The Order Book and Limit Orders at NASDAQ Stockholm

PART 1

The Limit Order Book and Flow at NASDAQ Stockholm

PART 2

Time Intervals Between Orders and Trades at NASDAQ Stockholm

Théodore Montel

Stockholm School of Economics / Università Bocconi

Acknowledgements

I thank Rolf Andersson at SHOF for spontaneously helping with the project without ever being asked, and by himself essentially rendering it possible. I also thank Marius Malinauskas, Cash equity operations at NASDAQ, for his patience and answers to my technical questions. I furthermore thank Michael Halling for taking the time to read through the thesis multiple times without having any such responsibility.

Finally, I thank my family and friends for supporting me throughout my studies.

I also thank my Bocconi supervisor Barbara Rindi.

START OF PART 1

Abstract

This study examines characteristics of the order book and flow, based on techniques established in (Biais, Hillion et al. 1995). The dataset is nanosecond-level ticks from July 2019 at the Stockholm stock exchange. The state of research with respect to seven key characteristics is established, and it is subsequently compared to this study's results. The research is conducted by reconstructing the order book and aggressive orders from messages disseminated through the NASDAQ Nordic Equity TotalView (NETV) ITCH system. The reconstruction is followed by a classification of all events (any market action, e.g., buying, cancelling an order, placing a passive order etc.) according to their aggressiveness. Important analyses include the average shape of the order book, order and trade volume conditional on the time of day, and the "diagonal effect", found in various papers. This study concludes that out of the seven key characteristics, four still hold at NASDAQ Stockholm. As such, a review of empirical results in general is suggested.

Table of Contents

Glossary

Term	Explanation
Aggressive order	Order that is executed immediately on arrival to the exchange
Passive order	Order that is not executed immediately on arrival to the exchange
Ask-price	Lowest price any market actor is willing to accept to sell the asset
Bid-price	Highest price any market actor is willing to accept to buy the asset
Mid-price	The average between the bid-price and the ask-price
Bid-ask spread	The difference between the ask-price and the bid-price
Displayed volume	Volume that is visible to all market actors
Undisplayed volume	Volume that is invisible to all market actors
Liquidity traders	Term used in modelling theory describing traders with a low degree of information and/or sophistication
Insiders	Term used in modelling theory describing traders with a high degree of information and/or sophistication
HFT	High frequency trader. Firms that trade at high speed using strategies such as cross-market arbitrage and market-making
Event	Any market action: buy, sell, passive order placement, cancellation, partial cancellation, hidden execution

1 Introduction

1.1 Context and motivation

More than half of the world's major stock exchanges rely on the limit order book mechanism (LOB) (Rosu 2009). It is, therefore, crucial to keep the understanding of limit order placements and their contribution to liquidity and price formation as current as possible.

However, this does not seem to occur as often as it ideally should. (Gould, Porter et al. 2013) point out that empirical studies make strong assertions regarding statistical regularities based on data from multiple years ago, of poor quality, describing only single stocks over short time periods, sometimes just a couple of days. For this reason, LOB models may be based on regularities in old and sometimes small samples, while traders' strategies and the rules governing stock exchanges change over time. Empirical observations from more than a decade ago may not accurately describe current LOB activity.

To perform analysis of the kind that uncovers regularities such as those mentioned above, high quality, high-frequency data is needed. The recent collaboration between the Swedish House of Finance and NASDAQ Stockholm provides such high-frequency data for research purposes in an accessible format.

This study capitalises on the opportunity to use that data to apply key parts of the framework developed by (Biais, Hillion et al. 1995) (BHC), and thereby provide insights into some of the empirical characteristics of a modern LOB. This study also contrasts the results with findings by BHC and, where it is relevant, more recent studies to emphasize statistical regularities that may need to be revised or further investigated.

1.2 Research question

This study seeks to answer the following research question:

Do the existing foundations for LOB modelling, regarding the order book and order flow, accurately reflect modern LOBs?

The seven subparts used to test the research question are presented in section 2.4.3.1.

Due to the nature of the data, an integral part of this study is the reconstruction of the order book and the reverse-engineering of aggressive, immediately executed, orders. It, therefore, also presents its techniques used for those purposes.

The study is organised as follows. Chapter 2 provides the background, with a definition of what a LOB is, basic facts about dark pools, relevant facts about the structure of NASDAQ Stockholm and a literature review. Then, chapter 3 describes the dataset, followed by chapter 4, providing an overview of the process used to reconstruct the order book and to reverse-engineer aggressive orders. Chapters 5 and 6 provide results on the order book and the order flow, respectively. Chapter 7 discusses the results and comments on the limitations of the study, and chapter 8 finally provides a conclusion and an answer to the research question.

2 Background

This chapter begins by describing the fundamental structure of the LOB to provide an overview of the subject of this study. It continues by describing general characteristics of dark pools and afterwards establishes relevant specifics of NASDAQ Stockholm to motivate later choices in the discussion. Finally, the chapter ends with a literature review.

The review consists of basic information about modelling attempts of the LOB, the results of BHC, as well as more recent findings in the area. Where more recent findings exist, they are used as the foundation, in reference to the research question. The review finally describes the concepts of impact, and strategic order splitting, which are crucial to understanding an important motivation of the main result in BHC.

2.1 The LOB

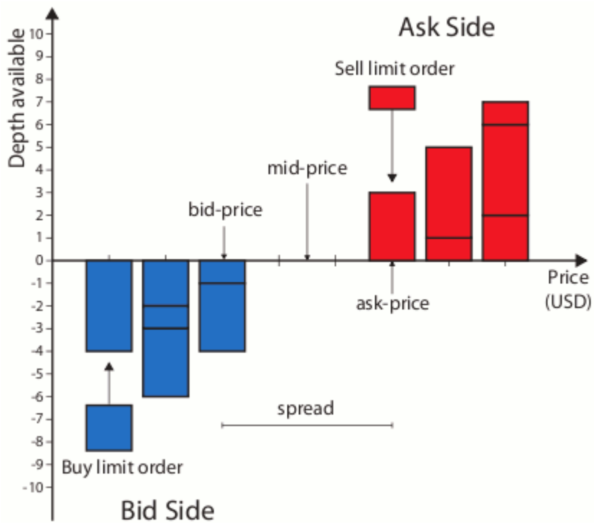
2.1.1 Fundamentals

The definition of a LOB is the following. “A record of unexecuted limit orders maintained by the specialist.” (NASDAQ 2020).

The easiest way to understand the LOB, however, is through a picture. Figure I shows an example of a LOB with the essential parts included.

Figure I

Graphical representation of a limit order book.



Note: Source (Gould, Porter et al. 2013).

The defining features of the LOB are evident in the representation above. Note that Figure I does not make any attempt to indicate undisplayed volume. It is important to remember that LOBs may also contain such volume.

Starting from the left and working towards the right. The bid side provides the price given to market actors wishing to sell the asset. The queues are formed by what is called limit orders. Limit orders are "passive" orders, not executing immediately, waiting for "aggressive" orders, executing immediately, to take liquidity from them. The limit orders stack up at different price levels, and these queues form the total available depth at each price level. The bid-price is the current best offer to buy the asset, i.e., the highest price anyone in the market is currently willing to pay for the asset.

The ask side provides the price given to actors wishing to buy the asset. The functioning is similar to the bid side in all regards. The ask-price is the current best offer to sell the asset, i.e., the lowest price anyone in the market is currently willing to accept for the asset.

When an aggressive order comes into the exchange, it is matched according to a priority rule. The most common by far is *price-time* (Gould, Porter et al. 2013). Execution following *price-time* means limit orders queue up as described in Figure I, and when an aggressive order comes in, ties are broken by selecting the limit order with the earliest submission time for execution against the aggressive order.

The mid-price is the average of the bid- and ask-prices. The bid-ask spread is the difference between the bid-price and the ask-price.

2.1.2 The tick size

The tick size in the LOB is one of the fundamental rules governing actions in the market microstructure; it is the smallest step by which one can increment an asset's price. The tick size dictates how much more expensive it is for a trader to gain the priority associated with a higher (lower) price when placing a buy (sell) order (Parlour, Seppi 2008). In Figure I, the small gaps between the bid side columns and ask side columns represent the tick size, and the pricing of orders can be chosen freely by market participants so long as the tick size is respected.

For example, consider a stock with a price of 90.0 and a valid price space of $90.0 \leq \ p \leq 91.0$. Without a specified tick size, an order with a price up to any decimal in p is valid, such as $p = 90.035$. However, if a tick size of 1.0 is specified, only prices in $p \ \in \{90.0. 91.0\}$ are acceptable. Consider now a tick size of 0.5 on the same stock. It is then possible to price the order as $p \ \in \{90.0. 90.5. 91.0\}$.

2.2 Dark pools

LOBs are not the only transaction medium employed by exchanges. Of interest to this study are also dark pools, as NASDAQ Stockholm has one. Executions from that dark pool are present in the order flow. In general, dark pools are characterised by limited or no trade transparency, anonymity and almost exclusively a mid-price peg. That is, dark pool orders almost always execute at the mid-price of the main book.

However, it should be noted that the rules governing them vary greatly (Gould, Porter et al. 2013). Dark pools have a substantial part of share volume, which is estimated at around 8 – 10% (Buti, Rindi et al. 2010). The data also differs quite a lot. Depending on the source used, dark pool trading share may be as high as ~30% (Phillips 2012).

2.3 NASDAQ Stockholm

This section describes the specificities of NASDAQ Stockholm that are relevant to this study.[1] The section motivates reconstruction, reverse-engineering as well as enables discussion of results later in the study.

2.3.1 Order types and displayed/undisplayed volume

The two main order types at NASDAQ Stockholm are limit orders and market orders. These order types are crucial to the reconstruction of the order book. Furthermore, the uncovering of market orders is part of the reason for performing the reverse-engineering described later.

Limit orders are in general passive, meaning that they lay waiting in the book forming the queues seen in Figure I. However, they do not have to be passive. Aggressive limit orders exist and are used by market actors, as will be shown later in the reverse-engineering. At NASDAQ Stockholm, it is possible to specify limit orders to be partially or fully undisplayed. Displayed volume is volume such that all market participants may observe it. Undisplayed volume is volume such that it is hidden to market participants, but still present in the LOB. A requirement for fully undisplayed orders in the main book is that they must be large-in-scale (LIS). MiFID II describes LIS requirements, and they depend on the traded instrument. The requirement places a floor on the volume that is allowed to be requested in an order by a market actor.

Market orders at NASDAQ Stockholm are IOC (Immediate-Or-Cancel). Therefore, automatic cancellation occurs of any volume exceeding what is on offer at the bid-ask spread.

To illustrate, consider the following example. Assume a LOB with 100 units offered on each side at the bid-ask spread, and 200 units offered on each side at the second price level. A market buy order for 150 units hits the ask-side and fills 100 at the bid-ask spread. The 50 left will be automatically cancelled by the system. To walk up the book on the ask side, executing on both the best and higher price levels on the

---

[1] Information in section 2.3 is taken from NASDAQ Nordic Market Model 2019:4 or NASDAQ Nordic Equity TotalView-ITCH v.3.03.4

ask side, one may use a limit buy order with a price higher than the ask-price and volume higher than what is offered at the ask-price.

2.3.2 Order execution

NASDAQ Stockholm follows *price-internal-display-time* priority, which is a variant of the standard *price-time* priority. Differences in priority rules can incentivise market actors in different ways. *Price-time* incentivises to place limit orders early, while alternatives such as for example *price-size* incentivises to place orders with higher volume. Having a *display* part will incentivise market actors to provide displayed volume. Disregard the internal part of the priority rule for the following example as it is irrelevant for this study. Consider aggressive order A with volume 100 that has just come into the exchange. Assume orders B and C were laying in the book at the price of A, and that B arrived before C. B has displayed volume 20, undisplayed volume 30. C has displayed volume 10, undisplayed volume 40. The execution flow is then the following.

A will be matched in the following steps: 1) 20 from B, 2) 10 from C, 3) 30 from B and 4) 40 from C. This shows the incentive to provide displayed volume, and to place orders early, for market actors, as it gives priority in execution.

2.3.3 Tick size

The NASDAQ Stockholm tick sizing grid is a function of the average of daily volume over one year, and the price of the security. The complete tick sizing grid is subject to regulatory oversight and comes from MiFID II, see appendix I. Currently, the OMX30 divides into five tick groups within this grid, see appendix II.

2.3.4 Nordic@Mid

NASDAQ Stockholm offers a dark pool solution for efficient matching of large orders with less risk of getting “picked off” or front-run by faster market actors and avoiding the LIS restrictions imposed by MiFID II. Nordic@Mid is typical in that no market actor can see the Nordic@Mid book. Nordic@Mid functions in the following way. A market actor sends a Nordic@Mid order, specifying only volume. The order is stored until another Nordic@Mid order comes in that matches the requirements of the first order. The process is thus quite opaque. There is no interaction between the main book and the Nordic@Mid.

Nordic@Mid orders are pegged to the mid-price, as in a typical dark pool. Market participants may, however, enter a limit price to protect their downside, as well as a minimum matching volume. For example, when entering a Nordic@Mid bid order, the market participant may specify to buy only if the mid-price is below a certain threshold, and only if the matched sell order offers more than a particular volume.

2.4 Literature review

2.4.1 Modelling approaches of the LOB

(Gould, Porter et al. 2013) detail three major approaches to modelling the LOB. These approaches provide explanations for different patterns under different assumptions and are used to explain multiple statistical regularities in the context of the research question.

First is the perfect rationality approach. Many perfect rationality models assume that there is one group of informed traders and another group of uninformed traders in the market at all times. The informed traders “know something” about what the traded asset is fundamentally worth, while the uninformed traders do not. This approach explains at least one common phenomenon; for example, the model developed by (Rosu 2009) predicts hump-shaped depth profiles. Various studies confirm these shapes empirically, for example, at the Stockholm stock exchange (Hollifield, Miller et al. 2004).

Second is the zero-intelligence approach. Zero-intelligence models assume that order arrival and cancellation rates are governed solely by stochastic processes, like Poisson processes (Smith, Farmer et al. 2003) or more recently Hawkes processes (Chakraborti, Patriarca et al. 2011). Historical data informs the parameters of the processes. It is then possible to compare model outputs to real data to verify the predictions. These models provide explanations to regularities such as the hump-shape, as well as the clustering of high order-arrival rates and low order-arrival rates.

Third is the agent-based approach. Agent-based modelling allows for heterogeneous actors, as well as some randomness to affect the market. In these models, a large number of traders interact according to pre-defined rules. Agent-based models lie in between perfect rationality and zero-intelligence perspectives and add flexibility to the previous approaches. However, they also come with drawbacks. Due to the complexity and large state-space of LOBs, it is difficult to track precisely how a specific event affects the outcome of the LOB in the model. Furthermore, it is difficult to encode rules for trader behaviour and to be sure that a set of rules producing an outcome is the unique set of rules producing that specific outcome.

Agent-based models can reproduce a large number of empirical regularities in LOBs. These include heavy-tailed return distributions, clustered volatility and aggregational Gaussianity (Challet, Stinchcombe 2003).

2.4.2 Findings in BHC

Since BHC use terms to describe their different order types, some terminology is explained before presenting the findings.

- "Large orders" are orders specifying a price that is above or below the bid-ask spread, and a volume that is higher than the volume offered at the bid-ask spread.
- "Market orders" are similar to the market orders at NASDAQ Stockholm, except that volume exceeding the volume available at the bid-ask spread is added as a limit order at the bid-ask spread instead of cancelled immediately. This category also includes limit orders with price at the bid-ask spread and volume larger than the volume offered at the bid-ask spread.
- "Small orders" are orders specifying a price at the bid-ask spread and a volume lower than or equal to what is available at the bid-ask spread.
- "Applications" are block trades that can execute either in or at the bid-ask spread.

2.4.2.1 Dataset

BHC use a dataset from the Paris Bourse, collected from October 29 to November 26, 1991. That date range includes 19 trading days. The studied stocks are the members of the CAC40; however, the dataset only includes the five top levels in the book. Furthermore, all information at the Paris Bourse was available in real-time to market participants.

2.4.2.2 Summary statistics

Reported statistics are per stock per day. BHC find the average number of trades to be 148.6, and the average number of passive orders to be 160.6. The difference compared to current data in this regard is striking. While the number of Applications is small, they represent a relatively large share of the total volume traded. Cancellations amount to ~10% of events recorded.

2.4.2.3 The order book

Regarding the slope of the order book, BHC find on average weak concavity in their shape profile.

Regarding the bid-ask spread, BHC find that it is larger than twice the difference between adjacent prices on any side of the book. In absolute terms, it is around 3x the tick size for stocks with tick FF 1, and 9x the tick size for stocks with tick FF 0.1. However, when moving up the book, the difference between adjacent prices on each side of the book is relatively constant within tick sizes. So, for example, given tick size FF 0.1 the difference between the price of the third and fourth level of the book will be more or less equal to the difference between the price of the fourth and fifth level. The result is independent of sides.

