UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"

CORSO DI LAUREA MAGISTRALE / SPECIALISTICA IN ECONOMICS AND FINANCE

TESI DI LAUREA

"THE INFORMATION CONTENT OF THE LIMIT ORDER BOOK ON ENERGY MARKETS"

RELATORE:

CH.MO PROF. ENRICO EDOLI

LAUREANDO/A: CARLO BENATO

MATRICOLA N. 1129530

ANNO ACCADEMICO 2017 – 2018

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

____

CONTENTS

1. *INTRODUCTION*

In the last twenty years the introduction of electronic trading platforms in derivative exchanges has played the principal role in the change of the financial markets' structure. The traditional open-outcry auctions have been substituted by the limit order book mechanism, whose main benefit has been to improve consistently the pre-trade transparency allowing traders to gain a real-time access to the top n levels of depth of the contracts listed. The main goal of the thesis is to investigate whether this improved pre-trade transparency facilitates price discovery and helps market participants in forecasting short-term price movements. In particular, the authors focus their attention on the incremental information content of the limit order book, over and above the information traditionally available in a dealer market. We construct imbalance measures to describe the current status of the book in the limit buy and sell orders beyond the best quote. Then, in a repetitive regression framework we demonstrate that these variables are suitable in explaining future short term returns. For our knowledge similar analyses have been performed only in stock or bonds markets, the paper instead test data on a power futures contract traded at the European Energy Exchange.

The thesis is composed as follows: section 2 and section 3 introduce the limit order book as the main trading tool in an order-driven venue and the importance of data visualization. The third section provides us with an overview of the literature on the book depth analysis and it presents the principal indicators used to describe the order book shape. An introductory presentation of electricity markets is assessed in section 5. The last section investigates the additional information content of the book, explaining the methodology and showing the empirical results of the simulation example.

2. THE LIMIT ORDER BOOK MECHANISM

Preliminaries

Until the mid-90s the majority of financial trades took place in quote-driven marketplaces where large market makers centralize buy and sell orders by publishing the prices at which they are willing to buy or sell the traded asset. They provide immediacy to the market, i.e. they continuously provide liquidity on demand to anyone. Of course, this liquidity service could be valuable only if they sell at prices higher than those at which they buy; the difference between the two best quotes is the well-known bid-ask spread. That profit counterbalances the risk of acquiring an undesirable inventory position or for being subject to the adverse selection problem, that is encountering other traders who are better informed about the asset value and who gain a lot by trading repeatedly with the market maker.

The entire process is nowadays become much more flexible thanks to the introduction of the limit order book mechanism, where every player has the option of posting buy and sell orders at his preferable price. We therefore move from quote-driven to electronic order-driven markets in which a trading algorithm tries to match all the upcoming orders as they come. A Limit Order Book (LOB) is defined as a set of queues, each of which consists of all the outstanding active buy or sell orders in a market at a specified price and at a particular point in time. An order is considered active if the algorithm is not able to find a match for it immediately. It remains active until it becomes matched to another incoming order or it is cancelled.

The majority of the most relevant trading venues has already adopted this dynamic: the Helsinki, Hong Kong, Shenzhen, Swiss, Tokyo, Toronto, and Vancouver Stock Exchanges all now operate as pure Limit Order Book. Other, such as the New York Stock Exchange, NASDAQ, and the London Stock Exchange operate with a hybrid Limit Order Book system.

Thanks to the technological progresses, the traders have almost real time access to the huge amount of information contained in a limit order book and these detailed historical time series pave the way to further researches to test economic theories or statistical regularities. Analysing the data can give us an insight into the best practises to follow in a given market situation. Furthermore, we can learn how to execute optimally an order strategy or how to minimize the market impact after an order has been posted. From the investment companies' point of view, studying the data patterns can help designing better automated trading algorithms or assessing market stability.

Modelling a Limit Order Book

Various models have been proposed that investigate the limit order book dynamics and it is possible to classify them in two main branches. Drawing from economics the Perfect Rationality approaches consider LOBs as a series of sequential games due to the behaviour of individual traders, who are driven by their own information and try to exploit it to maximize their utility. On the opposite side many mathematicians support the Zero Intelligence methodologies where aggregate order flows are treated as random, that is they are governed by stochastic processes. In both cases researchers have to deal with two main difficulties: the state-space complexity, where the state space is the set of values which a process can take, and the feedback and coupling. In addition to the traders' actions depending on the current order book state, the book status itself clearly varies according to the traders' choices. These mutual dependences induce feedback between the LOB and the agents' behaviour. Indeed, it has been empirically demonstrated that the current order flows depend on both the previous flows and on the set of all active orders in a market at time t. Also, the order flow creates a strong coupling between the best bid and the best ask. The top ask level determines a boundary condition for buy limit order placement because any buy order placed at or above the best ask partially matches immediately. A similar role is played by the best bid on the demand side. Therefore, the main goal is to design a model that defines appropriately this conditional behaviour and at the same time simplifies the evolving dimensionality of the state space.

Market and Limit Orders

Describing more in details the features that characterize a limit order book, we should start from the order types. Essentially a trader can submit a market order or a limit order: the former finds a match immediately upon submission, the latter instead does not and it becomes an active order. The term 'limit' is used basically only to emphasize whether an incoming order triggers an immediate matching or not. Consider a buy order x that is defined by the vector (p_x, w_x, t_x), where p_x is the submission price, w_x is the order size and t_x is the precise moment in time of submission.

The buy order price could be:

- $p_x < b(t)$, where $b(t)$ is the best bid price, i.e. the highest state price among active buy order at time t or mathematically $b(t) = \max p_x$, $x \in B(t)$ where $B(t)$ is the set of all limit buy orders. In this case the order x is a limit order, it does not change the best bid price $b(t)$ and it becomes active increasing the depth of the LOB;

- $b(t) < p_x < a(t)$, where $a(t)$ is the lowest sell price, $a(t) = \min p_x$, $x \in A(t)$ where $A(t)$ is the set of all limit ask orders. Then the limit buy order becomes active and it narrows the bid-ask spread;

- $p_x \geq a(t)$, the order is a market order and, depending to its size, immediately upon arrival it matches against one or more active sell orders. This match occurs at the highest price available in the market. If the order size is bigger than that of the best active sell limit order, any residual size is considering for matching to the next highest priority order.

Ultimately, the LOB bid-ask spread can be considered as the measure of how highly the market values the immediacy and certainty associated with the market orders versus the waiting and uncertainty associated with limit orders. Indeed, Copeland and Galai (1983) state that a limit order can be considered as a derivative contract written to all market participants. These options are free since the writer does not get any explicit premium, but rather receives an implicit gain if the limit price is hit and his contract is executed at better terms.

An order can also be classified on the basis of its aggressiveness. A limit order is considered aggressive whenever it improves the best bid or ask quote and accordingly when it shifts all depth levels on that side. Differently, a passive limit order is an order that is posted behind the market, i.e. its limit price is smaller than the current best quote. It does not change the prevailing quote and it affects only the depth. Such an approach gives the trader total control over the execution price, but it is completely uncertain when or even if the order will be executed. The investor shall then wait for the market to move in his or her favour and therefore a passive trading strategy cannot fulfil a requirement of immediacy or that the matching of a given shares size will be completed within a specified time period. Also, a market order can be defined as aggressive if it walks up the book, that is when its size is bigger than the size of the opposite prevailing quote and consequently all depth levels and the best quote shift. Hence, even if the agent is certain of the trade, this large market order causes a substantial negative impact on the market, worsening the execution price relative to the current market quote.

In accordance with the specific exchange rules, an order may be affected by more detailed information regarding the order execution, such as price- and quantity-related instructions, duration of commitment, time of expiration, visibility and the terms of any principal-agent obligation. For example, a pegged strategy involves orders that adjust their limit price according to the changes of the best quotes. Automatically correcting the price to the best bid or ask, these orders remains competitive relative to the other passive orders and therefore they increase the likelihood to be filled.

Once an order has been submitted, the matching algorithm assigns it a priority within each price level. The most common system is built on a price-time basis, that means a trader gets the precedence whenever his or her order is placed at a better price than the remaining active orders. If then two orders share the same price, whoever places it earlier gets the priority. Another mechanism is pro-rata, whenever a tie occurs and the aggregated bid size exceeds the aggregated offer size, each order is partially filled proportional to the fraction of depth available. Since placing order with a bigger size than the desirable one may help in gain priority, agents face a huge problem in deciding the optimal size. As for the pro-rata, there is also the price-size mechanism that rewards traders for providing liquidity to the market: among those at the best price, the active order with larger size is chosen.

Hidden Liquidity

A LOB does not always reveal fully the trading intentions of the whole market. Once a participant takes choices using the actual book status, it has to consider some issues like incomplete sampling and hidden liquidity. Or, more simply, it could be the case that some traders do not show their preferences by submitting orders only when it is strictly necessary. Some electronic venues do not show all the liquidity available in the market because the book displays only a set of prices within a predefined range. The displayed liquidity may also be affected by how fast new orders become visible, the frequency and the choice of parameter (every T-second, trade-by-trade, event-by-event) with which the LOB is updated are very relevant. Some other exchanges permit to personalize the own book. Through a bilateral agreement a trader can define a block-list of other traders with whom he or she is unwilling to trade. Therefore, his or her book shows only the active orders of the traders with whom it is possible to trade. Even if the book does not suffer from latency issue and it displays all active orders at whatever price, some exchanges permit the submission of iceberg orders, i.e. large volume orders that are publicly visible only partially. After the visible part has been matched by an incoming market order, depending on the venues another part of equal size becomes public or the hidden portion that has not been filled is cancelled.

In the last decade a phenomenon has increased over and over: the so-called Dark Pools, trading venues without pre-trade transparency. These exchanges allow traders to submit large orders completely anonymously and without revealing, sometimes also specifying, any price. The matching algorithm can vary a lot, some dark pools are essentially standard LOBs where all active orders are hidden. In other cases a traditional exchange is taken as reference and the dark pool simply replicates it, matching the hidden orders on the basis of the prices of the principal market. Most of the time transactions in these pools are carried out only for speculative reasons.

Another difficulty in modelling a LOB is to deal with the volatility, that is the measure of the variability of returns of the traded asset. Generally, an asset with high volatility is expected to register higher price changes in a given period of time than a low volatile asset. It is easy to understand why finding an unbiased estimate is fundamental in selecting assets for constructing a portfolio with a low risk exposure. This task has become more difficult since many traders started to place limit orders and immediately cancel them in order to discover the hidden liquidity in the market. This practice causes the prices to change very fast, but does not have a meaningful impact or any economic belief behind and that is why we can call this variability in prices microstructure noise. Nowadays there are completely automated algorithms expressly programmed to submit and cancel orders in order to discover the traders' intentions.

Resolution Parameters

A limit order book structure and the amount of completed transactions are highly defined by its resolution parameters, that are the lot size and the tick size. The lot size is the minimum amount of the tradable asset that can be exchanged. It directly affects a trader submission strategy, in particular when he or she decides to split a large order in smaller ones to impact as little as possible to the final trading price. Instead, the thick size represents the cost that an agent has to bear if she or he wants to gain priority, in other word it is the minimum price increment that an asset is allowed to move. In October 2016 NYSE launched the Tick Size Pilot Program that has widened the minimum quoting and trading increment for small capitalization stocks, which began trading in five-cent increments instead of one-cent increments. The goal of the pilot is to boost liquidity and to increase the percentage of order that get filled by increasing the spreads for thinly-traded stocks.

Opening and Closing Auctions

We conclude our fast review of the mode of operation that rules an electronic venue presenting another feature that characterizes many exchanges. The open and the close of markets are extremely relevant event for traders and regulators. The opening period absorbs all the information gathered overnight and it plays a fundamental role in aggregating them and in the price discovery; the closing price is the price at which investors make decisions regarding their investment portfolios. In fact, it represents a reference point for determining performance over a specific time frame. That is the reason why at the beginning and end of the trading day, many markets, for instance Paris, Frankfurt and London, suspend the continuous limit order trading to match orders and instead use an auction system. The main purposes of call action involve the sharpening of price discovery, the enhancement of quantity discovery, the reduction of trading costs and the avoidance of price manipulation. Despite the widespread use, the call auction mechanisms show remarkable differences on the world's major exchanges, showing disagreement about its optimal design. In general, in both open and close auction traders can view and place orders as usual but no match occurs. Therefore, it could also happen a cross market situation, that occurs when the best limit buy price is higher than the minimum limit ask price. After having collected all orders, for each price p with positive depth the matching algorithm calculates the total volume that can be traded between buy orders with a price equal to or greater than p and the corresponding sell orders with a price equal to or less than p. The algorithm then maximizes the total volume defining the uncrossing price at which all trades take place, in contrast to standard continuous limit order book trading.

3. *DATA VISUALIZATION*

Visualizations have often acted as means of communicating results, by translating and summarizing multi-dimensional data into a form visually accessible to users. This ease of communication comes from the ability of visuals to both help externalize the memory associated with the data, and to more closely represent a user's mental model of the data. This efficiency can free the user's memory to support further, necessary, cognitive operations or tasks. Indeed, the design goal for a data analysis system is to facilitate the knowing-learning-understanding process on users, freeing time and energy for them to pursue insights or iteratively explore the data.

In this chapter we present the market data structure and we show some simple examples on how data visualization is a fundamental tool for building intuition and enabling exploratory data analysis.

Market data are classified on the basis of the amount of information they provide.

- Basic data is known as Level I and it includes real-time best ask and bid, the volume available at these highest quotes, the information regarding the last transaction, the price and its size, and ultimately the high and low price for the day;

- Level II data provides much more comprehensive and in depth information, it indeed discloses the full limit order book;

- A third layer of data information, sometimes incorrectly described as Level III, is restricted to venue's member firms that operate as registered market makers. It gives the ability to enter quotes, execute orders and send attached confirmations or particular instructions.

Visualizing correcting the huge set of data held in a limit order book gives potentially enormous benefits to traders because it provides a balanced portrait of the dynamic and static components of the market system. Analysing the charts market participants can not only watch the evolution of the spread and find different liquidity at each price level, but also identify patterns such as intraday momentum or price changes due to the order book activity. On the following paragraph we review the development of data visualization taking into account that displaying Level II data requires at least three dimensions: price levels, total size at each level and time dimension. Only most recent tools allow to combine uniformly these three different sides.

