influences the securities returns.

In the market microstructure literature some authors focus their empirical studies modeling the orderbook events through a variable called 'Order flow Imbalance'. The orderflow imbalance variable captures the events on the orderbook and it is shown that has an explanatory power of the traders' intentions.

Chordia and Subrahmanyam [4] in their study developed a trading strategy based on the order imbalance. They focus on the daily relation between the order imbalance and stock returns. They compute the order imbalance variable signing every trade with the Lee and Ready algorithm. Bringing together all the trades, they obtained the daily order imbalance for each stock of the sample. Moreover, they defined two variables as order imbalance: OIBNUM, as number of buyer-initiated trades minus the seller-initiated trades scaled by the total number of trades, and OIBVOL, as the buyer-initiated dollar volume less the sell-initiated dollar volume scaled by the total dollar volume. The order imbalance exhibit a positive autocorrelation of the time-series and the authors explained this results supporting the idea that traders sent their orders splitting them over time to minimize price impact. This also entails a positive intertemporal correlation between the order imbalance and price changes. The imbalance in terms of number of transactions exhibited a stronger autocorrelation respect the dollar imbalance. The time-series return regression used market adjusted open-to-close returns in order to reduce cross-correlation in error term across stocks. The contemporaneous imbalance and four lags of order imbalance were used as explanatory variable.

The results showed that about 75% of the coefficients of the first lag were positive and statistically significant, instead the latter coefficients were smaller or negative. Given these results, they developed a trading strategy that buys a stock at ask price at open and close the position selling at bid price at close if the previous day's order imbalance was positive. The results present a statistically significant daily average return of 0.09% for the entire sample of stocks.

Another important study that deserves to be mentioned is made by Rama Cont [5], 'Order book price events', where a more complex and complete form of order imbalance was used to explain the price action, called 'Order flow Imbalance'. It represents the difference between the order flow at the best bid and the order flow at the best ask taking in account the changes in size of these first two levels. In this way, the market orders are also taken in account because a market sell order decreases the bid size and conversely a market buy order decrease the ask size. The computing of order flow imbalance variable is made by two part: one represents the events related to a buy pressure and the other one is related to a sell pressure. The buy pressure is made by the cancelation of a limit order on the ask, a submission of a limit order on the bid and the arrival of a market buy order.

On the other hand, the sell pressure is made by the cancelation of limit order on the bid, a submission of a limit order on the ask and the arrival of a market sell meritaorder.

The limit orders events can generate three situations and it is important to identifying them in order to highlight the different behaviour of the orderflow function.

If the bid price is lower than the previous bid price, this means that a market sell order arrived and filled all the previous buy limit orders or the previous limit order were cancelled. If the bid price is equal to the previous bid price, the function takes in account the difference between the current quantity on the bid and the previous quantity on the same level. If the bid price is greater than the previous bid price, this means that there was a submission of buy limit orders at a higher level. They divided the data in interval of 30 seconds and run the regression for clusters of 30 minutes.

They found, with an OLS regression, that the coefficients of the Order Flow Imbalance, considered as price impact coefficients, explaining the Price change be positive and the results were statistically significant in 98% of sample. Moreover, it was regressed the price change with Trade Imbalance, computed as the difference between the buyer-initiated trades and seller-initiated trades, but the results were much weaker than the regressions with Order Flow Imbalance as explanatory variable.

From this study conducted by Rama Cont that highlights a positive relation between

Another prominent study was done by Huang [7] that studied the relations between the order imbalance and the return and order imbalance and price volatility for a sample of 150 stocks in the tender offer announcement day.

The order imbalance is computed for three different timeframes, 5-10-15 minutes and the trade assignment is computed using Lee and Ready algorithm.

First, they employ a multi-period regression model to find the coefficients of contemporaneus and lagged order imbalances. These coefficients found in this studiy are negative for lagged order imbalance on current returns. They interpretate the results attribuiting the reason to the the market makers' behaviour to accomodate big price impacts caused by traders that react to the news. Consequently, the inventory levels of market makers are a fundamental aspect to consider to attribuite an interpretation to returns in these days. Instead, they found the coefficients of the contemporaneus order imbalance be positive and significant for the 59.30% of the results. As result of that, they developed a trading strategy based on this findings. The strategy buys a share at trading price when the order imbalance is positive and sell it at the trading price when it turns negative. The daily strategy returns outperform the daily returns of the underlying.

Another interesting aspect studied in this paper was the order imbalance-volatility relation for the three intervals. They employ a Garch (1,1) model to estimate it and they found that the order imbalance has negative impact on price volatility. As the time interval getting longer the order imbalance effect on volatility became weaker. The explanation in this case is given by the fact that the market makers have an obligation to reduce price volatilities and they reduce the effect of the news induced by the arrival of orders by traders. They face the variability of the stock after the release of the news increasing the bid ask spread in order to make more profit.

From the empirical studies emerges a strong interest from the academic community to study the impact on the prices of the orderbook events. The focus is on the orderflow imbalance considered one of the main driver for the price change. It captures the traders' intent to get in to a long or short position and consequently has an explanatory power to anticipate short-term movement.

2.5 High Frequency Trading strategies

The technological process enable some professional market partecipants to deploy the information flow in the microstructure and build algorithmic trading strategies characterised by high speed of execution and rapid-decay alpha, defined 'High Frequency Trading strategies'. The introduction of these sophisticated and low latency machines raised a strong interest in the academic researches and lead the financial regulator to face this new challenge.
The first definition was given in June 2012 by the U.S. Commodity Futures Trading Commission (CFTC) as follow:

"High-frequency algorithmic trading is a trading technique characterised by describing these strategies as algorithms for decision making, order initiation, generation, routing, or execution for each individual transaction without human direction. The high frequency trading firms benefit from low-latency technology that is designed to minimize response times, including proximity and co-location services, high speed connections to market for order entry and high message rate (orders, quotes or cancellations)."

The co-location services allow the high frequency trading firms to gain a speed advantage on the other partecipants thanks to the trading servers situated in the same facility of exchange's matching servers. Given this advantages of milliseconds, the information contained in this small time interval can be extremely remunerative if exploited with the right strategies and the the right models
The HFT become therefore some of the most important market partecipants, especially in the microstructure environment.

Their activity can be divided in two main ways: aggressive trading and passive trading.

Latency Arbitrage

Latency Arbitrage is the first strategy, born thanks to the use of the fastest technologies that allow these big partecipants to receive the signals from different markets before everyone else. Given the Law of One Price, when the same security is listed in more than one exchange, the difference between prices among different exchange produce an arbitrage opportunity for the fastest trader that realizes that. The fiber that the big players of Wall Street adopted allow them to have speed advantage and exploiting that. For example, if in one market the security has its ask price quoted at 74 usd and in another one the same security has its bid price quoted at 74.10 usd, the opportunity is exploited by an algorithm is to buy at ask price at 74 usd and sell it at the bid price of 74.10 usd.

Spread Capturing

This is a passive strategy and the high frequency trader acts as a market maker. It provides liquidity in the order book when is more convenient and it generates profits when the incoming market orders hit the pending limit orders. The profit is given by the spread between the best ask and the best bid. It is particularly profitable in high liquid markets when the probability of being executed is higher

than less liquid markets. Obviuosly, this strategy is subject to inventory and adverse selection risks linked to the market making activity risks.

Quote Matching

The arrival of limit orders can cause short term price movements and the High Frequency Trader is able to detect the market impact, takes advantage of this information and takes profit after few ticks of movements. The success of this strategy relies on HFT's ability to understand which limit orders will generate positive or negative market impact.

Spoofing

This strategy is considered market manipulation because it modify artificially the other trader's view in order to to obtain a better execution price. The high frequency trader submits a big limit order in the orderbook in order to induce other traders to get a distorted information from the orderbook. For example, the HFT intents to buy will send a sell limite order, with the intent of not being executed, and a hidden buy limit order. The other traders, seeing a big sell order will sell consequently causing a downturn in price. In this way, HFT will be able to get a better purchase price. After the execution, the pending sell order will canceled. This strategy is illegal in the United States.

Momentum Ignition

In this strategy, the High Frequency Trader sends market orders in order to get a large position. This behaviour induce other traders that see the incoming flow to take a position on the same side emphasizing the price movements. The HFT that have bought or sold at the beginning of the move, is able to liquidate the position and get some profits. This strategy is most profitable if the price gets to levels where a lot of stop orders are positioned. The activation of those orders accelerates the movement and increase the potential profit of the strategy.

Pinging

The way to benefit from other liquidity provision is the pinging strategy or Liquidity Detection. The HFT aims to discover hidden orders or price levels that trigger orders submitted by algorithmic strategies and to exploit the price movements consenquently to these incoming flow. Once in the price level, for example a bid price, is detected the presence of a hidden order (called also 'Iceberg') the HFT will buy just above in order to have a risk-reward trade favourable.

2.6 Inefficiencies of market microstructure

The market microstructure presents some cases of inefficiencies that cause large price movement and increments the short term volatility or deviations from the fair price. These efficiencies can be the reasons of losses or can be the alpha of trading strategies, especially for High Frequency Trading strategies.
The most famous kind of inefficiency is the so called 'Flash Crash', an event in which there is a deep withdrawal of limit orders on the bid size followed by a big flow of sell market orders, executed at prices much lower than expected. This causes a strong lowering of prices and a negative return for the security. The most famous case of flash crash happened the 6th May of 2010 on the Dow Jones Futures.

Some authors, Christensen and Kolokolov [2], analyzed the microstructure features such as trading volume, market depth, bid ask spread during the flash crashes on a sample of French Stock. They found an expected behaviour of these variables. While the price was dropping faster the trading volume increased, due to the fact that the market partecipants reacted immediately to the 'Panic Selling'. The market depth and bid ask spread increased during this event, due to the market makers' behaviour to protect their inventory levels from the high volatility.
The widening of the spread causes some 'jumps' on the orderbook, the price goes from on level to the next without passing for every tick of the movement.

Figure 4: Jump of the order book on the Future Dax Dec 19 at 11.932,5 – Volcharts Trading Platform

The figure x.x represents 1 minute chart of Future Dax. The grey square indicated by the arrow is an inefficiency of the market. Given the auction process that characterize the way which the trading activity is carried out,

all the price levels have to be traded before passing across. If the market let an empty space, it means that an auction has not been done and the price moves jumping some price levels. There is an increase in the spread between bid and ask and some sell market orders hit the bid. They execute all the buy limit orders and the best ask moved down.
There were not buy market orders that hit the ask and consequently the price moved down. Letting undefeated the first ask the first levels of the book, the price moved down letting the auction at the precedent price level incompleted.

Figure 5: Jumps of the order book on the Future Dax Dec 19 at 11.819,5 11.816 11.814 – Volcharts Trading Platform

The figure x.2 shows the same kind of inefficiences. The price, before going up, did not pass all the price levels, it let some price levels not traded causing an inefficiency in the market.