Also, the depth at the bid-ask spread and the depth at the other levels in the book differ. The book is deeper at higher levels. When excluding the bid-ask spread, the depth is not significantly different between the bid and ask sides. That is, if we freeze time and look at the bid side, it will offer approximately the same volume in aggregate on the first five levels as the ask side.

BHC suggest two interpretations of these results. From an auction-theoretic perspective, the shape may reflect correlation in the value of the security to various bidders on the same side of the market, the competition among bidders on the same side and the shading of bids compared to the underlying reservation values. If the shape instead arises from information asymmetries, the authors' findings suggest that the marginal information content of trades is decreasing with size.

BHC also find that the bid-ask spread is larger than the minimum tick size. Thereby the discreteness in the bid-ask spread is endogenous. When looking further up the book, however, the median number of ticks between two consecutive prices ranges from two to four stocks with tick FF 0.1. For stocks with tick FF 1, the median is one tick.

BHC suggest that this means the tick size is binding for large tick stocks and non-binding for small tick stocks, and that traders in the smaller tick stocks are exploiting the tick sizing grid for profit through strategic bidding.

They present the following example to support their point. Assume a tick size of SEK 1, current best ask at 155, second-best ask at 158. Assume also that 155 is the “fair” price for the best ask quote, where the liquidity supplier breaks even, and the “fair” price for the second-best ask is 156. Should a potential liquidity supplier post an ask at 156? If the supplier posts at 157, he can earn a surplus, so he obviously posts at 157.

What is interesting is that the next supplier does not have any incentive to post at 156 either, where he only breaks even. He should instead queue at 157, where he has at least a chance to earn positive profits. The example shows the first-mover advantage in liquidity supplying where there is a tick sizing grid, as well as the incentive to supply liquidity created by the discreteness of the tick grid, and time priority.

2.4.2.4 The order flow

Regarding the unconditional probability of orders and trades, the most common events are Small orders. BHC conjecture that this might reflect the impact of small investors, or the ability of larger investors to split their orders. Small orders are the most common order type, followed by new limit orders placed in or at the bid-ask spread. Most of the activity, therefore, takes place at the bid-ask spread.

Subsequently, the authors carry out a study of the probabilities of orders and trades conditional on the time of day. In general, orders and trades exhibit a U-shaped pattern where activity drops in the middle of the day. Market orders, Small orders and passive limit orders are most common in the morning, Large orders are frequent later in the day.

BHC present four interpretations of these results. First, the results could suggest that Small orders in the morning contribute to price discovery, and that placement of Large orders often occurs after price discovery. Second, fund managers, who are likely to initiate a significant fraction of Large orders, are evaluated based on the closing price. Of course, given this, they are incentivised to trade later in the day. Third, the high frequency of Small orders in the morning could reflect banks transmitting Small orders received but not processed before the market opening. Finally, the pattern may reflect strategic splitting of orders during the day. Such behaviour has been observed, for example, by (Roth, Murnighan et al. 1988) where during experimental bargaining, participants delayed reaching agreements until they were close to the deadline.

A presentation of results on probabilities conditional on the previous event follows. BHC find a “diagonal effect” in the data, meaning that events cluster. For example, the probability of a Large order following a Large order is high compared to the probability of seeing a bid in the bid-ask spread unconditionally. Suggested explanations for the phenomenon include strategic order splitting, mimicking or successive reactions to similar information.

However, not all patterns follow the general diagonal one. BHC give separate comments for these results.

Large orders often precede passive orders in the bid-ask spread on the same side of the book. For example, given a Large sale, an ask is likely to be placed in the bid-ask spread. The authors conjecture that this reflects information effects. A Large sale would shift the order book downward, as such, the ask side gets an opportunity for time priority and profit. The shifts are not observed after Market sales and Small sales.

Furthermore, Large orders also often precede cancellations on the opposite side of the book. Similarly, as before, the interpretation is in terms of information effects where Large sales (buys) suggest a negative (positive) sentiment about the stock, therefore, making it less appealing to purchase (sell), respectively.

Market orders often precede cancellations on the same side of the book. Bourse officials suggest that this is a practice of “liquidity pinging” where market actors submit Market orders to find undisplayed liquidity and cancelling the part of the order that is not executed. Recall that Market orders at this time submitted a limit order of the unexecuted part.

Market orders furthermore often precede Market orders on the opposite side of the book. BHC suggest this means following Market orders provide liquidity to the first ones, while restoring the initial bid-ask spread and mid-price. The market does not, however, appear to perceive Market orders as informationally motivated, since they attract liquidity on the opposite side. The same reasoning applies for Small orders on the opposite side of the book following Market orders.

Large orders and cancellations frequently precede Applications. The authors motivate this observation by comparing the costs and benefits of Applications with the block trading procedure. As clearing of the book must happen at the block price when performing a block trade, the parties are better off clearing the book first by submitting a Large order and then submitting an Application instead of using a block trade.

2.4.2.5 Summary of findings in BHC

As many findings are presented above, this section presents an overview of the ones this study wishes to highlight.

1) The order book is concave
2) The bid-ask spread characteristics differ from the characteristics of higher levels of the book
3) The size of the differences between adjacent prices higher up in the book is larger than one tick
4) The most common events are Small orders
5) Most of the activity takes place at the bid-ask spread
6) Conditional on the time of the day, orders and trades exhibit a U-shaped pattern
7) The diagonal effect, including related observations

Related observations are the results pointed out by BHC that do not follow the general diagonal pattern.

2.4.3 Later studies relating to the findings in BHC

Following the publication of BHC, several researchers have studied similar empirical phenomena. As (Gould, Porter et al. 2013) point out, however, the results are somewhat contradictory. That is attributed to a range of causes, amongst others differing matching algorithms, differing liquidity levels between markets, and researchers using different levels of data quality.

In any case, cancellations represent a larger share of orders in later studies than they did in BHC. (Challet, Stinchcombe 2001) find that approximately 70% of orders end in cancellation on one of the first electronic stock exchanges, the Island ECN. Later studies suggest even higher cancellation rates in a range of markets (Gould, Porter et al. 2013). (Cao, Hansch et al. 2008) find that priority considerations play a key role for market actors when deciding whether to cancel active orders and that a smaller tick size encourages cancellations because market actors can gain priority cheaply.

(Hasbrouck, Saar 2009) find the reason for the high rates of cancellation to be “pinging” for undisplayed liquidity by market actors, similar to the suggestion by BHC.

They call such orders “fleeting” and argue that four factors lie at the heart of their usage: 1) technology, 2) active trading, 3) market fragmentation and 4) the emerging importance of undisplayed liquidity.

Regarding the mean relative bid- and ask-side depth differences, (Bouchaud, Mézard et al. 2002), among others, corroborate the findings of no difference in BHC. However, (Gu, Chen et al. 2008) obtain a different result in their study of the Shenzhen Stock Exchange. Rules applying to order placement in Shenzhen during the data collection period were thought to explain the asymmetry in that case.

Regarding event clustering, (Ellul, Holden et al. 2003) find in essence the same results as BHC on the NYSE. Furthermore, (Bouchaud, Farmer et al. 2009) argue that clustering is mainly due to strategic order splitting. Strategic order splitting is an important part of market microstructure and is described in section 2.4.4.

2.4.3.1 Current regularities

This section presents the regularities in section 2.4.2.5 updated according to studies completed after BHC. The purpose of the regularities presented below is that they split the research question into discrete parts, possible to test against current data.

1) The order book is concave
2) The bid-ask spread characteristics differ from the characteristics of higher levels in the book
3) The size of the differences between adjacent prices higher up in the book is larger than one tick
4) The most common events are cancellations
5) Most of the activity takes place at the bid-ask spread
6) Over the time of the day, orders and trades exhibit a U-shaped pattern
7) “Diagonal effect”, including related observations. Clustering of orders is due to strategic order splitting

2.4.4 Impact: Strategic order splitting

Impact is composed of price and market impact (Gould, Porter et al. 2013). Price impact refers to changes in the bid-price or ask-price resulting from a market actor’s action. Market impact refers to the effect on the whole book from a market actor’s action. While the issue of impact is important and is studied extensively, it is not crucial to provide a complete detailed review of it for this study. What needs to be understood is that large orders move the market against the market actor that places that order, and a tool used to prevent this from happening is strategic order splitting.

Strategic order splitting refers to the practice of splitting large orders into smaller pieces. Through this, investors can lessen their orders' impact on security prices as well as on the market in general. It is a key motivation, supported by (Bouchaud, Farmer et al. 2009), for the diagonal effect observed by BHC.

Except for supporting BHC, (Bouchaud, Farmer et al. 2009) also find the practice of order splitting to be prevalent in most markets in their paper. They furthermore quote other research, e.g., (Vaglica, Lillo et al. 2008), showing that orders split by institutions may execute over multiple weeks or even months.

In a study of market impact, (Hautsch, Huang 2009) report that traders interpret the arrival of market orders as a strong information signal. Interestingly, this goes in the opposite direction of what BHC reported about their market orders.

3 Dataset

This chapter presents an overview of the raw data used in this study, further explanations and basic descriptive statistics to provide a basis for replication of the study.

The data used to answer the research question is virtually identical to the data distributed through the NASDAQ Nordic Equity TotalView-ITCH (NETV) system, which disseminates messages in real-time during the trading day to all market participants subscribed to the service.

The format of the data is messages in rows in .csv files, and the total size of the dataset is ~3.6 GB. It contains exclusively passive orders and related executions from July 1 to July 31, 2019 for the OMX30 index, or a total of 24 trading days. The resolution is nanosecond-level. The data does not contain the original aggressive orders as they appear when they come into the exchange. See Table I for a sample of the raw data.

Table I

Sample of ITCH messages. The firmId column contains a unique code used to identify the company. The msgType defines which operation the message instructs. The orderId is the unique Id of each passive order sent to NASDAQ Stockholm. The side indicates whether the order is a Buy or Sell order.

timeStamp	firmId	msgType	orderId	side	price	volume
20190701.070019.102723633	3966	A	3467192	S	187.85	24000
20190701.070019.105420265	3966	A	3467223	B	186.75	4000
20190701.070019.105447426	3966	D	3347607	_	0	0
20190701.070019.105454675	3966	E	14372	_	187	100

An explanation of Table I follows. An A-message signals adding a passive order to the book. A D-message signals deleting a passive order from the book. An E-message signals the execution of a passive order. As is apparent from the sample data, the messages do not show the complete book in each message, only the changes that are applied to it.

Thus, Table I tells us that 24 000 units were added to the ask side at 187.85, followed by 4 000 units at 186.75 added on the bid side. Then, the order with the orderId 3347607 was deleted. The E-message implies that an aggressive order came into the exchange and lifted 100 from the limit order with orderId 14372. As is apparent, no order was shown as added in the stream before the E-message showing 100 lifted from the limit order 14372. This is the reason for needing to reverse-engineer aggressive orders.

NASDAQ Stockholm is the main market for the OMX30 index, but the index is traded outside of the market too, for example through Depositary Receipts of different types. This volume is not reflected in the dataset used for this study, and this market fragmentation is discussed in the limitation and analysis parts.

Regarding accessibility, the information available in this dataset is theoretically available to all investors. However, if one were to act professionally on it, high fees apply. Furthermore, as the analysis in this study deals with data on the nanosecond-level, many market participants are shut out of trading not only because of fees but also because the technological sophistication required to use this data is high.

Table II

Summary statistics on daily message market activity. The table only concerns ITCH-messages, not actual orders. The statistics are averages per day per stock. For example, ~4 700 execution messages are disseminated any given day for any given stock in the OMX30 index.

	Mean	Min.	1st Quart.	Median	3rd Quart.	Max.
Return (%)	- 0.1	- 10.1	- 0.8	- 0.1	0.6	5.2
Hi-Lo (%)	2.0	0.5	1.3	1.7	2.3	12.3
Add order (k)	39.4	8.3	23.7	35.0	48.8	168.6
Displayed trades (k)	4.7	0.7	2.8	4.0	5.4	37.3
Undisplayed trades	150.8	0.0	45.0	100.0	183.0	1 723
Displayed volume traded (k shares)	1 580.4	53.3	477.6	951.9	2 038.2	22 407.2
Undisplayed volume traded (k shares)	86.1	0.0	15.0	42.3	98.9	1 582.1
Displayed value traded (SEKm)	212.0	33.9	111.0	176.9	271.3	1 874.7
Undisplayed value traded (SEKm)	11.9	0.0	30.3	7.7	14.7	148.5

Table II presents summary statistics on daily message activity per firm. The exchange receives an average of ~39k add order messages per stock on any given day. The trading day is 8.5 hours long; thus, any given stock receives ~77 new passive orders every minute.

As expected, considering the LIS rules imposed on undisplayed orders, the average volume per execution of undisplayed volume of ~0.57k shares is higher than the average volume per execution of displayed volume of ~0.34k shares. On average, executions of undisplayed volume make up ~3% of total trades, but ~5% of total volume. Finally, the share of cancelled messages on the Stockholm stock exchange is high. The current data shows ~88% of messages ending in some other way than trades.

The average numbers for the Return and Hi-Lo are quite close to the numbers found by BHC. It is, however, worth noting that there seems to be a larger difference between min-max values during the current period than there was in 1991. BHC report a minimum (maximum) return of - 0.77%. (0.4%) while the current data shows - 10.1%. (5.2%). The tails, therefore, seem more extreme in current data than during the original period.

4 Methodology

As explained previously, to perform the analysis required to answer the research question, the order book has to be reconstructed, and the aggressive orders need to be reverse engineered. When the reverse-engineering is done, aggressive and passive orders finally have to be classified. This chapter details the procedures used for these three tasks to ensure replicability.

4.1 Reconstruction of the order book

NASDAQ documentation[2] explains the different message types disseminated in the ITCH stream and how they are supposed to be handled. Consequently, one can easily proceed to use the messages to reconstruct the book. (Huang, Polak 2011) describe the reconstruction process in detail, so this section only provides a short overview.

The fundamental driver of the program is a loop that goes through each message chronologically. The order book is stored in a hashmap during the process, see appendix III for a visual description of the structure. The unique orderId assigned to each message needs to be stored to track the order throughout its life. An auxiliary hashmap stores the orderId in combination with the volume assigned to that order, see appendix III for a visual description of that data structure as well.

For each message adding volume, the program slots it into its corresponding price level, or initiates a new price level if one did not previously exist. For each message deleting or executing volume, the program finds the corresponding orderId in the auxiliary hashmap and performs the required operation on the main book. The program deletes empty price levels whenever they appear.

Now that the order book is reconstructed, the reverse-engineering can begin.

4.2 Reverse-engineering aggressive orders

The ITCH-stream presents information about the market in the form of discrete messages, and it only disseminates the A-message, add order, and its corresponding

[2] Nordic Equity TotalView ITCH version 3.03.4

orderId for passive orders. Therefore, to be able to perform the analysis in BHC, aggressive orders need to be reverse engineered.

The messages used for this task are mainly E- and P-messages. E-messages represent the execution of displayed volume, and P-messages represent the execution of undisplayed volume. This study assumes that timestamps come from matching engine event time, to enable the reverse-engineering. See appendix IV for a discussion of the validity of the assumption in the NASDAQ Stockholm case. Either way, the method suffers from a weakness which makes it impossible to consider the period immediately following the uncross, see appendix IV for an explanation of this flaw.

Streaks are useful to define to describe the reverse-engineering procedure. A *streak* is a queue starting with an E- or P-message, containing subsequently only E- or P-messages following within one ns of each other.

The reverse-engineering logic is a loop that runs whenever the program hits an E- or P-message, that “matchId” number has not yet been processed. The matchId is a number that uniquely identifies every execution; it is not included in Table I due to space. When the logic is activated, the book is frozen. Book volumes and prices referenced below are book volumes and prices at the beginning of the streak. The logic tracks total reverse-engineered order volume via summing the individual volume of messages included in the current streak.

Now that the order book is reconstructed and the reverse-engineering logic is explained, a description of the classification of orders can follow.

4.3 Order classes

This study follows the naming convention and definitions established in BHC for the order classes, for comparability. That is, however, not possible for the Application class due to structural differences in the market. The analogue in this study is called the Hidden order class to emphasize the difference between the two.

Trades and orders are differentiated depending on two factors, aggressiveness and position in the book. There are eight classes on the ask-side, and eight classes on the bid-side. The aggressive orders are classified according to their aggressiveness, while the passive ones are classified according to their position in the book.
The most aggressive class is the Large order class. A limit sell order with a price lower than the bid-price, or higher than the ask-price for a limit buy order, and volume higher than the volume offered at the best bid-price is a Large order.

The second aggressive class is the Market order class. A limit, or market, order with a price equal to the best bid- or ask-price and volume equal to or higher than the volume offered at the best bid- or ask-price is a Market order. Actual (IOC) market orders cannot be differentiated from limit orders through the ITCH messages. They will therefore by definition be included partly in this class and partly in the Small order class below. Market actors do not observe the difference between IOC market orders and limit orders through the ITCH-stream either, so this is not a big issue.

The third aggressive class is the Small order class. A limit, or market, order with price equal to the best bid- or ask-price and volume equal to or below the volume offered at the best bid- or ask-price is recorded as a Small order.

