5-14-2017

Trading Dynamics in a Fragmented Market

Krzysztof Herman
Syracuse University

Abstract

In this dissertation, we study the effect of recent regulatory and technological changes on trading dynamics. Advances in communication and computing technologies have made millisecond latencies as the new trading standard and have resulted in a new era of automated trading. The introduction of Reg-NMS (Regulation National Market System, implemented in 2007) has set strict rules for the access and removal of liquidity from the fragmented US equity market, de facto linking the trading activities across trading venues. These transformations have not only changed how equity markets function but also how market participants interact with the market and among themselves.

We begin this study by examining how the introduction of Reg-NMS has affected the trading strategies of fast, impatient traders. The implementation of Rule 611, which extends price priority across all the trading venues in the National Market System, forces them to monitor all trading venues in order to correctly asses the placement of their orders. We find evidence that because of their impatient nature, these traders react to all events that negatively affect the position of their orders, regardless of the venue of origin. This behavior results in an order flow that is made up of a high volume of very short-lived limit orders, which is consistent with a previously studied, but not yet fully explained, phenomenon of fleeting liquidity.

We then investigate whether fast, impatient traders are able to leverage their speed advantage to turn market fragmentation in their favor. We find evidence that their ability to anticipate the

order flow of the other market participants, allows them to engage in a trading strategy that relies on the simultaneous submission of multiple orders across exchanges. Such strategy, called *Overbooking*, aims at executing only one of these orders rather than all of them and uses the availability of multiple exchanges to increase the probability of execution while limiting the risk of over execution thanks to their speed advantage.

Overall, our findings show that a sub group of traders was not only able to adapt to a changing trading environment but actually take advantage of it. The *Overbooking* trading strategy is effective at increasing the probability of execution while also decreasing execution time and it is particularly effective for stocks with a high degree of competition for superior queue placement. This suggests that the ability to effectively trade on multiple venues simultaneously allows fast, impatient traders to avoid engaging in a costly algorithmic battle for a favorable queue placement, or submitting very aggressively priced limit orders, to attain quick execution.

Moreover, our findings show that the actions of these traders, driven by their fast, and impatient nature and constrained by the complex rules that regulate liquidity access and provision on the National Market System, result in the linking of order flow dynamics across trading venues. We find that the cancellation of an order can be determined by changes that have occurred elsewhere in the market and that to model correctly order flow dynamics it is necessary to include in the analysis the changes that occur on all trading venues. We also find that fast, impatient traders, actively monitor the state of all trading venues after order submission, and that they benchmark the present state of the market to the state at submission.

Trading Dynamics in a Fragmented Market

By

Krzysztof Herman

B.Sc. Mathematics, Universita di Bologna, Italy, 2009

Dissertation

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration

Syracuse University

June 2017

Acknowledgements

I dedicate this dissertation to my mother, Hanna Herman, my father, Michael Herman, my amazing brother, Rafael Herman, my 94 years old grandma, Krystyna Ozieblo, and my now gone grandpa, Leszek Ozieblo. Mom, thank you for all the encouragement over the years and for always being there when I needed help and support: I would not be here if it were not for your trust in me. Dad, thank you for always answering the phone and helping me with all of my questions: had you gone to Finance, rather than Nuclear Physics, we would be rich by now! Brother, thank you for providing me with a connection to the real world, were regressions, papers and conferences mean nothing: talking with you about books, movies, and politics was always a breath of fresh air. Grandma, thank you for always making sure that I had slept well, that I was eating healthy and for reminding me that, no matter what I do, you will always remember me as your crazy, tiny (not so much anymore) grandchild. Grandpa, I did not become an engineer like you but I hope you are proud of my many successes and I know that you are always by my side.

I would like to thank the numerous people who have helped me during my doctoral work at Syracuse University. Firstly, I would like to express my highest gratitude to Dr. Raja Velu for being such a knowledgeable and supportive adviser. I have become the researcher that I am thanks to your guidance, to the countless hours that you have dedicated to my research and to the many opportunities that you have opened for me. For this, I will always be grateful.

Secondly, I would like to thank my dissertation committee members for their invaluable feedback and for all their advice over the years: Dr. Anand, Dr. Easton, Dr. Harris and Dr. Weinbaum I cannot thank you enough for your time and help. I would also like to thank Dr. Ondrich for agreeing to be my Oral Chair: it was a pleasure to be your student and I have greatly enjoyed our conversations over the years.

Finally, I would like to thank Ann. As somebody else has once said: you complete me. If I had to do this all over again in order to meet you, I certainly would.

Contents

List of Figures

List of Tables

1 Introduction

Recent technological and regulatory changes have transformed how equity markets function. The competition enhancing spirit of Reg-NMS, combined with the significant developments in the technology used to transfer and analyze data, have allowed for a new breed of sophisticated traders to emerge. These traders, use a variety of different trading strategies, ranging from simple market making to more advanced statistical arbitrage, and they can vary considerably in terms of their size and resources. They do, however, have one thing in common: their reliance on algorithmic trading. These market participants increasingly use big data and machine learning techniques in their quantitative trading, and adopt algorithms in what has become mostly a computer-driven investing world. In a recent survey carried out by LCH Investments, "…four of the top 20 hedge funds that have generated the highest amounts of net returns are highly reliant on algorithmic trading." However, the spread of this computer and data driven approach to investing is not limited to the hedge fund industry: a growing number of investment funds have started to develop or strengthen their quantitative trading arms by investing in their analytical and data capabilities. Assessing the extent of the success of *quants* is not possible, given how difficult it is to collect data from some of the key players but the information that is available suggests that there will be a continuing trend in developing strategies that rely on automation.