Climbing The Market and Bar Chart

Figure 1 shows the most classical representation of a limit order book where it is illustrated a snapshot of current traders' intentions to trade taken at the last available time. The displayed numbers come from the dataset we have analyzed in the next chapters, it involves the Futures contract Germany Baseload 2018 traded at the European Energy Exchange. On the left the so-called "Climbing the Market" shows the different levels and the total volume lined up at each price to buy or sell, while on the right the bar chart displays more intutively the same data. Whenever a new order is submitted or an active one is filled/cancelled the graph is updated. The bid ask spread is calculated taking the difference between prices on the top of the book, therefore in this case the spread is 0.05.

Germany Baseload - 2018 - Last instant

BID		ASK	
Volume	Price	Price	Volume
5,000	31,05	31,10	20,000
30,000	31,00	31,15	2,000
10,000	30,97	31,17	5,000
20,000	30,95	31,20	15,000
61,000	30,90	31,25	5,000
5,000	30,88	31,28	5,000
16,000	30,85	31,30	5,000
11,000	30,80	31,40	2,000
10,000	30,75	32,05	5,000
5,000	30,70		
6,000	30,50		
10,000	30,45		
5,000	29,15		

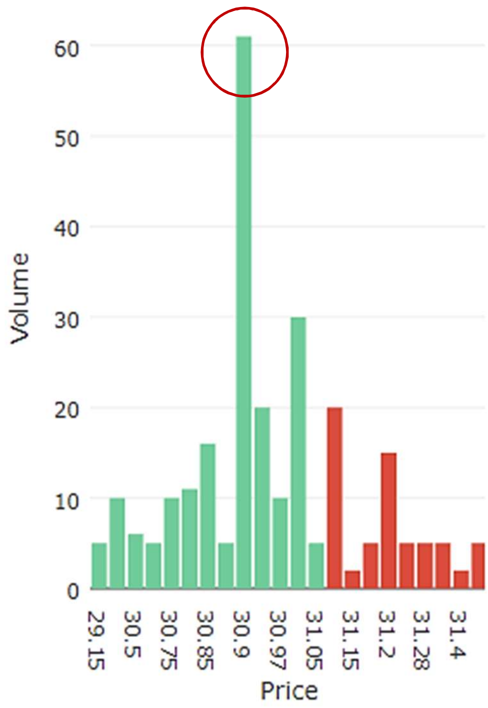

Figure 1. Climbing The Market and Bar Chart – Germany Baseload 2018, EEX

The shape of the market depth bars gives some hints about the conditions of the market. In an ordered and liquid market, for example, the sizes on both sides increase uniformly among throughout the first five levels, above and below the inside spread, forming a triangle shape as shown in Figure 2, chart 1. Other times it could be the case of an unbalanced market, where the number of buyers and sellers in the market are not equal. The fictional situation displayed in Figure 2,chart 2 shows a high selling pressure that will most likely cause prices to weaken. In the following sections of the thesis we investigate further and in a more quantitative way this possibility.

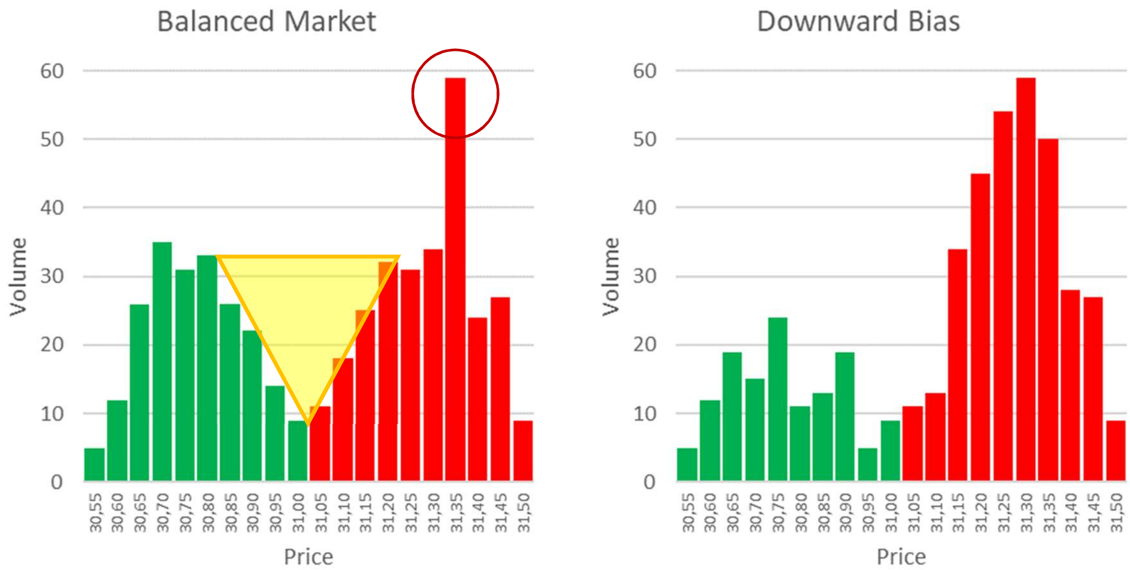

Figure 2. Ordered and Dawnward Bias Market – fictional situation

Support and Resistance Levels

Bar charts are very useful also in determining support or resistance levels which are particular market depth levels significantly larger than those surrounding them, respectively on the buy and ask side. In both Figure 1 and 2 we can notice a support and a resistance level, highlighted by the red circles, respectively at the price 30.90 and 31.35. The reason why these price levels are formed is not completely clear; the behavioural economics suggests that people have a psychological bias to place orders at numbers which they see as being more prominent than others, for instance at round numbers. Thus traders tend to place their stop loss[1] and take profit orders there and this causes the formation of support/resistance areas.

These outliers can signal areas where in the near future significant buying or selling will take place as active pending orders are matched, causing the price movement to slow down. Moreover, these high concetration zones will propel a market in either direction depending on the fact that the level is taken out or defended. Indeed, considering Figure 2, if prices reach the resistance at the 31.35 price level where many sellers are waiting, the price movement will pause and due to the new high selling pressure the price could get pushed down. Conversely, if prices touch a support level, they would probably bounce back. In fact, when the level is reached, the stop-loss/take-profit orders are triggered causing a pressure on the opposite direction. Also the short-termins of the profit-oriented traders can be accounted among the reasons why high concentration zones are formed and prices bounce back.

Before being broken, support and resistance levels are normally tested many many times causing the prices to vary inside that interval. Once such levels have been overtaken, they assume new roles. If prices go deeper than a support level, most likely that level will become a new resistance level on the ask side. The opposite holds as well. From a pure economic perspective the breaching through a support/resistance level can be caused by a new trigger in the market, such as corporate announcements of the underlying company or policy changes. In addition to this, we can consider also market participants' post-regret. For instance, the investors who sold the stock close to the resistance are in an awkward position once the price has cut through this level and then they are waiting to buy it back once the price falls back to the previous resistance level, which is become the new support level.

In a very liquid venue, such as forex or criptocurrency markets, these high concentration areas could be just attempts to manipulate the market. A common technique is to put a large bid order, a bid wall, in order to move market sentiment. Let us consider Figure 3 that represents three consecutive snapshots of the bid side of a bitcoin order book, taken from the dead Mt.Gox platform. Within the scope of affecting orders distribution a larger order to buy is set at the limit price $123. The traders whose limit orders lag below that level observe their execution probabibility to fall down and then they most likely move their limit orders ahead of the wall. After the market reaction the initial agent removes the bid wall, however the newer orders

[1] A stop loss order specifies a particular price. The order is executed at the best available price once a bid or offer is made at that particular price or a less-favourable price.

remain at levels above $123 and the buying pressure increases since new walls, composed by legitimate orders, have been created. There are many reasons to put into practice such a strategy, for instance reducing an inventory position or to create a buying support. In the former case the manipulator wants to obtain a higher average price in selling, while in the latter he wants to avoid a rude market downward due to the arrival of a high sell order.

Bid wall placed

Size	Bid
3,6241	126
177,2617	125
268,3822	124
5687,6996	123
480,1753	122
823,8882	121
1528,1493	120
534,7107	119
989,7167	118
926,0372	117

Orders move above wall

Size	Bid
3,6241	126
845,6244	125
1943,4425	124
5687,6996	123
80,1753	122
223,8882	121
528,1493	120
534,7107	119
989,7167	118
926,0372	117

Wall removed

Size	Bid
3,6241	126
845,6244	125
1943,4425	124
287,6996	123
80,1753	122
223,8882	121
528,1493	120
534,7107	119
989,7167	118
926,0372	117

Figure 3. Bid Wall – BTC/USD, Mt.Gox platform on April 2013

Depth and Candlesticks Charts

Additional to the "Climbing the market" and "Bar chart" tools, traditional trading platforms provide the client also with some other visualization methods such those illustrated in Figure 4 and Figure 5.

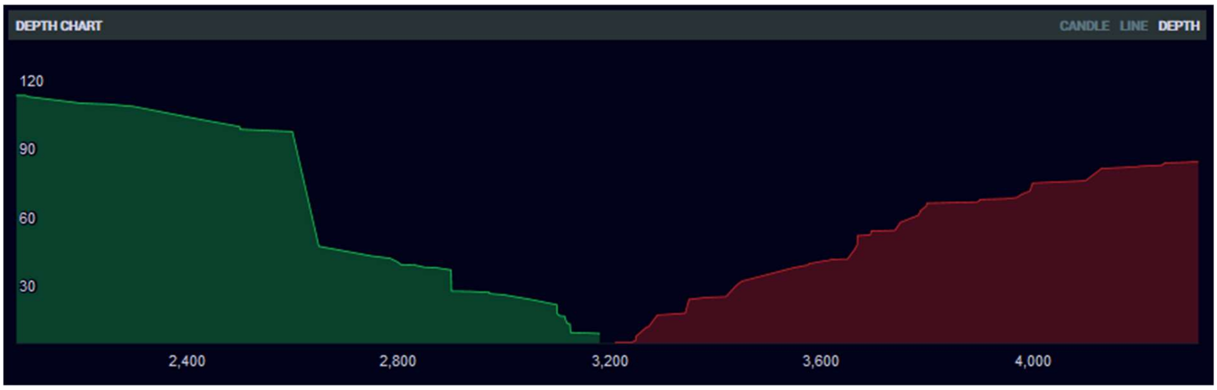

Figure 4. Depth chart – BTC/ZAR on 1st April 2015, Luno platform

The depth chart illustrates the cumulative sum of the volumes available at each price level. Again, it is helpful in determining the ability of a particular market to absorb large sell/buy orders without the underlying price moving in either direction, in discovering the next probable price move and in detecting support or resistance levels.

The solid candlesticks chart is a classical representation of the evolution over time of the underlying price. It does not show all the information contained in the order book, but it is a support tool valuable to keep in mind also the time dimension. In such a chart, each bar represents the four main pieces of information for the time interval considered (in Figure 5 each candlestick corresponds to one trading day): the open, high, low and close values. The filled portion of the candlestick is called "the body", while the long thin lines above and below the body are called "shadows" and they represent the high/low ranges. The high is marked by the top of the upper shadow, the low by the bottom of the lower shadow. The extremities of the body instead represent the open and the close values. As it can be easily seen in the figure below, the red colour stays for a selling pressure where the close is lower than the open. Vice versa, further the close is above the open longer the green area is. The grey bars on the background indicate the whole volume traded in the considered time interval.

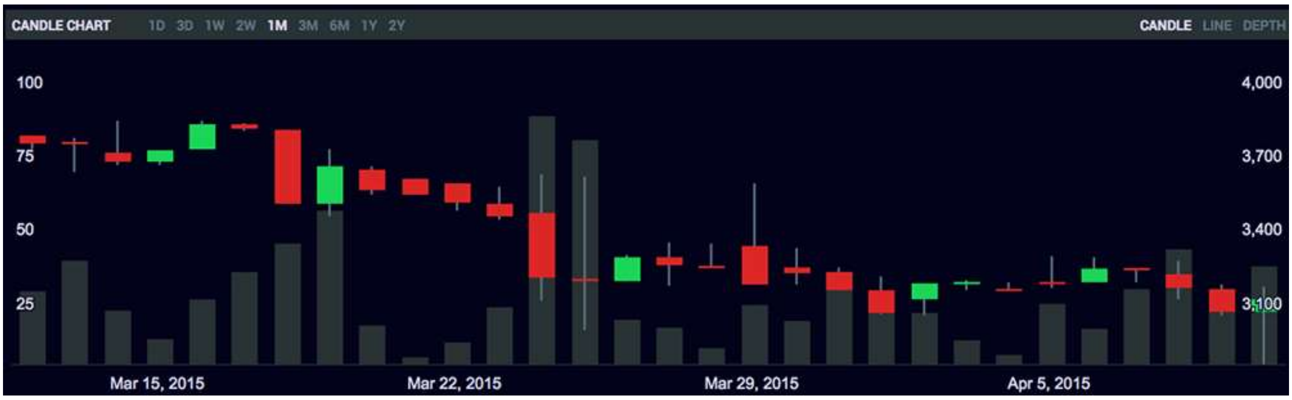

Figure 5. Candlesticks chart – BTC/ZAR, Luno platform

Heatmap

The charting tools we presented so far are limited in displaying instantaneous information and even combining them with the filled candlesticks chart they just provide a filtered or aggregated perspective. This lack of transparency does not allow traders to have a real insight into the market microstructure and to understand more complex dynamics involving price triggered strategies, algorithm activities, big players' moves or hidden liquidity discovery.

It is really important to highlight that the use of data visualization is not only fundamental in real-time sessions when a trader has to take immediate decisions or to test whether his or her own strategy is effective or not, but also in early stages when cleaning and checking the data accuracy is essential to develop a new strategy or to improve an existing one. Moreover, since the electronic trading has extensively depleted the ability of market participants to self-regulate themselves by hiding traders' identities and allowing manipulative practises to go more easily unrecognized, visualizations are also valuable tools for supporting a range of activities within the regulatory sphere including monitoring, enforcement and general oversight.

In order to solve the above mentioned issues in the last three years there has been developed a new visualization tool: the heatmap, whose real innovation has been to include a historical representation of the whole order book. In addition, previous tools have the common shortcoming to consider only a small set of the all data available in an order book, the heatmap overcomes it by giving the market participants a full insight to the whole market activity and it consists of 100% of the market data at a glance.