The fourth and final aggressive class is the Hidden order class. Hidden orders are identified when there is an exclusively hidden execution occurring. Exclusive streaks of P-messages or single P-messages are therefore used to identify this order class. These messages are, however, used by NASDAQ Stockholm to signal both matches of undisplayed volume in the main book and matches from Nordic@Mid. Usually, it is possible to distinguish between an undisplayed execution in the main book and a Nordic@Mid execution, using a flag in the ITCH messages. The current dataset, unfortunately, does not include that flag. The Hidden order class, therefore, combines the execution of undisplayed volume in the main book and volume on Nordic@Mid.

The passive order classes are differentiated depending on whether the orders were placed in the bid-ask spread, at the bid-ask spread or above/below the bid-ask spread.

Following passive order classes are the Cancel-replace and Cancel classes. A Cancel-replace is a cancellation of an order combined with the immediate submission of a new order in its place. The new order may contain updated volume and updated price, or either of the two separately.

A Cancel is a deletion of previously posted volume. The Cancel order class also includes partial cancellations, meaning cancellations where the order might have been executed in part earlier.

5 Results: Order book

This chapter presents findings based on the new data regarding regularities 1 – 4 in section 2.4.3.1. It starts by the shape of the order book, continues with investigating the bid-ask spread characteristics relative to other levels in the book. Finally, it presents the results on the mean number of ticks between adjacent prices away from the bid-ask spread and the mean number of ticks at the bid-ask spread. Thus, this chapter answers the research question regarding regularities 1 – 4. All ticks are in SEK. Each section additionally presents relevant methods.

5.1 Order book shape

BHC construct the graphs using the five best bid-ask prices and argue that this is sufficient to present an accurate picture of the book. With the available data, it is possible to go deeper into the book. Therefore, this study uses the ten best bid- and ask-prices to answer the research question with regard to regularity 1) with slightly higher precision.

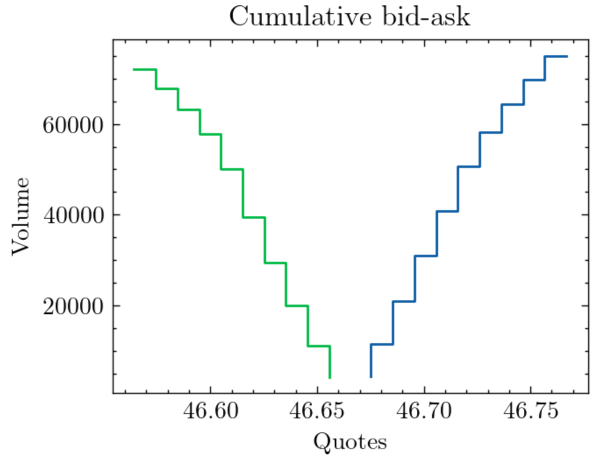
The method used to produce Figure II is as follows. The time-series averages of the prices at the first ten levels of the book are computed, followed by the average depth offered at these levels. Following that, grouping stocks into tick sizes, computing the cross-sectional means of the previous time-series means outputs the necessary data.

From ocular inspection of the graphs in Figure II, it appears as though the price schedule differs between tick sizes. At the smallest tick size, there is an average ask-side imbalance, and the rest of the books look balanced. In general, the price schedule looks weakly concave. The property is especially noticeable at tick size = 0.1.

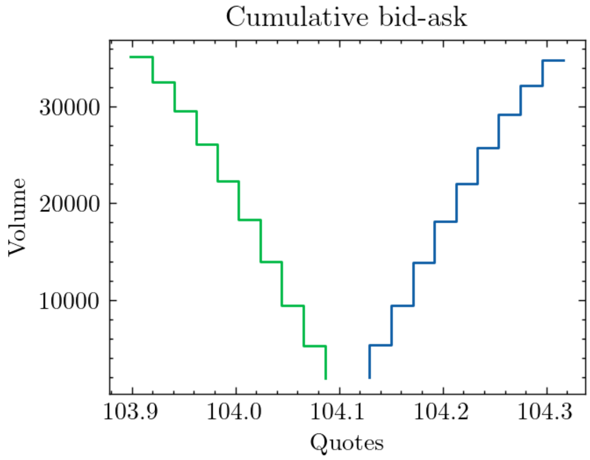
The exception to the weak concavity rule is the slope at tick size = 0.2, which looks closer to linear. However, there is only one stock in that tick size range in the OMX30, which might suggest not to take this result too seriously. See appendix II for a table detailing which of the stocks were in which tick ranges during the study period.

Figure II

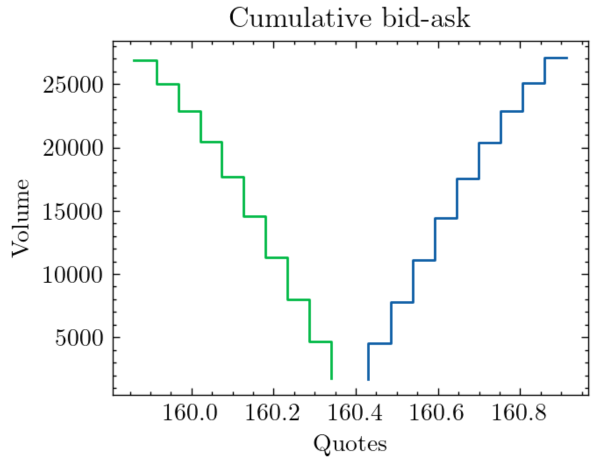
Limit order book shapes by tick sizes. One graph per tick size is presented. Ticks in SEK.



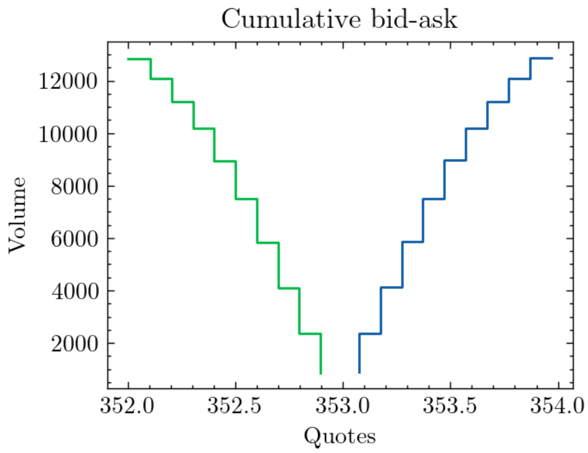
Tick = 0.01



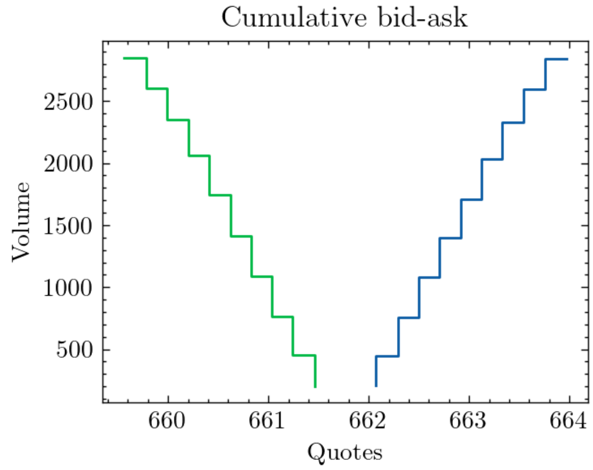
Tick = 0.02



Tick = 0.05



Tick = 0.1



Tick = 0.2

5.2 Bid-ask spread characteristics

To analyse the spreads and depths and thereby answer the research question regarding regularity 2), the Wilcoxon signed-rank test is applied to the sample of spreads and depths. The reason for using the Wilcoxon signed-rank test is as follows.

When looking at the spreads and depths, the data has values where it clusters significantly. This makes it inappropriate to assume normality. Furthermore, assuming independence of observations is not appropriate due to the complexity of the LOB system. Specifically, at least the tick size will represent a common factor between the bid-ask spread and the differences between prices away from the bid-ask spread, which cannot be assumed not to affect both. The Wilcoxon signed-rank allows for both of these specificities and is therefore suitable.

As the Wilcoxon signed-rank is a nonparametric test for dependent samples, it requires some pre-processing. The method is applicable if samples come in pairs, and there is a random and independent selection process of the pairs.

Therefore, all bid-ask spreads are paired with differences between adjacent prices away from the bid-ask spread through random sampling. The same process is followed for the depths. A random subset from the two samples of differences is then selected to perform the test on. The null hypothesis of the test is no difference in the compared samples.

Table III

Spreads and depths analysis using the Wilcoxon signed-rank method. Stocks are pooled in their respective tick groups before the test is carried out.

Hypothesis	Number of rejections (out of 30)
Equality of difference distributions including bid-ask spread	30
Equality of difference distributions excluding bid-ask spread	8
Equality of depth distributions including bid-ask spread	30
Equality of depth distributions excluding bid-ask spread	16

Table III presents the results from the Wilcoxon tests. The test rejects distribution equality for all stocks when including the bid-ask spread. However, it fails to reject the null hypothesis of equality of differences and depths for many stocks when excluding the bid-ask spread. There is, therefore, evidence pointing toward differences and depths away from the bid-ask spread being similar, while the bid-ask spread differs from differences and depths away from the bid-ask spread.

5.3 Price discreteness

This section aims to answer the research question with regard to regularity 3), which is that differences between price levels higher up in the book is larger than one tick.

Table IV presents results on the mean number of ticks in the bid-ask spread, as well as between prices away from the bid-ask spread. The perform the analysis, the study splits the stocks into their respective tick sizes, and then for each day, for each stock computes the mean number of ticks at the bid-ask spread, and away from the bid-ask spread for the ten first levels of the book. Then, these values were averaged within their respective tick sizes.

Table IV

Comparison bid-ask spread w. other levels of the book in terms of distance in ticks. All tick sizes are in SEK.

	Mean number of ticks at bid-ask spread	Mean number of ticks away from bid-ask spread
Tick size = 0.01	1.92	1.02
Tick size = 0.02	2.11	1.04
Tick size = 0.05	1.79	1.07
Tick size = 0.1	1.76	1.00
Tick size = 0.2	3.07	1.05

Clearly, the bid-ask spreads are wider than the tick size, with the mean being ~2-3x the tick. The mean number of ticks away from the bid-ask spread present noteworthy results. The tick is binding away from the bid-ask spread for all sizes. Using the median instead of the mean obtains similar results.

6 Results: Order flow

This chapter presents findings regarding regularities 4 – 7. It starts by showing unconditional frequencies of orders and trades, speaking to regularities 4) and 5). That is followed up by an investigation of frequencies of orders and trades conditional on the time of day, equivalent to regularity 6). Finally, the chapter ends with presenting the frequencies conditional on other events. That investigates whether the diagonal effect and the related observations present themselves in current data, and thereby addresses regularity 7).

Table V

Unconditional frequencies of trades and orders. Percentages and nominal values are averages per day per firm.

Type	Event	Percentage	Nominal
	Buy-side		
Trade	Large buy	0.03%	25.0
	Market buy	0.07%	63.8
	Small buy	1.63%	1 441.4
Order	New bid in spread	1.85%	1 628.0
	New bid at spread	8.37%	7 387.5
	New bid below spread	12.01%	10 597.4
Cancellation	Cancel-replace	4.39%	3 871.0
	Cancel	22.06%	19 460.0
	Sell-side		
Trade	Large sell	0.03%	23.7
	Market sell	0.07%	60.0
	Small sell	1.84%	1 351.1
Order	New ask in spread	1.84%	1 619.9
	New ask at spread	8.29%	7 311.1
	New ask above spread	11.76%	10 374.3
Cancellation	Cancel-replace	4.25%	3 749.1
	Cancel	21.71%	19 154.7
	Hidden	0.11%	100.4

6.1 Unconditional probabilities of orders and trades

This section presents results based on new data of regularities 4) and 5). Table V contains unconditional probabilities of different order classes. For example, ~1 500 aggressive orders came in for the average stock on any given trading day during the sample period. Table II shows different numbers because one aggressive event may contain multiple ITCH messages.

Cancels is the most frequent class, reflecting the importance of cancellations incorporated in the strategies of market participants. Recall that Cancels include partial cancellations, which represent ~90% of all Cancels. Therefore, the nominal frequency relative to the add orders seems high.[3] Orders away from the bid-ask spread follow Cancels in commonality. The high rate of orders away from the bid-ask spread suggests that market actors use the entirety of the book to a large extent.

6.2 Probabilities of trades and orders conditional on time of day

This section investigates regularity 6). Figure III presents the probability of an event occurring conditional on the time of day. The graphs present an overview of trades and orders, grouped in 10-minute bins over the trading day. The percentage on the y-axis is the share of the number of trades or orders in that particular 10-minute bin relative to the total number of trades or orders during the day. For example, Small orders grouped in the 10-minute periods around 12:00 represent ~1.5% of the total number of trades during the day on average in the dataset. In general, the results confirm the pattern of high activity in early mornings and late afternoons represented by the U-shaped pattern.

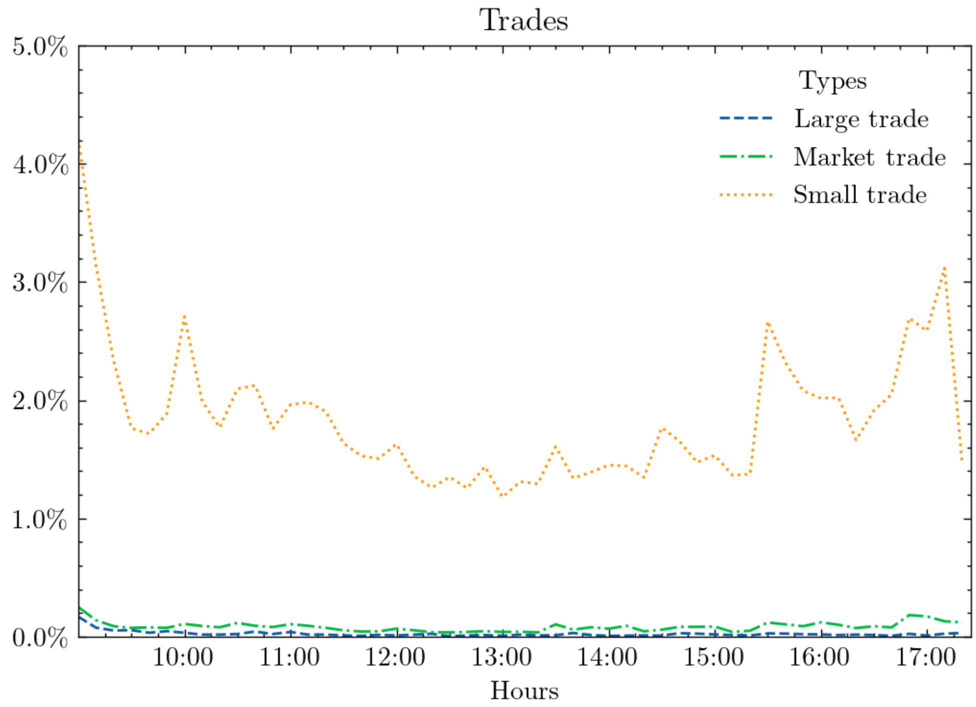
Three observations from the graphs are highlighted. First, the lines look quite spikey in general, implying variance between groups on the 10-minute level. Second, there are visibly few Large and Market orders as a share of the total number of aggressive orders. Third and final, a visible jump in activity both in trades and orders presents itself at around 15:30 in the trading day.

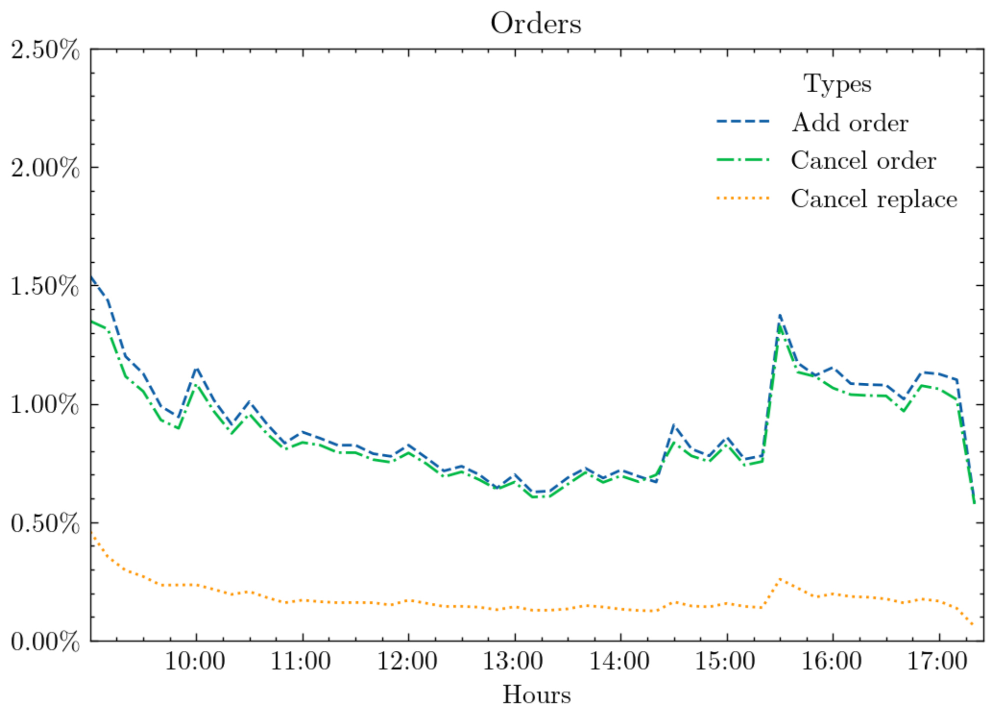
Activity, in general, tapers off after the 15:30-spike, although stays at a relatively high level compared to the rest of the day.