Traders that employ automation in their trading strategies can be characterized as fast and impatient. In fact, one of the primary objectives behind the automation of trading is that of increasing the speed with which a trader can react to market events. However, the use of powerful computing does not only allow algorithmic traders to react instantly to any changes in the market: it also allows them to base their decisions on the rapid analysis of the large amount of market data that is generated during a trading day. The extensive investments in computational resources made by traders, give them a considerable speed advantage over the other market participants, suggesting that making a distinction among traders based on their speed is, in fact, meaningful. On the other hand, the impatient nature of some traders is a result of their trading strategies. Some of their most common trading strategies are based on exploiting short-lived arbitrage opportunities across markets or across trading venues, or on trading from both sides of the market and making a profit on the difference between the two, or on creating the momentum that leads to favorable price changes. These are by no means the only ways in which traders can generate a profit but they do show how this sub-group of traders prioritizes the ability to quickly enter and exit the market, making them considerably less likely to hold any inventory or, patiently, wait for a favorable trading opportunity.

The significant role played by fast, impatient traders in today's equity market, with the extreme competition in the industry, make this sub group of traders particularly sensitive to changes in regulation and to the introduction of new technologies. In the first part of our study, we investigate how the implementation of Rule 611 from Reg-NMS has affected the trading dynamics of these impatient traders and, specifically, we investigate the role this regulation plays in explaining fleeting liquidity. In fact, several recent empirical studies have found a significant presence of short-lived limit orders in the market. Such orders do not fit the classical economic perspective that limit order submitters are patient liquidity suppliers who are willing to delay execution in exchange for a better price. This phenomenon of short-lived limit orders has received considerable attention yet there is no consensus on the rationale behind it. Anecdotal evidence suggests that the portion of fleeting liquidity that is submitted inside the spread could be explained as a trading strategy aimed at discovering hidden liquidity. “Pinging”, as it is often called by industry practitioners, the limit order book is supposed to help locate hidden liquidity and allow to trade at prices better than the quoted ones. Yet, it is unclear why such strategy would not be implemented with visible limit orders rather than with hidden ones or why is it not done with immediate-or-cancel orders which, if not filled immediately, would be cancelled automatically without the need to send an additional message to do so. Hasbrouck and Saar (2009) suggest that market fragmentation creates a coordination problem between traders since patient traders need to decide the target venue for their hidden limit orders while impatient ones need to signal the venue on which they will be conducting their search. In such a setup, fleeting liquidity would serve a signaling purpose aimed at attracting the other trader’s attention to a specific venue. Such an explanation raises a number of questions, most notably, whose attention is this signal supposed to attract since it is not possible for a human trader to notice a signal from such short-lived orders. Moreover, such explanation does not justify the fleeting liquidity submitted outside the spread, which recent studies find to be a significant portion of the fleeting orders.

We posit that fleeting liquidity is caused by market fragmentation but not as the result of a coordination problem between traders. The impatient nature of traders suggests that they value a quick execution, which is only possible if their limit orders are placed as close as possible to the best Bid or best Ask. Hence, attaining a superior queue placement is particularly relevant for them and any market event that sets back one of their orders could determine their decision to cancel that order. Given that Rule 611 of Reg-NMS enforces a strict price priority rule across trading venues, this implies that to correctly asses the position of an order, it is necessary to consider the state of all the trading venues rather than only the state of the venue where the order is submitted. Hence, we argue that algorithmic traders actively monitor all trading venues after order submission and that, because of their impatient nature, react to any market event that results in a less favorable placement of their orders, regardless of the venue. Consistent with this argument, it follows that market fragmentation affects fleeting liquidity since the determinant of cancellation for an order can now be traced to any of the trading venues in the National Market System.

We propose and test the hypothesis that changes that lead to adverse positioning of the orders submitted by fast, impatient traders are behind the phenomenon of fleeting liquidity. In this study, we define fleeting orders as orders that stay in the limit order book for no more than one second, even shortening the previous cut off value of two seconds employed by Hasbrouck and Saar (2009). However, when necessary, we will re-run the analysis using the two second cut off value in order to compare our findings to previous studies. We believe, though, that the considerable increase in the speed and market activity of today's equity markets, requires me to update the cut-off time in order to correctly capture the order flow generated by the fast, impatient traders whose behavior is the core of our study.

The evidence presented in this study confirms that a distinction in order flow based on the lifetime of orders is meaningful. In our analysis of Level III data for select eighty-five stocks, we find a significant presence of fleeting liquidity in today's equity markets. On average, 52.8% of all submitted limit orders are cancelled within two seconds of submission, compared to only

36.9% found in Hasbrouck and Saar (2009). Figure 1 clearly indicates that markets are now functioning at a higher speed and cancellation rates are higher than those observed in Hasbrouck and Saar (2009) and it re-affirms the necessity for selecting a shorter life-s[an to characterize the fleeting orders.

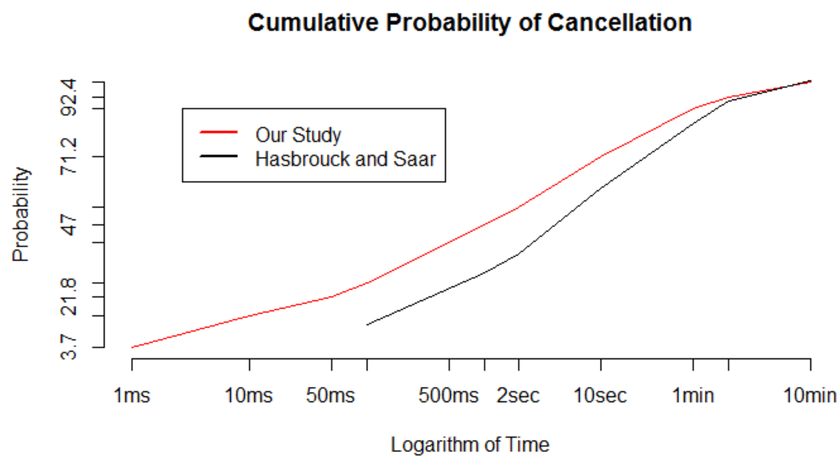

Figure 1: Cumulative Empirical Cancellation Probability. This figure compares the cancellation probability of outstanding limit orders (as a function of time) found in our study to that reported in Hasbrouck and Saar (2009). The values are obtained with a survival analysis approach where the survival functions are built using all non-marketable limit orders in our sample of eighty-five stocks. Moreover, in the estimation of the cancellation process, executions are taken as censoring events.