Figure 6: Flash Crash on Future Nasdaq Dec 19 – Volcharts Trading Platform

The last image, figure x.3, shows the most famous form of inefficiency, the "flash crash".

The chart represents the Nasdaq futures on 1 minute chart. The limit orders standing on the bid size were fastly removed leaving some price levels on the order book empty. All the sell market orders hit limit orders on the bid side much lower than expected (the investor who send a limit order is a price maker, who send a market order is a price taker). This fast removal caused a strong price downturn followed by a sharp recovery in the quotations.

CHAPTER 3

Empirical studies and trading strategy

3.1 Data and variables

In this work, the data used for the empirical analysis are relative to the Dax Futures (FDAX, with expiration on September 2019) quoted on EUREX Exchange. The Future Dax contract has a value of 25 EUR for each point and the tick value is 0.5 (12,5 EUR for one tick).

These data are downloaded by IQFeed Data Feed in comma separated values (CSV) format and they have tick-by-tick timestamp so it was recorded every trade with a microsecond precision. The information provided are:

- Trade identification number
- Trade time
- Trade volume
- Trade price
- Current Bid price
- Current Ask price

It has been taken into account 62 days of trading from the front future contract, being the most liquid and traded contract and so the order flow effects can be clearer and better than future contracts with longer expirations.

The trading hours of the contract is from 2.00 A.M. to 10.00 P.M. but the index is moved from the underlying stocks from 9.00 A.M. to 5.30 P.M. In the statistical study, it was considered only the data between the

9.00 A.M. and 5.00 P.M. in order to study only the part of the day where the liquidity is more present and the trading activity is intense.

Thanks to the granularity of the TAQ data (Trades and Quotes) it was possibile to compute the mid-quote between bid and ask price, variable used in computing the order flow variable with the Lee and Ready's algorithm.

It was assigned a positive sign to the trade volume if the trade occured above the mid-quote and negative sign if the trade occured below the mid-quote.

I separated all the data in days and I divided that in 30 seconds, 1 minute, 2 minutes and 5 minutes intervals in order to study how the order flow, used as explicative variable, has effect on the return of the instrument at high frequency level.

The percentage returns were computed as $R_t = (P_t - P_{t-1})/P_{t-1}$ considering the close price of the intervals. Consequently, I deleted the first element of the order flow vector in order to work with two vectors with the same length. To obtain an order flow variable for the intervals, all the trades occured in the same time interval are summed. If the result is positive, it means that there was more buying pressure on the price (more and bigger buy market orders than sell market orders). The data used for the empirical study are the returns and the orderflow. In the trading strategy, there is an additional variable used as trigger for the trade: the trading volume. The trading volume gives the information about the intensity of trading activity and associated with orderflow can be useful to exploit the short-term momentum.

3.2 Theory and statistical methods

Quantile Regression

To investigate the features of a random variable we need to know the moments of its distribution. These measures give each a singular information on the shape of the probability function and they can be divided for the charateristic that explain. The location measures are the mean and the median and they characterize the average and the center of the distribution. The dispersion measures as the variance indicate how far the observations are spread out from their average value. The third moments as skewness and kurtosis show the asymmetry of the distribution about its mean and the thickness of the tails of the distribution.

The conventional methods that describe the behaviour of y conditional on x, are the Least Square, that minimizes

$\sum_{t=1}^{T}(y_t - x_t'\beta)^2$ to obtain $\widehat{\boldsymbol{\beta_t}}$

and the Least Absolute Deviation, that minimizes

$\sum_{t=1}^{T}|y_t - x_t'\beta|$ to obtain $\widehat{\boldsymbol{\beta_t}}$

These methods approximates respectively the conditional mean and the conditional median of y given x.

They assume that the distribution of the dependent variable is not affected by the values of the covariates but only the central tendency of the dependent variable. The estimate provides a partial description of the relationship bewteen variables and it may be more useful to know the relationship at different level of the conditional distribution of the dependent variable. To investigate about the potential effects of the independent variable on the shape of the probability distribution we need a regression method robust to changes in the shape of distribution. The idea to model the quantiles of a conditional distribution $f(y|x)$ of a variable y, given the covariates x, was developed and introduced by Koenker and Basset in 1978. Quantile regression provides a more complete description of the distribution of the conditionated quantiles of y, given the independent variable x.

This statistical tool is especially useful when the relationship between the variables is asymmetric, for example at the tails of the distribution, and it cannot be captured properly by the OLS method.

Before moving forward in the model description, it is essential to define a statistical value, called Quantile, that divides the distribution in equal parts.

Quantiles are a position indexes distribution, the τ quantile is the value such that

$P(Y \leq y) = \tau$ for any $0 < \tau < 1$.

Starting from the cumulative distribution function (CDF)

$$F_Y(y) = F(y) = P(Y \leq y)$$

we define the quantile function of Y for the τth quantile as the inverse of the CDF

$$Q_Y(\tau) = Q(\tau) = F_y^{-1}(\tau) = \inf\{y \in \mathbb{R} : F_y(y) \geq \tau\}.$$

The quantile that divides the distribution in two simmetrical parts is the median.

Hao and Naiman (2007) define the quantiles as a center of the distribution c, obtained minimizing a weighted sum of absolute deviations. The quantile q is:

$$q_\tau = \underset{c}{\operatorname{argmin}}\, E\,[\rho_\tau(Y - c)]$$

Where ρ_τ is a loss function defined as:

$$\rho_\tau(y) = [\tau - \mathrm{I}(y < 0)]y = [(1 - \tau)\,\mathrm{I}\,(y \leq 0) + \tau \mathrm{I}\,(y > 0)]|y|$$

This is an asymmetric loss function; it is a weighted sum of absolut deviation where a weight (1- τ) is given to negative deviations and a weight τ is used for positive deviations.

The optimization problem, in a discrete case, becomes:

$$q_\tau = \underset{c}{\operatorname{argmin}}\, \{ (1 - \tau) \textstyle\sum_{y \leq c} |y - c| f(y) + \tau \sum_{y > c} |y - c| f(y) \}$$

In the continuous case, it becomes:

$$q_\tau = \underset{c}{\operatorname{argmin}}\, \{ (1 - \tau) \int_{-\infty}^{c} |y - c| f(y) d(y) + \tau \int_{c}^{\infty} | y - c\,| f(y) d(y) \}$$

This formulation is useful to go straight in the application of quantile regressive models.

In reference to the i-th sample unit, it is possible to write a linear model:

$$y_i = x_i^T \beta_q + e_i \quad \text{with i=1,…,n}$$

Where n is the sample numerousness, y_i is the response variable in the i-th unit, x_i is the vector of p explicative variables with first component equal to 1 to ensure the presence of the intercept, β is the vector of coefficients of the model and e_i is the error term.

The assumption $Q_q(e_i|y_i, x_i) = 0$ is a necessary condition to get the errors distribution centered on the q-th quantile ($0 \leq q \leq 1$).

The linear model for the q-th quantile of response variable y_i conditioned to the explicative variables x_i can be written as:

$$Q_q(y_i|\, x_i) = x_i^T \beta_q$$

The parameter vector $\widehat{\boldsymbol{\beta q}}$ is estimated minimizing the following object function

$$\widehat{\boldsymbol{\beta q}} = argmin_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_q \ (y_i - x_i^T \beta_q)$$

where ρ_q(u) is the loss function

$$\rho_q(u) = \begin{cases} qu & u < c \\ (1-q) & u > c \end{cases}$$

$$\widehat{\boldsymbol{\beta q}} = \underset{\beta_i}{\arg\min} \sum_{i:y_i \geq x_{i\beta_q}^T} q \mid y_i - x_i^T \beta_q \mid + \sum_{i:y_i \leq x_{i\beta_q}^T} (1-q) \mid y_i - x_i^T \beta_q \mid$$

Computing the parameters β for every quantile q can be expressed as a problem of linear programming. They can be computed with the simplex algorithm or the interior points method when the sample size is big. Quantile regression presents some advantages respect to the conventional methods. The Ordinary Square Estimator can be inefficient if the errors are not normal and the presence of outliers can significantly affects the results of the linear regression method. The quantile regression estimator is more robust because the model avoids some assumptions on the errors distribution.

An important aspect of this method is the equivariance. If there is the need to reparametrize the data to study the effect from a different perspective, the estimates change in the same in way of the reparametrization leaving the results invariant.

The estimator has four equivariance property:

- Scale Equivariance:

  for any $\alpha > 0$ and $\tau \in [0,1]$

  $\widehat{\boldsymbol{\beta}}(\tau; \alpha Y, X) = \alpha\widehat{\boldsymbol{\beta}}(\tau; Y, X)$

  $\widehat{\boldsymbol{\beta}}(\tau; -\alpha Y, X) = -\alpha\widehat{\boldsymbol{\beta}}(1-\tau; Y, X)$

- Shift Equivariance

  for any $\gamma \in R^k$ and $\tau \in [0,1]$

  $\widehat{\boldsymbol{\beta}}(\tau; Y + X\gamma, X) = \widehat{\boldsymbol{\beta}}(\tau; Y, X) + \gamma$

- Equivariance to reparametrization of design

  Let A be any $p \times p$ nonsingular matrix and $\tau \in [0,1]$

  $\widehat{\boldsymbol{\beta}}(\tau; Y, XA) = A^{-1}\widehat{\boldsymbol{\beta}}(\tau; Y, X)$

- Invariance to monotone transformations

  If h is a non decreasing function on R

  $h\,(Q_{Y|X}(\tau)) = Q_{h(Y)|X}(\tau)$

3.3 Empirical studies

The aim of this work is to detect the effect of the orderflow at different level of the distribution of returns. Firstly, we used the percentage returns of the underlying over 30 seconds, 1 minute, 2 minutes and 5 minutes period, denoted as *r*, as dependent variable. The variable OrderFlow, denoted as *OF*, is used as independent variable.

The tables describes summary statistics of the returns, the orderflow and the trading volume for the four time intervals.