[3] Results have been analysed splitting pure and partial cancel categories. They are not reported for brevity as no significant differences in conditional patterns were found between Cancel and Partial Cancel

Figure III

Frequency of events conditional on time of day. The percentages are averages per day per firm. The trades and orders are grouped in 10-minute bins, with the y-axis representing the fraction of the total number of the type of trade or order during the day attributable to the 10-minute bin at the relevant point in time.





6.3 Probabilities of orders and trades conditional on the last order or trade

This section answers the final subpart 7) of the research question. It first deals with the diagonal effect, then with the separate observations from BHC and finally comments on some noteworthy results unrelated to the diagonal effect and the separate observations.

The orders are, after applying reverse-engineering, placed in succession and looped through to obtain conditional probabilities of orders and trades. For each order, the program checks the class of the next order and adds it to a count, see appendix IV for the data structure used. The count is then divided by the number of total orders following each order to find the share corresponding to each class.

Furthermore, since this section compares unconditional and conditional probabilities, a χ^2-test for statistical significance of difference is performed. The method used averages the conditional class results per firm for each trading day. For each group of 24 values, it runs the test comparing them to the unconditional probability. Thereby, the statistical significance of difference is established on the aggregate level per day. All results that are commented on are statistically significant at the 5%-level at least.

Table VI (1) and Table VI (2) detail the results. The subsections with comments treat them as a function of their prior order.

6.3.1 The diagonal effect

While there is some evidence of the diagonal effect in current data, the relationship is not strong. It is a question of interpretation when considering whether to classify current data as possessing the “diagonality”. While the events do *tend* toward clustering around the diagonal, the effect is by no means clear. This is discussed further in chapter 7.

Except for the diagonal effect, BHC also report separate observations that are not in accordance with it. These are presented below.

6.3.2 Conditional relationships separate from the diagonal effect

This section treats the results in new data based on the specifically mentioned cases in BHC. The cases not following the diagonal effect in BHC were:

1) Order in the spread on the same side following Large orders

2) Cancellations on the opposite side following Large orders

3) Cancellations on the same side following Market orders

4) Market and Small orders on the opposite side following Market orders

5) Applications following Large orders

Regarding 1), the effect still holds. Following Large orders, there is a high probability of bids in the spread following in current data.

Also, for 2), the effect still holds. Cancellations on the opposite side following Large orders are among the most common observations there are.

When it comes to 3), this effect is not observed. There are no signs of Cancellations being common after Market orders; in fact, the conditional probability is much lower than the unconditional one, pointing toward an opposite effect.

Regarding 4), the effect still holds. Although now, Large orders are also included in the common orders following Market orders and Small orders.

Finally, regarding 5). While Applications are not in definition similar to Hidden orders in this study, some similarities in conditional probabilities do exist. These are discussed further in chapter 7. For this specific case, however, the relationship does not hold.

6.3.3 Other conditional relationships

Two observations are worth noting in this regard. First, Large orders, Market orders and orders in the bid-ask spread present a high conditional frequency following Hidden orders. Second, there is a significant frequency of Small orders on the opposite side following orders in the bid-ask spread.

Table VI (1)

Conditional probabilities of orders based on an order at time t – 1. Table VI (1) presents conditional orders on the bid side. Table VI (2) presents conditional orders on the ask side and the Hidden order column. Each row corresponds to an event at time t – 1, each column corresponds to an event at time t. The three largest probabilities in each column are bolded. Each row of each table is a probability vector. It adds up to 100% if combined with the equivalent row in Table VI (2). Place Table VI (2) to the right of Table VI (1) when reading to ease interpretation.

t - 1	**Bid**							
	Large	Market	Small	In s.	At s.	Bel. s.	Replace	Cancel
Large buy	**0.15**	0.08	1.34	**14.96**	**21.71**	**19.68**	0.30	2.78
Market buy	0.01	0.03	0.15	2.17	12.03	**20.26**	0.87	7.72
Small buy	0.05	**0.61**	**12.08**	**8.09**	**19.12**	9.15	1.14	7.20
Bid in spread	0.01	0.03	1.08	1.37	16.30	**21.77**	3.21	16.08
Bid at spread	0.02	0.09	2.44	3.98	**17.36**	13.42	2.84	19.89
Bid bel. spread	0.03	0.05	1.61	1.66	8.07	17.29	4.29	23.08
Cancel-replace	0.03	0.05	1.27	1.16	5.19	13.44	**16.74**	24.48
Cancel bid	0.02	0.06	1.16	1.47	7.84	13.57	**5.83**	30.38
Large sell	0.01	0.01	2.82	1.44	1.64	1.31	0.54	**38.95**
Market sell	**0.09**	**0.31**	**8.44**	7.56	4.66	2.49	0.62	**31.18**
Small sell	0.01	0.02	0.32	3.43	3.72	3.71	1.26	**31.39**
Ask in spread	0.04	0.21	**3.58**	0.77	3.49	3.59	1.19	28.05
Ask at spread	0.01	0.02	0.67	1.20	4.43	5.86	1.91	27.01
Ask ab. spread	0.04	0.07	1.38	1.03	4.71	11.13	3.16	21.64
Cancel-replace	0.05	0.05	1.23	0.85	4.10	7.47	**7.01**	17.23
Cancel ask	0.02	0.06	1.63	2.00	9.84	11.13	3.07	13.24
Hidden	**0.13**	**0.92**	2.54	**12.15**	9.27	5.16	0.68	10.75
Unconditional	0.03	0.07	1.63	1.85	8.37	12.01	4.39	22.06

Table VI (2)

Conditional probabilities of events based on an order at time t – 1.

t - 1	**Ask**								
	Large	Market	Small	In s.	At s.	Ab. s.	Replace	Cancel	Hidden
Large buy	0.01	0.01	2.58	1.43	1.46	1.25	0.61	**31.62**	0.03
Market buy	**0.09**	**0.31**	**7.78**	7.76	4.37	2.15	0.60	28.39	**5.30**
Small buy	0.01	0.02	0.37	3.09	3.84	3.79	1.26	**29.00**	1.18
Bid in s.	0.04	0.20	**3.75**	0.78	3.23	3.23	1.18	27.70	0.03
Bid at s.	0.01	0.02	0.57	1.12	4.61	5.30	1.72	26.51	0.11
Bid below s.	0.03	0.06	1.17	0.90	4.43	13.86	2.67	20.76	0.04
Cancel-rep.	0.04	0.05	1.08	0.88	3.97	6.90	**8.39**	16.32	0.02
Cancel bid	0.02	0.05	1.51	1.91	9.77	10.28	2.70	13.38	0.05
Large sell	**0.12**	0.07	1.50	**13.95**	19.75	15.01	0.44	2.39	0.05
Market sell	0.00	0.03	0.18	2.37	10.99	**17.46**	0.70	7.63	**5.27**
Small sell	0.03	**0.60**	**12.20**	**8.71**	**17.84**	8.02	1.05	6.67	1.03
Ask in s.	0.01	0.03	1.12	1.43	**16.46**	**20.72**	3.32	15.95	0.04
Ask at s.	0.02	0.08	2.36	4.19	**17.55**	12.65	2.70	19.23	0.11
Ask above s.	0.03	0.05	1.56	1.71	8.18	**17.20**	**4.36**	23.71	0.04
Cancel-rep.	0.03	0.05	1.25	1.19	5.01	13.11	**17.70**	23.65	0.02
Cancel ask	0.02	0.06	1.08	1.50	7.77	12.92	5.59	**30.02**	0.06
Hidden	**0.15**	**0.98**	2.35	**14.31**	7.95	4.63	0.63	11.90	**15.50**
Unconditional	0.03	0.07	1.84	1.84	8.29	11.76	4.25	21.71	0.11

7 Discussion

This chapter compares regularities observed in prior research with the ones found in current data and discusses their significance, as well as suggesting future research based on them. The chapter ends with a discussion of the limitations of the study.

The regularities referenced in this section are the following:

1) The order book is concave
2) The bid-ask spread characteristics differ from the characteristics of higher levels in the book
3) The size of the differences between adjacent prices higher up in the book is larger than one tick
4) The most common events are cancellations
5) Most of the activity takes place at the bid-ask spread
6) Over the time of the day, orders and trades exhibit a U-shaped pattern
7) "Diagonal effect" including related observations. Clustering of orders is due to strategic order splitting

7.1 The order book: regularities 1 – 3

Regarding regularity 1) BHC find a weakly concave shape of the book, for 2) they reject equality of differences and depths, including the bid-ask spread. They are not able to reject in most cases when excluding the bid-ask spread, meaning that their evidence points toward there being no difference between the bid and ask side characteristics.

Current data exhibits a weakly concave book shape and bid-ask spreads are larger than the differences between adjacent prices away from the bid-ask spread in general. The depth at the bid-ask spread is also different from the depth at the levels away from the bid-ask spread.

It is therefore concluded that regarding regularities 1) and 2), current LOB data shows similar characteristics as has been observed previously.

Regarding regularity 3), BHC find the bid-ask spread to be 3x the tick for stocks with tick FF 1, and 9x the tick for stocks with tick FF 0.1. This is not supported in current data. The bid-ask spread now seems to be closer to 2x the tick.

Furthermore, the OMX30 book, in general, seems tighter than the book BHC reported. The average bid-ask spread is about two times the tick for all tick sizes, while the difference between adjacent prices away from the bid-ask spread is one tick for all tick sizes, where BHC find that adjacent prices higher up in the book have differences substantially larger than one tick. These results suggest that the tick size may hinder a tighter book from materialising for all stocks in the OMX30, rather than for just the larger tick size stocks as in BHC.

BHC motivate the lack of a tight book for its small tick stocks by showing by example that a stock with a small tick size may not have a tight book due to it not being profitable to compete away the holes in the book, see section 2.4.2.3.

A hypothesis for why no holes present themselves today could be the increased computerisation of the market. Given that inputting and cancelling orders today is fast and requires no human input, the competition for liquidity provision should be tougher. Furthermore, the perfect rationality model was discussed and critiqued after the publication of BHC. Specifically, there are questions regarding whether a fair price exists for a given asset. The example quoted in BHC relies heavily on the existence of such a fair price, and the ability of the market actors to calculate it. If it does not exist, or is impossible to calculate, the example is not relevant to justify how market actors think when pricing orders.

7.2 The order flow: regularities 4 – 7

When it comes to regularity 4), BHC find that most of the activity takes place at the bid-ask spread, since Small orders is the most common type, followed by passive orders in or at the bid-ask spread. This is not the case in current data. Orders below the spread alone now make up about half of the total passive order placement, and this share is almost unchanged when considering trades.

However, BHC only use the first five levels of the book for their analysis, so the explanation to this difference may be natural given that this study considers the entirety of the book. It would be interesting to investigate the order placement in these higher levels of the book more closely, given that they represent such a large share of activity in the order book. In any case, the previous results do not conform with current data, and so regularity 4) is not confirmed by this study.

Regarding regularity 5), later studies have already observed high rates of cancellations. It is safe to say that market actors have evolved in their usage of cancellations from the time of BHC, but compared to what was known before this study, no major change is noticed. The rate of cancellations is around what is expected given the later studies in the field.

Regularity 6) is interesting. BHC find a U-shaped pattern in order placement with high activity in the morning and evening. Their hypothesis for this observation is price discovery in the morning and unwinding of positions in the afternoon. Current data support this finding, and furthermore present a definite increase in activity around the 15:30-mark. The reason for the spike is not clear.

A hypothesis for this observation is interdependence between NASDAQ Stockholm and the U.S. stock markets. The U.S. markets open at 15:30 Stockholm time, and we know that some OMX30 stocks trade via Depositary Receipts, as per chapter 3. This could, in any case, be a topic for future research, to understand whether this occurs in other markets and if so, why.

7.3 The diagonal effect

The last regularity 7) is wide and includes more detailed observations, except for the diagonal effect, presented in chapter 6.

Regarding the diagonal effect, as mentioned in section 6.3.1, this study believes the question of whether it exists or not is one of interpretation. However, it would be difficult to see a new study arguing that there is such a diagonal effect only observing current data, and without having read BHC. Therefore, this study would tend toward suggesting the diagonal effect is not appropriate to use in the modelling of LOBs. If one looks to include conditional relationships in one’s model, it would be more appropriate to use a couple of the separate observations that have been proven to hold over a long time instead of the full diagonal effect.

Separate from this conclusion, some clustering appears. While the suggested explanation for this has been strategic order splitting, it could be discussed whether this is reasonable given the current data. One should remember that the clustering observed in this study occurs on millisecond timescales. Since strategic order splitting’s purpose is to minimize market impact, it would be strange if market actors send their orders in immediate succession on time intervals that are this short. That would probably not minimize market impact, rather the opposite.

7.3.1 Conditional relationships separate from the diagonal effect

The detailed observations are the following:

1) Order in the spread on the same side following Large orders
2) Cancellations on the opposite side following Large orders
3) Cancellations on the same side following Market orders
4) Market and Small orders on the opposite side following Market orders
5) Applications following Large orders

With respect to 1) and 2) the effect still holds.

Point 3) is interesting. The observed effect of cancellations following Market orders is gone. There could be a logical explanation for this disappearance. BHC motivate their finding by saying that the observation is a sign of market actors hunting for undisplayed liquidity by placing Market orders and then cancelling the unexecuted part. It is unlikely that the behaviour has actually disappeared, since (Hasbrouck, Saar 2009) find similar effects and explain them by, partly, suggesting undisplayed liquidity plays an increasingly important role in market actor's strategies. A hypothesis for why the effect does not appear at NASDAQ Stockholm is, therefore, the following. Market actors may use market orders that are IOC for this purpose; this allows them to avoid having to cancel any outstanding liquidity after execution. No IOC orders existed at the Paris Bourse when BHC studied the subject.

Result 4) also persists but is wider. Current data suggests Large orders, Market orders and Small orders all follow Market orders.

BHC motivate the finding by suggesting the Small and Market orders provide liquidity to the preceding Market order and interpret the effect as giving a hint about what the market thinks of the information value in Market orders. The information value is low since it attracts lots of interest from the opposite side.

Indeed, the effect seems even stronger today, and if the interpretation by BHC is correct, then the data suggests the market places even less informational value on Market orders today compared to the time of BHC. (Hautsch, Huang 2009) present conflicting results in this regard. They find that market orders were interpreted as possessing high information value. However, their definition of market orders were all orders taking liquidity from the book, which is different from the definition of Market orders in this study and BHC alike. An interesting further study in this area would be to define market orders as streaks corresponding to IOC orders, and see how the market responds in terms of conditional probabilities. This could be more relevant for today's markets, as "market orders" today usually are IOC.

7.3.2 Other conditional relationships

This section discusses some observations in current data differing from the ones in BHC. It first comments on the different orders following Hidden orders, and then the behaviour of Small orders on the opposite side following orders in the bid-ask spread.

The Hidden order class in this study differs in definition from the Application class due to data limitations. However, it is interesting to note that the two categories present similarities in terms of conditional events. While effect 5), in section 7.3.1, does not show up in current data, a similar effect does appear where Hidden orders often follow Market orders instead. This suggests that there possibly is a similar kind of relationship between larger orders, undisplayed liquidity and the market's actions as there was during the time of BHC, but that it has shifted location.

Furthermore, in current data, Large orders, Market orders, and orders in the spread are common following Hidden orders. In BHC, no order type was clearly common following Applications. An interpretation of the Large order and Market order part of the observation that fits with the findings by (Hasbrouck, Saar 2009) is that market actors exploit undisplayed liquidity more aggressively today than they previously did.

The orders in the bid-ask spread following Hidden orders could instead suggest market actors shifting the book according to where undisplayed liquidity is found, which would be a further sign of weight put on undisplayed liquidity. An interesting further study in this regard would be to compare events conditional on Nordic@Mid executions with those conditional on executions in the main book, which was not possible given the data in this study.

When it comes to the result of Small orders often being placed following orders in the bid-ask spread, if we stay in the framework of BHC, this could suggest a type of liquidity supplying similar to the one observed for Market orders. Since Market orders often act in essence as an order in the bid-ask spread following their "aggressive part" by which they take liquidity, this interpretation is reasonable.