Before testing our main research hypothesis, we investigate a competing argument proposed in Hasbrouck and Saar (2009) and find empirical evidence to support their claim that the fleeting liquidity submitted inside the spread can serve a signaling purpose to attract the attention of other traders to a specific venue. We then test our assumption that adverse changes in queue placement are behind fleeting liquidity, and investigate the determinants of cancellations of fleeting orders. We find evidence to support the hypothesis that fast, impatient traders actively monitor all exchanges in order to correctly asses the placement on their orders and that fleeting order flow dynamics can be properly explained only when accounting for the changes that occur on all trading venues. The empirical findings provide support of the research hypothesis and confirm that the implementation of Reg-NMS, combined with the impatient nature of algorithmic traders, is behind the puzzling phenomenon of fleeting liquidity.

In the second part of our study, We focus on how the latest technological developments have affected the trading strategies of fast, impatient traders. In particular, we investigate whether these traders are able to exploit their considerable speed advantage since a number of recent theoretical papers (van Kervel (2015) or Baldauf and Mollner (2015)), suggest that it should allow them to anticipate the order flow generated by the other, slower market participants. If that is the case, we argue that these fast, impatient traders should be able to leverage market fragmentation and make use of the availability of market data from multiple trading venues by engaging in an overbooking trading strategy. Such strategy would be based on the simultaneous submission of multiple limit orders across different trading venues with the objective of executing only one of them. In fact, we argue that such strategy would allow them to increase the probability of execution without increasing their risk of over execution, because of their speed advantage that allows them to anticipate the other market participants and cancel the remaining orders as soon as the execution of one is attained.

In order to test this hypothesis, we first develop a procedure to identify those orders that we believe are submitted simultaneously by the same trader, and we call them *clustered orders.* Such first step is necessary since Level III data does not provide any information about the identity of the order submitter, making it impossible to match directly individual orders to specific traders. Then, we carry out an analysis of the nature and composition of these *clusters* of limit orders and we compare their performance to remaining orders. Finally, we test the notion that they belong to the same *cluster* and we investigate when such trading strategy is most effective. The results provide evidence to support the claim that the proposed clustering procedure is able to recognize correctly those limit orders that are submitted simultaneously across trading venues and that belong to the same overbooking trading strategy. We also find evidence that such strategy is able to bring higher execution probability and shorter execution times and that it is particularly effective for stocks for which there is a lot of competition to attain a favorable queue placement or that have a high level of trade fragmentation.

The remainder of our study is organized as follows. In Section 2, we give an overview of Reg-NMS and we discuss the dynamics of order placement in the queue. In Section 3, we review the literature related to fast trading and market fragmentation while in Section 4 we describe the data used in our study and provide some summary statistics. In Section 5, we test the signaling hypothesis proposed in Hasbrouck and Saar (2009) by investigating the relation between fleeting orders submitted inside the spread and the execution of hidden liquidity. In Section 6, we test our hypothesis on the role of adverse changes in queue placement on fleeting liquidity by studying the impact of the events on all trading venues on the determinants of cancellations for fleeting orders. In Section 7, we investigate how fast, impatient traders are able to leverage their speed advantage over the other market participants by studying the effectiveness of a trading strategy based on the simultaneous submission of multiple orders across trading venues. Finally, the conclusions are presented in Section 8.

2 Market Structure, Superior Queue Placement and Reg-NMS

2.1 Market Structure

In this study, we focus on US equity markets and analyze data from four major trading venues: BATS-Z (the main BATS exchange), EDGE-X & A and NASDAQ. If a market participant wants to trade on either of these venues, he has two ways of doing so: he can use a *Market Order*, which allows him to immediately sell or buy the stock by paying the bid or ask price or, he can use a *Limit Order* which allows him to set the highest (lowest) price that he is willing to pay in order to buy (sell) the stock. This means that the *Limit Order* allows to trade at a better price, in exchange for a delay in execution, while the *Market Order* allows to obtain immediate ownership, in exchange for a higher price. The collection of all the outstanding *Limit Orders* at any point in time is called the *Limit Orders Book*, which is the platform used to operate all Order Driven Markets.

Trading venues can differ greatly in terms of their fee and rebate structure and in terms of the rules that regulate liquidity access and provision. However, the vast majority of them, including the four used in our study, enforce a strict time and price priority. This means that Limit Orders that are more aggressively priced, that is higher prices for buy orders and lower prices for sell orders, have priority over those with less aggressive pricing, and that if multiple orders are submitted at the same price, those that were submitted first will have priority over those that came in after. This mechanism makes it very convenient to think of order placement in a Limit Order Book following a queueing system, in with each queue represents the Limit Orders that are submitted at a given price. Figure 2 is a simple example of a Limit Order Book.

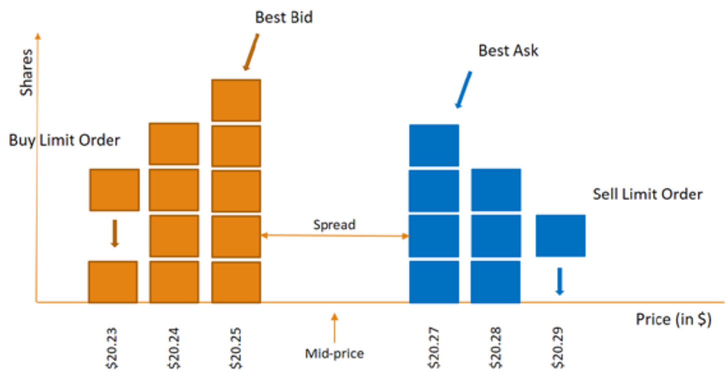

Figure 2: Structure of a Limit Order Book. The collection of all outstanding Limit Orders submitted to an exchange is called a Limit Order Book. The vast majority of trading venues, including the four in our study, enforce a time and price priority rule to regulate access, and provision of liquidity. Price priority implies that Limit Orders submitted at more aggressive prices, higher (lower) for buy (sell) orders, have priority over those at less aggressive prices, while time priority means that if multiple orders are submitted at the same price, those that were submitted first have priority over those that came later.