30 seconds time interval			
	Returns	Orderflow	Volume
Mean	0.000	0.02	75
Median	0.000	0.00	55
Standard deviation	0.0002	20.60	73.48
Skewness	0.21	0.09	3.85
Kurtosis	23.52	14.65	32.7
1^{st} Percentile	-0.000675	-58	7
5^{th} Percentile	-0.000365	-31	14
25^{th} Percentile	-0.000121	-9	32
75^{th} Percentile	0.000121	9	94
95^{th} Percentile	0.000363	31	204
99^{th} Percentile	0.000659	59	367

Table 3.1: Summary statistics of return, orderflow and trading volume in the 30 seconds time interval

1 minute time interval			
	Returns	Orderflow	Volume
Mean	0.000	0.06	98
Median	0.000	0.00	115
Standard deviation	0.0004	29.22	111.92
Skewness	0.7	0.05	3.22
Kurtosis	24.87	10.42	22.5
1^{st} Percentile	-0.000954	-81	21
5^{th} Percentile	-0.000517	-45	35
25^{th} Percentile	-0.000165	-14	70
75^{th} Percentile	0.000164	16	188
95^{th} Percentile	0.000511	45	390
99^{th} Percentile	0.000943	83	649

Table 3.2: Summary statistics of return, orderflow and trading volume in the 1 minute time interval

2 minutes time interval			
	Returns	Orderflow	Volume
Mean	0.000	-0.11	302
Median	0.000	0.00	239
Standard deviation	0.0005	41.65	234
Skewness	0.8	0.11	2.79
Kurtosis	21.52	7.4	18.2
1^{st} Percentile	-0.00133	-112	54
5^{th} Percentile	-0.00072	-65	82
25^{th} Percentile	-0.00024	-21	152
75^{th} Percentile	0.00024	21	377
95^{th} Percentile	0.00072	65	741
99^{th} Percentile	0.00130	116	1193

Table 3.3: Summary statistics of return, orderflow and trading volume in the 2 minutes time interval

5 minutes time interval			
	Returns	Orderflow	Volume
Mean	0.000	-0.31	759
Median	0.000	0.00	624
Standard deviation	0.0008	66.36	511
Skewness	0.48	0.02	2.29
Kurtosis	14.90	5.8	13.3
1^{st} Percentile	-0.00217	-181	166
5^{th} Percentile	-0.00115	-105	233
25^{th} Percentile	-0.00036	-36	416
75^{th} Percentile	0.00010	35	957
95^{th} Percentile	0.00034	105	1712
99^{th} Percentile	0.00213	187	2664

Table 3.4: Summary statistics of return, orderflow and trading volume in the 5 minutes time interval

The returns over short time intervals are close to zero for the sample considered. This result means that the intraday returns does not change substantially over time. The orderflow presents a similar feature; the mean and the median of the buying and selling pressure are close to zero, meaning that the market, after strong movements, comes back to its balance. This result is supported by the activity of the HFT that often trade in extreme situations. Moreover, the returns, orderflow and volume have high kurtosis, meaning that there is a good number of extreme values over the sample and consequently their distribution has bigger tails than a normal distribution. In particular, the orderflow and the trading volume have higher kurtosis for the smaller time intervals. This finding highlights that the market activity has less severe observation respect its distribution in higher time frames. Contrastly, at lower frequencies there are more extreme values.

The Figure 3.1 represents the autocorrelation function of the orderflow variable for the four time intervals considered in the study.

These time-series exhibit a particular feature: the lags-1 are positive and significant. This result indicates that positive orderflow (buying pressure) is followed by positive orderflow and viceversa. This characteristic is more accentuated for the smaller time intervals. The 30 seconds, 1 minute and 2 minutes time-series have also the second and the third lags positive and significative.

The 5 minutes time-series instead has only the first lag positive and significative. These results explain that the orderflow has a short term persistence, more evident as we move through smaller time intervals.

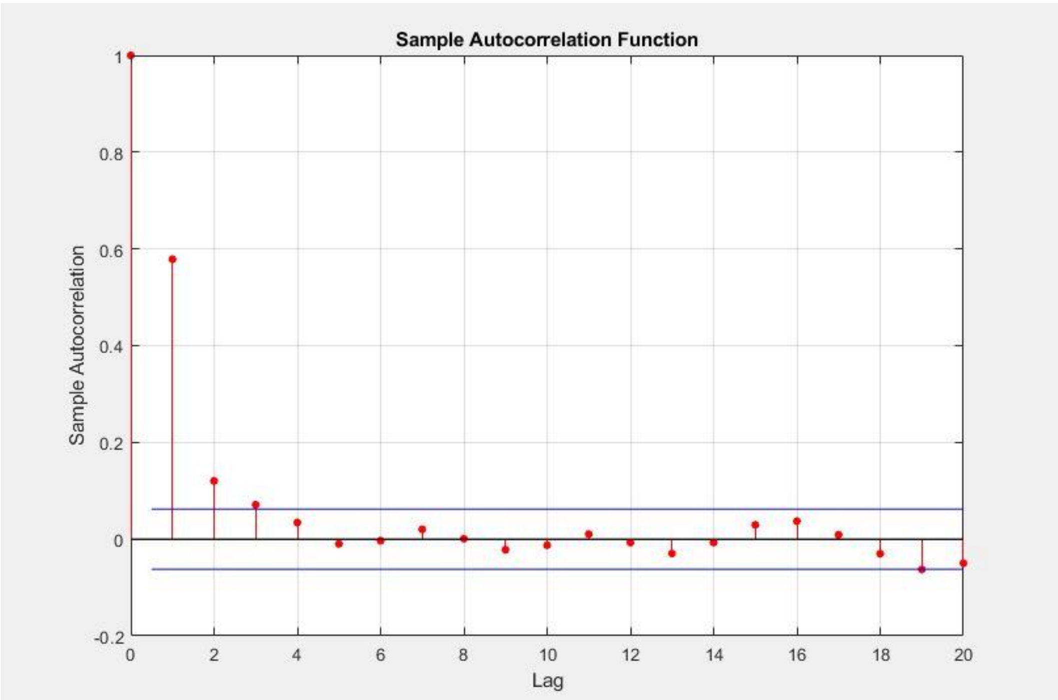


Figure 3.1: Autocorrelation function of the orderflow for the 30 seconds time intevals

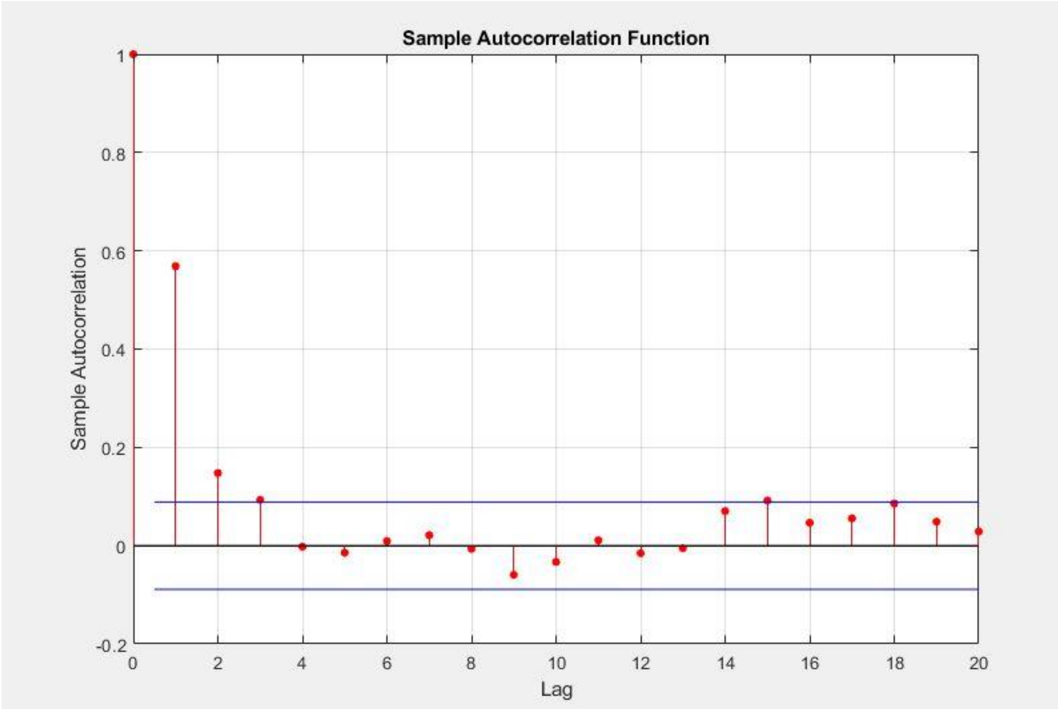


Figure 3.2: Autocorrelation function of the orderflow for the 1 minute time intevals

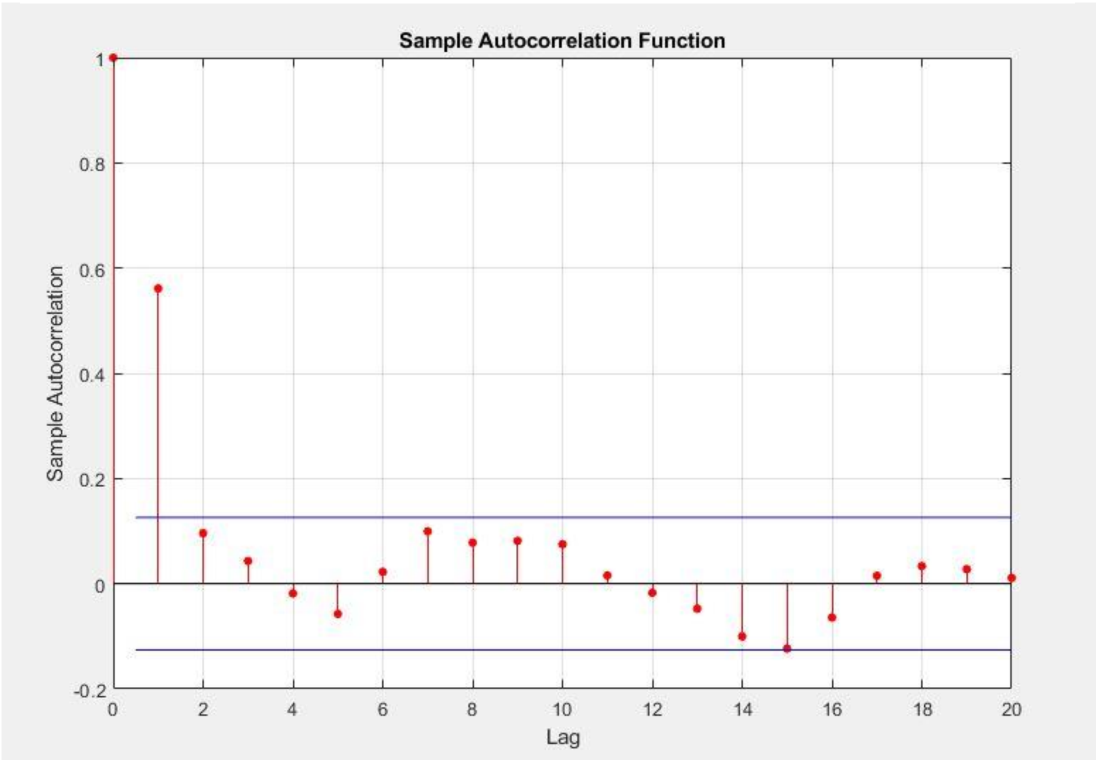


Figure 3.3: Autocorrelation function of the orderflow for the 2 minutes time intevals