7.3 Limitations

This section discusses potential limitations to the accuracy of this study and thereby to the answer to the research question. These limitations are the analysis method, the restricted sample, the classification issues in terms of Applications versus Hidden orders, differences in market characteristics at the sampling time and finally, strategic splitting of orders.

LOBs are difficult to study for multiple reasons. These include differing priority rules, incomplete sampling of data, undisplayed liquidity issues and the perfect rationality versus zero intelligence discussion (Gould, Porter et al. 2013). This causes the problem that the chosen analysis methods for this study may not capture the properties that are interesting or useful to understand with regard to modern LOBs.

The study uses the OMX30 index to provide relatively comparable data to the CAC40 through including similar-sized companies and trading conditions. Furthermore, both samples of companies are part of their respective national indices. However, this study does not comprise flow data from other exchanges. According to previous research, other exchanges appear to exhibit different characteristics. Therefore, results from NASDAQ Stockholm may not be applicable to other exchanges.

Furthermore, the research question is difficult to answer with regard to Applications versus Hidden orders. Due to their difference in definition, it is not clear whether the results from BHC still apply.

When it comes to different market characteristics, two things are worth pointing out.

First, trading of OMX30 companies occurs outside of NASDAQ Stockholm via different vehicles. The off-exchange trading occurs, for example, through American Depositary Receipts. The practice was illegal for the CAC40 at the time of BHC. The current market fragmentation can harm comparability to BHC as they presumably captured a large share of the order flow in their study, although trading did occur outside of the Paris Bourse at the time as well, something that is mentioned by BHC.

Second, this study uses data from July 2019 due it to being a good volatility match with the volatility during the period of the original paper as measured by the VIX. The VSTOXX did not exist at the time of BHC. The descriptive results of today, however, suggest that tail events are more extreme in the current period than in the period of

BHC. If volatility was higher during the period selected for this study, that might impact the conclusion of the research question.

Lastly, strategic order splitting. That practice poses problems when trying to conduct studies such as this one, as there is a risk that some of the observed conditional effects stem from a single larger order that is in fact split into smaller pieces rather than different market actors responding to one another. The reconstruction method used in this study only considers realised orders, as accounting for large orders that are split requires proprietary data. If market actors executed significant metaorders in the markets during the period of the study, this might harm results. Although, not necessarily comparability as this practice also seemed relatively common at the time of BHC.

8 Conclusion

The purpose of this study is to understand whether a selection of characteristics found in BHC and later research regarding the order book and order flow accurately reflected modern markets. The motivation is to provide updated empirical information to the research field following a request by a review paper. The review paper points out that many LOB models are currently based on old or small empirical datasets. The research question is the following.

Do the existing foundations for LOB modelling, regarding the order book and order flow, accurately reflect modern LOBs?

To do so, it uses high-frequency data from the NASDAQ Stockholm ITCH-stream. The method requires reconstruction of the order book and reverse-engineering of aggressive orders, as well as classification of the orders as a function of their aggressiveness and position in the book. The study starts by analysing the order book shape, bid-ask spread characteristics and price discreteness, and then presents results on unconditional and conditional probabilities of different order classes. The conditional probabilities are set to be dependent on the time of day, as well as the previous order. Given the results, the answer to the research question is no.

The results are meaningful to enable more accurate and up-to-date modelling of LOBs, as well as to enhance the understanding of order flow patterns on modern exchanges in general. A better understanding of these phenomena may be helpful when, for example, choosing how to best execute orders.

For future research, this study suggests three subjects to be of interest based on its findings. First, a deeper study to uncover the reason for the tightness of the book at higher levels would provide insight into whether the results at NASDAQ Stockholm are applicable in other markets and uncover some of their drivers. Second, the shape of the graphs of trades and orders conditional on the time of day suggests something happens around 15:30 at NASDAQ Stockholm; a future study could establish the reason for this activity spike. Third and last, the Hidden order class includes Nordic@Mid orders. It would be interesting to investigate conditional relationships of executions tied to undisplayed liquidity in the main book exclusively, and executions tied to Nordic@Mid exclusively to see which type of orders drives which behaviour.

References

BIAIS, B., HILLION, P. and SPATT, C., 1995. An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse. *Journal of Finance,* **50**(5), pp. 1655,1689.

BOUCHAUD, J., FARMER, J.D. and LILLO, F., 2009. *CHAPTER 2 - How Markets Slowly Digest Changes in Supply and Demand.* San Diego: North-Holland.

BOUCHAUD, J., MÉZARD, M. and POTTERS, M., 2002. Statistical properties of stock order books: empirical results and models. *Quantitative Finance,* **2**(4), pp. 251-256.

BUTI, S., RINDI, B. and WERNER, I., 2010. Diving into Dark Pools. *IDEAS Working Paper Series from RePEc,* .

CAO, C., HANSCH, O. and WANG, X., 2008. ORDER PLACEMENT STRATEGIES IN A PURE LIMIT ORDER BOOK MARKET. *Journal of Financial Research,* **31**(2), pp. 113-140.

CHAKRABORTI, A., PATRIARCA, M. and ABERGEL, F., 2011. Econophysics: agent-based models. *IDEAS Working Paper Series from RePEc,* .

CHALLET, D. and STINCHCOMBE, R., 2003. Non-constant rates and over-diffusive prices in a simple model of limit order markets. *Quantitative Finance,* **3**(3), pp. 155-162.

CHALLET, D. and STINCHCOMBE, R., 2001. Analyzing and modeling 1+1 d markets. *Physica A: Statistical Mechanics and its Applications,* **300**(1-2), pp. 285-299.

ELLUL, A., HOLDEN, C. and JAIN, P., 2003. Determinants of order choice on the New York Stock Exchange. *Journal of Finance,* .

GOULD, M.D., PORTER, M.A., WILLIAMS, S., MCDONALD, M., FENN, D.J. and HOWISON, S.D., 2013. Limit order books. *Quantitative Finance,* **13**(11), pp. 1709-1742.

GU, G., CHEN, W. and ZHOU, W., 2008. Empirical regularities of order placement in the Chinese stock market. *Physica A: Statistical Mechanics and its Applications,* **387**(13), pp. 3173-3182.

HASBROUCK, J. and SAAR, G., 2009. Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets,* **12**(2), pp. 143-172.

HAUTSCH, N. and HUANG, R., 2009. The Market Impact of a Limit Order. *IDEAS Working Paper Series from RePEc,* .

HOLLIFIELD, B., MILLER, R.A. and SANDÅS, P., 2004. Empirical Analysis of Limit Order Markets. *Review of Economic Studies,* **71**(4), pp. 1027-1063.

HUANG, R. and POLAK, T., 2011. LOBSTER: Limit Order Book Reconstruction System. *SSRN Electronic Journal,* .

NASDAQ, 2020-last update, NASDAQ Glossary. Available: https://www.nasdaq.com/glossary/l/limit-order-book [16 June, 2020].

PARLOUR, C. and SEPPI, D., 2008. Limit Order Markets. A survey. *Handbook of Financial Intermediation and Banking,* .

PHILLIPS, M., 2012. Where has all the stock trading gone? *Bloomberg Businessweek,* .

ROSU, I., 2009. A Dynamic Model of the Limit Order Book. *The Review of Financial Studies,* **22**(11), pp. 4601.

ROTH, A., MURNIGHAN, J., SCHOUMAKER, F., NEELIN, J., SONNENSCHEIN, H., SPIEGEL, M., BINMORE, K., SHAKED, A. and SUTTON, J., 1988. The Deadline Effect In Bargaining: Some Experimental Evidence. *The American Economic Review,* **78**(4), pp. 806.

SMITH, E., FARMER, J.D., GILLEMOT, L. and KRISHNAMURTHY, S., 2003. Statistical theory of the continuous double auction. *Quantitative Finance,* **3**(6), pp. 481-514.

VAGLICA, G., LILLO, F., MORO, E. and MANTEGNA, R., 2008. Scaling laws of strategic behavior and size heterogeneity in agent dynamics. *Physical review.E, Statistical, nonlinear, and soft matter physics,* **77**, pp. 036110.

Appendices

Appendix I

Liquidity band						
Price ranges	0 ≤ ADNT < 10	10≤ ADNT < 80	80 ≤ ADNT < 600	600 ≤ ADNT < 2000	2000 ≤ ADNT < 9000	9000 ≤ ADNT
0 ≤ price < 0.1	0.0005	0.0002	0.0001	0.0001	0.0001	0.0001
0.1 ≤ price < 0.2	0.001	0.0005	0.0002	0.0001	0.0001	0.0001
0.2 ≤ price < 0.5	0.002	0.001	0.0005	0.0002	0.0001	0.0001
0.5 ≤ price < 1	0.005	0.002	0.001	0.0005	0.0002	0.0001
1 ≤ price < 2	0.01	0.005	0.002	0.001	0.0005	0.0002
2 ≤ price < 5	0.02	0.01	0.005	0.002	0.001	0.0005
5 ≤ price < 10	0.05	0.02	0.01	0.005	0.002	0.001
10 ≤ price < 20	0.1	0.05	0.02	0.01	0.005	0.002
20 ≤ price < 50	0.2	0.1	0.05	0.02	0.01	0.005
50 ≤ price < 100	0.5	0.2	0.1	0.05	0.02	0.01
100 ≤ price < 200	1	0.5	0.2	0.1	0.05	0.02
200 ≤ price < 500	2	1	0.5	0.2	0.1	0.05
500 ≤ price < 1 000	5	2	1	0.5	0.2	0.1
1 000 ≤ price < 2 000	10	5	2	1	0.5	0.2
2 000 ≤ price < 5 000	20	10	5	2	1	0.5
5 000 ≤ price < 10 000	50	20	10	5	2	1
10 000 ≤ price < 20 000	100	50	20	10	5	2
20 000 ≤ price < 50 000	200	100	50	20	10	5
50 000 ≤ price	500	200	100	50	20	10

Tick sizing grid according to MiFID II. ADNT is shorthand for Average Daily Number of Transactions and is calculated and published by ESMA. ESMA is the European regulator responsible for financial markets oversight.

Appendix II

0.01	0.02	0.05	0.1	0.2
Nordea Bank	Ericsson B	Alfa Laval	Assa Abloy B	Autoliv
SSAB A	H & M B	Getinge B	Atlas Copco A	
Telia Company	SCA B	Sandvik	Atlas Copco B	
	SEB A	Securitas B	Boliden	
	SHB A	Skanska B	Electrolux B	
		SKF B	Essity B	
		Swedbank A	Hexagon B	
		Tele2 B	Investor B	
		Volvo B	Kinnevik B	
		ABB	Swedish Match	
			AstraZeneca	

Tabulation of tick sizes (SEK) as of July 1, 2019, for the OMX30 index. Tick sizes are subject to MiFID II and may change without forewarning from the exchange after new data is received from ESMA.

Appendix III

For completeness and pedagogy, the initial attempt and following optimisation are both described. To track the message impact, one needs to store the unique orderId assigned to each message by NASDAQ. Initially, the book was constructed through nested hashmaps with the structure: {side: {price: {orderId: volume}}}. For example, given two limit orders registered on the bid side at 102.2, the book object looked as follows:

```
{
    bid: {
        102.2: {
            12345: 100,
            12346: 200
            }
    }
}
```

The book object was for each message stored in an outer hashmap with timestamps as keys. The structure proved impossible to use as the RAM required to store the outer hashmap proved too high.

The reconstruction software was then optimised. The book object is still a nested hashmap. Consider the same hypothetical situation as before. The structure now looks as follows:

{

bid: {

102.2: 300

}

}

An auxiliary hashmap on the format {orderId*:* volume} was added to keep track of the orderIds. Again, for the same hypothetical situation:

{

12345: 100,

12346: 200

}

So, when there is, for example, an incoming execution message relating to order *12345*, the software looks up the corresponding volume in the auxiliary hashmap. It then proceeds to perform the necessary operations on the book object, given the information. That entails a drastically reduced book object.

Appendix IV

The reverse-engineering was completed in cooperation with NASDAQ staff and following suggestions from one of the founders of LOBSTER[4]. According to the research performed for this thesis, no standard procedure for this type of reverse-engineering has been described in the literature.

As said in section 4.2, timestamps are assumed to come from matching engine event time. That means that the matching engine at NASDAQ timestamps messages when the match, or message registration, occurs.

[4] The Limit Order Book Reconstruction System, https://lobsterdata.com/

The alternative would be for ITCH messages to be timestamped somewhere else. To make the assumption invalid, messages would there have to be stamped in bulk and sent out in intervals. Thus, it would entail multiple different types of messages being timestamped simultaneously and sent out in packets.

Because multiple order types would sometimes share timestamps if they are not timestamped according to matching engine event time, the validity of the assumption is easy to check empirically through, for example, add order (A) messages outside of the seconds immediately following cross periods. Those seconds are special because the engine places all uncrossed orders on the book at the same time and timestamps them within one ns of each other.

Therefore, if there are any streaks of exclusively A-messages within one ns, then timestamps do not come from matching engine event time.

It turns out that in the current dataset, no A-messages have timestamps within one ns of each other outside of cross periods. Thus, we may assume that timestamps come from matching engine event time.

This assumption leads us to the conclusion that the only way for messages to have timestamps within one ns from each other is for them to originate from the same (aggressive) order. This logic allows for the accurate reconstruction of these aggressive orders.

END OF PART 1

START OF PART 2

Abstract

This study compares time interval relationships between events (any market action, e.g., buying, cancelling an order, placing a passive order etc.) on the NASDAQ Stockholm exchange today with the relationships found on the Paris Bourse in 1995 in a study by (Biais, Hillion et al. 1995), to provide insights into modern limit order book characteristics. The research is conducted by reconstructing the order book and aggressive orders from messages disseminated through the NASDAQ Nordic Equity TotalView (NETV) ITCH system. This is followed by classifying the events according to their aggressiveness. Time intervals are then analysed conditional on previous time interval lengths, previous spread sizes, previous event classes, and by creating event sequences. Results point towards patterns in (Biais, Hillion et al. 1995) still being generally true, with market interactions happening several orders of magnitude faster today. Furthermore, results show a tendency towards greater differences between conditional intervals and the unconditional interval, and evidence of an increased information value of passive orders.

Table of Contents

Glossary

Term	Explanation
Aggressive order	Order that is executed immediately on arrival to the exchange
Passive order	Order that is not executed immediately on arrival to the exchange
Ask-price	Lowest price any market actor is willing to accept to sell the asset
Bid-price	Highest price any market actor is willing to accept to buy the asset
Mid-price	The average between the bid-price and the ask-price
Bid-ask spread	The difference between the ask-price and the bid-price
Displayed volume	Volume that is visible to all market actors
Undisplayed volume	Volume that is invisible to all market actors
Liquidity traders	Term used in modelling theory describing traders with a low degree of information and/or sophistication
Insiders	Term used in modelling theory describing traders with a high degree of information and/or sophistication
HFT	High frequency trader. Firms that trade at high speed using strategies such as cross-market arbitrage and market-making
Event	Any market action: buy, sell, passive order placement, cancellation, partial cancellation, hidden execution

1 Introduction

1.1 Context and motivation

More than half of the world's major stock exchanges rely on the limit order book mechanism (LOB) (Rosu 2009). It is, therefore, crucial to keep the understanding of limit order placements and their contribution to liquidity and price formation as current as possible.

However, this update process does not seem to occur as often as it ideally should. (Gould, Porter et al. 2013) point out that empirical studies make strong assertions regarding statistical regularities based on data from up to 20 years ago, of poor quality, describing only single stocks over short time periods. For this reason, LOB models may be based on regularities found in old, sometimes small and generally insufficient, samples, while traders' strategies and the rules governing stock exchanges change over time. As such, today's LOB activity may not be accurately reflected by those models.

To perform analysis that uncovers regularities such as those mentioned above, high quality, high-frequency data is needed. Such data is usually expensive, and difficult to come by for researchers. Furthermore, it is high in volume and usually formatted in binary, requiring significant pre-processing before being ready for use. However, the recent collaboration between the Swedish House of Finance and NASDAQ Stockholm provides such high-frequency data for free, and in an accessible format. This study capitalises on the opportunity to apply the part regarding time intervals between orders and trades of the framework developed by (Biais, Hillion et al. 1995) (BHC) to modern data, and thereby provide insights into the empirical characteristics of a modern LOB.

1.2 Research question and hypothesis

This study seeks to answer the following research question:

Are previous findings regarding time intervals between orders and trades valid given modern LOB data?

Here follows some expectations and reasoning regarding how the financial markets' evolution could show in this study's results. This study presents a hypothesis regarding time intervals conditional on passive orders. These intervals should be relatively shorter, compared to the unconditional mean time interval, than they were at the time of BHC.

Before large-scale trading automation, passive orders were believed to contain a lower information value than aggressive orders (e.g., (Glosten, Milgrom 1985), see section 2.3.1) since aggressive orders were viewed as the ones important for price discovery. This view has been questioned in modern literature, as the importance of passive orders for price discovery has been found to be increased (e.g., (Brogaard, Hendershott et al. 2019), see section 2.3.3), especially due to HFT. Meanwhile, reaction speed is taken as a proxy for information value assigned by the market to the specific event, in BHC. This leads to the conclusion that the relative reaction speed to passive order placement today should increase, in accordance with the higher information value that passive orders have been found to carry.