Execution of a Limit Order occurs when a counterparty willing to pay the asking price enters the market and submits a Market Order. The price priority rule implies that Market Orders will first get executed against the most aggressively priced Limit Orders. From Figure 2, it follows that a Limit Order to Buy submitted at $20.24 will be executed only after the execution of all the Limit Orders at $20.25. Thus, a clear trade-off between the limit price of an order and the waiting time until execution emerges: in fact, better prices for a trader, lower for buys and higher for sells, might lead to longer execution times, as more Limit Orders might need to be executed before this order. It is also important to realize that a delay in execution may also negatively affect the overall probability of attaining the execution since, while waiting, new information might arrive in the market and change the fundamental value of the asset moving the market away from the limit order. An interesting consequence of the price and time priority rules is that all Limit Orders on each side of the book belong, de facto, to the same queue that starts with the first order submitted at the most aggressive price and ends with the last order submitted at the least aggressive one.

2.2 Superior Queue Placement

In our study, we focus on the analysis of the order flow generated by fast, impatient traders. Such traders can be loosely thought of as traders who use some kind of automation in their trading strategies. The fast nature of their orders implies that they can react to changing market condition in a millisecond time scale while their impatience suggests that their main priority is to attain a quick execution. The combination of these two features, speed and impatience, is what makes the queue position of their limit orders particularly relevant for them. A fast, impatient trader will not be able to make the most of a short-lived trading opportunity unless he is able to place his limit orders as close as possible to the top of the queue. In fact, only if such superior queue placement is attained, the trader will be able to find, in a short time, a counter party willing to trade via a Market Order and not have to wait for a large number of Limit Orders in front to clear.

A superior queue placement, however, comes with some disadvantages as well as it exposes to a higher risk of being picked off by an informed trader. In fact, if new information that alters the fundamental value of the asset arrives, the limit prices of the outstanding orders need to be reevaluated. This means that the closer the order to the top of the queue, in less time the order will have to be re-submitted at a new price before an informed trader picks it off at the now stale and outdated, price. Hence, if on one hand being at the top of the queue allows for a shorter execution time on the other it requires constant monitoring of all information generated in the market to decrease the risk of being picked off by an informed trader.

2.3 Regulation National Market System

When the Securities and Exchange Commission proposed Reg-NMS in February 2004, it described it as “a series of initiatives designed to modernize and strengthen the National Market System for equity securities.” The declared objective of this financial regulation was to assure that investors received the best price for their executions by encouraging competition and by strictly regulating the access and provision of liquidity in the marketplace. In fact, by then regulators had become concerned with the extent of the fragmentation of the equity markets and with its possible implications on market quality. Thus, it became apparent that new set of rules had become necessary in order to regulate trading across all the different venues on which a market participant could buy, or sell, shares.

In Reg-NMS, two rules in particular play an important role in understanding our argument about how the actions of fast, impatient traders affect fleeting liquidity: Rule 610, also called the “Access Rule” and Rule 611, also known as the “Order Protection Rule”. The “Access Rule” is broad in scope and addresses the issue of accessing market data by market participants. Beyond prohibiting the imposition of discriminatory terms that would prevent access to an exchange’s quotations, it also requires that members of the National Market System avoid quotations that would “lock” or “cross” the marketplace. This means that market participants are not free to submit a Limit Order at any price they might want: when setting their limit price, an exchange needs to make sure that it does not match, or exceed, the best price on the opposite side of the book on another, competing exchange. Figure 3 illustrates an example when such rule is, for a very short period of time, violated. On January 25th, 2011, between

10:42.545 am and 10:44.547 am, the best Ask Price for AOL on BATS-Z was above the best Buy Price on the NASDAQ. However, shortly afterwards, the NASDAQ accepted a Limit Order to Buy shares of AOL at the same price at which one could sell those shares on BATS-Z. Such event is called a “locked” market as the two opposite best prices overlap. A further analysis of data in Figure 3 shows how shortly afterwards, the best Ask Price on BATS-Z actually drops below the initial value, resulting in what is called a “crossed” market during which the best Ask Price on one exchange is below the best Buy Price on another one. Such situations, even though not permitted, can occur several times during a trading day but are, generally, quickly corrected.

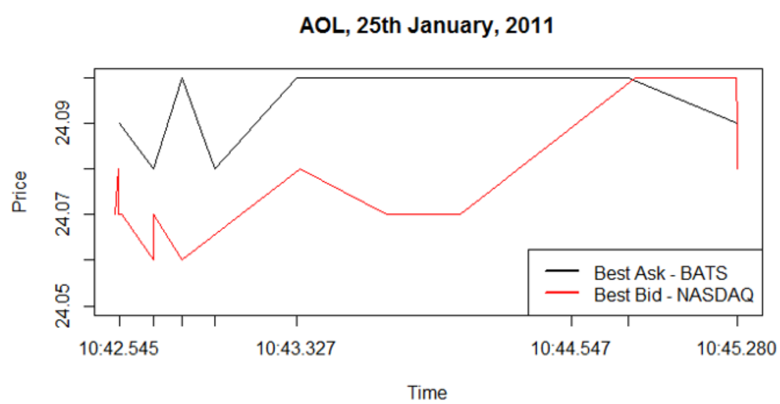

Figure 3: Locked and Crossed Markets. This figure provides a good example of what it means for the National Market System to be “Locked” or “Crossed”. Rule 610 in Reg-NMS, mandates that an exchange should not accept limit orders at prices that match, or exceed, the best price on the opposite side of the book on a competing trading venue. In practice, instances of locked or crossed markets can occur regularly during a trading day but they are usually immediately corrected. In our example, we see how on January 25th, 2011, between , 10:42.545 am and 10:44.547 am, the best Ask Price for AOL on BATS-Z was above the best Buy Price on the NASDAQ. However, shortly afterwards, the NASDAQ accepts a Limit Order to Buy shares of AOL at the same price at which one could sell those shares on BATS-Z. Such event is called a “locked” market as the two best prices overlap. Moreover, shortly afterwards, the best Ask Price on BATS-Z actually drops below the initial value, resulting in what is called a “crossed” market in which the best Ask Price on one exchange is below the best Buy Price on another on.