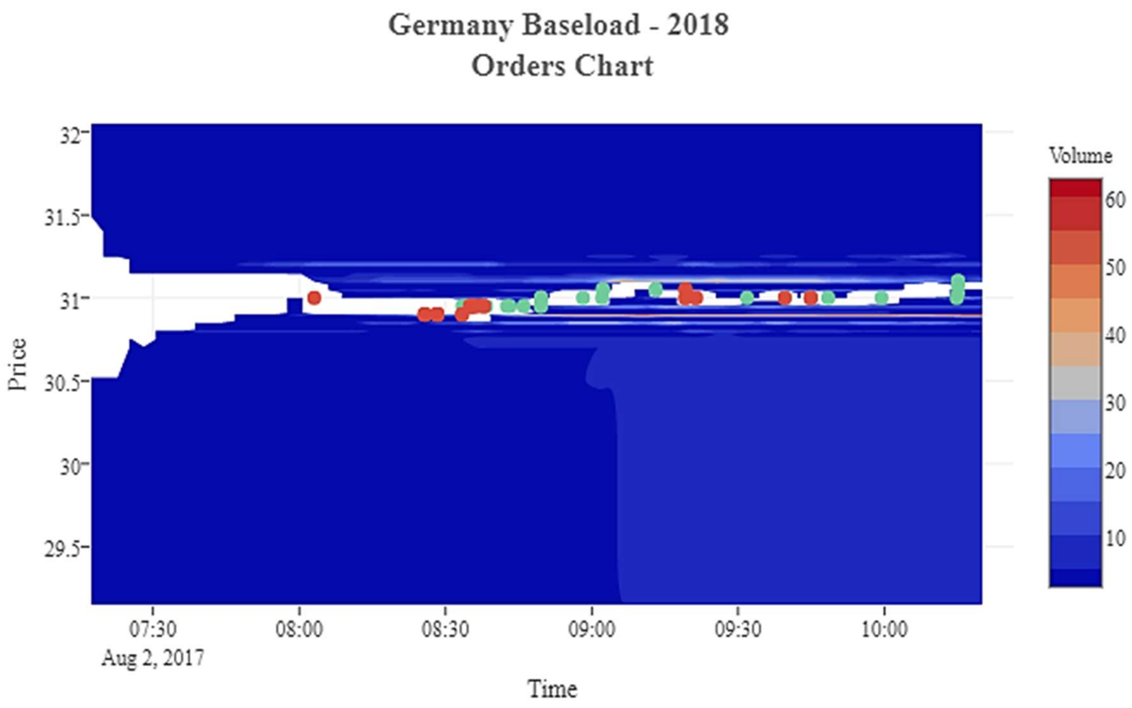

Figure 6. Heatmap – Germany Baseload 2018, EEX

It is constructed using a rectangular tiling of a data matrix. This tiling eases the inspection of three dimensional data using rows, columns, and a third attribute, colour, as it changes with respect to the first two. This allows large data matrices to be displayed effectively on a high-resolution colour images.

The heatmap showed in Figure 6 allows a trader to scrutinize liquidity expansion and contraction over a selected period of time. It is updated very frequently, indeed it records and visualizes every change in the order book by displaying it on a scale of different colours. Those colours are applied at specific price levels: the darker blue shades mark areas with a low number of resting orders while brighter (or warmer) shades mark price levels of higher liquidity. The green and red dots indicate the occurrence of a match, respectively initiated by a buy or a sell action.

Moreover, the traditional depth view can only provide continuously updated numerical representations of the market. For all those traders that use the depth of the market for trading, watching tick by tick market changes can be tiresome, time-consuming and prone to errors. This technology gives a clear view of how the entire limit order book and traded volume evolve over time, enabling a faster and deeper understanding of the market dynamics.

Also, when a user sees an abundant size liquidity in the individual price levels, he or she cannot gauge with old-fashioned platforms how long is it in the market or how it was changed during the past 1 hour or how it reacts to price movements. Most traders are focused on the volume that is being traded, which is the past.

Consider now an upward trending market as illustrated in Figure 7 and notice how the heatmap helps in following and illustrating the context of volume and liquidity evolution. The chart shows the situation of the E-mini Nasdaq 100 Index Dec 2017 on 17th October. In this case the heatmap uses a grey colour scale for indicating liquidity, where brighter shades indicate more resting orders. The dots instead represent the volume that has been traded in a pre-selected time period: the bigger the dot the much more trades have occurred. Again the colours green and red stay for buy and sell orders.

In an up-trending market we usually see more aggressive volume on the buy side trading at higher highs. Looking at the 5 minutes interval after 10.45 a.m. we can notice more aggressive buying pulling the market up into newer highs. Indeed the dots are larger than previous and next intervals and they are mainly green coloured. At the same time there is less volume and less aggressive selling pressure at the higher lows. The increasing path is followed by a limited sideways move. Zooming further, the same pattern, buying volume and sideways action, is repeated many times as illustrated by yellow circles. This is indicating that the trend has started lifting the offer into new levels. Also, it can be interpreted as a signal of a near-future price increase as it actually does between 11.15 and 11.20 a.m. It interesting how this behaviour is related with liquidity. The targets are high liquidity areas. Once the price reaches those zones, large transactions take place and the market goes sideways for a bit. Meanwhile the liquidity flips from the offer over the bid supporting the breakout to the upside.

Figure 7. E-mini Nasdaq 100 Index 100 Dec 2017, CME

4. BOOK DEPTH ANALYSIS

4.1 LITERATURE REVIEW

In the literature there are three closely related branches which concern with the main topic we are presenting. The first involves the market transparency, that is how much information traders can retrieve from the process of trading. In particular, the degree of pre-trade transparency determines how much order-flow information is available. The second considers the choice of order type, market or limit order, and how the structure of the limit order book and agents' preferences affects it. The last connected theme examines the information content of the LOB, that is whether its structure is informative regarding the future price movements.

Pre-trade Transparency

Over the last two decades global financial markets have been widely improved the disclosure of limit order book, increasing considerably the pre-trade transparency. This higher availability has significantly affected the information contained in the market depth. Indeed, the accessibility of pre-trade information on the trading intention of other participants influences the pattern of market reactions and agents' behaviour such as return and volatility. That is the reason why substantial attention has been given to higher disclosure of data in academic research. Regarding the question of whether it facilitates price discovery, mixed evidence have been found by the extant researches. Baruch (2005) designed a theoretical model showing that an open limit order book enhances both liquidity and price efficiency. Confirming those findings, Boehmer, Saar, and Yu (2005) discovered that transaction prices started to deviate less from the efficient prices once New York Stock Exchange adopted the open book mechanism. Bortoli et al. (2006) investigated the effect of the increased market disclosure, from the best price level to the top three levels, for the four most actively traded future contracts in the Sidney Futures Exchange in January 2001. Their attention was on the trading behaviour and liquidity. They showed that the depth decreases importantly at the best bid and ask level while the very small change in BBO spread was not significant at all. Their result suggested also a relevant increase in the use of aggressive market orders. On the other hand, Madhavan, Potter, and Weaver (2005) stated that a greater pre-trade transparency does not guarantee an improvement in the overall market quality. Indeed, they found that the disclosure of the top four price levels in April 1990 at the Toronto Stock Exchange caused larger spreads and higher volatility, increasing the transaction costs and eroding liquidity.

The Choice of Order Type

As seen also before, one of the fundamental functions of a financial market is the price discovery process. Private information is incorporated into prices through the choices of the informed traders, who can decide to submit market or limit orders in their dynamic strategies. If they trade employing limit orders, their superior knowledge is likely reflected in the book, especially when the orders are passive (away from the inside market). If, however, informed agents use market orders, the levels of the book beyond the best bid and best ask may not include any additional useful information.

Early microstructure models such as Glosten (1994), Rock (1996) and Seppi (1997) implicitly assumed that the limit order book cannot possibly be informative for the evolution of future prices. They indeed considered limit orders posted by uninformed market participants as free options susceptible to being picked off by later better-informed investors. They argued that informed traders enter the market aggressively to exploit their short-lived private information favouring market orders in order to guarantee an immediate execution. Furthermore, since the direction of prices is conditional on the private information, the execution likelihood of limit orders designed to exploit a trader's advantage is almost null. For instance, an informed investor who knows that the actual price at which the security is traded is too low will expect the price to increase in the near future, especially when the same information benefit is disposable to other investors. Therefore, the probability that a limit buy order will match is relatively low in this scenario.

Other theoretical papers allow informed traders to decide between market and limit orders. The majority of these studies contradicts the prediction of the previous literature, stating that informed investors post also limit orders in a rational expectation world. For example, Chakravarty and Holden (1995) demonstrated in their work that an optimal order placement strategy involved the submission of both types of orders. Parlour (1998) showed a dynamic model where the combined effects of the current liquidity situation and of the place in the order book queue impact on the decision between a limit order or a market order. Foucalt (1999) studied a market with heterogenous asset valuation among agents with no private information. His dynamic order placement model showed the existence of a trade-off between strategies that involve limit orders and strategies that involve market orders, depending on the volatility of asset returns and bid-ask spreads. In addition, Handa and Schwartz (1996, 2003) examined the role of asymmetric knowledge on the order placement strategies. Their findings stated that a limit order strategy can be profitable whenever the expected gain from limit order matching is higher than the cost of being picked-off by an informed participant.

In addition, Biais, Hillion, and Spatt (1995) showed that price revisions tend to move in the direction of previous limit order flows. Anand and Martell (2001) empirically demonstrated that, after controlling the characteristics of the orders, the price trend after the matching was beneficial to limit order traders. Moreover, they showed that limit orders submitted by individual investors (uniformed/liquidity traders) performed poorer than limit orders submitted by institutional (informed) traders. Therefore they suggested that, once institutional agents have learnt an extremely beneficial private information, they maximize their trading profits and reduce the risk of uncertain trading prices using limit orders. Contrary, Griffiths et al. (2000) showed that in the Toronto Stock Exchange limit orders are subject to be picked off by better informed traders, suggesting that they tend to impact negatively on prices and hence they may not contain much relevant information.

More recently, Ranaldo (2004) studied how an investor’s strategy is conditional on the state of the limit order book. He suggests that patient traders submit more market orders and hence they are more aggressive when their side of the book is thicker, the spread is wider and the volatility increases. Kaniel and Liu (2006) built a simple equilibrium model where private information is considered long-lived and available to a small number of agents. They showed that informed traders may prefer limit orders, and hence the book contains useful information. In this case, posting market orders signals impatience and reveals too much information. Therefore market orders may cause higher transaction costs, although their execution is certain and immediate.

The Information Content of the Limit Order Book

Regardless of the channel through which information gets embedded in the limit order book, the common result of the above mentioned works is that market depth should convey relevant information for the true value of the underlying asset traded and, thus, it should be indicative of future price movements. This information content and in particular how to exploit it is the main objective of third branch of the literature on the order book trading. Even in the absence of asymmetric information, the shape of the limit order book (i.e. the number of shares on each price level and how far away price steps are from each other) gives market participants an instantaneous picture of the market supply and demand. More in detail, the asymmetries between the buy and sell side of the book signal shifts in the demand and supply curves caused by unobservable exogenous factors which impact on the stock prices. Studying and analysing the market depth, agents have more chance of guessing what these determinants are and of forecasting the evolution of future price.

Among first to investigate further this concept, Harris (1990) studied two types of limit order traders: value-oriented and pre-committed agents. The former assign their asset valuation through their choice of limit price, while the latter use limit orders to reduce transaction costs, but will opt for market orders if their instructions stay unfilled for too long. Both actions may be helpful in predicting future movements of the stock price. A book imbalance caused by value-oriented agents will signal how they value the underlying asset and then this value will be incorporated in prices. In the other cases, an informative imbalance may also depend on the fact that pre-committed traders switch their unfilled orders into market orders. For instance, a thicker sell than buy side would reveal a future price decrease. The same author (Harris, 1990) stated that price evolution and asymmetries in the shape of the two sides of the limit order book may be joint by the behaviour of quote-matchers. The "quotematching" strategy consists in extracting options values from the standing active orders by trading ahead of the heavier side of the book.

More recent literature has studied the empirical determinants of the order aggressiveness. In an experimental paper, Bloomfield, O'Hara and Saar (2005) showed that informed traders exploit their information advantage early in the trading period to look for mispriced standing orders moving the market toward the right price, and therefore progressively reduce their knowledge value. As the closing of the trading period comes closer, they switch increasingly to limit orders, as the value of their informational advantage falls away. Analysing SuperDot limit orders in the TORQ database of the NYSE, Harris and Panchapagesan (2005) reaffirmed that specialists exploit the ability of the order book shape to indicate short term changes in prices. Evans and Lyon (2006) found that the predicting power of clients' order flow is able to outperform a random walk benchmark: result that has been contested by two studies, Danielsson et al. (2002) and, Sager and Taylor (2008), which documented limited and no evidence of superior forecasting ability of order imbalance strategies over random walk models at different forecast horizons. However, Latza and Payne (2010) showed that both limit and market orders can be indicative on stock returns. They also documented how this power was greater for limit orders flows, in particular in presence of a liquid market.

Based on the open book from the Australian Stock Exchange, Cao, Hansch and Wang (2009) documented the incremental information content of the price levels behind the best bid and offer. According to their work, the contribution of the LOB to price discovery accounts approximately for the 22%, while most of the variation in future returns comes from the best bid, ask and transaction prices. They then found that order imbalances between demand and supply schedules along the book are statistically significant to predict future short-term returns. Hautsch and Huang (2011) estimated the impulse response functions for the thirty stocks listed at Euronext Amsterdam: the limit orders, especially for orders posted on up to two steps behind the market price, have a significant effect on quote adjustments. Lin et al. (2012) studied the TAIEX futures, a benchmark of the Taiwan equity market and a market capitalization weighted index composed of all the ordinary stocks listed. Using the best five quotes of the limit order book, they proposed a trading strategy, which was able to earn positive returns even when transaction costs were taken into account.

Overall, the above-studies support the hypothesis that the order book contains useful information, in particular in the price levels immediate behind the best quotes. However they all are focused on equity or forex markets, where normally there is not any liquidity issue. Moreover, almost all of them analysed spot markets.