Of course, smaller interval magnitudes in general are also expected due to trading automation. With the exception of the above differences, the patterns found in BHC are should persist. This is because important market microstructure factors that affect those patterns, like tick size and priority rules, are similar today to what they were at the time of BHC. This in the sense that ticks still exist, although smaller, and that the priority rule in general is price-time, see section 2.2.2, as it was at the time of BHC. As such, this study can see no obvious incentive for market actors to significantly alter their trading behaviour with respect to the relevant patterns.

The rest of the study is organised as follows. Chapter 2 gives a background regarding the basics of LOBs, and a literature review of relevant papers in the context of the current research question. Chapter 3 presents a description of the data used in the thesis. Chapter 4 contains the methods used in the reconstruction of the order book and creation of the order classes. Chapter 5 presents the results, and chapter 6 and 7 discussion and conclusion respectively.

2 Background

This chapter begins by describing the structure of the LOB, to provide an overview of the fundamental subject of the study. A similar description can be found in Part 1 of the thesis. It continues by establishing relevant specifics of the NASDAQ Stockholm LOB, and ends with a literature review.

2.1 The LOB

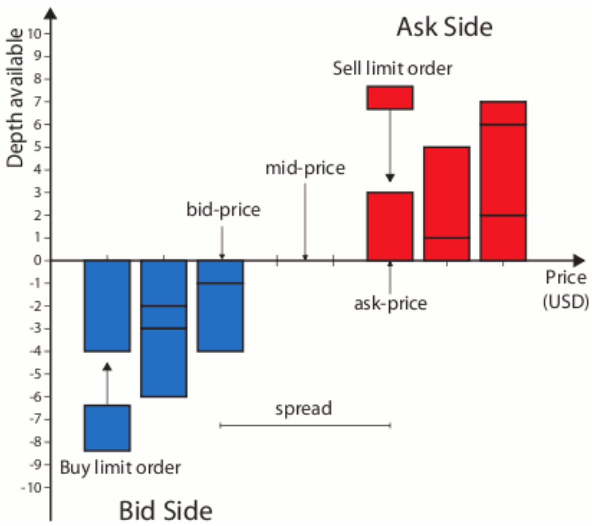
2.1.1 Fundamentals

The definition of a LOB is the following. “A record of unexecuted limit orders maintained by the specialist.” (NASDAQ 2020). The specialist in this case is a term coming from MiFID II, meaning the firm providing exchange services.

Figure I shows an example of a LOB with the essential parts included.

Figure I

Graphical representation of a limit order book.



Source: (Gould, Porter et al. 2013)

Starting from the left and working towards the right. The bid side provides demand for market actors wishing to sell the asset. The stacks are formed by passive limit orders, explained below, and they form the depth available at each price level. The gaps between the columns represent the tick size, explained in section 2.1.3. The bid-price is the current best offer to buy the asset, i.e., the highest price anyone in the market is currently willing to pay for the asset. The spread is the difference between the bid-price and ask-price, in this study referred to as the bid-ask spread. The mid-price is the average of the bid- and ask-prices.

The ask side is the opposite of the bid side, and so provides supply for market actors wishing to buy the asset. The functioning is otherwise similar to the bid side in all regards, except for being inversed. The ask-price is of course the current best offer to sell the asset, i.e., the lowest price anyone in the market is currently willing to accept to sell the asset.

Limit orders are usually passive, as shown in Figure I, in the sense that they do not execute immediately when coming into the exchange. However, they do not have to be, and this is commented on in section 2.2.1.

Aggressive orders, instead, are orders that execute immediately when coming into the exchange. The most common example of an aggressive order is a market order. Market orders do not have a specified price, as they execute at the best price available on arrival to the exchange.

Matching between buyers and sellers is an important aspect of market microstructure. When an aggressive order comes into the exchange, the matching against passive limit orders occurs according to a priority rule. The most common such rule by far is *price-time* (Gould, Porter et al. 2013). Matching by *price-time* means passive orders stack up as described in Figure I, those offering the best price first, and when an aggressive order comes in, ties at the best price are broken by selecting the passive order with the earliest submission time for execution against the aggressive order.

2.1.2 Undisplayed volume and dark pools

What is referenced when colloquially talking about the order book is usually displayed volume. That is, volume that all market actors can observe on their screens. Undisplayed volume does not show up on market actors' screens, but still exists in the LOB. Figure I does not make any attempt to show this undisplayed volume, but it plays a big role in most of today's LOBs. (O'Hara 2015) mentions that ~30% of the volume available in a typical LOB is undisplayed. "Hiding" volume as such is used for a range of reasons, including when market actors do not wish to disclose their full open interest.

A separate construction that can be used for similar, but even better, execution privacy is a dark pool. A dark pool functions essentially as a normal order book but is completely hidden, such that no one knows who wants to buy or sell units, how many, or at what price. It is kept apart from the main LOB, so that priority rules of orders in the main book does not affect it (Buti, Rindi et al. 2010). In contrast, undisplayed volume in the main book is still bound by the priority rules of the rest of the main book. See section 2.2 in Part 1 of the thesis for further details on dark pools.

2.1.3 Tick size

The tick size in the LOB is one of the fundamental rules governing actions in market microstructure; it is the smallest step by which one can increment an asset's price. Put another way, the tick size dictates how expensive it is for a trader to gain the priority associated with a higher (lower) price when placing a buy (sell) order (Parlour, Seppi 2008). In Figure I, the small gaps between the bid side columns and ask side columns represent the tick size, the pricing of orders can be chosen freely by market actors so long as the tick size is respected.

For example, consider a stock with a price of 90.0 and a valid price space of $90.0 \leq\ p \leq 91.0$. Without a specified tick size, an order with a price up to any decimal in p is valid, such as $p = 90.035$. However, if a tick size of 1.0 is specified, only prices in $p\ \in \{90.0. 91.0\}$ are acceptable. Consider now a tick size of 0.5 on the same stock. It is then possible to price the order as $p\ \in \{90.0. 90.5. 91.0\}$.

2.2 NASDAQ Stockholm

This section describes the specificities of NASDAQ Stockholm that are relevant to this study.[5] The section also motivates the reconstruction and reverse-engineering sections and enables the discussion of results.

2.2.1 Order types

The main order types at NASDAQ Stockholm are limit orders and market orders. They are crucial for the reconstruction of the order book. While there are a range of order attributes available to market actors, which change order behaviour, these are not identifiable through the ITCH-stream and therefore disregarded.

Limit orders are usually passive, but they do not have to be as said before. Aggressive limit orders are created by inputting a price that is higher/lower than the bid-ask spread. Aggressive limit orders at NASDAQ Stockholm are used whenever a market actor has a higher execution urgency than that which a market order can fulfil.

---

[5] Information in section 2.2 is taken from NASDAQ Nordic Market Model 2019:4 or NASDAQ Nordic Equity TotalView-ITCH v.3.03.4

Market orders at NASDAQ Stockholm are “immediate-or-cancel” (IOC), meaning that they either execute at the current best price, or are cancelled.

Consider an example of the IOC attribute. If one were to submit a limit buy order with volume higher than what is offered at the spread, that additional volume would be added to the book after execution of the first part of the order, leaving opportunity for further execution. If one were to use a market buy order with the same specifications, the additional volume would be cancelled.

2.2.2 Order execution

NASDAQ Stockholm follows *price-internal-display-time* priority, which is a variant of the *price-time* priority mentioned in section 2.1.1. Differences in priority rules can incentivise market actors in different ways. *Price-time* incentivises to place limit orders early, while alternatives such as for example *price-size* incentivises to place orders with higher volume. Having a *display* part incentivises market actors to provide displayed volume.

An example is provided of the dynamic that results from the *price-internal-display-time* priority that NASDAQ operates. Please note the *internal* part is disregarded, as it is complicated and does not add much value for this study’s purpose.

Assume orders A and B are the only ones in the book at current best bid, and that A arrived before B. A has displayed volume 20, undisplayed volume 30. B has displayed volume 10, undisplayed volume 40. Now consider a market sell order C with volume 100 that comes into the exchange. The execution flow is then the following:

1) 20 from A (*display-time* priority)
2) 10 from B (*display* priority)
3) 30 from A (matched after B due to undisplayed)
4) 40 from B (matched last due to undisplayed and incoming after A)

This shows the incentive to provide displayed volume and to place orders early to gain execution priority that is created by the *price-internal-display-time* priority rule.

2.2.3 Undisplayed volume and the Nordic@Mid dark pool

At NASDAQ Stockholm, it is possible to specify limit orders to be partially or fully undisplayed. A requirement for fully undisplayed orders in the main book is that they are large-in-scale (LIS). MiFID II describes LIS requirements, which depend on the traded instrument. The requirement mostly places a floor on the volume allowed to request in an order, so in general fully undisplayed orders should be larger than displayed orders.

NASDAQ Stockholm furthermore offers a dark pool solution, Nordic@Mid, for efficient matching of large orders with less risk of getting “picked off” or front-run by faster market actors. Nordic@Mid also allows for avoiding the LIS restrictions imposed by MiFID II. The Nordic@Mid dark pool is a typical dark pool as described in section 2.1.2, with a book that is separate from the main book and hidden from market actors. See section 2.3.4 in Part 1 of the thesis for more details on Nordic@Mid.

2.2.4 Tick size

The NASDAQ Stockholm tick sizing grid is a function of the average of daily volume over one year, and the price of the security. As such, the tick size differs between firms. The complete sizing grid is subject to regulatory oversight and comes from MiFID II, see appendix I.

2.3 Literature review

This section presents literature relevant to the current research question, and empirical findings in BHC similarly relevant to the current research question. For further literature regarding e.g., modelling of the LOB and a description of order splitting, please refer to section 2.4 in Part 1 of the thesis. Section 2.4 in Part 1 of the thesis also contains separate results from BHC.

2.3.1 Theory of time intervals between orders and trades

In the theoretical literature, at least three determinants of time intervals between orders and trades have been suggested.

An important contribution comes from (Easley, O'Hara 1992). The authors build a model describing the impact of time on security prices. In it, insiders rely heavily on market orders, as they have conviction in their knowledge, while liquidity traders are less willing to use market orders since they have less information. The intuition behind why time should impact security prices is, according to the authors, that a market actor will interpret the lack of action as an information event, just as they interpret an action as an information event. For example, a lack of action may suggest no new information has come into the market, therefore it would be reasonable to expect e.g., a market maker to be willing to quote a tighter spread following a certain period of inaction. This results in a range of predictions, among which one is especially important for this study. Considering that given a trade occurring, the probability of the trader being an insider is high, a high trading frequency should signal a high share of insiders in the market. This result will henceforth be referred to as the “asymmetric information theory” or “asymmetric information”.

A second contribution comes from (Harris 1994), who provides predictions regarding the dynamics of the market reacting to tick size changes in combination with time priority. The paper mainly suggests that a reduction in tick size implies decreased spreads, decreased quote sizes and increased trading volume. A separate, and important for this study, part of the results is that there is a first-mover advantage when the tick sizing grid is discrete and time priority is enforced. Such conditions will lead to competing traders having an incentive to place orders rapidly, when supplying liquidity is profitable. This result will henceforth be referred to as the “time priority theory” or “time priority”.

The third and final contribution comes jointly from (Pagano 1989) and (Admati, Pfleiderer 1988), they analyse the clustering of trades that is found in e.g., (Biais, Hillion et al. 1995) and recently in Part 1 of the thesis. The first paper examines the dynamics of having two similar markets, for example one OTC and one open, accessible for a certain type of asset. It predicts that thin markets will absorb large orders only in combination with an adverse price effect, and that traders for this reason will prefer to combine two similar markets into one.

However, it also argues that with differences in transaction costs, two markets may coexist if the cost to participate in one is too great for a certain type of trader. For example, assume a low liquidity market and a high liquidity market where the same asset is traded. The low liquidity market is free, and the high liquidity market has a high entrance fee. Then, small traders will trade in the low liquidity market while large traders trade in the high liquidity market, and no one wants to combine the two since market dynamics would change with this action. This explains the clustering incentives of traders in markets, and the existence of multiple markets for the same asset to some extent.

The second paper investigates why trading is clustered in particular periods of time during the day, why returns are more variable in those periods compared to others, and why more trading often implies higher return volatility. The important part of the results is that, according to the authors, as liquidity traders have discretion with regard to timing of trades, they choose to trade when the market is thick. This is because that would be when their trading has the smallest effect on price. This dynamic creates an incentive for clustering of liquidity traders in time within the same market and is in fact found to also incentivise insiders to trade in these same times. More insiders then incentivise liquidity traders to cluster even more, because insiders compete with each other and in the process provide better prices for the liquidity traders and so the clustering process continues. Naturally, if trades cluster in time then so should other events, so this result does not only concern trades but events in general. This result will henceforth be referred to as the "clustering theory" or "clustering".

2.3.2 Findings in BHC

This section describes the dataset, the results and the significance testing procedure used by BHC to arrive to the results regarding time intervals between orders and trades.

2.3.2.1 Dataset and relevant microstructure specifics

BHC used a dataset from the Paris Bourse collected from October 29 to November 26, 1991 including 19 trading days. The studied stocks were the members of the CAC40; however, the dataset only included the five top levels of the book. Please see appendix IV for summary statistics presented by BHC.

All information at the Paris Bourse was available in real-time to market actors, and the Paris Bourse was the only centralised trading venue for French stocks by law. Strict price and time priority were enforced at the Bourse. However, some trades were still executed in London, and as such outside of the priority rules. In (de Jong, Nijman et al. 1995), the authors find that ~10% of the total number of trades was executed abroad, representing 30%-50% of the total value in French stocks was executed abroad.

The CAC40 was split on two tick sizes. The tick size was decided by the price of the stock in question. For stocks in the FF (French francs) 100-500 range, the tick size was FF 0.1. For stocks above FF 500, the tick size was FF 1.

2.3.2.2 The time interval between orders and trades

The first issue investigated by BHC is whether the frequency of events reflects the clustering of events. The second issue is to find out whether the asymmetric information theory is true. Finally, the third issue is to find out whether the time priority theory is true. Key results are presented in short summary in this section. Please see appendix IV for full results tables.

Initially, BHC note that they find support for the first issue of frequency of events reflecting event clustering. The authors show that a time interval that follows one that is larger than its median tends to be large, and vice versa. This leads to the conclusion that if an event occurred rapidly previously, then it is likely that also the next event occurs rapidly. As such, these results document clustering in a way that is different from previous analyses, confirming that the frequency of events does reflect event clustering.

To investigate the second issue of asymmetric information, and the third issue of time priority, BHC examine 1) The average time interval between events, conditional on the bid-ask spread, 2) The average time interval between events, conditional on the previous event and 3) The average time interval between specific events.

The authors find support for the second issue of asymmetric information when reporting the results in 2). This is because time intervals following Large and Market orders are short on average. That is likely to be due to Large and Market orders being reactions to strong information signals, or themselves being strong information signals. Since insiders are likely to use Large and Market orders, and the conditional time intervals following are short, a high trading frequency could be indicative of a high share of insiders in the market.

Regarding the third issue of time priority, the authors find that in the results from 1), events occur relatively faster following a Large spread than a Small spread. This suggests that time priority plays a role in market actor considerations. Put differently, a fixed pricing grid combined with time priority entails a first mover advantage to supplying liquidity.

Results from 3) provide further illustration and support to 1) and 2).

2.3.2.3 Significance testing

The authors note that a test is needed to see whether results are significant. To this end, they conduct a χ^2-test. The test is motivated as follows.

The conditional average interarrival times computed to find the results in section 2.3.2.2 are the maximum likelihood estimates of Poisson processes with changing parameters. The states *S* are the different characterisations of the conditioning variable, for example in the case of the clustering analysis there are two states, a large spread state and a small spread state. The estimate of the parameter is the empirical average time interval until the next event in *s*.

The test statistic is the ratio of the unconstrained likelihood of the observations to the likelihood under the constraint that parameters are equal across states. The ratio asymptotically follows a χ^2-distribution, with $S - 1$ degrees of freedom, where S is the total number of states. See appendix IV for numerical results.

2.3.3 Recent developments

With the development of financial markets between the time of BHC and this study, a perspective that seems to be shifting in the literature is worth highlighting for the purpose of this study.

Traditionally, price discovery was modelled as occurring mainly through trading, e.g., (Glosten, Milgrom 1985). This was due to trades being viewed more or less analogously to insiders revealing their private information.

A relatively recent paper challenging this view is (Hautsch, Huang 2012). The authors find evidence, using VAR models, suggesting that limit orders do contribute to price discovery and move markets. This perspective finds further support in (Brogaard, Hendershott et al. 2019), where similar VAR models are used to show that in fact, a majority of the price discovery process on modern exchanges seems to occur through limit orders. This is due to HFTs being active in today's markets, and those firms having a high order-to-trade ratio. As such, the aggregate information value of their limit orders dominates the information value gained from their trading.