The second most relevant rule to our study is Rule 611, also called the “Order Protection Rule”, which requires that members of the National Market System prevent “trade-throughs” in the marketplace. What this means is that those Limit Orders that are submitted at the most aggressive prices will always have priority over all others, regardless of the venue on which they are submitted. Figure 4 provides an example that will help to better understand the implications of Rule 611. If the best Bid on the NASDAQ is at $24.05, while that on BATS-Z is $24.06, any limit order on the best Bid of the NASDAQ will be executed after the execution, or removal, of all the buy Limit Orders on BATS-Z at $24.06. Hence, even though limit orders to buy shares on the NASDAQ at $24.05 are the most aggressively priced ones on that venue, they are not the most aggressively priced ones across the entire National Market System since more aggressive orders exist on BATS-Z. This rule protects the best price across the entire market, regardless of how small the venue quoting it might be, guaranteeing that the most aggressive orders will be executed first. Such rule allows to attain two important results: first, it allows even the smaller venues to compete with the larger ones, as long as they are willing to provide a better price. In fact, a growing concern for regulators before the introduction of Reg-NMS was that the large exchanges, such as the NASDAQ or NYSE, were able to attract traders based on the large depth provided in their limit order books (which offered the opportunity to trade fast and in larger quantities) rather than because of their competitive quotes. This made it very hard for the smaller trading venues to compete more effectively with the leading exchanges, since they were not able to match the level of liquidity provided by their larger competitors. Second, it protects the traders who submit their Market Orders by assuring that they will be always able to buy at the lowest price or sell at the highest one. In fact if, in the example in Figure 4, a trader were to submit a *Market Order* to sell shares on the NASDAQ, open receiving such order the NASDAQ would have to automatically re-route it to BATS-Z where the seller would be able to receive a higher price for his shares.

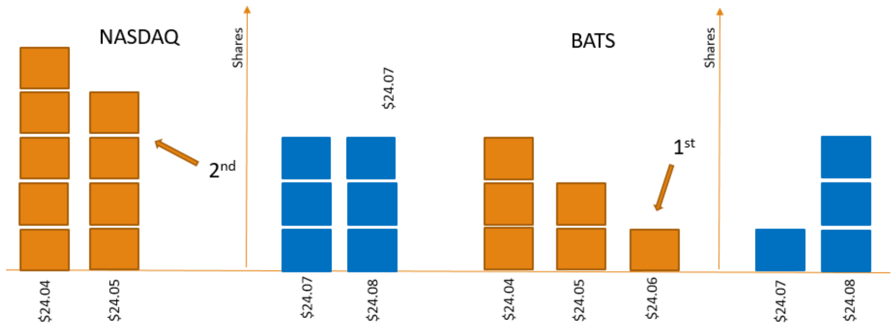

Figure 4: Trade-Throughs on the National Market System. Rule 611, called the "Order Protection Rule", mandates that the most aggressive price always receives priority regardless of the venue quoting it. In the following example, the most aggressive Bid price on the NASDAQ is $24.05, while that on BATS-Z is $24.06. Given that the most aggressive price is on BATS-Z, it will not be possible for a Buy Limit Order on the NASDAQ at $24.05 to be executed until all outstanding Limit Orders on BATS-Z at $24.06 are removed or executed. Such mechanism attains two purposes: first, it allows even the smaller venues to compute more effectively with the larger ones, as long as there are willing to provide a better price. Second, it guarantees that every trader will always receive the best price possible regardless of his target venue. In fact, a Market Order will be always automatically re-directed to the venue offering the best price if that does not match the venue it was submitted to.

The implementation of Rule 611 can result in an interesting consequence: it extends the price priority rule across trading venues. In fact, if after the submission of my Limit Order a new, more aggressive one is submitted on another venue, its effect on the priority of my order will be the same as if this new order had been submitted on my own venue. This, in essence, creates one large queue for each side of the book across the entire Market System since Rule 611 aggregates the orders submitted on all trading venues. This implies that the overall position of a Limit Order will now depend from the distribution of Limit Orders on all trading venues and not only from the shape of the venue where my order resides. One important clarification is in order: even though Rule 611 extends the price priority across trading venues, the same is not true for time priority. If two Limit Orders are submitted at the same price but on different trading venues, it will not matter which one was submitted first: time priority will always only apply to those Limit Orders that are submitted at the same price and on the same venue. What this implies is that when looking across venues in order to better understand the placement of an order in the overall queue, one has only to consider those orders that are submitted at prices that are more aggressive than the order's limit price.

3 Literature Review

The scope of our research, calls for reviewing those areas of market micro-structure that are relevant to market fragmentation, to the interaction between trading venues and to the issue of fast and slow trading.

The increased competition resulting from the implementation of Reg-NMS brought an explosion of trading venues leading to today's fragmented market with over 30 lit venues. These venues differ in terms of trading rules, pricing structure for posting and removing liquidity and for the level of pre-trade transparency while competing on execution speeds, system latencies and order types. This proliferation of venues has, apparently, brought several improvements to the equity markets: O'Hara and Ye (2011) find that a higher degree of market fragmentation results in lower transaction costs and shorter execution times and that prices appear to follow more closely a random walk, suggesting that the market has become more efficient. Boehmer and Boehmer (2003) find that after the NYSE started trading ETFs listed on the American Stock Exchange there has been a substantial decrease in trading costs across market centers and a considerable increase in quoted depth as a result of declined estimated price impact of trades. Hendershott and Mendelson (2000) demonstrate that the addition of an alternative trading venue reduces inventory risk, narrows spreads and attracts new liquidity which allows to increase liquidity-based order flow and, in the case of fundamental, value-based information trading, provides another venue for informed trading for individual dealers.

Foucault and Menkveld (2008) find that the entry of a new trading venue increases consolidated limit order book depth enhancing liquidity supply while Fong, Madhavan, and Swan (2001) find positive effects on trading costs for large Australian stocks when executed off-exchange. Finally, a number of earlier empirical work also find positive effects of fragmentation. Most notably, Battalio (1997) finds that spreads narrowed on the NYSE after a major third market broker, Bernard L. Madoff Investment Securities, started trading.