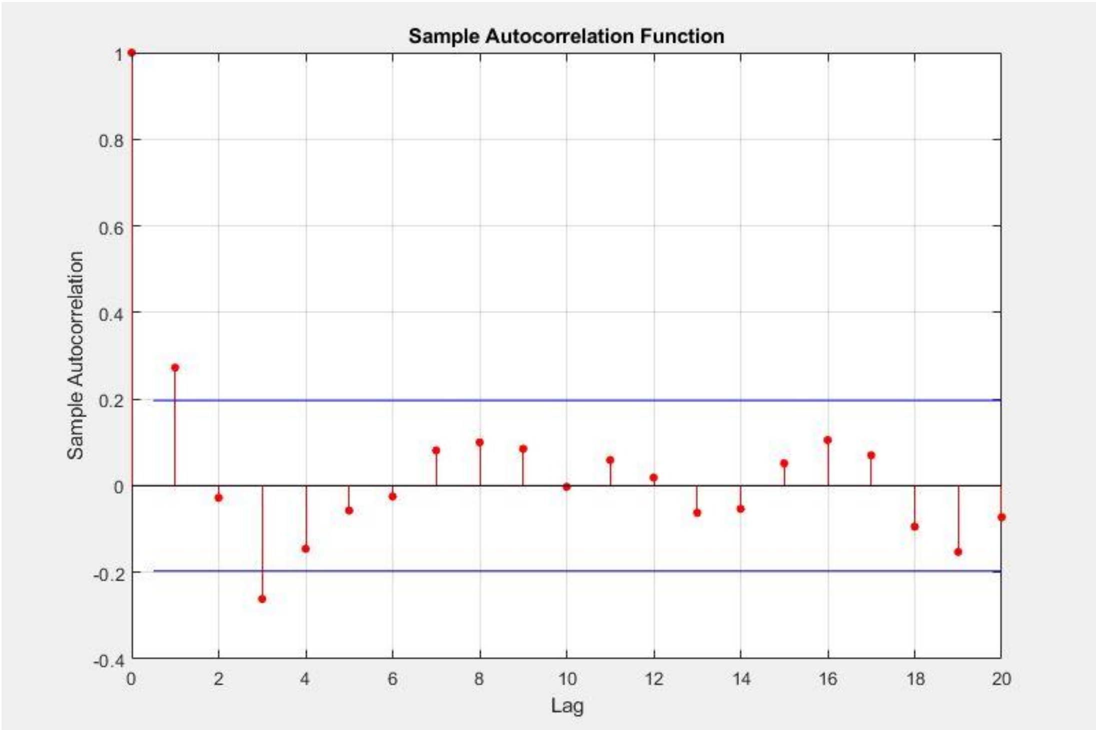


Figure 3.4: Autocorrelation function of the orderflow for the 5 minutes time intevals

Moreover, we computed the first difference of the orderflow variable and I found that it has a significant lag-1 negative autocorrelation for every time intervals considered. These results are consistent with the results by Chordia [3].

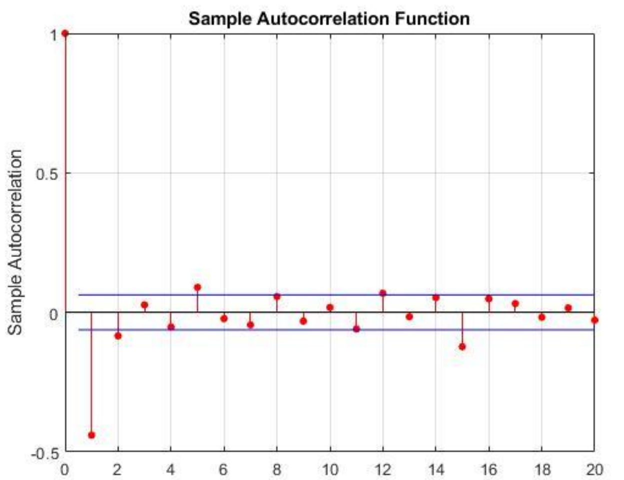


Figure 3.5 Autocorrelation function of the first-difference of the orderflow time-series for the 30 seconds time interval

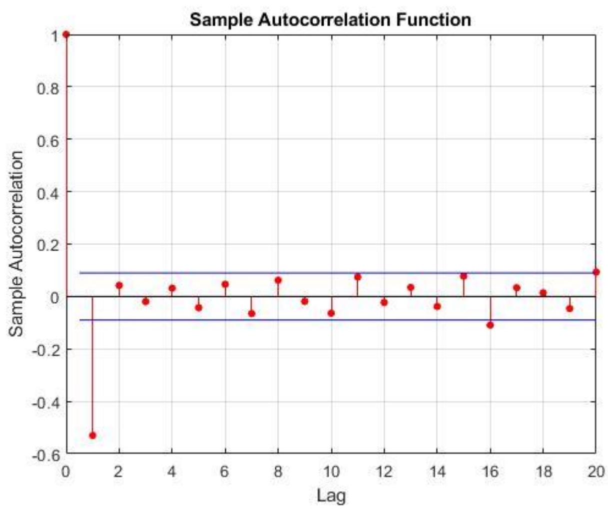


Figure 3.6 Autocorrelation function of the first-difference of the orderflow time-series for the 1 minute time interval

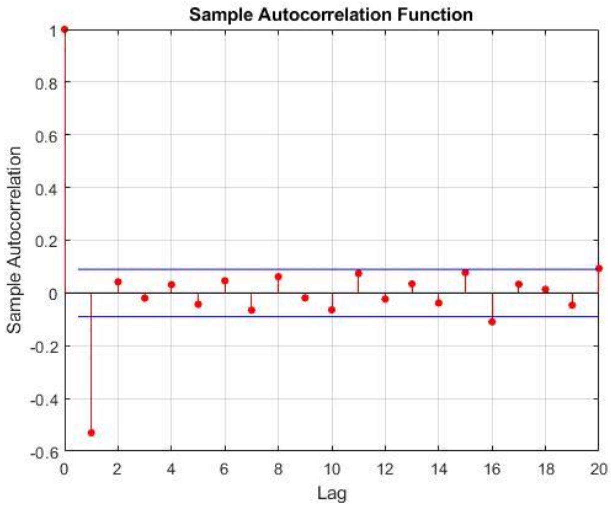


Figure 3.7 Autocorrelation function of the first-difference of the orderflow time-series for the 2 minutes time interval

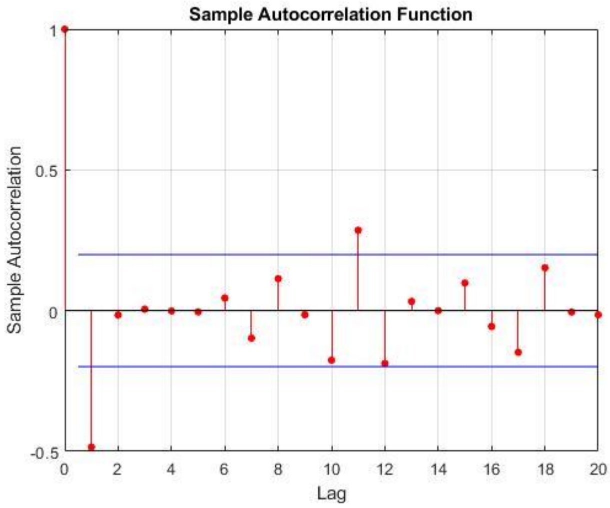


Figure 3.8 Autocorrelation function of the first-difference of the orderflow time-series for the 5 minutes time interval

Furthermore, I found a positive correlation between contemporaneous return and orderflow variable. These interesting findings indicate that more we move through smaller time intervals, stronger the correlation between these two variables.

The correlation for the 30 seconds, 1 minute, 2 minutes and 5 minutes time intervals is illustrated in the next table.

	30 seconds	1 minute	2 minutes	5 minutes
Correlation	0.5739	0.5513	0.5320	0.4781

Table 3.5: Correlation between orderflow and market returns

Before I move in the quantile regression, I fit a linear regression on the returns and the orderflow

$$r_t = a + \beta_t OF_t + e_t$$

to highlight the difference between the results of the two statistical models and the benefits given by studying the relation for different levels of the distribution. This linear model, given the positive values of β describes the positive relation between orderflow and returns. In addition, also this model gives a R^2 higher as we move through smaller time intervals.

Linear Regression				
	30 seconds	1 minute	2 minutes	5 minutes
Intercept	-0.000985	-0.00025	-0.00047	-0.00233
Beta	0.006912	0.00665	0.00637	0.00553
R^2	0.329	0.304	0.283	0.229

Table 3.6: Results of linear regression

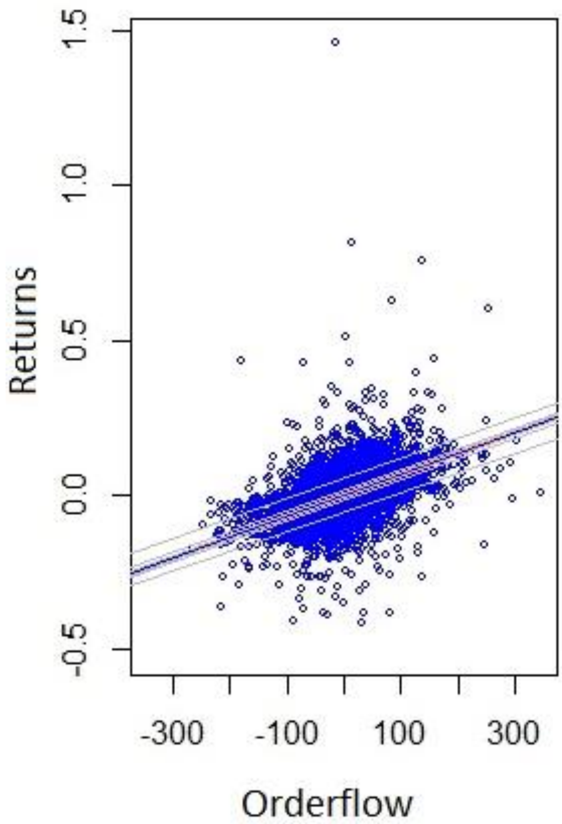


Figure 3.9: Linear regression plot between orderflow and market returns for the 30 seconds time intervals

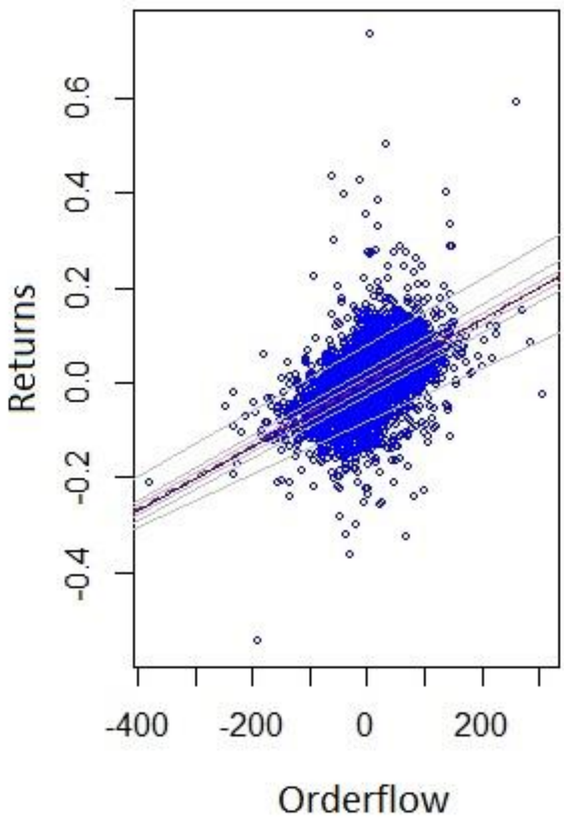


Figure 3.10: Linear regression plot between orderflow and market returns for the 1 minutes time intervals