3 Dataset

This chapter presents an overview of the raw data used in this study and basic descriptive statistics. See chapter 3 in Part 1 of the thesis for more details.

3.1 Basic information

The data used to answer the research question is virtually identical to the data distributed through the NASDAQ Nordic Equity TotalView-ITCH (NETV) system, which disseminates messages in real-time during the trading day to all market actors subscribed to the service.

The format of the data is messages in rows in .csv files, and the total size of the dataset is ~3.6 GB. It contains exclusively passive orders and related executions from July 1 to July 31, 2019 for the OMX30 index, or a total of 24 trading days. The market opens at 09:00 and closes at 17:30 Stockholm time, thus the trading day is 8.5 hours long. The period was chosen due to being a good volatility match to the period in the BHC paper as measured by the VIX. The resolution is nanosecond-level. The data does not contain the original aggressive orders as they appear when they come into the exchange. See Table I for a sample of the raw data.

Table I

Sample of ITCH messages. The firmId column contains a unique code used to identify the company. The msgType defines which operation the message instructs. The orderId is the unique Id of each passive order sent to NASDAQ Stockholm. The side indicates whether the order is a Buy or Sell order. No side indication is given when an order is Executed or Deleted due to redundancy.

timeStamp	firmId	msgType	orderId	side	price	volume
20190701.070019.102723633	3966	A	3467192	S	187.85	24000
20190701.070019.105420265	3966	A	3467223	B	186.75	4000
20190701.070019.105447426	3966	D	3347607	_	0	0
20190701.070019.105454675	3966	E	14372	_	187	100

An explanation of Table I follows. An A-message signals adding a passive order to the book. A D-message signals deleting a passive order from the book. An E-message signals the execution of a passive order. As is apparent from the sample data, the messages do not show the complete book in each message, only how it changes.

The timeStamp column shows trading occurring at 07:00 due to messages being timestamped at GMT, Swedish summertime is GMT + 2. Table I tells us that 24 000 units were added to the ask side at 187.85, followed by 4 000 units at 186.75 added on the bid side. Then, the order with the orderId 3347607 was deleted. The E-message implies that an aggressive order came into the exchange and lifted 100 from the limit order with orderId 14372. Clearly, no order was shown as added in the stream before the E-message showing 100 lifted from the limit order 14372. This is the reason for the need to reverse-engineer aggressive orders.

3.2 Summary statistics

Table II presents summary statistics of daily message activity per firm. The exchange receives an average of ~39k add order messages per stock on any given day. With the trading day being 8.5 hours, any given stock receives ~77 new passive orders every minute. Appendix IV contains summary statistics presented in BHC for comparison.

Table II

Summary statistics of daily message activity. The table only concerns ITCH-messages, not actual orders. Statistics are averages per day per stock. For example, ~4.7k execution messages are disseminated on any given day for any given stock in the OMX30 index.

	Mean	Min.	1st Quart.	Median	3rd Quart.	Max.
Return (%)	- 0.1	- 10.1	- 0.8	- 0.1	0.6	5.2
Hi-Lo (%)	2.0	0.5	1.3	1.7	2.3	12.3
Add order (thousands)	39.4	8.3	23.7	35.0	48.8	168.6
Displayed trades (thousands)	4.7	0.7	2.8	4.0	5.4	37.3
Undisplayed trades	150.8	0.0	45.0	100.0	183.0	1 723
Displayed shares traded (thousands)	1 580.4	53.3	477.6	951.9	2 038.2	22 407.2
Undisplayed shares traded (thousands)	86.1	0.0	15.0	42.3	98.9	1 582.1
Displayed value traded (SEKm)	212.0	33.9	111.0	176.9	271.3	1 874.7
Undisplayed value traded (SEKm)	11.9	0.0	30.3	7.7	14.7	148.5

As expected, considering the LIS rules imposed on undisplayed orders, the average volume per execution of undisplayed volume of ~0.57k shares is higher than the average volume per execution of displayed volume of ~0.34k shares. On average, executions of undisplayed volume make up ~3% of total trades, but ~5% of total volume. Finally, the share of cancelled messages on the Stockholm stock exchange is high. Current data shows ~88% of messages ending in some other way than trades.

4 Methodology

To perform the analysis required to answer the research question, the order book has to be reconstructed and aggressive orders have to be reverse-engineered. Following this, events are classified according to their aggressiveness. This chapter details the procedures used for these operations, see chapter 4 in Part 1 of the thesis for further details.

4.1 Reconstruction of the order book

NASDAQ documentation[6] explains the different message types disseminated in the ITCH-stream and how they are supposed to be handled. Consequently, one can easily proceed to use the messages to reconstruct the book. All levels of the order book are reconstructed for the purpose of this study. (Huang, Polak 2011) describe the reconstruction process in detail. See section 4.1 in Part 1 of the thesis for an overview of the program loop that is used for reconstruction.

4.2 Reverse-engineering of aggressive orders

The main messages used for this task are E- and P-messages. E-messages represent the execution of displayed volume, and P-messages represent the execution of undisplayed volume or dark pool volume. For the "linking" of separate messages into a complete order, this study assumes that timestamps come from matching engine event time. That assumption has a weakness which makes it impossible to consider the period immediately following the uncross. Appendix III presents a discussion of the assumption's validity. See section 4.2 in Part 1 of the thesis for details on the reverse-engineering.

4.3 Order classification

This study generally follows the naming conventions and class definitions established in BHC, with two differences. First, the Application class is approximated by the Hidden class but the two cannot be compared. Second, this study has a Cancel-replace class that did not exist in BHC.

Aggressive orders are classified according to their aggressiveness, while passive ones are classified according to their position in the book. There are eight classes on the ask side, and eight classes on the bid side. The Hidden class is special because it includes events from both sides. It is the final 17th order class.

---

[6] Nordic Equity TotalView ITCH version 3.03.4

The most aggressive class is the Large order class. A limit sell/buy order with a price lower/higher than the bid/ask-price, and volume higher than the volume offered at the best bid/ask-price is a Large order.

The second most aggressive class is the Market order class. A limit, or market, sell/buy order with a price equal to the best bid/ask-price and volume higher than the volume offered at the best bid/ask-price is a Market order.

Actual (IOC) market orders cannot be differentiated from limit orders through the ITCH messages. They will therefore by definition be included partly in this class and partly in the Small order class below. Market actors do not observe the difference between IOC market orders and aggressive limit orders through the ITCH-stream.

The third most aggressive class is the Small order class. A limit, or market, sell/buy order with price equal to the best bid/ask-price and volume equal to or below the volume offered at the best bid/ask-price is a Small order.

The fourth and final aggressive class is the Hidden order class. Hidden orders are identified when an exclusively hidden execution occurs. This means streaks of P-messages or single P-messages only are used to identify this order class.

P-messages are, however, used by NASDAQ Stockholm to signal both matches of undisplayed volume in the main book and matches from the Nordic@Mid dark pool. Usually, it is possible to distinguish between an undisplayed execution in the main book and a Nordic@Mid execution by a flag in the ITCH messages. The current dataset however does not include that flag.

The Hidden order class, therefore, combines the execution of undisplayed volume in the main book and volume on Nordic@Mid. For the reasons above it is also impossible to identify the side of the execution, therefore, the Hidden order class does not differentiate between these. The Hidden class is different from the Application class in BHC due to structural changes, comparability is therefore limited in this regard.

The passive order classes are differentiated depending on whether the orders were placed in the bid-ask spread, at the bid-ask spread or above/below the bid-ask spread.

Following passive order classes are the Cancel-replace and Cancel classes. A Cancel-replace is a cancellation of an order combined with the immediate submission of a new order in its place. The new order may contain updated volume and updated price, or either of the two separately. The Cancel-replace class did not exist in BHC.

A Cancel is a deletion of a posted order. The Cancel class includes partial cancellations, where only part of a posted order is deleted. For example, consider a bid order for 100 units in the book. A Cancel event is recorded if the relevant market actor either deletes the order fully, or deletes only e.g., 50 units from it, leaving 50 in the book to be executed.

5 Results

This chapter presents findings regarding the time interval metrics described in section 2.3.2.2.

The chapter first handles results related to the clustering theory, and then presents findings related to the asymmetric information and time priority theories.

5.1 The clustering theory

Table III presents results aiming to show whether the frequency of events reflects the clustering of events. The unconditional interval time to the next event is ~0.45 seconds.

Put differently, if we were to randomly pick a firm in the sample and consider its events, without any further information provided, we would expect to see about two events every second. An event could be any market action, e.g., a trade, an added passive order, a cancellation or any other type.

Table III

Expected time interval until the next event, conditional on the previous time interval. Conditioning variables are calculated per firm. A Large previous interval is an interval larger than the firm's time series median interval. A Small previous interval is an interval smaller than the firm's time series median interval.

Conditioning variable	Expected time (s)	% to Unconditional	χ^2-test	BHC expected time (s)
Unconditional interval	0.45			98.0
Large previous time interval	0.65	+44.4%	$\chi^2 > 10,000$	128.5
Small previous time interval	0.20	-55.6 %	*p-value = 0.00, d.f = 1*	68.3

Table III shows that given a large previous time interval, defined as being larger than the time series median for the particular stock, the expected time until the next event is larger than the unconditional interval. Given a small previous time interval, defined in a corresponding way, the opposite is true. These results in effect provide an alternative way of documenting the clustering found in BHC, and Part 1 of the thesis.

To see why this is the case, consider a firm where e.g., a passive order was just placed, and we have observed a time interval larger than the firm's time series median before that order came in. We will then expect the next event to follow within 0.65 seconds as per Table III. If we instead observe a time interval smaller than the firm's time series median, we will on average see an event occurring within 0.20 seconds. This creates the "clustering" effect of events following in a tight succession.

So, the frequency of events reflects the clustering of events in this study, similarly to in BHC.

Furthermore, the results are strongly significant, as seen in the likelihood ratio test results. The test is conducted with a null of identical mean interval time over conditional classes, where the identical mean is the unconditional interval, against non-identical mean intervals over conditional classes, where the different means are the conditional ones. See appendix V for a detailed description of the process followed in the testing procedure.

5.2 The time priority and asymmetric information theories

Table IV presents results related to the theories of time priority and asymmetric information by providing results on time intervals conditional on prior spread, and prior event respectively.

Starting with spread results, to provide some intuition, consider a firm where we observe a Large spread in the firm's book, defined as being larger than the 75th percentile of spreads for that firm. Following that observation, we expect a time interval of 0.34 seconds to the next event. As such, the effect of a Large spread is similar to the Small time interval mentioned in section 5.1 above. Large spreads have a clustering effect on events, or equivalently, events are likely to occur faster following the observation of a Large spread than unconditionally.

The above supports the theory of time priority incentives, as a Large spread does indeed seem to present an opportunity for profitable liquidity provision, which is why market actors act faster than otherwise to gain time priority. The difference to the unconditional interval is less pronounced when it comes to other spread categories. This is also expected from theory, which states that a spread close to normal will not provide the same opportunity for profit as a Large one, for obvious reasons. Results are strongly significant according to the likelihood ratio test.

Regarding results conditional on events in Table IV, it is notable that Large and Market buys/sells see on average 0.00 second intervals until the next event, meaning market actors react rapidly to such events. Other event classes see conditional interval lengths of varying degrees, but they all have in common that they see longer intervals to the next event than Large and Market buys/sells.

Bids/Asks below/above quotes and Cancel-replace events are special in that they are the only ones to see a conditional interval of ~0.5 seconds, which is larger than the unconditional interval of ~0.45 seconds. Equivalently, market actors respond with less urgency to such events. In general, a majority of the events see conditional time intervals that are shorter than the unconditional interval. Results are strongly significant according to the likelihood ratio test.

Table IV

Expected time interval until the next order or trade, conditional on the spread or previous event. Spreads are divided into quartiles to get large, medium-large, medium-small and small cutoffs. Conditioning variables are calculated per stock. Classes that did not exist in BHC, or are not comparable for other purposes, are n.a.

Conditioning variable	Expected time (s)	% to Unconditional	χ^2-test	BHC expected time (s)
Unconditional interval	0.45			98.0
Large spread	0.34	-24.4%		77.4
Medium-large spread	0.45	~0.0%	$\chi^2 > 10,000$	99.7
Medium-small spread	0.47	4.4%	*p-value = 0.00, d.f = 3*	107.1
Small spread	0.45	~0.0%		107.6
Large buy	0.00	-		72.8
Market buy	0.00	-		80.5
Small buy	0.15	-66.7%	$\chi^2 > 10,000$	107.6
Bid in quotes	0.04	-91.1%	*p-value = 0.00, d.f = 16*	93.0
Bid at quotes	0.32	-28.9%		92.8
Bid below quotes	0.48	+6.7%		98.6
Cancel-replace bid	0.49	+8.9%		*n.a*
Cancel bid	0.29	-35.6%		82.5
Large sell	0.00	-		70.5
Market sell	0.00	-		68.5
Small sell	0.09	-80.0%		105.6
Ask in quotes	0.03	-93.3%		104.0
Ask at quotes	0.31	-31.1%		114.3
Ask above quotes	0.50	+11.1%		100.9
Cancel-replace ask	0.50	+11.1%		*n.a*
Cancel ask	0.29	-35.6%		73.6
Hidden	0.11	-75.6%		*n.a*

In Table V, the time interval referred in the “Expected time” column is the mean of all time intervals between the events in the “Event sequence” column. To provide some intuition about what a sequence could signify, consider the following example. Consider the Large buy (sell) – Large buy (sell) case. This is the mean of the intervals between a Large buy followed by a Large buy, or a Large sell followed by a Large sell. It does not consider any other combination, such as a Large buy followed by a Large sell. The mean interval in such a sequence is 0.00 seconds, so, if market actors place a Large buy/sell after having observed a Large buy/sell previously, they are likely to do so rapidly. Other sequences with the same short interval are Market buy (sell) – Market buy (sell) and Market buy (sell) – Cancel bid (ask).

In light of this example, it becomes clear that market actors express less urgency on average for e.g., the Small buy (sell) – Small buy (sell) sequence, as the mean interval between those events is 0.10 seconds.

Table V

Expected time interval between selected event sequences. Sequences are calculated per stock.

Event sequence	Expected time (s)	% to Unconditional	BHC expected time (s)
Unconditional interval	0.45		98.0
Large buy (sell) – Large buy (sell)	0.00	-	67.1
Small buy (sell) – Small buy (sell)	0.10	-77.8%	95.2
Market buy (sell) – Market sell (buy)	0.00	-	73.1
New bid (ask) in spread – New bid (ask) in spread	0.02	-95.6%	86.0
Large spread – New bid (ask) in spread	0.23	-48.9%	74.6
Cancel bid (ask) – Cancel bid (ask)	0.20	-55.6%	55.5
Market buy (sell) – Cancel bid (ask)	0.00	-	66.3

It is also worth noting that Table V shows short time intervals, shorter than the unconditional interval, for all sequences. This is also the case in BHC, although those numbers are less different to the unconditional interval relatively speaking. The difference to the unconditional interval is commented on in section 6.2. The likelihood ratio test performed previously is not applicable in this case.

6 Discussion

This chapter discusses the results presented in the previous chapter, and how they relate to the original BHC findings. The chapter also presents some limitations to the results.

6.1 The clustering theory

First of all, it should be noted that current time intervals are smaller by, in most cases, several orders of magnitude when comparing to BHC results. The clearest example of this is the unconditional interval today of 0.45 seconds, versus 98.0 seconds in BHC. This is expected from the technological evolution as mentioned in section 1.2.

Section 5.1 presents an explanation as to why the time intervals suggest clustering of events. As said, the frequency of events in this study reflects the clustering found in BHC and Part 1 of the thesis. The results are highly significant, even more so today than they were at the time of BHC. As the number of events in this dataset vastly surpasses the number of events of BHC, the strong significance is quite natural, given of course that the differences in means are persistent. This is also expected considering the findings in Part 1 of the thesis, where clustering is apparent.

In sum, there is evidence of frequency of events reflecting clustering of events in today's market.

6.2 The time priority and asymmetric information theories

BHC results indicating asymmetric information and time priority hold up well too. For example, Large and Market orders get fast responses, whereas other less significant actions are less likely to be followed as rapidly by another event.

In general, the relative differences to the unconditional interval appear to be accentuated in today's data, when comparing with relative differences in BHC results. Large and Market orders in this study see almost instant reactions (0.00 seconds), nearly a 100% difference to the unconditional interval. The difference for the same event to the unconditional interval in BHC was only about -26%. The trend in this regard for more extreme differences to the unconditional interval is clear.

The results are consistent with the asymmetric information theory, as large volume trades seem to follow rapidly on each other, and since insiders are likely to conduct such trades, a high trading frequency does make it more likely that insiders are in the market.

Regarding time intervals conditional on spreads, events follow Large spreads faster (0.34 seconds) than they do more normal spread conditions (~0.45 seconds). This corresponds to results in BHC where Large spreads have a conditional interval of 74 seconds, with the unconditional interval being 98 seconds. This finding is related to time priority considerations by market actors. A Large spread should entice market actors to react quickly to the supposed opportunity for profit, according to the theory, and they seem to do so. The profit opportunity stems from the spread being "too large" for a short while, allowing market actors to gain time priority while keeping profit margin.