Yet other empirical research contradicts those findings. Bennett and Wei (2006) find that order flow becomes consolidated when stocks switch from a dealer market, NASDAQ, to an exchange, NYSE, and that the post switching improvements of market quality are directly related to the degree of order flow fragmentation on the NASDAQ. Gajewski and Gresse (2007) find reduced trading costs when moving to a more consolidated market while Amihud, Lauterbach and Mendelson (2003) test the value effect of consolidation and find that the liquidity of the stock improves after consolidation of trading on the Tel-Aviv Stock Exchange. In addition, a number of earlier theoretical papers have shown results in favor of consolidation. Chowdry and Nanda (1991) argued that adverse selection costs increased with the number of markets trading a given asset. Pagano (1989) demonstrated that equilibrium with multiple venues is inherently unstable as traders eventually gravitate towards the venue with more liquidity and Madhavan (1995) found that dealers can benefit from fragmentation by being less competitive and that informed and large traders could benefit by being able to hide their trades.

However, even if the fundamental issue of how market fragmentation affects market quality is still debated, and even if we account for a number of potential venue consolidations in the near future (going beyond those observed this far of BATS and Direct Edge or ICE and NYSE in 2013) or the introduction of new federal regulation aimed at consolidating the market (as recently declared by SEC Commissioner Dan Gallagher, http://www.cnbc.com/id/102050921), a fragmented market with several points of entry appears to be the new standard in the US equity market.

A growing yet still very limited body of literature addresses the issue of the interaction between trading venues. Significant empirical evidence suggests that sophisticated investors employ cross-venue trading strategies in order to maximize their execution probabilities and minimize trading costs. Cont and Kukanov (2013) show that, market participants consider order flow characteristics and queue size on all trading venues when making their order routing decisions. Moreover, they find that traders have the tendency to place, across multiple exchanges, more orders than they need to fill and argue that such strategy is aimed at reducing the risk of non-execution. van Kervel (2015) shows that cross-venue strategies create highly interlinked markets since trades on one venue are followed by sizeable cancellations of limit orders on competing ones. These cancellations are explained in a simple model of competition between two limit order markets with fast and slow traders. O'Hara (2015) crafts a careful analysis of market microstructure since the high frequency trading revolution, which developed in parallel to the process of trading automation that has followed the introduction of Reg-NMS and brings up the issue of the effect of sophisticated cross-venue trading strategies on the interpretation of market data and, more importantly, of the information content of some well-established measures of market performance. The evidence points to the fact that events occurring on one venue can be the result, or the determinants, of events occurring on other venues which suggests that studying the interaction in order flow across venues can help us better understand how market participants interact in a fragmented market environment.

Finally, a number of recent theoretical and empirical studies investigate the coexistence of fast and slow traders on lit venues. Baldauf and Mollner (2015) study the consequences of high frequency trading in a multi-venue framework and find that the ability of fast traders to anticipate orders from other market participants reduces the incentive to carry out costly information acquisition about the fundamental value of an asset. This, in turn, leads to less information available to incorporate into prices, which hurts the price discovery function of lit venues. Hoffman (2014) finds that the ability by fast traders to revise their limit orders soon after the arrival of new information allows them to reduce the risk of being picked off by an informed trader and increases trading. On the other hand, their presence also induces slow traders to strategically submit limit orders with a lower execution probability, which can actually reduce trading and hurt social welfare. In van Kervel (2015) the author develops a simple model of competition between two limit order markets with fast and slow traders and finds that the presence of fast traders prompts market makers to reduce liquidity supply on all venues which results in highly interlinked markets. In these studies it is assumed that the two types of traders differ in their reaction speed to the release of new information and that fast traders are able to trade on multiple venues while slower ones are either unable to do so or are unable to do so successfully.

4 Data

4.1 Sample Summary Statistics

In order to select our sample, we perform a double sort based on equity market capitalization and volatility[1] of all domestic common stocks with primary listing on the New York and NASDAQ exchanges. We then obtain a one hundred stocks sample by taking one stock for each decile of the sort, resulting with fifty-four stocks with primary listing on the NASDAQ and forty-six on the NYSE. In our empirical analysis, however, we drop all stocks that have average price below $2, resulting in a final sub-sample of eighty-five stocks. The sample period is the twenty trading days of January 2011 and we collect data from four major US exchanges: BATS-Z (the main BATS exchange), EDGE-X & A and NASDAQ. The data was purchased from a low latency data vendor and the choice of venues was dictated by the sellers' availability.

Our dataset contains almost every message generated by each exchange and Table 1 presents a summary of the types of messages that are reported. The only message that is not included is the one reporting the submission of a hidden limit order. Given that our dataset is comprised of a time-stamped sequence of almost all the messages generated by each exchange, we are able to rebuild completely the visible portions of the four limit order books. Whenever a new limit order is submitted, it is reported in our dataset with a "B" if it is a buy order and "S" if it is a sell one. If an outstanding order is cancelled, it is reported as a "D" for a full cancellation and as a "C" for a partial one. On the other hand, if the outstanding order is executed, it will be reported

[1] Equity market capitalization and volatility are computed based on monthly data for 2010 collected from the CRSP dataset.

as an “F” for a full execution and as an “E” for a partial one. For each message reported in the dataset, a time-stamp, precise to the millisecond, is added together with a unique order ID identifying the limit order affected by the message. When a newly submitted limit order is marketable, it is matched (fully or in part) against (one or multiple) outstanding limit orders and a message of partial or full execution is generated. If, however, the limit order is not marketable, a new order submission message is generated specifying the side (buy or sell), price and size of the order. When an outstanding limit order is cancelled (fully or in part) a cancellation message is generated specifying the order being removed. Even though the submission of a partially or completely hidden order is not reported in our dataset, a message reports when a marketable limit order is executed against a hidden one, event reported in our dataset with a "T". In our analysis we will only use messages that are generated during regular trading hours and we will only investigate those limit orders that are submitted at the top ten price levels of the limit order book.