In the following chapters we present and examine the information content of an open order book, using data about Germany Baseload 2018, a power futures contract traded in the European Energy Exchange. One of the main characteristics of this futures market is that many market participants operate for the purpose of hedging and risk management. This issue, plus the peculiarity of the contract considered, makes our dataset very low liquid and volatile.

4.2 ORDER BOOK IMBALANCE MEASURES

The relationship between trade activity (volume) and price change and volatility has been one of the prominent field of study in the finance literature. At any point in time, a limit order book is populated by a large number of buy and sell orders. As traders submit limit orders to buy or sell, they impact on the bid or ask volume of the book and hence they give other participants a perspective of their trading intentions. Classifying trade volume as either taking the bid or ask would permit us to obtain a better understanding of the upcoming price changes. The *Order Imbalance* measure is used to identify this price direction and in its simplest form can be defined as the difference between the bid and ask volume. The Order Imbalance is a relevant indicator that allows us to gain a deeper insight into the general sentiment and into the next probable move of the market. Whenever informed agents obtain news that have not been incorporated into the asset price yet, they can take a long or short position depending on the grade of information positivity and consequently affecting the imbalance on the asset. Who, instead, is merely looking at this descriptor in the LOB would exploit it and enhance his or her strategy to generate positive returns.

Volume Order Imbalance Ratio

Many different imbalance measures have been proposed over years. The most common representation is given by the *Volume Order Imbalance Ratio (VOI)*, that is the difference between buy side and ask side volumes normalized for the total volume available at a specified lowest depth having non-zero volume

$$VOI(t) = \frac{V_b(t) - V_a(t)}{V_b(t) + V_a(t)}$$

where $VOI(t) \in (-1; +1)$, and both V_a and V_b are computed as the weighted average volumes at the n-th best price levels. More positive (negative) values indicate an imbalance in favour of the bid (ask) side.

Trade Imbalance

The *Trade Imbalance (TI)*, instead, is measured as the imbalance between the executed buy and sell volumes in a certain time interval (Δt). Buy volume and sell volume are classified when marketable buy/sell orders are matched to active limit orders sitting on the other side of the book. Hence, Trade Imbalance summarizes the net transaction volume in the considered time interval.

$$TI_{\Delta t} = \sum (b_{\Delta t} - s_{\Delta t})$$

where $b_{\Delta t}$ and $s_{\Delta t}$ are buyer- and seller-initiated trades.

Order Flow Imbalance

Focusing on the Level I order book, Cont et al. (2011) introduced the *Order Flow Imbalance*, defined as the imbalance between supply and demand at the best level, which encompasses trades, limit orders and cancelations. Each observation of the bid and ask consists of the bid price P^B and the size q^B of the bid queue (in number of shares), the ask price P^A and the size q^A. Enumerating them by n and comparing two subsequent observations, one of the following events can occur:

- $P_n^B > P_{n-1}^B$ or $q_n^B > q_{n-1}^B$ signifying an increase in demand;
- $P_n^B < P_{n-1}^B$ or $q_n^B < q_{n-1}^B$ signifying a decrease in demand;
- $P_n^A < P_{n-1}^A$ or $q_n^A > q_{n-1}^A$ signifying an increase in supply;
- $P_n^A > P_{n-1}^A$ or $q_n^A < q_{n-1}^A$ signifying a decrease in supply.

Hence, they have defined the variable e_n which measures the contribution of the n-th event to the size of the bid and ask queues:

$$e_n = I_{\{P_n^B \geq P_{n-1}^B\}} q_n^B - I_{\{P_n^B \leq P_{n-1}^B\}} q_{n-1}^B - I_{\{P_n^A \leq P_{n-1}^A\}} q_n^A + I_{\{P_n^A \geq P_{n-1}^A\}} q_{n-1}^A$$

If P^B does not vary, the variable gets the value $e_n = q_n^B - q_{n-1}^B$, regardless of an increase or a decrease in q^B. If q^B increases, e_n indicates the size that was added at the bid. If instead q^B decreases, the variable measures the size that was removed from the bid, whether due to a market sell or a cancel buy order. When P^B increases, $e_n = q_n^B$ represents the size of a price-improving limit order. If P^B decreases, $e_n = q_{n-1}^B$ indicates the size that was removed, whether due to a market order or a cancellation. With signs reversed, the same classification holds for the ask side.

The Order Flow Imbalance (*OFI*) is then defined as the sum of individual event contributions e_n over a specified time interval

$$OFI = \sum e_n$$

Limit Order Book Imbalance

Another proxy of the order book shape, the *Limit Order Book Imbalance (LOI)* can be simply obtained by taking the difference between the volume weighted price on the top n price levels and the mid-quote.

$$LOI = \frac{\sum(P_k^B q_k^B + P_k^A q_k^A)}{\sum(q_k^B + q_k^A)} - MID$$

where $MID = \frac{bestAsk+bestB}{2}$ and k indicates the price level.

Step-wise scaled imbalance in the length and in the height

In accordance with Cao and al. (2009), the aggregate market demand and supply are represented by limit orders as step functions of the accumulated number of shares offered at each price level. They then calculated the height and the length of each step for the demand and the ask side of the book respectively.

- The $Height$ of a particular price level k on the demand side is defined as follows $H_k^B = P_k^B - P_{k-1}^B$. The height at the first step is obtained by taking the difference between the price at the best first level and the mid-quote;

- the $Length$ at a step k on the demand side of the book, Q_k^B, is the aggregate number of shares across all orders at price P_k^B;

- the heights and lengths of steps on the supply side are defined analogously.

In order to capture the aggressiveness of the orders submission, these last two indicators are then used to calculate the *step-wise scaled imbalance in the length (QR) and in the height* (*HR*) using the following formulae:

$$QR_k = \frac{Q_k^A - Q_k^A}{Q_k^A + Q_k^A}$$

$$HR_k = \frac{(P_k^A - P_{k-1}^A) - |P_k^b - P_{k-1}^b|}{(P_k^A - P_{k-1}^A) - |P_k^b - P_{k-1}^b|}$$

where the HR indicates the revealed differences in price between two subsequent quotes. When there exists a greater competition among the buyers, they are likely to submit more aggressive orders. Hence the price difference should be smaller. If it is the situation, the HR parameter has a positive value forecasting a higher likelihood of an upward move in the short-term future prices. The contrary, that is negative value of HR, predicts a greater chance of a decline in prices. QR is measured from the perspective of market demand and supply. A positive value means that the buying side has lower depth, that is less investors want to buy the underlying contracts. In these instances, more shares are submitted to the ask side, this drives the price down due an excess in the supply. In the opposite scenario the price will move upward.

Order Book Slope

We conclude our brief list of order book shape indicators introducing the *Order Book Slope (SLOPE)* measure. Following Naes and Skjeltorp (2006), for the asset i in the interval t the predictor is defined as follows:

$$SLOPE_{i,t} = \frac{DE_{i,t} + SE_{i,t}}{2}$$

where $DE_{i,t}$ is the slope of the demand (bid) side and $SE_{i,t}$ stays for the slope of the supply (ask) side. The order book slope for the demand side for asset i in the interval t is given by:

$$DE_{i,t} = \frac{1}{N_B}\left\{\frac{v_1^B}{|\frac{p_1^B}{p_0} - 1|} + \sum_{k=1}^{N_B-1} \frac{\frac{v_{k+1}^B}{v_k^B} - 1}{|\frac{p_{k+1}^B}{p_k^B} - 1|}\right\}$$

Similarly, the order book slope for the supply side can be calculated as:

$$SE_{i,t} = \frac{1}{N_A}\left\{\frac{v_1^A}{\frac{p_1^A}{p_0} - 1} + \sum_{\tau=1}^{N_A - 1} \frac{\frac{v_{k+1}^A}{v_k^A} - 1}{\frac{p_{k+1}^A}{p_k^A} - 1}\right\}$$

where N_A and N_B are total number of offer and buy price levels with non-zero size, respectively. k indicates the price level, with $k = 0$ denoting the best bid-ask midpoint and $k = 1$ denoting the best bid/ask quote with positive share volume. p_o represents the best bid-ask midpoint, while v_k^A and v_k^B are the natural logarithm of the aggregated total volume at the price level k (p_k). In other words, v_k^A (v_k^B) is the natural logarithm of total the share volume supplied (demanded) at the quote level p_k. Essentially, the Order Book Slope is an average elasticity measure ${}^{\delta q}/_{\delta p}$ that describes how the supplied quantity q changes with respect to the price p across all price levels with non-zero size in the limit order book. The more concentrated (widely distributed) the shares size in the LOB are, the steeper (more gentle) the slope will be. Empirical findings (Duong and Kalev, 2008) demonstrated the negative significant relation between the slope variable and the future volatility, that is the volatility decreases the more steeper the slope is.

5. *INTRODUCTION TO ELECTRICITY MARKETS*

5.1 ELECTRICITY AS A COMMODITY

A commodity is a tradable economic asset that has an intrinsic value. It has to be sufficiently standardized in relation to some fundamental characteristics in order to be completely fungible. In other words, the quality of a given commodity may differ slightly, but it is essential uniform across different producers. Hence, commodities are undifferentiated goods with equivalent quality standards that are exchanged without any additional added value. If traded on a trading venue, they must also meet specified minimum standards, known as a basis grade.

Electricity is an energy commodity with unique economic and technical characteristics. It is used for a very wide range of applications because it is easy to control, non-polluting at the location of its usage and convenient. As a secondary energy source, it is generated from the conversion of natural gas, coal, hydropower, nuclear power and other renewable sources. This implies that electricity markets and prices are fundamentally linked to markets for primary fuels and environmental conditions. Hence, it is essential to consider both the power generation processes and the primary fuel markets in order to understand the mechanism that drives electricity markets.

From the perspective of the consumption side, the energy demand shows a high variability, both on the short and medium-long term. During daytime there is an alternation between peak hours (from 8 a.m. to 8 p.m.), when there is the maximum need, and off-peak hours, when the power need is relatively low. Considering instead a long-term perspective, the energy demand of each country shows different seasonality patterns related to the social and climate characteristics of that region. A second relevant property is the demand rigidity, that is the low elasticity with respect to a price change due to the fact that electricity is an essential and hardly replaceable good. Moreover, it is considered a homogeneous product because electricity does not have qualities that can change the consumer’s liking. In contrast, from the perspective of the supply side, the technologies and production costs are heterogeneous and they depend significantly on the type of fuel used. A mix of diversified power generation techniques allows to change the utilization factor of each industrial plant in order to meet the demand at the lowest possible cost.

From the technical point of view, the storage impossibility implies that the energy commodity cannot be purchased in periods of excess production to be resold in those of shortage. This causes a huge impact on the market infrastructure and organization because the electric power is useful if and only if it can be delivered in one period of time. Indeed, on energy markets, market participants exchange contracts which provide the delivery of the underlying in a pre-specified time interval. The unit of measurement is the Megawatt-Hour (MWh): the energy produced by supplying the power of 1 MW for 1 hour. A related characteristic resulting from the non-storability is the high volatility of spot[2] prices in case of a tight or excessive supply situation. In a forward/futures[3] market the price movements are much smaller because the availability of power plants and the weather-dependent demand are still unknown.

Moreover, the lack of storability requires an exact matching of supply and demand at all times. The energy must be supply whenever it is requested, and consequently complex dispatching operations are needed in order to ensure the continuous balance between demand and supply. The necessity of a dispatching network in turn prevents a global market. The task to balance the system is performed by the Transmission System Operator (TSO), who charges the merchant directly or the retail clients via a transmission fee for its service. The TSO specifies a balancing period, that is the granularity of the measured electric energy supply. The continuously volatile needs of the retail customers are integrated over the balancing period and the average power is the forecasted size that should be delivered by the supplying merchant.

---

[2] A spot contract is a financial agreement between two counterparties, the seller V and the buyer A, that establishes the quantity of the commodity traded and the amount to be paid in exchange. V has the obligation to physically deliver to A a certain amount x of the commodity on the stipulation date t, meanwhile A must pay V a fixed sum $S(t)$, known as spot price of the commodity in t.

[3] A futures contract is a standardised financial agreement between two counterparties, the seller V and the buyer A, that establishes the price and the quantity of the commodity traded that the former must deliver to the latter on the T date, called expiration date, later than the stipulation date t. V has the obligation to deliver to A a certain amount x of the commodity in T, meanwhile A must pay V a fixed amount $F(t,T)$, the futures price. $F(t,T)$ is the commodity price in t with delivery in T; the time interval (t,T) is defined as Time-To-Maturity.

Consequently, the merchant delivers electricity as a discrete time series with time intervals according to the balancing period and constant power during this time periods.

All the principal products traded in the electricity markets are delivery schedules in a granularity not finer than the balancing period. Figure 8 illustrates the different categories of an electricity market and the corresponding time flow.

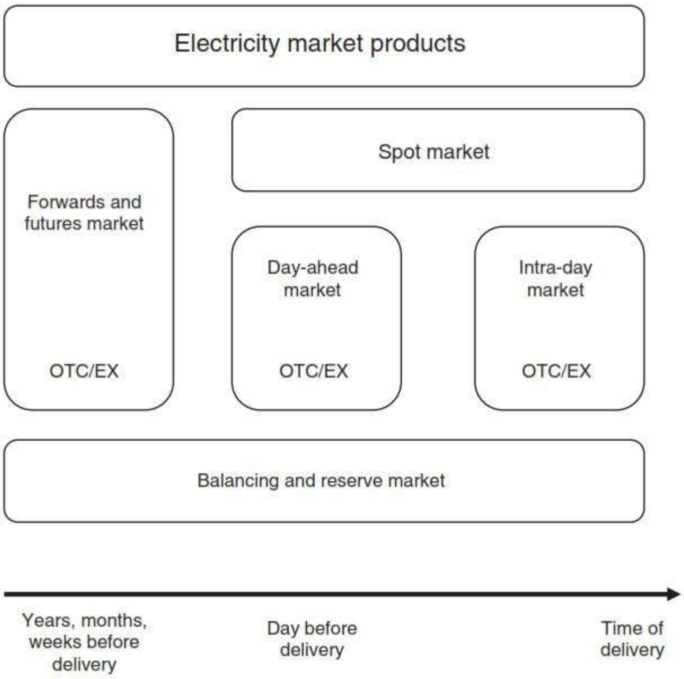

Figure 8. Categories of the electricity market

In general, the electricity markets can be divided into:

- Forward and Futures markets are used both for risk management, hedging and speculative purposes. It is also the relevant venue for traders who actively take positions and thereby provide liquidity for hedgers.