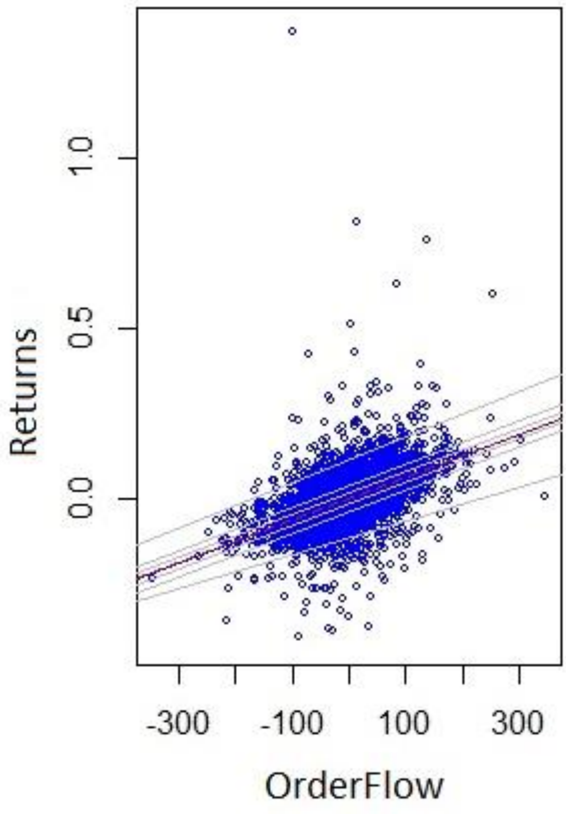


Figure 3.11: Linear regression plot between orderflow and market returns for the 2 minute time intervals

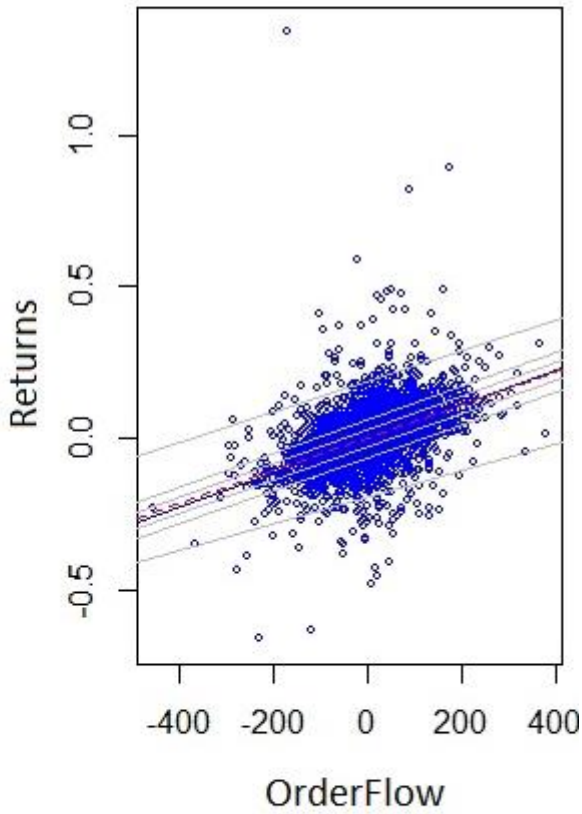


Figure 3.12: Linear regression plot between orderflow and market returns for the 5 minute time intervals

The order flow variable, denoted as *OF,* is used as independent variable. Having obtained a first image of the relation between the orderflow and the market returns, the effect was studied introducing an other regressor, the trading volume VOL , in order to have a more complete picture among these variables. A multiple linear regression has been used to discover the potential explanatory power. It can be defined as:

$$r_t = a + \beta_1 OF_t + \beta_2 VOL_t + e_t$$

This statistical method was applied to all the four time intervals. the coefficients found of trading volume are positive and significative different from 0 in all samples. The R^2 are slightly higher than the previous regression, as expected.

Multiple Linear Regression				
	30 seconds	1 minute	2 minutes	5 minutes
Intercept	-0.00016	-0.00080	-0.00170	-0.00430
B_1	0.00069	0.00066	0.00042	0.00023
B_2	0.00038	0.00011	0.00007	0.00004
R^2	0.367	0.349	0.316	0.248

Table 3.7: Results of Multiple Linear Regression

Market returns have a clear feature that characterize them: they exhibit heteroscedasticity, or autocorrelation in the squared residuals, meaning that the variance is a dynamic process. To identify such charateristic, an Engle's Arch (Autoregressive Conditional Heteroscedasticity) test has been used on the returns time-series of all time intervals of the sample. Starting from the basic time-series theory, the ARCH is illustrated as follows:

$$y_t = \mu_t + \epsilon_t$$

where μ_t is the conditional mean, ϵ_t is the innovation with mean 0.

$$\epsilon_t = \sigma_t z_t$$

where z_t is an i.i.d. process with mean 0 and variance 1 and the innovations are uncorrelated across time.

Now define the residual series

$\epsilon_t = y_t - \widehat{\mu_t}$

To test the possibility of a serial correlation between residuals we emply an Engle's Arch test

$H_a : \epsilon_t^2 = a_0 + a_1\epsilon_{t-1}^2 + \ldots + a_m\epsilon_{t-m}^2 + \mu_t$

Where μ_t is a white noise error process. The null hypothesis is

$H_0 : a_0 = a_1 = a_m = 0$

The null hyphotesis of the test is that there is no autocorrelation in the residuals time-series.

Moreover, I analyzed the market returns time-series in order to detect some arch-effect. The Table 3.8 represents, as expected, the results. For all the time intervals the market returns time-series exhibit heteroscedasticity and the volatility clustering is well shown in these outcomes.

Engle's Arch Test on Market Returns				
	30 seconds	1 minute	2 minutes	5 minutes
Test results P-value	Reject H_0 $< .001$	Reject H_0 $< .001$	Reject H_0 $< .001$	Reject H_0 $< .001$

Table 3.8: Results of Engle's Arch Test

The final empirical study of this work aims to investigate the effect of the orderflow to different levels of market returns.

In order to do that, I conduct a set of quantile regressions to examine the relationship between contemporaneous orderflow and returns. The model can be expressed as:

$$r_q = \beta_q OF_q + e_q$$

30 seconds time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.001) -0.055	(0.000) 0.000624
0.05	(0.000) -0.025	(0.000) 0.000620
0.10	(0.000) -0.016	(0.000) 0.000672
0.25	(0.000) -0.07	(0.000) 0.000678
0.50	(0.000) 0.000	(0.000) 0.000676
0.75	(0.000) 0.007	(0.000) 0.000676
0.90	(0.000) 0.016	(0.000) 0.000667
0.95	(0.000) 0.025	(0.000) 0.000659
0.99	(0.001) 0.55	(0.000) 0.000648

Table 3.9: Results of quantile regression on 30 seconds time interval

1 minute time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.002) -0.082	(0.000) 0.000556
0.05	(0.001) -0.040	(0.000) 0.000657
0.10	(0.000) -0.027	(0.000) 0.000659
0.25	(0.000) -0.012	(0.000) 0.000664
0.50	(0.000) 0.000	(0.000) 0.000669
0.75	(0.000) 0.012	(0.000) 0.000669
0.90	(0.000) 0.027	(0.000) 0.000684
0.95	(0.000) 0.040	(0.000) 0.000679
0.99	(0.002) 0.08	(0.000) 0.000696

Table 3.10: Results of quantile regression on 1 minute time interval

2 minutes time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.004) -0.113	(0.000) 0.000556
0.05	(0.001) -0.058	(0.000) 0.000657
0.10	(0.001) -0.039	(0.000) 0.000659
0.25	(0.000) -0.017	(0.000) 0.000664
0.50	(0.000) 0.000	(0.000) 0.000669
0.75	(0.000) 0.017	(0.000) 0.000669
0.90	(0.001) 0.038	(0.000) 0.000684
0.95	(0.001) 0.056	(0.000) 0.000679
0.99	(0.003) 0.114	(0.000) 0.000696

Table 3.11: Results of quantile regression on 2 minutes time interval

5 minutes time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.012) -0.194	(0.000) 0.000439
0.05	(0.003) -0.092	(0.000) 0.000580
0.10	(0.001) -0.063	(0.000) 0.000548
0.25	(0.001) -0.027	(0.000) 0.000553
0.50	(0.001) 0.001	(0.000) 0.000564
0.75	(0.001) 0.028	(0.000) 0.000552
0.90	(0.001) 0.061	(0.000) 0.000560
0.95	(0.003) 0.090	(0.000) 0.000532
0.99	(0.009) 0.188	(0.000) 0.000501

Table 3.12: Results of quantile regression on 5 minutes time interval

The Tables reports the regression results. The coefficients of the orderflow are positive and consistent in the quantile around the mean for all the time intervals.

In the 30 seconds and 5 minutes charts, it is possible to notice a reverse-U values for the orderflow coefficients. This results lead us to say that the orderflow has less explanatory power for the extreme values of the return distribution.

In the 1 minute and 2 minutes time intervals there is an interesting feature. The orderflow coefficients are higher for the return quantiles above the mean, with maximum values in the quantiles that reflect the most positive values of the entire distribution of returns. These finding leads us to say that the orderflow has stronger effect in the positive returns. These results will be exploit to build the trading strategy. In the most negative return quantiles the orderflow effect seems disappear at all the frequencies considered. Not only the coefficients have valuable information. The intercepts of the model are negative in the return quantiles below the mean and positive in the return quantiles above the mean.

These results are present in all the time intervals of the study.

Consequently, considering only above the mean part of the return distribution, it can be possible to support the idea that the orderflow variable has strong effect.

We attribute this effect to the fact that if the investors are willing to buy the Futures contract, they want to increase their long exposure in the market causing an upturn in the prices. On the contrary, the sell of a Futures contract can be done to hedge a portfolio with long exposure the underlying. The effect in this case is much weaker on the price, as shown in the results of the quantile regressions.