Since NASDAQ today has similar priority rules to the Paris Bourse at the time of BHC, it is not surprising that both results point in the same direction.

When it comes to time intervals conditional on events, one observation is worth mentioning. BHC presented results where seven event classes had a conditional interval longer than the unconditional interval. Today, only three event classes have a conditional interval longer than the unconditional interval. This is noteworthy, as it could suggest today's data has outliers in the form of long intervals affecting mean calculations. The distribution of the time intervals is unfortunately not provided by BHC, making an explicit comparison impossible. A glance at the summary statistics, section 3.2 for this study's numbers and appendix IV for BHC's, could however point towards more volatility in today's dataset. Consider for example the Max/Min return in this study of 5.2% / -10.1%, versus in BHC 2.9% / 0.7%.

Moving on to the conditional event sequences. In general, the time intervals follow a similar pattern as in BHC, albeit at a different scale. For example, all conditional intervals are below the unconditional interval, as they were in BHC. Large and Market orders on the same side following each other almost instantly (0.00 seconds) as mentioned in section 5.2, while Small orders on the same side follow within 0.10 seconds. According to a theory presented by BHC, "shorter" time intervals are indicative of traders imitating each other, or successively reacting to the same information. If instead "longer" time intervals are observed, they could suggest order splitting behaviour. No guidance is provided as to what should be considered short or long.

Therefore, it would be fitting to interpret this study's results in the way that successions of Small orders are to a greater extent created by order splitting algorithms. However, it should be noted that this hypothesis cannot be verified, as data is lacking.

6.3 Initial hypothesis

To verify whether the hypothesis that reactions to passive order placement are relatively faster today when comparing with the unconditional interval, the intervals to consider are the ones following Bids in/at/below the quotes and Asks in/at/above the quotes. If the hypothesis is false, we expect to see the relative differences between time intervals conditional on passive orders, and the unconditional interval to be virtually unchanged, using BHC as benchmark. All BHC relative differences are available in appendix IV.

For Bids in the quotes, the relative difference to the unconditional interval in this study is ~-90%. In BHC it was ~-5%. For Bids at the quotes this study's result is ~-30%. In BHC it was ~-5% again. For Bids below the quotes, this study's result is ~+7%, while BHC had ~+1%. Results are similar for the Ask side.

As such, the hypothesis is confirmed for passive orders in and at the quotes, which yield a violently faster market response today than what was observed by BHC. However, for orders above the quotes, the same trend does not appear. This since the difference in relative relationship is not that great, only ~+7% for this study versus ~+1% for BHC.

So, if the hypothesis is true, and passive orders contain a higher information value today than at the time of BHC, why does this not apply to passive orders above/below the quotes? One reason could be that market actors have not awarded passive orders a higher information value across the board, instead focusing on the most important ones, which are the ones in and at the quotes.

6.4 Limitations

This section discusses potential limitations of this study in terms of answering the research question, and general results applicability. The identified limitations are the restricted sample, differences in market characteristics and order splitting.

When it comes to the restricted sample issue, this study uses the OMX30 index to provide relatively comparable data to the CAC40. While, obviously, the indexes are not the same and as such do present some fundamental differences, both samples of companies consist of their respective national indices. Therefore, they should include a similar class of companies and general trading conditions. As such comparability with BHC is quite likely acceptable. However, in general, extrapolating these results to stocks that are not in national indices, or not so liquid ones, might be problematic.

When it comes to differing market characteristics, the main issue is that trading of OMX30 companies occurs outside of NASDAQ Stockholm via different vehicles. Off-exchange trading occurs, for example, through depositary receipts and dual listings.

As said previously, NASDAQ Stockholm is estimated to have ~70% of daily volume in OMX30 companies. The practice of trading French stocks outside of the Paris Bourse was illegal at the time of BHC, however, it still occurred. The current market fragmentation might or might not harm comparability to BHC, as the data is quite unclear on the exact extent of off-exchange trading at the time. In any case, this is negative for general applicability, since the “full picture” of the OMX30 cannot be seen in the order book data gathered solely from NASDAQ Stockholm. Most similar studies suffer from the same issue.

Lastly, order splitting. The practice poses problems as there is a risk that some of the observed conditional effects stem from a single larger order being split into smaller pieces, rather than different market actors responding to one another. The order book method in this study reconstructs only realised orders, as accounting for order splitting requires proprietary data. A consequence of this is that disentanglement of order splitting effects and market actor response effects is difficult, as is noted with regards to Small orders for example. This harms general applicability and might do so for comparability with BHC, as the extent of this practice is unclear today and was at the time as well.

7 Conclusion

The purpose of this study has been to understand whether time interval patterns established in BHC regarding the order book and order flow accurately reflect modern markets. The motivation is to provide updated empirical information to the research field following the remarks on the need for research on new data made in (Gould, Porter et al. 2013). The research question is the following:

Are previous findings regarding time intervals between orders and trades valid given modern LOB data?

In general, the question can be answered such that the studied patterns seem to persist in modern markets, albeit at a significantly smaller scale and with larger relative differences to the unconditional interval. The increasing importance put by market actors on passive order placement found in recent studies also shows up in time interval data.

The results in this study are meaningful to enable more accurate and up-to-date modelling of LOBs, as well as to enhance the understanding of time interval patterns on modern exchanges in general. A better such understanding may be helpful when, for example, choosing how to best execute orders and evolve exchanges.

For future research, this study suggests to further examine the market impact and importance for price discovery of passive orders, while differentiating between the levels on which the orders are placed.

References

ADMATI, A.R. and PFLEIDERER, P., 1988. A Theory of Intraday Patterns: Volume and Price Variability. *The Review of financial studies,* **1**(1), pp. 3-40.

BIAIS, B., HILLION, P. and SPATT, C., 1995. An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse. *Journal of Finance,* **50**(5), pp. 1655,1689.

BROGAARD, J., HENDERSHOTT, T. and RIORDAN, R., 2019. Price Discovery without Trading: Evidence from Limit Orders. *The Journal of finance (New York),* **74**(4), pp. 1621-1658.

BUTI, S., RINDI, B. and WERNER, I., 2010. Diving into Dark Pools. *IDEAS Working Paper Series from RePEc,* .

DE JONG, F., NIJMAN, T. and RÖELL, A., 1995. A comparison of the cost of trading French shares on the Paris Bourse and on SEAQ International. *European Economic Review,* **39**(7), pp. 1277-1301.

EASLEY, D. and O'HARA, M., 1992. Time and the Process of Security Price Adjustment. *The Journal of finance (New York),* **47**(2), pp. 577-605.

GLOSTEN, L.R. and MILGROM, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics,* **14**(1), pp. 71-100.

GOULD, M.D., PORTER, M.A., WILLIAMS, S., MCDONALD, M., FENN, D.J. and HOWISON, S.D., 2013. Limit order books. *Quantitative Finance,* **13**(11), pp. 1709-1742.

HARRIS, L.E., 1994. Minimum Price Variations, Discrete Bid--Ask Spreads, and Quotation Sizes. *The Review of financial studies,* **7**(1), pp. 149-178.

HAUTSCH, N. and HUANG, R., 2012. The market impact of a limit order. *Journal of economic dynamics & control,* **36**(4), pp. 501-522.

HUANG, R. and POLAK, T., 2011. LOBSTER: Limit Order Book Reconstruction System. *SSRN Electronic Journal,* .

NASDAQ, 2020-last update, NASDAQ Glossary. Available: https://www.nasdaq.com/glossary/l/limit-order-book [16 June, 2020].

O'HARA, M., 2015. High frequency market microstructure. *Journal of Financial Economics,* **116**(2), pp. 257-270.

PAGANO, M., 1989. Trading Volume and Asset Liquidity. *The Quarterly journal of economics,* **104**(2), pp. 255-274.

PARLOUR, C. and SEPPI, D., 2008. Limit Order Markets. A survey. *Handbook of Financial Intermediation and Banking,* .

ROSU, I., 2009. A Dynamic Model of the Limit Order Book. *The Review of Financial Studies,* **22**(11), pp. 4601.

Appendices

Appendix I

Liquidity band						
Price ranges	0 ≤ ADNT < 10	10≤ ADNT < 80	80 ≤ ADNT < 600	600 ≤ ADNT < 2000	2000 ≤ ADNT < 9000	9000 ≤ ADNT
0 ≤ price < 0.1	0.0005	0.0002	0.0001	0.0001	0.0001	0.0001
0.1 ≤ price < 0.2	0.001	0.0005	0.0002	0.0001	0.0001	0.0001
0.2 ≤ price < 0.5	0.002	0.001	0.0005	0.0002	0.0001	0.0001
0.5 ≤ price < 1	0.005	0.002	0.001	0.0005	0.0002	0.0001
1 ≤ price < 2	0.01	0.005	0.002	0.001	0.0005	0.0002
2 ≤ price < 5	0.02	0.01	0.005	0.002	0.001	0.0005
5 ≤ price < 10	0.05	0.02	0.01	0.005	0.002	0.001
10 ≤ price < 20	0.1	0.05	0.02	0.01	0.005	0.002
20 ≤ price < 50	0.2	0.1	0.05	0.02	0.01	0.005
50 ≤ price < 100	0.5	0.2	0.1	0.05	0.02	0.01
100 ≤ price < 200	1	0.5	0.2	0.1	0.05	0.02
200 ≤ price < 500	2	1	0.5	0.2	0.1	0.05
500 ≤ price < 1 000	5	2	1	0.5	0.2	0.1
1 000 ≤ price < 2 000	10	5	2	1	0.5	0.2
2 000 ≤ price < 5 000	20	10	5	2	1	0.5
5 000 ≤ price < 10 000	50	20	10	5	2	1
10 000 ≤ price < 20 000	100	50	20	10	5	2
20 000 ≤ price < 50 000	200	100	50	20	10	5
50 000 ≤ price	500	200	100	50	20	10

Tick sizing grid according to MiFID II. ADNT is shorthand for Average Daily Number of Transactions and is calculated and published by ESMA. ESMA is the European regulator responsible for financial markets oversight.

Appendix II

For completeness and pedagogy, the initial attempt at order book reconstruction and the following optimisation are both described.

To track the message impact, one needs to store the unique orderId assigned to each message by NASDAQ. Initially, the book was constructed through nested hashmaps with the structure: {side: {price: {orderId: volume}}}. For example, given two limit orders registered on the bid side at 102.2, the book object looked as follows:

```
{
    bid: {
        102.2: {
            12345: 100,
            12346: 200
            }
    }
}
```

The book object was for each message stored in an outer hashmap with timestamps as keys. The structure proved impossible to use as the RAM required to store the outer hashmap proved too high.

The reconstruction software was then optimised. The book object is still a nested hashmap. Consider the same hypothetical situation as before. The structure is now as follows:

```
{
    bid: {
        102.2: 300
    }
}
```

An auxiliary hashmap on the format {orderId*:* volume} was added to keep track of the orderIds. Again, for the same hypothetical situation:

```
{
    12345: 100,
    12346: 200
}
```

So, when there is, for example, an incoming execution message relating to order *12345*, the software looks up the corresponding volume in the auxiliary hashmap. It then proceeds to perform the necessary operations on the book object, given the information. This entails a drastically reduced book object.

Appendix III

The reverse-engineering was completed in cooperation with NASDAQ staff and following suggestions from one of the founders of LOBSTER[7].

As noted in section 4.2, timestamps are assumed to come from matching engine event time. That means that the matching engine at NASDAQ timestamps messages when the match, or message registration, occurs.

The alternative hypothesis would be for ITCH messages to be timestamped somewhere else. For the alternative hypothesis to be true, messages would there have to be stamped in bulk and sent out in intervals. Thus, it would entail multiple different types of messages being timestamped simultaneously and sent out in packets.

Because multiple order types would sometimes share timestamps if they are not timestamped according to matching engine event time, the validity of the assumption is easy to check empirically through, for example, add order (A) messages outside of the seconds immediately following cross periods. Those seconds are special because the engine places all uncrossed orders on the book at the same time and timestamps them within one ns of each other. This is also the reason for why the reconstruction method cannot be used in these intervals.

Therefore, if there are any streaks of exclusively A-messages within one ns, then timestamps do not come from matching engine event time.

It turns out that in the current dataset, no A-messages have timestamps within one ns of each other outside of cross periods. Thus, we may assume that timestamps come from matching engine event time.

---

[7] The Limit Order Book Reconstruction System, https://lobsterdata.com/

This assumption leads us to the conclusion that the only way for messages to have timestamps within one ns from each other is for them to originate from the same (aggressive) order. This logic allows for the aggressive order reconstruction.

Appendix IV

Original summary statistics in BHC paper.

Table VI

For the 19 trading days in the period between October 29 and November 26, 1991, for each stock included in the CAC 40 index at that time, BHC compute the daily mean return, difference between highest and lowest price divided by the lowest price (hi-lo), number of trades, number of orders (that were not immediately executed), number of "applications", trading volume in shares, value of shares traded (in FFm), value of "applications", and the number of times a hidden order is hit. The table reports summary statistics about the cross sectional distribution of these 9 daily averages across the 40 stocks.

	Mean	Min.	1st Quart.	Median	3rd Quart.	Max.
Return (%)	-0.19	-0.77	-0.4	-0.17	-0.03	0.4
Hi-Lo (%)	1.8	0.7	1.4	1.8	2.2	2.9
Number of trades	148.6	43.7	76.8	113.7	196.4	448
Number of orders	160.6	76	106.8	140.5	198.8	429.8
Number of "applications"	6.8	1.6	3.5	6	10	15.5
Trading volume in thousands of shares	55.6	1.3	16	29.1	66.7	310.4
Value of shares traded (FFm)	29.7	2.9	8.8	21	40.5	145.9
Value of "applications" (FFm)	4.7	0.4	1.4	2.7	5.5	18.4
Number of times a hidden order is hit	18.3	2	8.2	18.2	25.1	54.6

"Applications" is a special class of hidden order executions. No similar event exists today. As such, summary statistics of this study instead presents statistics on undisplayed liquidity, as this is closest to "applications" in today's context.

Original results on time intervals in BHC paper.

Table VII

Expected time interval until the next order or trade conditional on different variables.

Conditioning variable	Expected time (s)	% to Unconditional	χ^2-test
Unconditional interval	98.0		
Large spread	77.4	-21.0%	χ^2 = 4,557
Large medium spread	99.7	+1.7%	p = 0.00, d.f = 3
Small medium spread	107.1	+9.3%	
Small spread	107.6	+9.8%	
Large previous time interval	128.5	+31.1%	χ^2 = 24,326
Small previous time interval	68.3	-30.3%	p = 0.00, d.f = 1
Large buy	72.8	-25.7%	
Market buy	80.5	-17.9%	χ^2 = 31,890
Small buy	107.6	+9.8%	p = 0.00, d.f = 14
New bid within	93.0	-5.1%	
New bid at	92.8	-5.3%	
New bid below	98.6	+0.6%	
Cancel bid	82.5	-15.8%	
Large sell	70.5	-28.1%	
Market sell	68.5	-30.1%	
Small sell	105.6	+7.8%	
New ask within	104.0	+6.1%	
New ask at	114.3	+16.6%	
New ask below	100.9	+3.0%	
Cancel ask	73.6	-24.9%	
Application	112.9	+15.2%	

Table VIII

Expected time interval between two specific events.

Sequence of events	Expected time (s)	% to Unconditional
Unconditional interval	98.0	
Large buy (sell), large buy (sell)	67.1	-31.5%
Small buy (sell), small buy (sell)	95.2	-2.9%
Market sell (buy), market buy (sell)	73.1	-25.4%
New ask (bid) within, new ask (bid) within	86.0	-12.2%
Large spread, new ask (bid) within	74.6	-23.9%
Cancel bid (ask), cancel bid (ask)	55.5	-43.4%
Market buy (sell), cancel bid (ask)	66.3	-32.3%

Appendix V

The steps followed for the likelihood ratio test are described hereunder, with the two-state test for large and small previous time intervals used as an example. Steps are followed for each firm, all statistics below are firm-specific:

1. Gather the:
    a. Time intervals
    b. Threshold (median) for large and small time intervals
    c. Conditional means for each state
    d. Unconditional mean time interval
2. The time intervals between events are assumed to follow an exponential distribution. Calculate the unconditional log-likelihood of the observations using the exponential log-likelihood function, with the unconditional interval as the maximum likelihood estimate
3. Split the time intervals into conditional on large and small previous intervals respectively. Calculate the log-likelihood of each state using the corresponding conditional mean as the maximum likelihood estimate
4. Sum the conditional log-likelihoods to get a total conditional log-likelihood. Subtract the unconditional log-likelihood, multiply total by 2 following the formula for the chi-squared statistic

5. Find the p-value of the chi-squared statistic using S – 1 degrees of freedom, S being the total number of states. In this case, the total number of states are 2 and thus the degrees of freedom are 1

END OF PART 2 AND THESIS