- Day-ahead market, where products are exchanged with a delivery on the next trading day. Day-ahead products are common spot contracts and can be traded on a power exchange or as bilateral agreement (Over-The-Counter[4]). Normally the spot market is taken as underlying for the reference price for the forward and futures market.

- Intra-day market whose products are exchanged with a delivery on the same day. This type of market allows the power producers a short-term load dependent optimisation of their generated quantity and typically it is not used for pure trading purposes. As for day-ahead contracts, intra-day products are traded OTC or on a power regulated exchange.

- Reserve and Balancing market. The first allows the transmission system operator to purchase at short notice the quantity needed for compensating temporary imbalances between the demand and supply in the electricity system. In a balancing market a merchant sells or buys the additional energy needed for taking in equilibrium his accounting grid. The balancing market can be considered a market only in a broad sense due to the fact that the balancing operations are provided by the TSO and it gets a fee or pays the merchant for the additional energy. Only in some countries the merchant has the choice to buy or sell this additional amount of energy from or to someone else.

[4] Over-The-Counter (OTC) operations involve transactions of securities that are not listed in a regulated, formal exchange. They represent non-standardized contracts that are traded through a private dealer network which negotiates directly with buyers and sellers.

5.2 FUTURE POWER CONTRACTS

Since our analysis is based on a futures energy contract, we analyse deeper its characteristics. Energy contracts differ from futures agreements on other more generic commodities because they are settled into a delivery period, instead of an expiration date, during which the long counterparty receives the specified quantity and pays the price established on the stipulation. Normally, delivery periods may have a weekly, monthly, quarterly or yearly duration and each contract defines the provision of a constant electrical power of 1 MW over the entire delivery period. Moreover, energy futures are divided into base-load and peak-load, depending on the daytime interval in which the power delivery is guaranteed. The definition of peak hours depends on the trading venue, but often the interval is (8 a.m. – 8 p.m.) from Monday to Friday. Considering a monthly base-load futures, the underlying is defined as follows:

1 MW * 24 h/day * 31 days = 744 MWh

More generally, a base-load contract over a n-days delivery period will oblige to provide electricity equal to 24n MWh, constantly delivering the power of 1 MW. Generally, the minimum price fluctuation is equal to 0,01 €/MWh.

As in the most common exchanges, the energy futures can be distinguished into those physically delivered and those settled in cash based on a reference price. The determination of the final settlement price is based on an index which is the mean value of all auction prices of the hourly Day-Ahead contracts traded for the respective market area and delivery time (Base/Peak) of the respective delivery period. Since financial energy futures do not require the effective electricity production and consumption, they are useful both for hedging purposes and for arbitrage or speculative reasons.

In addition, all futures on energy markets are characterised by two fundamental properties:

- the Samuelson effect (Samuelson, 1965) states that the futures volatility is decreasing with respect to the time to delivery, hence it is lower the longer is the Time-To-Maturity (TTM). Moreover, the decay is exponential and the volatility tends to a constant strictly positive value when the expiration period converges to infinity. Samuelson suggests that the most relevant information on the underlying asset becomes common knowledge only in the proximity of the futures maturity; therefore, it does not impact on contracts with distant delivery. For instance, climate changes that affect the natural gas demand or temporary problems in the electricity transmission network will cause adjustments only in spot prices and shorter futures. Conversely, those contracts with a bigger TTM will show a long-term volatility which is the result of other macro-factors, such as technology innovations or geopolitical issues that can influence the commodity value.

- Seasonality: the climate and behavioural characteristics of the specific region, where the commodity is needed, are among the main drivers of its price. Those factors show time cycles both on the short and medium term. For example, the natural gas demanded by residential users depends strongly on the season temperature. Since these cycles are mirrored into the information flow that determines the reference prices, the volatility is expected to manifest them too.

As for more generic commodities, the margins system is used in order to guarantee the transaction success. On the stipulation date the counterparties deposit to the Clearing House the initial margin and, later, at the end of each trading day the party, who has accrued a loss, pays the variation margin to restore the previous equilibrium. In the case of contracts with delivery time longer than one month, the cascading mechanism is required. In the case of a quarter future, the cascading process implies that on the third business day before the beginning of the delivery period, each open position in a Quarter Future is replaced by equivalent positions in the three Month Futures whose delivery months together correspond to the delivery quarter. An example would explain clearly how it works. At the end of the last trading day, a long position in a quarter future Q1-2018 will be replaced by three long positions in the futures Jan-2018, Feb-2018 and Mar-2018. Considering instead a yearly future, each open position will be converted into equivalent positions in the three Month Futures for the delivery months from January through to March and the three Quarter Futures for the second through to the fourth delivery quarter whose delivery periods together correspond to the delivery year.

5.3 THE EUROPEAN ENERGY EXCHANGE

The European Energy Exchange (EEX) is the leading energy trading venue in central and western Europe. It develops, operates and connects secure, liquid and transparent markets for energy and related products on which power, natural gas, CO_2 emission allowances and coal are traded. It involves both spot and futures transactions and beside them, also Over-The-Counter operations are allowed to be registered via its fully electronic platform. It was born in Leipzig in 2002 as the result of the merger between two previously existing exchanges: the Leipzig Power Exchange and the European Energy Exchange located in Frankfurt. Since then, it has grown into a global market group through a series of partnership and acquisitions throughout Europe, Asia and US. Besides its flagship market in Germany, the group consists of the following companies:

- the European Power Exchange (EPEX SPOT), that operates physical short-term electricity markets in Central Western Europe and the United Kingdom;
- the Powernext, a market facilitator based in Paris that develops tailor-made solutions and operates the Pan-European Gas Platform PEGAS;
- the Prague-based Power Exchange Central Europe (PXE), that is committed to further developing products and services for the Czech, Slovak, Polish, Hungarian and Romanian market;
- the Danish gas exchange Gaspoint Nordic that offers short-maturity products for physical trade of natural gas;
- the regulated futures venue Cleartrade Exchange (CLTX), based in Singapore. It specializes in Freight, Ferrous Metals, Agricultural and Energy markets in Asia;
- the Virginia-based Nodal Exchange, that provides price, credit and liquidity risk management solutions to participants in the North American energy markets;
- two clearing companies, the European Commodity Clearing (ECC) and Nodal Clear.

Currently, the group holding company is a subsidiary of Eurex, the largest European derivatives and options market, which in turn is under the Deutsche Börse umbrella of companies. Overall the entire group has extended its commodities variety, offering also contracts on oil, freight, metals, environmental and agricultural products. In 2017 it has served more than 500 trading participants from more than 30 countries.

Spot transactions are conducted both in the trading type of continuous trading and auctions, while futures derivatives contracts are traded only via the former one. The opening price is determined during the opening auction on the basis of both limit and market orders contained in the trading system and it shall be the price at which the largest possible number of contracts of such orders and quotes may be executed (principle of maximizing executions). During continuous trading instead, the exchange prices are determined by matching orders at the best possible bid and ask limits indicated in the order book. In the event that prices are identical, orders and quotes are matched in the order in which they were entered into the EEX system (price-time priority); unlimited orders are executed first. All available orders shall be displayed cumulatively at the respective limits (open order book).

Focusing on spot markets, an open auction is generally divided into the call phase and the price determination. During the call phase, exchange participants may enter, change or delete orders. If there are orders that could be executed against one another, in auctions with a closed order book, a potential execution price is displayed during the call phase. If this is not the case, the best buy and/or sell limit is displayed. In auctions with an open order book, the cumulated order volumes of each of the buy and/or sell limits are also displayed. Neither a potential execution price nor order volumes are displayed in a closed auction.

All the contracts offered by the European Energy Exchange are traded throughout the T7 trading infrastructure. This is a proprietary technology of the Deutsche Börse and indeed it is also being used by Xetra (the reference market for German equities and ETFs) and the Eurex Exchange (the largest European futures and options market); in addition to the american International Securities Exchange (ISE) and the Bombay Stock Exchange (BSE).
The EEX products can also be exchanged throughout the Trayport trading platform, specifically designed to meet the particular needs of the energy market for physical and financial trading. The London-based company does not own or operate markets, it is instead an experienced OTC and Exchange software provider. Initially used by energy inter-dealer brokers as a price dissemination platform, their product range has evolved significantly and nowadays it offers also trading and clearing solutions in a multiple markets configuration.

6. *TESTING THE LIMIT ORDER BOOK INFORMATIVENESS*

6.1 DATA

The data used in this study are extracted from the Trayport Trading System. The dataset considers the period from 3rd October 2016 to 28th February 2017 and it accounts for all OTC and EEX transactions that occurred on the contract Phelix DE Baseload 2018. The futures specifications are designed by the European Energy Exchange and it defines the provision of electricity over the full year 2018 (365 days) for each hour of the day, constantly delivered at the power of 1 MW, for the market area Germany. It is traded throughout continuous trading mechanism from 8 a.m. to 6 p.m. CET without any opening or closing auction. The minimum tick size is equal to 0,01 €/MWh. As very simple classification, we can consider that OTC transactions involve the physical delivery, whereas EEX transactions are settled in cash. The underlying reference price for the final settlement is based on the hourly Day-Ahead auction prices determined by the EPEX spot for the corresponding market area.

The open limit order book makes no distinction between OTC and EEX transactions and it gives information until the best twentieth quotes. For each order, we have at our disposal the order type (market or buy), the order arrival date and time, the order direction (buy or sell), the order price and quantity. In the trades file instead, we collect information on the transaction date and time, and the transaction price. Any other information about the quantity traded or transaction directions, that is whether the transaction was buyer-/ or seller-initiated, was not available.

To avoid confounding effects due to the opening or the closing of the trading session, the authors restrict their attention to the period from 9 a.m. to 5 p.m.. Moreover, only the top five levels are considered in the empirical model; this is due to the fact that most of the time the fifth quote represents the deepest level with non-zero volume.

6.2 METHODOLOGY

As discussed in previous chapters, there has been empirical evidence in favour of the information content of the limit order book in equity and stock markets. More precisely the public information of an order book, determined by the number of shares on each price level and by how far away price levels are from each other, gives investors a better chance to guessing the unobservable exogenous factors that have an impact on returns. What we would like to test now is whether these findings are valid also for a completely different scenario like that provided by an energy market. Using some of the indicators introduced previously, we want to examine the predictive power of imbalance along the limit-order book on the future short-term returns.

From now on, we will explicitly refer to the model proposed by Cao, Hansch and Wang (2009) to verify the following hypothesis:

Limit orders behind the best bid and ask prices contain information about short-term future price movements.

A rejection of the hypothesis would imply that the make-up and shape of the limit order book are not related to, and hence do not convey any additional information about, short-term price movements beyond those contained in the inside spread and depth.

The methodology involves the examination of the relation between the returns of futures contracts and lagged order-book statistics that are constructed from the demand and supply schedules. In order to avoid any issues about missing data and/or transactions mispriced by investors, the returns have been obtained using the buying and selling price midpoints for the best quote instead of the futures transaction prices. The formula for the return is as follows:

$$MID_t = \frac{bestAsk + bestBid}{2}$$

$$r_t = \ln(\frac{MID_t}{MID_{t-1}})$$

Since our dataset showed a very low number of average transactions per day, this study experimented snapshots of the order book at different time intervals. We investigated samples taken every one-minute, five-minute, ten-minute and fifteen-minute. The five-minute interval has been proved to be the best balance between the need to have a sufficient large number of observation points and the need to let the futures price to experience a meaningful change between two subsequent observations. As short-term returns may exhibit autocorrelation, the authors run the Akaike Information Criterion (AIC) test to find the autoregressive model that fits best with our dataset. Based on the smallest AIC value, we used the AR(5) model to obtain innovation in returns as follows:

$$r_t = \alpha_0 + \sum \alpha_i r_{t-1} + \varepsilon_t^*$$

where r_t is the five-minute midprice return at t and ε_t^* is the return innovation.

Since the residual ε_t^* captures the unpredictable component of the returns, it has been then used as the dependent variable in the regression analysis. As independent variables, the authors took the relative inside spread, which is defined using the following formula:

$$Spread = \frac{bestAsk - bestBid}{bestAsk + bestBid}$$

and the step-wise scaled imbalance indicators in the length, QR_k, and in the height, HR_k, that we have defined previously in subsection 4.2. The relative spread variable and the scale imbalance in the quantity at level 1, QR_1, have been included to control for the information contained at the top of the order book that traders normally had access to at all times.

Hence, the regression model is defined as follows:

$$\varepsilon_t^* = \alpha_0 + \beta_0 Spread_{t-1} + \gamma_t QR_{1,t-1} + \sum_k \beta_k HR_{k,t-1} + \sum_k \gamma_k QR_{k,t-1} + \eta_t \quad (1)$$

During the empirical analysis, variables from different quotes, k, are progressively included into the model. Each regression studies the book information up to step n where n = 1,2,3,4,5. It has not been possible to include a deeper price level because most of the time the fifth quote represents the maximum depth quote that our futures contract had. As the main objective is to investigate the predictive power of the limit order book imbalances, a primary focus goes to the change in the adjusted R-squared as more steps are included in the regression. It, indeed, represents the proportion of the variance in the dependent variable that is predictable from the independent variables, while controlling for the number of predictors. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. Conversely, it decreases when a predictor improves the model by less than expected by chance. The results report also the F-statistic for the null hypothesis that all coefficients are jointly zero.

Intuitively, return is predicted to be positively related to $HR_{k,t-1}$ because smaller price increments on the demand side would imply a higher aggressiveness in buying rather than selling shares. Conversely, return is supposed to decrease with $QR_{k,t-1}$ as more volume on the offering side should have a negative impact on prices due to the excess supply.

The relation between future order submissions and the status of the order book has been documented by several studies conducted on order placement strategies (for instance Griffiths, Smith, Turnbull, & White, 2000; Hollifield, Miller, & Sanda, 2004). Their findings documented that more market sell (buy) orders are due to a large depth at the sell (buy) side of the book. These marketable orders consume the active limit orders on the other side of the LOB and they cause a decrease (increase) in futures prices. Hence, it is the feedback effect between the order placement strategies and the book that links future price movement and the current book status.