Exchange Generated Message	Limit Order Book Event
B	Submission of visible limit order on Buy side
S	Submission of visible limit order on Sell side
E	Partial execution of visible limit order
F	Full execution of visible limit order
C	Partial cancellation of visible limit order
D	Full cancellation of visible limit order
T	Execution of hidden limit order

Table 1: List of Messages Generated on Each Exchange. The dataset contains almost every messages generated on each exchange, with the only exception of the submission of a hidden order. Each event has a timestamp, precise to the millisecond, and a unique order ID that specifics the order that it refers to.

In Table 2 we present some summary statistics for the stocks in our sample. We do not resort to any sampling, but rather use the entire messaging data. By construction, there is large variation in market cap, with values ranging from $18m to $326b and with average value of $29.1b. Similarly, stock price also shows considerable variation with the cheapest stock in our sample having an average closing price of $2.1 while the most expensive one $494.12. The average turnover, defined as the ratio of traded shares out of all outstanding, is 28% but we also include stocks with very low and very high turnover ratios (respectively, 1% and 133.6%). In the lower panel of Table 2, we look at the extent of order flow fragmentation, which allows me to get a better idea of the variation in market share of each venue across stocks.

	Mean	Median	SD	Min	Max
Stock Characteristics					
Market Cap ($mln)	29,127	2,059	59,822	18	325,931
Price ($)	31.52	16.82	60.68	2.10	494.12
Turnover (%)	28	21	27	1	134
Volatility	0.20	0.14	0.19	0.04	1.06
Order Flow Fragmentation					
Executions (Percentages)					
EDGE-A	9.5	8.4	7.8	0	38.2
EDGE-X	17.3	13.6	11.1	5.1	55.1
NASDAQ	52.4	52.6	11.4	24.2	78.1
BATS – Z	20.8	22.2	9.5	1.0	37.1
Submissions (Percentages)					
EDGE-A	9.3	9.2	7	0	27.9
EDGE-X	11.3	10.4	5.4	1.3	27.5
NASDAQ	57.2	55.9	15.4	18.6	98.1
BATS – Z	22.3	24.3	9.4	0.6	63.7

Table 2: Sample Summary Statistics. In the upper panel we present the sample distribution of market cap (in $billions), price (in US dollars), turnover (measured as the ratio of traded shares out of all outstanding) and volatility (measured as the standard deviation of monthly returns). The data necessary to build this portion of the table was collected from COMPUSTAT and CRISP for years 2009 and 2010. In the lower panel, we show the sample distribution of order flow fragmentation across venues, measured as the percentage of executions or submissions occurring on each venue out of total. This portion of the table was built using all the messages in our sample period of January 2011.

The NASDAQ appears to be the dominant exchange for both executions and new order submissions as, on average, we see that every other execution (52.4% of market share) and new order submission (57.2% of market share) occurs on it. BATS-Z is a distant second with, around, one execution or new order submission every five (respectively, 20.8% and 22.3% of market share). Not surprisingly, EDGE-A (the only inverted pricing[2] venue in our sample) has, on average, a very small market share confirming the niche role played by venues utilizing this type of fee and rebate structure.

We will now summarize the Limit Order Book characteristics of the stocks in our sample: in the specific, we will look at the level of market activity and outstanding depth imbalance. In fact, given our ultimate goal of studying the cancellation determinants of fleeting orders, it is clear that we will be working on a microstructure level and, hence, it is important to verify that our sample is representative of a wide range of microstructure behaviors.

In the upper panel of Table 3, we present the distribution of the average number of trades per day on each trading venue for the stocks in our sample. Once again, we see a fair degree of variation suggesting that the sample includes both very heavily traded stocks, with a maximum average number of daily trades of almost 20,000, and not very active ones, that average only a handful of executions per day. Consistent with the results reported in the lower panel of Table 2, we see that the NASDAQ is the venue that has the largest number of trades: the most actively traded stock in our sample reports almost 20,000 trades every day on the NASDAQ while on BATS-Z the most actively traded stock has only about 9,000. If we look at the mean or median number of trades per day, we see that, once again, the NASDAQ is the venue were

[2] An inverted pricing venue is a venue on which market participants are charged a fee when posting a limit order and collect a rebate when removing it. Their name is due to the fact that most venues implement the opposite fee and rebate structure, paying a rebate for order submissions and collecting a fee for their removal.

most of the trades occur. BATS-Z is a fairly distant second with about half the executions reported on the NASDAQ while the two EDGES, combined, have a similar number of trades to that of BATS-Z.

Another way of characterizing the level of market activity for the stocks in our sample, is to look at the number of limit order book modifications that occur during a trading day. Such approach gives a more complete view of limit order book dynamics than that obtained by simply looking at the number of trades report during a day. The results for the average number of daily messages generated on each exchange are presented in the lower panel of Table 3. Once again, we see that some of the stocks in our sample are considerably more active than others, with an average daily number of messages that ranges from 0 to over 600,000. Moreover, consistent with what was seen those far, the NASDAQ is still the venue with the largest average daily number of limit order book modifications, 117,837, while EDGE-A is the one with the smallest one, only 28,049. It is also interesting to point out that the results in Table 3 suggest that our sample includes a stock that is not at all traded on EDGE-A.

We now investigate whether there is any systematic imbalance between the Bid and Ask sides during our sample period. The results in Table 4 present the summary statistics for the depth available to traders on each venue for the two sides of the Limit Order Book. In the specific, we look at the time-weighted average aggregate value, in US dollars, of all the outstanding limit orders on the Bid (Ask) side at prices between the mid-point and the mid-point minus (plus) ten basis points.

	Mean	Median	SD	Min	Max
Trades					
EDGE-A	1,026	169	1,664	0	6,902
EDGE-X	1,149	191	1,777	0.31	7,991
NASDAQ	2,983	785	4,023	4	19,465
BATS – Z	1,830	415	2,570	0.05	9,064
Limit Order Book Events					
EDGE-A	28,049	7,260	49,168	0	307,081
EDGE-X	35,461	5,950	55,940	5	245,862
NASDAQ	117,837	34,348	147,898	263	603,896
BATS-Z	61,310	14,371	89,679	0.3	451,881

Table 3: Limit Order Book Sample Summary Statistics. The upper portion of the table presents the summary statistics for the average number of trades reported each day on each venue for the stock in our sample. The lower portion of the table, presents the distribution of the average number of limit order book events generated one each venue. Looking at the number of messages generated by each stock on each venue, as opposed to only counting the number of executions, allows for a more complete picture of the level of activity for the stocks in our sample.