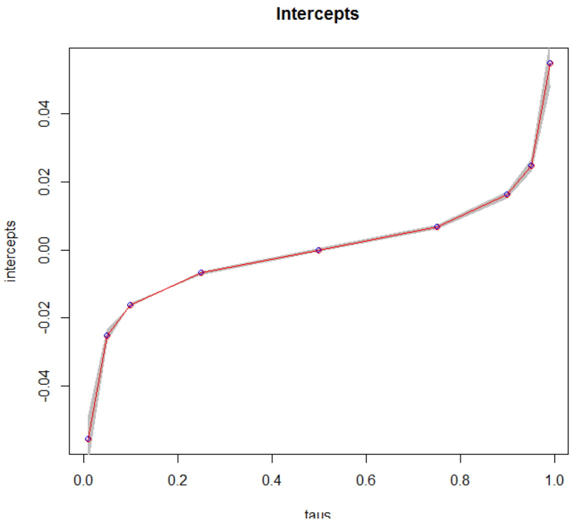


Figure 3.13: Intercepts for the 30 seconds time interval

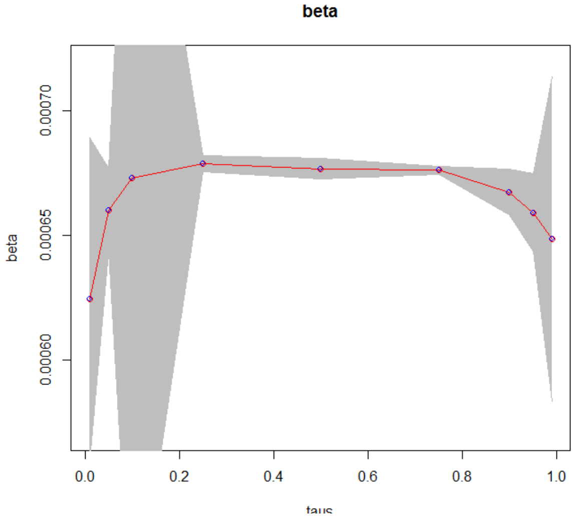


Figure 3.13: Beta for the 30 seconds time interval

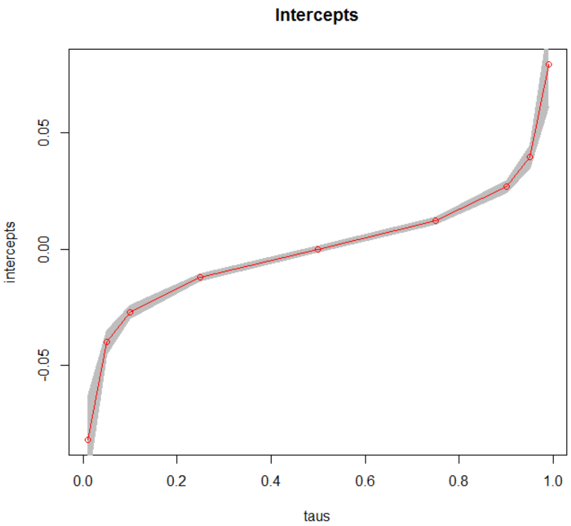


Figure 3.14: Intercepts for the 1 minute time interval

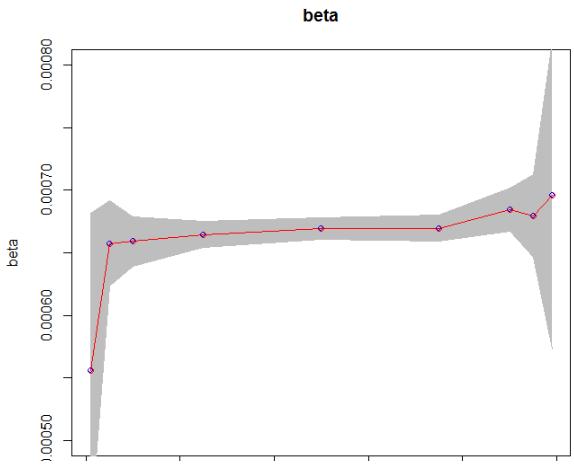


Figure 3.15: Beta for the 1 minute time interval

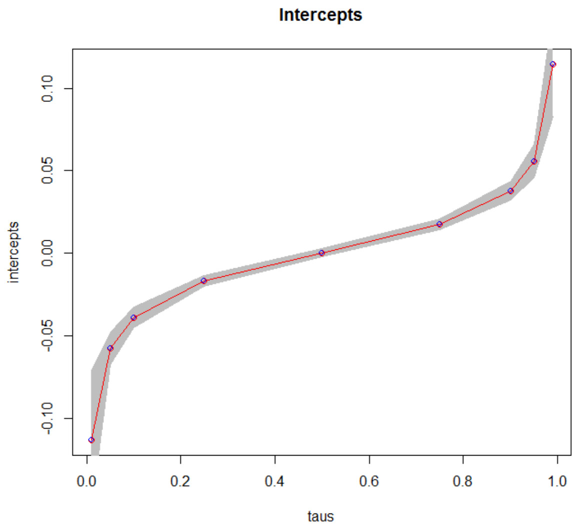


Figure 3.16: Intercepts for the 2 minutes time interval

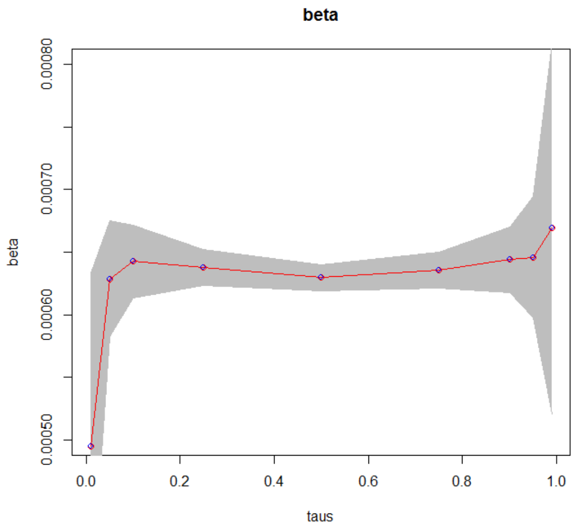


Figure 3.17: Beta for the 2 minutes time interval

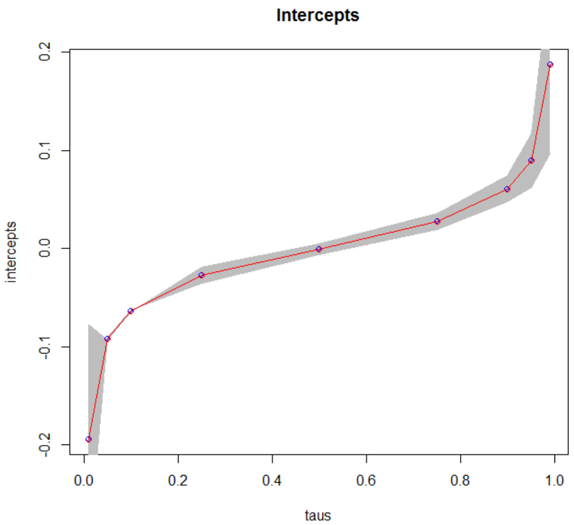


Figure 3.18: Intercepts for the 5 minutes time interval

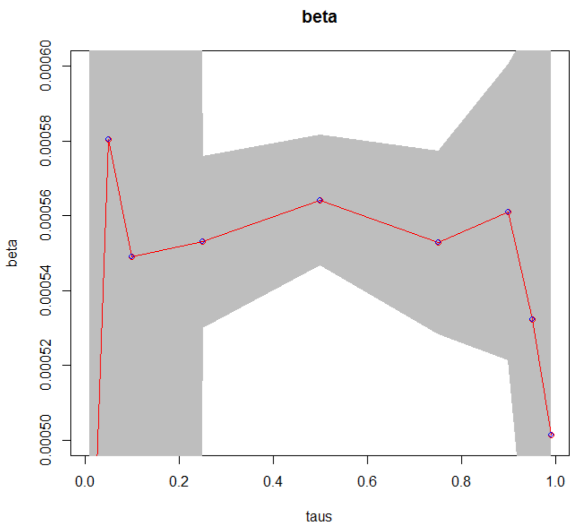


Figure 3.19: Beta for the 5 minutes time interval

3.4 Trading Strategy

The goal of the previous statistical studies was to highlight the features of both the orderflow and the trading volume, together with their short-term explanatory power for the analyzed returns.

This trading strategy is based on two particular aspects emerged in the precedent paragraph: the orderflow has a positive and significant first lag in its autocorrelation function. The quantile regression on orderflow and returns illustrates that the effect of the orderflow is stronger for the positive quantile of returns' distribtuion. Moreover, with the use of the multiple regression it is shown that trading volume has a positive relation with returns.

In the light of these outcomes, the strategy will only open long positions due to the fact that the coefficients of orderflow for positive returns are higher, meaning that the effects are stronger. The trading volume is included as variable of the strategy because we choose to identify the situations when the market activity is intense and with a positive buying pressure. The strategy will perform only on 1 time interval because the statistical results show clearer effects than results based on the other time intervals.

It is necessary specify the assumptions underlying:

- When the strategy opens a trade, it sends a buy market order and there is always a counterparty on the ask side.
- There is no latency between the arrival of market data and the execution time.
- The strategy trade only 1 contract of Future Dax (12,5 euro/tick)
- There is not spread between bid and ask
- There are no commission or other transaction costs.

The trigger to open a trade has been choosen considering the orderflow and trading volume distributions in order to identify situations with strong buying pressure and intense trading activity.

A market buy order will be sent only if the orderflow and trading volume in the 1 minute time interval are bigger than specific thresholds. The position will be opened at the beginning of the next time interval and closed at the end.

It was conducted an analysis of 17 percentiles of both distributions, moving by 5% from 10% to 90%.

To choose the most performing parameters it was conducted an investigation for all the combinations of percentiles.

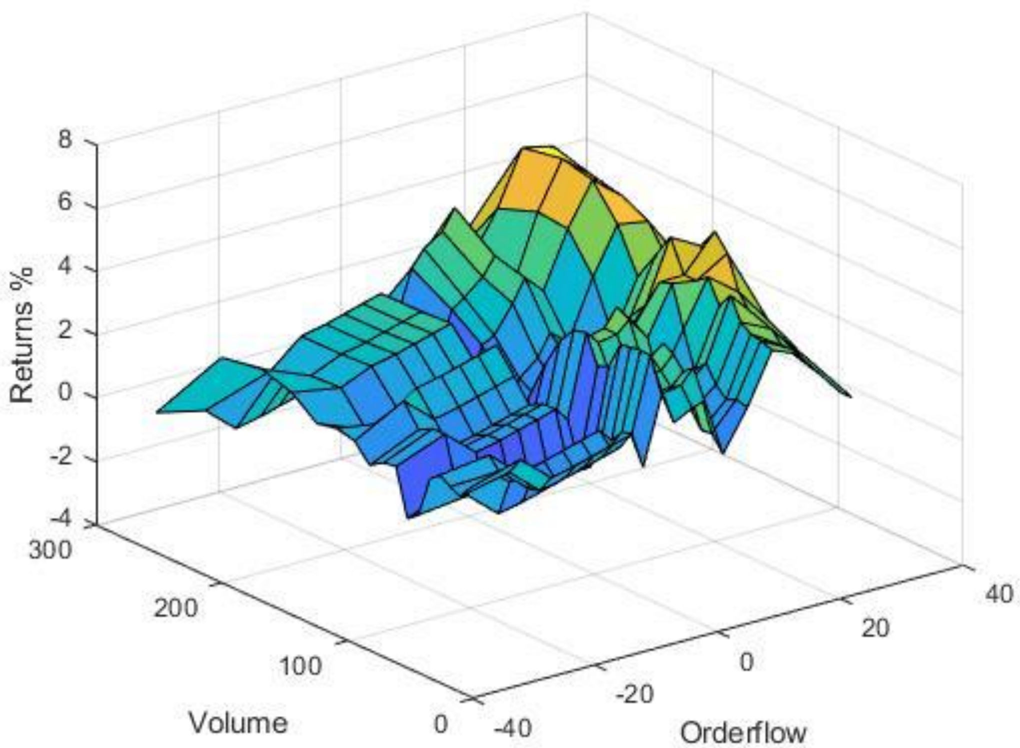


Figure 3.20: Trading results with all the combinations of parameters

The most performing parameters are the 85^{th} percentile of the orderflow distribution and the 85^{th} percentile of the trading volume distribution, equal respectively to 24 and 242.