Equation (1) has then rewritten considering the supply and demand side of the book separately; that is the length and the height of steps on the supply and demand side are taken as independent variables. The new empirical model is

$$\varepsilon_t^* = \alpha_0 + \beta_0 Spread_{t-1} + \sum_k \gamma_{d,k} Q_{k,t-1}^d + \sum_k \gamma_{s,k} Q_{k,t-1}^s + \sum_k \beta_{d,k} H_{k,t-1}^d + \\ + \sum_k \beta_{s,k} H_{k,t-1}^s + \eta_t \quad (2)$$

In addition to the abovementioned analysis, this study measures the scope of imbalance using the price impact indicators as independent variables. Indeed, the chances for informed trading are maximum when significant imbalances exist between the demand and supply. Conversely, the opportunities are minimal in the opposite scenario. The price impact indicator is obtained ex ante as a cost of trading for a hypothetical trade size of q shares. Since the average trade size was not available in our dataset, the authors set q as multiples of the average bid/ask volume at the top level. For the demand side, the price impact measure $LD(q)$ is calculated as the discount per share that a market-seller gets below the midpoint of the best bid and ask:

$$LD(q) = \frac{P_1^s + P_1^d}{2} - \frac{\sum_{j=1}^{m_1-1} P_j^d Q_j^d + P_{m_1}^d Q_{m_1}^{ld}}{q}$$

and $q = \sum_{i=1}^{m_1-1} Q_i^d + Q_{m_1}^{ld}$

where the step m_1 is defined according to $\sum_{i=1}^{m_1-1} Q_i^d < q \leq \sum_{i=1}^{m_1} Q_i^d$ and Q^{ld} is the number of shares on step m_1 to fulfil the order q shares after the first $m_1 - 1$ steps are filled.

In a similar way, the price impact $LS(q)$ on the supply side is calculated as the premium per share a market-buyer needs to pay above the midpoint of the best quotes:

$$LS(q) = \frac{\sum_{j=1}^{m_2-1} P_j^s Q_j^s + P_{m_2}^s Q_{m_2}^{ls}}{q} - \frac{P_1^s + P_1^d}{2}$$

and $q = \sum_{i=1}^{m_2-1} Q_i^s + Q_{m_2}^{ls}$

where, again, the step m_2 is determined according to $\sum_{i=1}^{m_2-1} Q_i^2 < q \leq \sum_{i=1}^{m_2} Q_i^s$ and m_2 is not necessary equal to m_1.

Considering the price impact and to predict the movement of the mid-quote, a model for buying and selling orders is obtained as follows:

$$\varepsilon_t^* = \alpha_0 + \beta_0 Spread_{t-1} + \sum_{j=1}^{3} \beta_{d,j} LD(jq)_{t-1} + \sum_{j=1}^{3} \beta_{s,j} LS(jq)_{t-1} + \eta_t \qquad (3)$$

The $LD(q)$ and $LS(q)$ represent inverse measures of liquidity. As smaller price impact implies better liquidity, the return at t should be positively related to $LS(q)_{t-1}$ and negatively related to $LD(q)_{t-1}$. When there is more liquidity on the demand side, more limit buy orders are submitted at crowded buy steps. These limit orders attract more market buy orders that consequently drive the price up. Since it is highly improbable that a trader would submit such a large market order to consume all and more than the cumulated quantity available at step 3 on the opposite side, we restricted last regression model considering only the top three levels of the limit order book.

6.3 EMPIRICAL RESULTS

Descriptive Statistics

Table I

Descriptive Statistics of the Limit-Order Book Shape

	Height (%)		*Length (%)*	
Steps	*Buy*	*Sell*	*Buy*	*Sell*
1	17,98	19,72	14,97	16,22
2	19,39	22,39	19,75	20,83
3	18,55	19,56	21,67	21,43
4	19,60	18,94	21,84	20,98
5	24,49	19,39	21,76	20,54

Table I reports the summary statistics of the average heights and lengths in each step for the futures contract Phelix DE Baseload 2018. The length is defined as the number of shares at each step as a fraction of the total number of shares. The height is calculated as the price difference between each step and its previous step, as a fraction of the spread between mid-quote and final step. For the first step, the price of its previous step is set to be the mid-quote.

Table I reports the summary statistics of the average lengths and heights in each step for the futures Phelix DE Baseload 2018. Both indicators are presented in percentage terms. The length is obtained as the number of shares at each step as a fraction of the total number of shares; whereas the height is calculated as the price difference between each step and the previous one, as a fraction of the spread between the mid-quote and the final step. For the demand side the steps closer to the top of the book are shorter, whereas for the supply side all steps are uniform in the length, excluding the first one at the best price level. In both cases we find that the third and fourth level offer more depth than any other quotes. Indeed, for both ask and buy side more than 84% of the total volume is located beyond the best bid and ask price level. This feature highlights the importance of testing the incremental information content of the limit orders behind the best bid and ask level.

This statistics are partially consistent with Cao et al.'s findings (2009) and Duong et al.'s paper (2013), respectively in the equity and bond market. As in our research, their studies underline the fact that most of the shares are present below the top of the limit order book. However, unlike to our energy market situation, in both equity and bond market the steps closer to the top of the book are generally longer. A possible explanation of this discrepancy could be the fact that we are analysing a futures contract with a very long delivery period, in particular the data available covers a period that is on average a year far away from the start of the power delivery. Hence the competition is much lower.

As far as the height of the demand side is concerned, the top steps are lower (price increments are smaller) and they increase monotonically along the order book. On the ask side, once again, the step heights are uniform, with the exception of the second quote.

Overall, we notice a stronger imbalance situation on the buy side with respect to the ask side. Hence, we should expect a higher statistical significance of the indicators that summarize the shape of the demand side. Moreover, the feature that on average price ticks are smaller on the buy side suggests a buying pressure. This is confirmed from the upward trend that characterizes the mid-price in the five-month period we studied.

Shot-term Return Predictability of the Limit Order Book

Table II reports the results of the first set of regressions based on equation (1). We obtain the innovation in returns from the AR(5) model. The return innovation is then regressed against the unbalanced order book height (HR) and length (QR). The model has been estimated five times; each regression includes the limit-order book information up to step k, where k = 1,2,3,4,5.

The adjusted R-squared indicates that when the variables from the top level are used, the value is 5,33% with respect to the 4,90% when only the spread is taken as regressor. However, we registered the biggest increase when the quotes of the second step are added, the explanatory power increases by 0,89% to 6,22%. Instead, when the lagged book imbalances from steps 3 to 5 are included, we notice a decrease in fitting the results. The value is even smaller than the simple case with the spread as single predictor.

To test for the joint significance of the coefficients added in each step, a F-test is conducted using a 5% significance level. It tests the empirical model versus the constant model; in other words, for a given step k the null hypothesis is that the coefficients of HR_k and QR_k are jointly zero. The F-statistics affirms that the book imbalance coefficients beyond the top price quote are jointly significantly different from zero. This confirms the informativeness of the limit orders behind the best price level, even if it appears to be little.

From the perspective of the coefficients of the regression model, both the QR and HR indicators produce the expected sign in each level, except for HR_3 in the fifth quote.

Overall, we notice that the majority of the parameters from the third steps onwards is not statistically significant. This could explain the decrease in the adjusted R-squared when the step 3, 4 and 5 are added.

Table II

Regression analysis of the order book imbalance and returns

j		*1*	*2*	*3*	*4*	*5*
α_0	-0,044***	-0,041***	-0,044***	-0,030***	-0,013***	-0,006**
$Spread_{t-1}$	24,999***	24,814***	25,662***	17,278***	7,656***	4,214***
$HR_{2,t-1}$			0,062***	0,061***	0,060***	0,048***
$HR_{3,t-1}$				0,010	0,005	-0,003
$HR_{4,t-1}$					0,008	0,014***
$HR_{5,t-1}$						0,014***
$QR_{1,t-1}$		-0,035***	-0,040***	-0,039***	-0,041***	-0,04***
$QR_{2,t-2}$			-0,011**	-0,014***	-0,015***	-0,011**
$QR_{3,t-1}$				-0,001	-0,002	-0,002
$QR_{4,t-1}$					-0,005	-0,009*
$QR_{5,t-1}$						-0,008*
Adj-R ^{2}	0,0490	0,0533	0,0622	0,0337	0,0178	0,0160
F-stat	509	279	164	58,1	23,1	16,3

Table II uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t^* = \alpha_0 + \beta_0 Spread_{t-1} + \gamma_0 QR_{1,t-1} + \sum_{j=2}^{n} \beta_j HR_{j,t-1} + \sum_{j=2}^{n} \gamma_j QR_{j,t-1} + \eta_t$$

where ε^* is the innovation in returns estimated by using an AR(5) model, $Spread$ is the inside spread, QR_j is the scaled imbalance in quantity at step j, and HR_j is the scaled imbalance in price at step j. The $Adjusted\ R^2$ and the $F-statistics$ are reported above. For a given n step, the null hypothesis of the $F-test$ is that coefficients of HR_n and QR_n are jointly zero.
Note: the coefficient is the result multiplied by 100.

The results above do not allow us to confirm or reject our hypothesis. The descriptive statistics, previously illustrated, show that in the dataset there is an imbalance in the demand side, whereas the supply side is quite uniform. Hence, to better understand the potential information content of the LOB, we run the second set of regressions, where the height and length of the steps on the demand and supply sides are considered separately. Table III reports the results.

Table III

Relationship between the order book height, length, and the return

j		*1*	*2*	*3*	*4*	*5*
α_0	-0,044***	-0,047***	-0,048***	-0,044***	-0,029***	-0,009
$Spread_{t-1}$	24,999***	24,897***	27,833***	23,255***	12,307***	0,985
$Q^d_{1,t-1}$		0,002***	0,002***	0,001***	0,002***	0,002***
$Q^d_{2,t-1}$			0,001***	0,001***	0,001***	0,001***
$Q^d_{3,t-1}$				0,000	0,000	0,000
$Q^d_{4,t-1}$					0,000	0,000
$Q^d_{5,t-1}$						0,000
$Q^s_{1,t-1}$		-0,002***	-0,002***	-0,002***	-0,002***	-0,002***
$Q^s_{2,t-1}$			0,000	0,000	0,000	0,000
$Q^s_{3,t-1}$				0,000	0,000	0,000
$Q^s_{4,t-1}$					0,000	0,000
$Q^s_{5,t-1}$						0,000
$H^d_{2,t-1}$			-0,572***	-0,0597***	-0,707***	-0,688***
$H^d_{3,t-1}$				-0,0192***	-0,140***	0,058***
$H^d_{4,t-1}$					0,020	0,021
$H^d_{5,t-1}$						-0,010
$H^s_{2,t-1}$			0,0414***	0,601***	0,626***	0,504***
$H^s_{3,t-1}$				0,113**	0,222***	0,132**
$H^s_{4,t-1}$					-0,004	0,000
$H^s_{5,t-1}$						0,026
Adj-R^2	0,0490	0,0521	0,0998	0,0868	0,0785	0,0501
F-stat	509	182	157	85,9	56,1	27,3

Table III summarizes the results for the following regression model:

$$\varepsilon_t^* = \alpha_0 + \beta_0 Spread_{t-1} + \sum_{j=2}^{n} \beta_{d,j} H^d{}_{j,t-1} + \sum_{j=2}^{n} \beta_{s,j} H^s{}_{j,t-1} + \sum_{j=1}^{n} \gamma_{d,j} Q^d{}_{j,t-1} + \sum_{j=1}^{n} \gamma_{s,j} Q^s{}_{j,t-1} + \eta_t$$

where ε^* is the innovation in returns estimated by using an AR(5) model, $Spread$ is the inside spread, $Q_j^d\,(Q_j^s)$ is the length of step j on the demand (supply) side, and $H_j^d\,(H_j^s)$ is the height of step j on the demand (supply) side of the limit order book. The $Adjusted\ R^2$ and the $F-statistics$ are reported above. For a given n step, the null hypothesis of the $F-test$ is that coefficients added in step n are jointly zero.
Note: the coefficient is the result multiplied by 100.

These findings are in line with those of Table II. When only the best level is considered, the adjusted R-squared is 5,21%. When the second quote is added, this value registers a huge increase of 4,77% and it accounts for 9,98%. Once again, this is the highest explanatory power reached in the regression model. From step 3 onwards the value starts to decrease. However, contrary to the previous results, the explanatory power at level 3 and 4 is considerably higher than the case where only the best bid and ask level is used. Consistent with the values in Table II, the full set of regression pass the F-test and almost the total majority of coefficients estimates has the anticipated signs. Indeed, a greater order quantity on the demand (supply) side, Q_d (Q_s), suggests the possibility of attracting more aggressive orders on that side, and, thus, it predicts a positive (negative) relationship with the returns. Smaller price difference in selling (buying) indicates more aggressive selling orders, which lead to a positive (negative) relationship between H_s (H) and the returns. In both height and length parameters, the demand coefficients' estimates are statistically more significant than those of the supply. This confirms the inference we made from the descriptive table: in our dataset buying orders reveal more information than selling orders.

The results of the third set of regressions are provided in Table IV. The imbalances between the price impacts of market supply and demand are used as regressors to predict the return for the next quote. Using the supply side as an example, the price impact is defined as follows: if the demand is hypothesized to provide a quantity q of market orders, which subsequently consume the limit orders from the supply side and produce changes in transaction prices, and if the supply volume is significant enough during the best level, the transaction price should be consistent with the supply price for the best quote. Conversely, the transaction cost increases if the supply is insufficient. In other words, the price impact measures the liquidity of the demand and supply: the greater the price impact, the poorer the liquidity. Hence in the example it means a greater chance for a price increment and consequently a positive relationship between the price impact and the returns. The opposite scenario produces a negative relationship between the price impact on the buy side and the returns.

Table IV

Relationship between the price impact measures and the return

j		*1*	*2*	*3*
α_0	-0,044***	-0,035***	-0,022***	-0,022***
$Spread_{t-1}$	24,999***	41,190***	53,470***	53,499***
$LD(1q)_{t-1}$		-0,0529***	-0,378***	-0,380***
$LD(2q)_{t-1}$			-0,55***	-0,546***
$LD(3q)_{t-1}$				-0,026
$LS(1q)_{t-1}$		0,000	0,000	0,000
$LS(2q)_{t-1}$			0,001***	0,001**
$LS(3q)_{t-1}$				0,000
$Adj\text{-}R^2$	0,0490	0,0691	0,0914	0,0914
F-stat	509	245	200	143

Table IV summarizes the results for the following regression model:

$$\varepsilon_t^* = \alpha_0 + \beta_0 Spread_{t-1} + \sum_{j=1}^{3} \beta_{d,j} LD(jq)_{t-1} + \sum_{j=1}^{3} \beta_{s,j} LS(jq)_{t-1} + \eta_t$$

where ε^* is the innovation in returns estimated by using an AR(5) model, $Spread$ is the inside spread, $LD(jq)$ and $LS(jq)$ are the price impact measures on the demand and supply side for a hypothetical trade size q. The $Adjusted\ R^2$ and the $F-statistics$ are reported above. For a given trade size q, the null hypothesis of the $F-test$ is that $\beta_{d,j}$ and $\beta_{s,j}$ are jointly zero. Note: the coefficient is the result multiplied by 100.

The adjusted R-squared is 6,91% when the hypothetical trade size is q and it increases to 9,14% at $2q$. Afterwards the value remains the same; indeed none of the additional coefficients' estimates are statistically significant. All the parameters have the anticipated sign and we can reject the null of the F-test. These findings suggest that the returns decrease when the supply side is more liquid than the demand side and vice versa. When the imbalance between the market demand and supply becomes more severe, due to a higher quantity traded, the chance of informed trading increases. Subsequently price changes are more drastic and thus the model can fit better the results.

Discussion of the results

Our findings confirm previous evidence on short-term informativeness of the Limit Order Book on future price movements, for instance Bloomfield et al. (2005), Harris and Panchapagesan et al. (2005), and Kaniel and Liu (2006).
Consistent with our hypothesis, the active limit orders beyond the best level, especially those submitted from steps 2 to 3, are useful and do provide additional explanatory power. Our results are also in line with the analysis of Cao et al. (2009) and Duong et al's (2013), who implemented the same methodology respectively on the equity and bonds market.

Nevertheless, there is a main difference that could mislead us from affirming the incremental information content of the LOB: the adjusted R-squared reaches the highest value at step 2 and then it starts to decrease. This feature could be due to the presence of hedgers, traders whose main goal is to manage the risk of an unexpected large price movement. Electricity retailers are uncertain about how much electricity their residential customers will use at any time of the day until they actually turn switches on. Retailers are exposed to joint quantity and price risk on an hourly basis given the physical singularity of electricity as a commodity. Hedging allows them to lock in prices and margins in advance and it reduces the potential for unanticipated loss. This strategy reduces their exposure by shifting that risk to those with opposite risk profiles or to investors who are willing to accept the risk in exchange for profit opportunity. Hence, a hedge involves establishing a position in the futures that is equal and opposite to a position at risk in the physical market. Consequently, it is very likely that the limit orders immediately beyond the top level are submitted by informed traders, who are focused on short term profits and have a superior knowledge on the next price move. Whereas, at deeper levels, the book is populated by limit orders posted by hedgers, who are uninformed and are trying to obtain the best favourable price. Hence their limit orders do not provide any valuable information.

Further Work

The summary statistics reported in Table III show a different shape between the demand and supply side. In an empirical study, Ranaldo (2004) demonstrated that in an aggressive order submission strategy traders behave asymmetrically on the buying and selling side, i.e. their choices differ depending on whether the price is decreasing or increasing. This characteristic is related to the disposition effect, presented for the first time by Shefrin and Statman (1985). In their work, they explain that investors are willing to sell as quickly as possible when they are gaining to retain the profit; however, in the case of a loss, they are averse to sell and tend to wait for a price reversal. To better investigate whether the agents in the energy futures market have different responses to buying and selling behaviours, our empirical analysis could be repeated after having divided the return series into upward and downward trend.

Another interesting supplementary analysis is to find a way to exploit the additional information contained in the LOB. Starting from the same methodology we tested, Lin, Tsai, Zheng and Lung (2012) proposed a simple trading strategy based on the signals the trader retrieves form imbalance measures. In their model they used the HR and QR indicators, along with two new variables for changes in limit order book height and length in previous and following quotes, to predict the return on the following quote and establish the strategies. Analysing the Taiwan Futures Market they found a positive return even considering transaction costs.

7. CONCLUSION

In the first sections of the thesis we presented the limit order book mechanism and the evolution of its visualization. We then moved on reviewing quickly the literature on the book depth analysis and in the subsection 3.2 a list of variables, that summarize the order book shape, has been presented. Section 4 offers the reader an introduction to electricity market and explains in detail how a power futures contract works.

In the last part we tested the limit order book informativeness. In particular, this study investigates the incremental information conveyed by the limit orders beyond the best level. The empirical model is based on the revealed order quantity (length) and quote (height) in a futures energy market limit order book. The objective is to use these imbalance measures to forecast the price movement in the following quote. The length and depth of each price level were individually measured and repetitively integrated into the regression model. The increment in the explanatory power of the model is used to test the amount of additional information provided by deeper quotes. Considering the imbalances between the height and length of the demand and supply side together, the findings show that only the second best level increases the value of the adjusted R-squared of the model. However, if we consider separately the imbalance impacts of the demand and supply, we found stronger results.

The thesis then investigates whether the chance of informed trading is bigger whenever the market imbalance becomes more severe. Exploiting the price impact caused by an hypothetical trade of size q, the explanatory power of the regression analysis demonstrated that the forecasting ability of the model increases significantly when the market is more imbalanced.

Overall our findings are consistent with the current literature which states that informed traders submit limit orders to exploit their information advantage. The study indeed shows that short term future return is significantly affected by the information content of the limit order book.

APPENDIX A: MATLAB CODE

Innovation in returns

```
%load data at the top level
load basetime.mat

%setting timeorders format
timeorders.Format='dd/MM/yyyy hh:mm:ss';
t=datenum(timeorders);

%round timeorders at 5 min interval
T = datevec(t);
T(:,5) = floor(T(:,5) / 5) * 5;
T=datenum(T);

%calculate Midquote
MID=(bestAsk+bestBid)/2;

%create Midquote financial time series
MIDts=fints(T,MID);

%create 5 min interval from 8am to 16pm
dv = 8/24:1/288:16/24;
c=datestr(dv);
c=cellstr(c);

%extract Midquote time series from 8am to 16pm
MIDts5=fetch(MIDts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);

%calculate returns with a 5 min interval
MIDts5_vec = fts2mat(MIDts5.series1);
r=diff(log(MIDts5_vec));

% Define AR() model
EstMdl1 = arima('ARLags',1);
EstMdl2 = arima('ARLags',1:2);
EstMdl3 = arima('ARLags',1:3);
EstMdl4 = arima('ARLags',1:4);
EstMdl5 = arima('ARLags',1:5);

% Preallocate loglikelihood vector
logL = zeros(5,1);

% Estimate the AR() models
[~,~,logL(1)] = estimate(EstMdl1,r,'print',false);
[~,~,logL(2)] = estimate(EstMdl2,r,'print',false);
[~,~,logL(3)] = estimate(EstMdl3,r,'print',false);
[~,~,logL(4)] = estimate(EstMdl4,r,'print',false);
[~,~,logL(5)] = estimate(EstMdl5,r,'print',false);

% AIK criterion
[aic,bic] = aicbic(logL, [3; 4; 5; 6; 7], 9873*ones(5,1))

% Extract innovation in return
m=estimate(EstMdl5,r);
[E,V]=infer(m,r);
```

Summary Statistics

```
%load data
load dataset.mat
load InnovationInReturn.mat

%avg Height step on the demand side
H1d=nanmean(abs(bestBid-MID));
H2d=nanmean(abs(bidl2-bestBid));
H3d=nanmean(abs(bidl3-bidl2));
H4d=nanmean(abs(bidl4-bidl3));
H5d=nanmean(abs(bidl5-bidl4));
Hd=H1d+H2d+H3d+H4d+H5d;

h1d=100*(H1d/Hd);
h2d=100*(H2d/Hd);
h3d=100*(H3d/Hd);
```

```
h4d=100*(H4d/Hd);
h5d=100*(H5d/Hd);
hd=[h1d; h2d; h3d; h4d; h5d];

%avg Height step on the ask side
H1s=nanmean(bestAsk-MID);
H2s=nanmean(askl2-bestAsk);
H3s=nanmean(askl3-askl2);
H4s=nanmean(askl4-askl3);
H5s=nanmean(askl5-askl4);
Hs=H1s+H2s+H3s+H4s+H5s;

h1s=100*(H1s/Hs);
h2s=100*(H2s/Hs);
h3s=100*(H3s/Hs);
h4s=100*(H4s/Hs);
h5s=100*(H5s/Hs);
hs=[h1s; h2s; h3s; h4s; h5s];

%avg Lenght step on the demand side
Q1d=nanmean(bidvol1);
Q2d=nanmean(bidvol2);
Q3d=nanmean(bidvol3);
Q4d=nanmean(bidvol4);
Q5d=nanmean(bidvol5);
Qd=Q1d+Q2d+Q3d+Q4d+Q5d;

q1d=100*(Q1d/Qd);
q2d=100*(Q2d/Qd);
q3d=100*(Q3d/Qd);
q4d=100*(Q4d/Qd);
q5d=100*(Q5d/Qd);
qd=[q1d; q2d; q3d; q4d; q5d];

%avg Lenght step on the ask side
Q1s=nanmean(askvol1);
Q2s=nanmean(askvol2);
Q3s=nanmean(askvol3);
Q4s=nanmean(askvol4);
Q5s=nanmean(askvol5);
Qs=Q1s+Q2s+Q3s+Q4s+Q5s;

q1s=100*(Q1s/Qs);
q2s=100*(Q2s/Qs);
q3s=100*(Q3s/Qs);
q4s=100*(Q4s/Qs);
q5s=100*(Q5s/Qs);
qs=[q1s; q2s; q3s; q4s; q5s];

%table
Steps = {'step 1';'step 2';'step 3';'step 4';'step 5'};
Height=[hd, hs];
Lenght=[qd, qs];
format bank
SummaryStatistic=table(Height, Lenght, 'RowNames', Steps)
writetable(SummaryStatistic,'SummaryStatistics.xlsx','WriteRowNames',1);
```

First set of regressions

```
%load data
load SummaryStatistics.mat

%calculate spread time series at 5 min interval
spread=(bestAsk-bestBid)./(bestAsk+bestBid);
spreadts=fints(T,spread);
spreadts5=fetch(spreadts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
spreadts5_vec = fts2mat(spreadts5.series1(1:end-1));

%imbalance in the height
HR2=((askl2-bestAsk)-abs(bidl2-bestBid))./((askl2-bestAsk)+abs(bidl2-bestBid));
HR2ts=fints(T,HR2);
HR2ts5=fetch(HR2ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
HR2ts5_vec = fts2mat(HR2ts5.series1(1:end-1));

HR3=((askl3-askl2)-abs(bidl3-bidl2))./((askl3-askl2)+abs(bidl3-bidl2));
HR3ts=fints(T,HR3);
HR3ts5=fetch(HR3ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
HR3ts5_vec = fts2mat(HR3ts5.series1(1:end-1));
```

```
HR4=((askl4-askl3)-abs(bidl4-bidl3))./((askl4-askl3)+abs(bidl4-bidl3));
HR4ts=fints(T,HR4);
HR4ts5=fetch(HR4ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
HR4ts5_vec = fts2mat(HR4ts5.series1(1:end-1));

HR5=((askl5-askl4)-abs(bidl5-bidl4))./((askl5-askl4)+abs(bidl5-bidl4));
HR5ts=fints(T,HR5);
HR5ts5=fetch(HR5ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
HR5ts5_vec = fts2mat(HR5ts5.series1(1:end-1));

%imbalance in the lenght
QR1=(askvol1-bidvol1)./(askvol1+bidvol1);
QR1ts=fints(T,QR1);
QR1ts5=fetch(QR1ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
QR1ts5_vec = fts2mat(QR1ts5.series1(1:end-1));

QR2=(askvol2-bidvol2)./(askvol2+bidvol2);
QR2ts=fints(T,QR2);
QR2ts5=fetch(QR2ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
QR2ts5_vec = fts2mat(QR2ts5.series1(1:end-1));

QR3=(askvol3-bidvol3)./(askvol3+bidvol3);
QR3ts=fints(T,QR3);
QR3ts5=fetch(QR3ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
QR3ts5_vec = fts2mat(QR3ts5.series1(1:end-1));

QR4=(askvol4-bidvol4)./(askvol4+bidvol4);
QR4ts=fints(T,QR4);
QR4ts5=fetch(QR4ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
QR4ts5_vec = fts2mat(QR4ts5.series1(1:end-1));

QR5=(askvol5-bidvol5)./(askvol5+bidvol5);
QR5ts=fints(T,QR5);
QR5ts5=fetch(QR5ts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
QR5ts5_vec = fts2mat(QR5ts5.series1(1:end-1));

%first set of regressions
X = [spreadts5_vec, QR1ts5_vec];
lm1 = fitlm(X,E);
X2 = [spreadts5_vec, HR2ts5_vec, QR1ts5_vec, QR2ts5_vec];
lm2 = fitlm(X2,E);
X3 = [spreadts5_vec, HR2ts5_vec, HR3ts5_vec, QR1ts5_vec, QR2ts5_vec, QR3ts5_vec];
lm3 = fitlm(X3,E);
X4 = [spreadts5_vec, HR2ts5_vec, HR3ts5_vec, HR4ts5_vec, QR1ts5_vec, QR2ts5_vec, QR3ts5_vec,
QR4ts5_vec];
lm4 = fitlm(X4,E);
X5 = [spreadts5_vec, HR2ts5_vec, HR3ts5_vec, HR4ts5_vec, HR5ts5_vec,...
    QR1ts5_vec, QR2ts5_vec, QR3ts5_vec, QR4ts5_vec, QR5ts5_vec];
lm5 = fitlm(X5,E);
```