As expected, when the orderflow is relatively high and positive and trading volume assumes a big value respect to its distribution, the 1 minute ahead return is often positive.

This situation meaning that the majority of investors are willing to open long positions driving the prices up.

This strategy yields a positive returns of 4.1% in the 62 days of sample, it has an average win ratio of 69% and an average risk-reward ratio of 1.18.

Considering the results deriving from the quantile regression on contemporaneuos orderflow and returns, the trading strategy is improved introducing the trailing stop. The same parameters are used as threshold for the trades and the open positions will not be closed at the end of the time interval, but it will be held until the orderflow variable is positive in order to exploit the positive momentum.

The strategy opens less trades than the strategy without trailing stop but the win-ratio is approximately the same.

The total return increase by a 1.1% for the sample considered and the driver is the improvement of the risk-reward ratio. Now, the strategy obtains bigger returns than the previous backtest and the size of stop losses are the same.

The results are illustrated in the table as follow:

Trading Strategy Results			
Return	Win-Ratio	Risk:Reward	Nr. Trades
4.12%	68.5%	1.18	496
5.27%	69.7%	1.37	462

Table 3.13: Results of the trading strategy and the trading strategy improved with the trailing stop

4 Conclusion

In this paper was introduced the area of empirical market microstructure. Starting from the description of the functioning of the limit orderbook, it was done a literature review, were described some high frequency trading strategy and showed some inefficiencies. The empirical study aims to show that using the orderflow and the trading volume, a profitable trading strategy can be developed. The multiple regression highlights the significativity of these variables in the explanation of market returns. The market returns, as expected, present some volatility clustering. This information can be derived with a statistical method that tests the presence of arch effects in the residuals time-series. The quantile regression shows a particular feature. The orderflow has a stronger effect on the positive quantile of the returns' distribution. Starting from these results, a trading strategy that is able to produce consistent returns was developed exploiting the outcomes of the previous statistical relations.

We can conclude that it is possible to developed an alpha generating strategy exploiting the order flow that comes to the market.

Bibliography

[1] Irene Aldridge. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. John Wiley & Sons, Inc, 2010.

[2] Mario Bellia, Kim Christensen, Aleksey Kolokolov, Loriana Pelizzon, and Roberto Renò. High-frequency trading during flash crashes: Walk of fame or hall of shame? *Working Paper*, 2018.

[3] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Order Imbalance, liquidity, and market returns. *Journal of Financial Economics*, 65:111-130, 2002.

[4] Tarun Chordia and Avanidhar Subrahmanyam. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 72:485-518, 2004.

[5] Rama Cont, Arseniy Kukanov, and Sasha Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47-88, 2014.

[6] Joel Hasbrouck and Gideon Saar. Low-Latency Trading. *Journal of Financial Markets*, 16(4), 2013.

[7] Han-Ching Huang, Yong-Chern Su, and Yi-Chun Liu. The performance of imbalance-based trading strategy on tender offer announcement day. *Investment Management and Financial Innovations*, 11(2):38-46, 2014.

[7] Van Kervel, V. and Menkveld, A. J. High-frequency trading around large institutional orders. *SSRN Working Paper*, 2016

[8] Charles M.C. Lee and Mark J. Ready. Inferring trade direction from intraday data. *The Journal of Finance*, 46:733-747, 1991.

[9] Jaeram Lee, Geul Lee, and Doojin Ryu. The difference in the intraday return-volume relationships of spot and futures: a quantile regression approach. *Economics: The Open-Access, Open-Assessment E-Journal*, 13:1–37, 2019.

[10] Rahul Ravi and Yuqing Sha. Autocorrelated Order-Imbalance and Price Momentum in the Stock Market. *International Journal of Economics and Finance*, 6(10):39-54, 2014.

[11] Koenker, Roger, and Gilbert Bassett. "Regression Quantiles." *Econometrica 46*, no. 1 (1978): 33-50. Accessed February 6, 2020.

SUMMARY

The technological develpment and the innovations in the financial markets made possible to collect and study the information descending from the market activity in the microstructure.

The information obtained can be extremely useful in order to explain the price movements in the short-term.

The first empirical studies highlight the relevance of trading volume and successively the attention shifted to the orderflow. The orderflow is a more advanced concept respect to the trading volume because it attributes an additional feature to each trade executed making possible to identify where a trade is buyer-initiated or seller-initiated. Consequently, an investor able to get and exploit this information may have an advantage respect the other partecipants who do not, at least in the short-term. The partecipants who, for definition, exploit and trade relying on these signals are the High Frequency Trader (called 'HFT').

They trade at extreme speed against the other investors and aim to gain small profits countless times.

In order to support the thesis that the market activity is a driver of market returns, we investigated with statistical tools the relationships between returns, orderflow and trading volume for four time intervals (30 seconds, 1 minute, 2 minutes, 5 minutes) and we were able to develop a trading strategy based on signals deriving from these variables.

In the first chapter it is illustrated the market framework, the evolution during this last decades and the way how investors trade.

The second chapter describes with the market activity from a more technical view. There are described the methods use to assign the sign to every trade, the first model that riassumes the relations between

trading volume, returns with information flow that comes to the market. In this part some studies has been mentioned in order to understand how previous authors deal with these concepts and what they found. Moreover, we describes the most common high frequency trading strategies and analyzed some market inefficiencies.

The third chapter is the main part of this thesis. We studied the distributions of the three variables taken into account, the autocorrelation function of the orderflow and the correlation between returns and orderflow. The first tool used to get a description of these relationships is a multiple linear regression that shows that the effects of the independent variables (orderflow and trading volume) are significant. To examine more deeply the relation between returns and orderflow we availed of quantile regression and found that positive returns are more influenced by orderflow than negative returns.

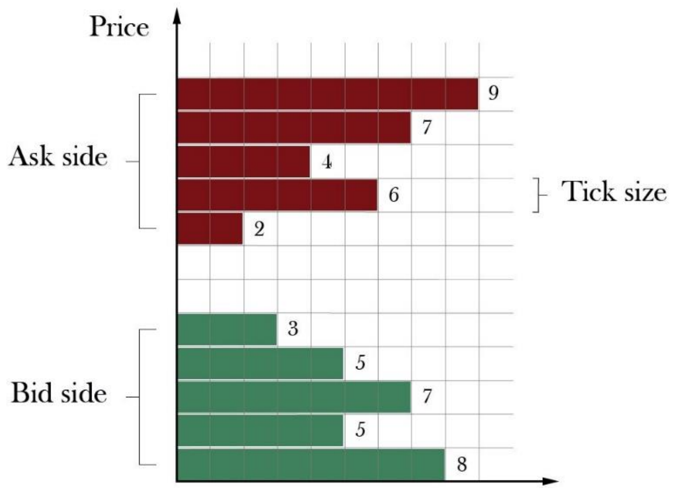
In the light of these findings a profitable trading strategy has been developed exploiting the orderflow and trading volume statistical features and the relationship with returns. This strategy considers these variable and takes in account the returns of one period ahead.

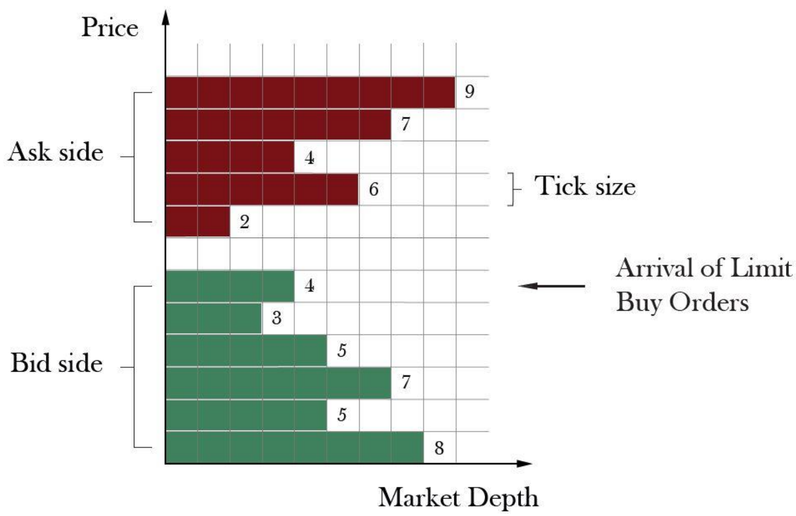
After having shown that choosing specific thresholds of orderflow and volume can be profitable in the sample considered, another rule is introduced in the strategy deriving from the relation between orderflow and returns. Until the orderflow variable remains positive the trade will not be closed in order to exploiting the buying momentum until it ends. This improves the results of the strategy and shows how the information deriving from orderflow and trading volume can be used in a proper way.

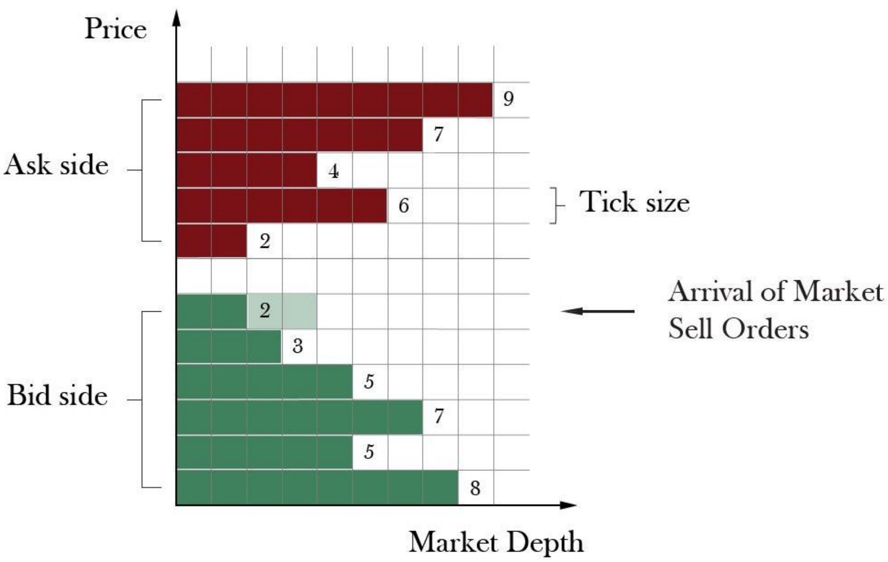
Starting from the definition of the limit order book It is a centralized database of pending limit orders for a specified price or better and with a given size. It is a transparent system for the market partecipants and

the matching algorithm for limit orders with market orders works with a price time priority algorithm: it works with matching priority for price, time and lastly visibility.
A buy limit order with the highest price represents that buyer is willing to pay more than the other traders and he gets the priority of execution when a sell market order will arrive. Consequently, a market order is executed on the best proposal, best bid or best ask (in the US market the first levels are guaranteed by NBBO, National Best Bid Offer). After that, it follows the FIFO principle: First in First out.
The limit order that is submitted earlier will be executed before the later arrival.
Therefore the queue of limit orders follows this chronological rule.
In most markets, there is the possibility to put hidden orders, that stands on the book without being displayed. They are the last limit orders to be executed because the precedence is given to the visible orders.
The ordebook events can be mainly identified in three types: submission of limit orders, cancellation of limit order, execution of market orders.
The empirical studies in the financial literature focus their attention on studying the price impact of the orderbook events. The outstanding limit orders, called Market Depth, concept related to the market liquidity, is associated with the magnitude of price changes. In a thin orderbook, where the market depth is low given by the small size and the low number of limit orders, it is easier to cause a bigger price change given to the fact that there is less need of market orders, respect a thick orderbook, to buy or sell all the first level proposals and than to move the price.

The price dynamics are also conditioned by the arrival, the size and the direction of the market orders. The funcioning can be summarized by the following three images. (Figure 2.1, Figure 2.2, Figure 2.3)







Before, starting with the empirical studies with the regressions, I analyze the statistical features of the three variables considered for the work: market return, orderflow and trading volume. The feature are illustrated in the following tables. (Table 2.1, Table 2.2, Table 2.3, Table 2.4)

30 seconds time interval			
	Returns	Orderflow	Volume
Mean	0.000	0.02	75
Median	0.000	0.00	55
Standard deviation	0.0002	20.60	73.48
Skewness	0.21	0.09	3.85
Kurtosis	23.52	14.65	32.7
1^{st} Percentile	-0.000675	-58	7
5^{th} Percentile	-0.000365	-31	14
25^{th} Percentile	-0.000121	-9	32
75^{th} Percentile	0.000121	9	94
95^{th} Percentile	0.000363	31	204
99^{th} Percentile	0.000659	59	367

1 minute time interval			
	Returns	Orderflow	Volume
Mean	0.000	0.06	98
Median	0.000	0.00	115
Standard deviation	0.0004	29.22	111.92
Skewness	0.7	0.05	3.22
Kurtosis	24.87	10.42	22.5
1^{st} Percentile	-0.000954	-81	21
5^{th} Percentile	-0.000517	-45	35
25^{th} Percentile	-0.000165	-14	70
75^{th} Percentile	0.000164	16	188
95^{th} Percentile	0.000511	45	390
99^{th} Percentile	0.000943	83	649

2 minutes time interval			
	Returns	Orderflow	Volume
Mean	0.000	-0.11	302
Median	0.000	0.00	239
Standard deviation	0.0005	41.65	234
Skewness	0.8	0.11	2.79
Kurtosis	21.52	7.4	18.2
1^{st} Percentile	-0.00133	-112	54
5^{th} Percentile	-0.00072	-65	82
25^{th} Percentile	-0.00024	-21	152
75^{th} Percentile	0.00024	21	377
95^{th} Percentile	0.00072	65	741
99^{th} Percentile	0.00130	116	1193

5 minutes time interval			
	Returns	Orderflow	Volume
Mean	0.000	-0.31	759
Median	0.000	0.00	624
Standard deviation	0.0008	66.36	511
Skewness	0.48	0.02	2.29
Kurtosis	14.90	5.8	13.3
1^{st} Percentile	-0.00217	-181	166
5^{th} Percentile	-0.00115	-105	233
25^{th} Percentile	-0.00036	-36	416
75^{th} Percentile	0.00010	35	957
95^{th} Percentile	0.00034	105	1712
99^{th} Percentile	0.00213	187	2664

After having described the characteristics of these variables we exploit a multiple linear regression in order to found the relationships between trading volume, orderflow with the market return.

The outcomes show a positive relation meaning that these variables have an explanatory power on the retunrs.

Multiple Linear Regression				
	30 seconds	1 minute	2 minutes	5 minutes
Intercept	-0.00016	-0.00080	-0.00170	-0.00430
B_1	0.00069	0.00066	0.00042	0.00023
B_2	0.00038	0.00011	0.00007	0.00004
R^2	0.367	0.349	0.316	0.248

Table 3.7 shows another interesting finding is that the R^2 of the regression is bigger if we move through smaller time intervals.

After having ensure the positive relation between these variables we investigated these effects with a more deep examination exploiting the quantile regression. The results show a clear situation. The coefficients for the orderflow are bigger in the positive percentile of the returns distributions. We attribute this effect to the fact that if the investors are willing to buy the Futures contract, they want to increase their long exposure in the market causing an upturn in the prices. On the contrary, the sell of a Futures contract can be done to hedge a portfolio with long exposure on the underlying. The effect in this case is much weaker on the price, as shown in the results of quantile regression.

The table 3.9, table 3.10, table 3.11 and table 3.12 summary the outcomes of the regression.

30 seconds time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.001) -0.055	(0.000) 0.000624
0.05	(0.000) -0.025	(0.000) 0.000620
0.10	(0.000) -0.016	(0.000) 0.000672
0.25	(0.000) -0.07	(0.000) 0.000678
0.50	(0.000) 0.000	(0.000) 0.000676
0.75	(0.000) 0.007	(0.000) 0.000676
0.90	(0.000) 0.016	(0.000) 0.000667
0.95	(0.000) 0.025	(0.000) 0.000659
0.99	(0.001) 0.55	(0.000) 0.000648

1 minute time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.002) -0.082	(0.000) 0.000556
0.05	(0.001) -0.040	(0.000) 0.000657
0.10	(0.000) -0.027	(0.000) 0.000659
0.25	(0.000) -0.012	(0.000) 0.000664
0.50	(0.000) 0.000	(0.000) 0.000669
0.75	(0.000) 0.012	(0.000) 0.000669
0.90	(0.000) 0.027	(0.000) 0.000684
0.95	(0.000) 0.040	(0.000) 0.000679
0.99	(0.002) 0.08	(0.000) 0.000696

2 minutes time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.004) -0.113	(0.000) 0.000556
0.05	(0.001) -0.058	(0.000) 0.000657
0.10	(0.001) -0.039	(0.000) 0.000659
0.25	(0.000) -0.017	(0.000) 0.000664
0.50	(0.000) 0.000	(0.000) 0.000669
0.75	(0.000) 0.017	(0.000) 0.000669
0.90	(0.001) 0.038	(0.000) 0.000684
0.95	(0.001) 0.056	(0.000) 0.000679
0.99	(0.003) 0.114	(0.000) 0.000696

5 minutes time intervals		
Quantiles	Intercept	Orderflow
0.01	(0.012) -0.194	(0.000) 0.000439
0.05	(0.003) -0.092	(0.000) 0.000580
0.10	(0.001) -0.063	(0.000) 0.000548
0.25	(0.001) -0.027	(0.000) 0.000553
0.50	(0.001) 0.001	(0.000) 0.000564
0.75	(0.001) 0.028	(0.000) 0.000552
0.90	(0.001) 0.061	(0.000) 0.000560
0.95	(0.003) 0.090	(0.000) 0.000532
0.99	(0.009) 0.188	(0.000) 0.000501

The goal of the previous statistical studies was to highlight the features of both the orderflow and the trading volume, together with their short-term explanatory power for the analyzed returns.

This trading strategy is based on two particular aspects emerged in the precedent paragraph: the orderflow has a positive and significant first lag in its autocorrelation function. The quantile regression on orderflow and returns illustrates that the effect of the orderflow is stronger for the positive quantile of returns' distribtuion. Moreover, with the use of the multiple regression it is shown that trading volume has a positive relation with returns.

In the light of these outcomes, the strategy will only open long positions due to the fact that the coefficients of orderflow for positive returns are higher, meaning that the effects are stronger. The trading volume is included as variable of the strategy because we choose to identify the situations when the market activity is intense and with a positive buying pressure. The strategy will perform only on 1 time intervals because the statistical results show clearer effects than results based on the other time intervals. There are some underlying assumptions to take in account for the evaluation of the strategy: no transaction costs are taken in account, the bid-ask spread is always 1 tick and there is no delay in the data received.

The trigger to open a trade has been choosen considering the orderflow and trading volume distributions in order to identify situations with strong buying pressure and intense trading activity.

A market buy order will be sent only if the orderflow and trading volume in the 1 minute time interval are bigger than specific thresholds. The position will be opened at the beginning of the next time interval and closed at the end.

It was conducted an analysis of 17 percentiles of both distributions, moving by 5% from 10% to 90%.

To choose the most performing parameters it was conducted an investigation for all the combinations of percentiles.

The most performing parameters are the 85^{th} percentile of the orderflow distribution and the 85^{th} percentile of the trading volume distribution, equal respectively to 24 and 242.

As expected, when the orderflow is relatively high and positive and trading volume assumes a big value respect to its distribution, the 1 minute ahead return is often positive.

This situation meaning that the majority of investors are willing to open long positions driving the prices up.

This strategy yields a positive returns of 4.1% in the 62 days of sample, it has an average win ratio of 69% and an average risk-reward ratio of 1.18.

Considering the results deriving from the quantile regression on contemporaneuos orderflow and returns, the trading strategy is improved introducing the trailing stop. The same parameters are used as threshold for the trades and the open positions will not be closed at the end of the time interval, but it will be held until the orderflow variable is positive in order to exploit the positive momentum.

The strategy opens less trades than the strategy without trailing stop but the win-ratio is approximately the same.

The total return increase by a 1.1% for the sample considered and the driver is the improvement of the risk-reward ratio. Now, the strategy obtains bigger returns than the previous backtest and the size of stop losses are the same.
The results are illustrated in the table 3.13 and in the figure 3.16 as follow:

Trading Strategy Results			
Return	Win-Ratio	Risk:Reward	Nr. Trades
4.12%	68.5%	1.18	496
5.27%	69.7%	1.37	462