Second set of regressions

```
%load data
load ImbalanceMeasures.mat

%Height and Length parameters
HD2=abs(bidl2-bestBid);
h2=fints(T,HD2);
hd2=fetch(h2,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h2v = fts2mat(hd2.series1(1:end-1));

HD3=abs(bidl3-bidl2);
h3=fints(T,HD3);
hd3=fetch(h3,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h3v = fts2mat(hd3.series1(1:end-1));

HD4=abs(bidl4-bidl3);
h4=fints(T,HD4);
hd4=fetch(h4,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h4v = fts2mat(hd4.series1(1:end-1));

HD5=abs(bidl5-bidl4);
h5=fints(T,HD5);
hd5=fetch(h5,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h5v = fts2mat(hd5.series1(1:end-1));
```

```
HS2=askl2-bestAsk;
h2s=fints(T,HS2);
hs2=fetch(h2s,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h2sv = fts2mat(hs2.series1(1:end-1));

HS3=askl3-askl2;
h3s=fints(T,HS3);
hs3=fetch(h3s,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h3sv = fts2mat(hs3.series1(1:end-1));

HS4=askl4-askl3;
h4s=fints(T,HS4);
hs4=fetch(h4s,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h4sv = fts2mat(hs4.series1(1:end-1));

HS5=askl5-askl4;
h5s=fints(T,HS5);
hs5=fetch(h5s,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
h5sv = fts2mat(hs5.series1(1:end-1));

QD1=fints(T,bidvol1);
q1d=fetch(QD1,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q1v = fts2mat(q1d.series1(1:end-1));

QD2=fints(T,bidvol2);
q2d=fetch(QD2,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q2v = fts2mat(q2d.series1(1:end-1));

QD3=fints(T,bidvol3);
q3d=fetch(QD3,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q3v = fts2mat(q3d.series1(1:end-1));

QD4=fints(T,bidvol4);
q4d=fetch(QD4,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q4v = fts2mat(q4d.series1(1:end-1));

QD5=fints(T,bidvol5);
q5d=fetch(QD5,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q5v = fts2mat(q5d.series1(1:end-1));

QS1=fints(T,askvol1);
q1s=fetch(QS1,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q1sv = fts2mat(q1s.series1(1:end-1));

QS2=fints(T,askvol2);
q2s=fetch(QS2,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q2sv = fts2mat(q2s.series1(1:end-1));

QS3=fints(T,askvol3);
q3s=fetch(QS3,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q3sv = fts2mat(q3s.series1(1:end-1));

QS4=fints(T,askvol4);
q4s=fetch(QS4,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q4sv = fts2mat(q4s.series1(1:end-1));

QS5=fints(T,askvol5);
q5s=fetch(QS5,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
q5sv = fts2mat(q5s.series1(1:end-1));

%second set of regression
X6 = [spreadts5_vec, q1v, q1sv];
lm6 = fitlm(X6,E);
X7 = [spreadts5_vec, q1v, q2v, q1sv, q2sv, h2v, h2sv];
lm7 = fitlm(X7,E);
X8 = [spreadts5_vec, q1v, q2v, q3v, q1sv, q2sv, q3sv, h2v, h3v, h2sv, h3sv];
lm8 = fitlm(X8,E);
X9 = [spreadts5_vec, q1v, q2v, q3v, q4v, q1sv, q2sv, q3sv, q4sv, h2v, h3v, h4v, h2sv, h3sv,
h4sv];
lm9 = fitlm(X9,E);
X10 = [spreadts5_vec, q1v, q2v, q3v, q4v, q5v, q1sv, q2sv, q3sv, q4sv, q5sv,...
    h2v, h3v, h4v, h5v, h2sv, h3sv, h4sv,h5sv];
lm10 = fitlm(X10,E);
```

LD and LS measures

```
%load data
load ImbalanceMeasures.mat

cum_bvol=cumsum(bidvolume,2, 'omitnan');

for ii=1:size(bidvol1)
    if cum_bvol(ii,1)>7
        f(ii)=1;
    else f(ii)=0;
    end
end
f=f';

for ii=1:size(bidvol1)
    if cum_bvol(ii,2)>7
        n(ii)=2;
    else n(ii)=0;
    end
end
n=n';

for ii=1:size(bidvol1)
    if cum_bvol(ii,3)>7
        o(ii)=3;
    else o(ii)=0;
    end
end
o=o';

for ii=1:size(bidvol1)
    if cum_bvol(ii,4)>7
        p(ii)=4;
    else p(ii)=0;
    end
end
p=p';

A=[f,n,o,p];
A=sum(A,2);

k = find(isnan(bidvol1))'; bidvol1(k) = 0; bidvol1(isnan(bidvol1)) = 0;
k = find(isnan(bidvol2))'; bidvol2(k) = 0; bidvol2(isnan(bidvol2)) = 0;
k = find(isnan(bidvol3))'; bidvol3(k) = 0; bidvol3(isnan(bidvol3)) = 0;
k = find(isnan(bidvol4))'; bidvol4(k) = 0; bidvol4(isnan(bidvol4)) = 0;
bl1=bidvol1;
bl2=bl1+bidvol2;
bl3=bl2+bidvol3;
bl4=bl3+bidvol4;

for ii=1:size(bidvol1)
    if A(ii)== 7
        ql(ii)= 7 - bl2(ii);
    elseif A(ii)== 9
        ql(ii) = 7 -bl1(ii);
    else
        ql(ii)=0;
    end
end
ql=ql';

for ii=1:size(A)
    if A(ii) == 4
        LD(ii) = MID(ii) - [(bidvol3(ii)*bidl3(ii) + bidvol2(ii)*bidl2(ii) +
bidvol1(ii)*bestBid(ii))/(bl3(ii))];
    elseif A(ii)==7
        LD(ii) = MID(ii) - [(ql(ii)*bidl3(ii) + bidvol2(ii)*bidl2(ii) +
bidvol1(ii)*bestBid(ii))/(bl2(ii)+ql(ii))] ;
    elseif A(ii)==9
        LD(ii) = MID(ii) - [(ql(ii)*bidl2(ii) + bidvol1(ii)*bestBid(ii))/(bl1(ii)+ql(ii))] ;
    elseif A(ii)== 10
        LD(ii)= MID(ii) - [(7*bestBid(ii))/7];
    else
        LD(ii)= 0;
    end
end
LD=LD';
```

```
LDts=fints(T,LD);
LDts5=fetch(LDts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
LD7 = fts2mat(LDts5.series1(1:end-1));

cum_avol=cumsum(askvolume,2, 'omitnan');

for ii=1:size(askvol1)
    if cum_avol(ii,1)>21
        w(ii)=1;
    else w(ii)=0;
    end
end
w=w';

for ii=1:size(askvol1)
    if cum_avol(ii,2)>21
        g(ii)=2;
    else g(ii)=0;
    end
end
g=g';

for ii=1:size(askvol1)
    if cum_avol(ii,3)>21
        u(ii)=3;
    else u(ii)=0;
    end
end
u=u';

for ii=1:size(askvol1)
    if cum_avol(ii,4)>21
        z(ii)=4;
    else z(ii)=0;
    end
end
z=z';

B=[w,g,u,z];
B=sum(B,2);

k = find(isnan(askvol1))'; askvol1(k) = 0; askvol1(isnan(askvol1)) = 0;
k = find(isnan(askvol2))'; askvol2(k) = 0; askvol2(isnan(askvol2)) = 0;
k = find(isnan(askvol3))'; askvol3(k) = 0; askvol3(isnan(askvol3)) = 0;
k = find(isnan(askvol4))'; askvol4(k) = 0; askvol4(isnan(askvol4)) = 0;
al1=askvol1;
al2=al1+askvol2;
al3=al2+askvol3;
al4=al3+askvol4;

for ii=1:size(askvol1)
    if B(ii)== 7
        ql(ii)= 21 - al2(ii);
    elseif B(ii)== 9
        ql(ii) = 21 -al1(ii);
    else
        ql(ii)=0;
    end
end
ql=ql';

for ii=1:size(bidvol1)
    if B(ii) == 4
        LS(ii) = [(askvol3(ii)*askl3(ii) + askvol2(ii)*askl2(ii) +
askvol1(ii)*bestAsk(ii))/(al2(ii)+ql(ii))] - MID(ii);
    elseif B(ii)==7
        LS(ii) = [(ql(ii)*askl3(ii) + askvol2(ii)*askl2(ii) +
askvol1(ii)*bestAsk(ii))/(al2(ii)+ql(ii))] - MID(ii);
    elseif B(ii)==9
        LS(ii) = [(ql(ii)*askl2(ii) + askvol1(ii)*bestAsk(ii))/(al1(ii)+ql(ii))] - MID(ii);
    elseif B(ii)==10
        LS(ii)= [(21*bestAsk(ii))/21] - MID(ii);
    else
        LS(ii)= 0;
    end
end
LS=LS';

LSts=fints(T,LS);
```

```
LSts5=fetch(LSts,'3-Oct-2016','08:00','28-Feb-2017','16:00',1,'d',c);
LS21= fts2mat(LSts5.series1(1:end-1));
```

Third set of regressions

```
%load data
load ImbalanceMeasures.mat
load LRImb.mat

X14 = [spreadts5_vec, LD7, LS7];
lm14 = fitlm(X14,E)

X15 = [spreadts5_vec, LD7, LD14, LS7, LS14];
lm15 = fitlm(X15,E)

X16 = [spreadts5_vec, LD7, LD14, LD21, LS7, LS14, LS21];
lm16 = fitlm(X16,E)
```

REFERENCES

Anand, A., & Martell, T. (2001). 'Informed' limit order trading. Working paper

Baruch, S. (2005). Who benefits from an open limit-order book? Journal of Business, 78, 1267–1306

Biais, B., Hillion, P., & Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris bourse. Journal of Finance, 50(5), 1655-1689

Bloomfield, R., O'Hara, M., & Saar, G. (2005). The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. Journal of Financial Economics, 75, 165–199

Boehmer, E., Saar, G., & Yu, L. (2005). Lifting the veil: An analysis of pre-trade transparency at the NYSE. Journal of Finance, 60, 783–815

Bortoli, L., Frino, A., Jarnecic, E., Johnstone, D. (2006). Limit order book transparency, execution risk, and market liquidity: evidence from the Sydney Futures Exchange. Journal of Futures Markets, 26, 1147-1167

Burger, M., Graeber, B., Schindlmayr, G. (2014). Managing Energy Risk: A Practical Guide for Risk Management in Power, Gas and Other Energy Markets, Second Edition

Cao, C., Hansch, O., & Wang, X. (2009). The information content of an open limit-order book. Journal of Futures Markets, 29(1), 16-41

Chakravarty, S. and Holden, C.W. (1995). An integrated model of market and limit orders. J. Financ. Intermed., 4, 213–241

Cont, R., Kukanov, A., & Stoikov, S. (2011). The price impact of order book events. arXiv:1011.6402

Copeland, T.E., & Galai, D. (1983). Information effects on the bid-ask spread. J. Finance, 38, 1457–1469

Danielsson, J., Payne, R.G. & Luo, J. (2002). Exchange Rate Determination and Inter-Market Order Flow Effects. London School of Economics, mimeo.

Duong, H.N., & Kalev, P.S. (2008). Order Book Slope and Price Volatility. Working Paper

Duong, H.N., Kalev, P.S. & Sun, Y. (2013). Pre-trade Transparency and the Information Content of the Limit Order Book. Market Microstructure and Nonlinear Dynamics - Keeping Financial Crisis in Context, edited by Dufrénot, G., Jawadi, F. & Louhichi, W. (2014). Springer-Verlag

Evans, M.D.D. & Lyons, R.K. (2006). Understanding Order Flow. International Journal of Finance and Economics 11, 3-23

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. Journal of Financial Markets, 2, 99–134

Glosten, L. (1994). Is the electronic open limit order book inevitable? Journal of Finance, 49, 1127–1161

Gould, M.D., Porter, M.A., Williams, S., McDonald, M., Fenn, D. & Howison, S. (2013). Limit Order Books. Quantitative Finance 13 (11), 1709-1742

Griffiths, M., Smith, B., Turnbull, A., & White, R. (2000). The costs and determinants of order aggressiveness. Journal of Financial Economics, 56, 65–88

Harris, L. (1990). Liquidity, trading rules, and electronic trading systems, New York University. Monograph series in finance and economics, Monograph 1990–4

Handa, P. & Schwartz, R.A. (1996). Limit Order Trading. Journal of Finance, vol. 51, issue 5, 1835-61

Handa, P., Schwartz, R.A. & Tiwari, A. (2003). Quote setting and price formation in an order driven market. Journal of Financial Markets, vol. 6, issue 4, 461-489

Harris, L., & Panchapagesan, V. (2005). The information-content of the limit order book: Evidence from NYSE specialist trading decisions. Journal of Financial Markets, 8, 25–67

Hautsch, N. & Huang, R. (2011). The market impact of a limit order. J. Econ. Dyn. Control, 36, 501–522

Hollifield, B., Miller, R., & Sandas, P. (2004). Empirical analysis of limit order markets. Review of Economic Studies, 71, 1027–1063

Kaniel, R., & Liu, H. (2006). So what orders do informed traders use? Journal of Business, 79

Latza, T. & Payne, R. (2010). Measuring the Information Content of Limit Order Flows. Working paper

Lin, W., Tsai. S, Zheng, Z. & Lung, P. (2012). The Information content of the limit-order book and the corresponding trading strategy. Working Paper

Madhavan, A., Potter, D., & Weaver, D. (2005). Should securities markets be transparent? Journal of Financial Markets, 8, 265–287

Næs R., Skjeltorp J. (2006). Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market, Journal of Financial Markets, vol. 9, issue 4, 408-432

Parlour, C. (1998). Price dynamics in limit order markets. Review of Financial Studies, 11, 789–816

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. Journal of Financial Markets, 7, 53–74

Rock, K. (1996). The specialist's order book and price anomalies. Working Paper. Harvard University

Seppi, D. (1997). Liquidity provision with limit orders and a strategic specialist. Review of Financial Studies, 10, 103–150

Sager, M. & Taylor, M. (2008). Commercially available order flow data and exchange rate movements: Caveat emptor. Journal of Money, Credit and Banking 40(4), 583–625

Samuelson, P. (1965). Rational theory of warrant pricing. Indust. Management Rev., 6:13–32

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and losers too long: Theory and evidence. Journal of Finance, 40(3), 777-790

www.eex.com – European Energy Exchange