The choice of ten basis points guarantees that only depth that is close to the mid-point is considered which assures that we only consider the most relevant price levels. This is important as depth far away from the best bid and ask is very unlikely to get executed and hence it is of little interest to traders. Moreover, the choice of time-weighting the aggregate values further helps to give an idea of the true depth available to traders without giving too much importance to phantom liquidity that cannot be effectively accessed by most traders. Based on these requirements we have developed the following measures of liquidity for the two side of the book and for each one of the *N* trading venues.

If we define P_j with $j = 1, 2, \dots, J$ the price grid of prices available within 10 basis points from the mid-price, Q_j as the outstanding depth available at each price level and *time* as the amount of time during which that level if outstanding depth is available in the book then for venue *N*

$$Depth\ Ask_N = \sum_{j=1}^{J} P_{j,N}^{Ask} Q_{j,N}^{Ask} * time,$$

$$Depth\ Bid_N = \sum_{j=1}^{J} P_{j,N}^{Bid} Q_{j,N}^{Bid} * time.$$

indicate the time-weighted value, in US dollars, of all the outstanding depth within 10 basis points from the mid-quote on each side of the book.

The upper part of Table 4 present the results for the Ask side, while the lower one those for the Bid side: the entries represent US dollars. With the exception of EDGE-X which present some degree of asymmetry in the depth available on the two sides of the book, with more depth available on the ask side, the other three venues have similar values of outstanding depth. Moreover, if we compare the results in Table 4 to those in Table 3, we see that the higher the executed volume in a venue, the higher the depth available. In fact, we can see how the NASDAQ, which executes twice the trades of BATS-Z has also, roughly, twice the depth available. A similar pattern holds when comparing BATS -Z with EDGE-X and EDGE-X with EDGE-A. This is in contrast with the findings of van Kervel (2015) who, using data for a sample of FTSE 100[3] stocks, found that venues with a smaller share of executions offered a comparable level of outstanding depth to the dominant ones.

Overall, the results in Table 2, Table 3 and Table 4 underlined the extent of the variation in our sample. The selected eighty-five stocks vary considerably in terms of size, volatility, price and turnover. Moreover, they also differ considerably in terms of market activity, measured by

[3] The Financial Times Stock Exchange 100 Index, also called the FTSE 100, is a share index of the 100 companies with the highest market capitalization on the London Stock Exchange.

either the daily number of executions or, more completely, by the daily number of limit order book modification.

Depth Ask	Mean	Median	SD	Min	Max
EDGE-A	96,107	37,232	151,254	3,740	477,241
EDGE-X	232,785	172,210	233,030	6,564	642,507
NASDAQ	599,704	391,232	624,766	22,088	1,841,260
BATS – Z	250,377	228,665	247,364	16,624	793,033
Depth Bid					
EDGE-A	143,624	42,343	185,282	3,782	485,634
EDGE-X	188,801	99,655	195,172	7,036	545,765
NASDAQ	572,210	403,318	587,792	21,686	1,712,756
BATS – Z	236,487	195,753	248,975	7,888	803,096

Table 4: Summary of Outstanding Limit Order Book Depth. This table presents the results for the time-weighted average aggregate value (in US dollars) of all the Limit Orders submitted within ten basis points from the mid-quote. Such choice for the cut-off value allows us to correctly represent only the depth that is of interest to traders as limit orders that are deep in the book are unlikely to be executed, hence are of limited interest. Moreover, by computing a time-weighted average, rather than a simpler arithmetic one, we make our measure resistant to possible short-lived peaks of phantom liquidity that cannot be effectively accessed by traders.

However, our preliminary findings also show some remarkable differences between the four trading venues. The NASDAQ is the dominant exchange in terms of executions and new order submission and it is also considerably more active than other venues. On the other hand, EDGE-A, the only inverted pricing venue, is the smallest and least active one in our sample. Given the extent of the variation across exchanges reported in these first results, and in light of the objective of our study, we will dedicate the remainder of this section to further investigate the key differences between the four venues in our sample.

First, we look at the cancellation statistics across trading venues since they allow to understand whether venues differ considerably in terms of market quality. In fact, a common measure used, and mandated by the SEC, to evaluate the quality of the liquidity provided by an exchange is the cancellation rate, defined as the ratio of Limit Order cancellations to the total number of Limit Order submissions. The intuition behind this is that a high cancellation rate signifies that a high proportion of Limit Orders is cancelled rather than executed. This suggests that it is not possible to trade against a high proportion of limit orders, which, in turn, implies that the liquidity provided on that exchange cannot be relied upon. The results in Table 5 show how, on average, there is very little variation in the cancellation rates across the four venues. Consistent with the findings of Hasbrouck and Saar (2009), we see that the vast majority of Limit Orders is cancelled. We find that the overall cancellation rate across is 95% with little variation. These findings are not surprising and support the commonly held belief that liquidity dynamics can change dramatically in a very short period of time.

	EDGE.A	EDGE.X	BATS.Z	NASDAQ	Total
Submitted LO	24,493,611	30,752,292	52,498,530	99,175,346	206,919,779
Cancelled LO	22,743,830	28,845,117	49,643,881	94,478,106	195,710,934
% of cancelled LO	93%	94%	95%	95%	95%
% of total cancellation on each venue	12%	15%	25%	48%	100%

Table 5: Cancellation Rates. This table presents the total number of Limit Order submissions and cancellations reported on each exchange. It further presents the cancellation rates for each venue, defined as the proportion of cancelled Limit Orders out of the total of all orders submitted. In the last row of the table, we present the percentage of cancellations on each venue out of the total.

Second, we look at the difference across trading venues in terms of the cost of trading. In fact, as documented in a recent study by Battalio, Corwin and Jennings (2016), traders do take into consideration the fee and rebate structure of an exchange when executing their trading strategies. Hence, it is important to know whether any such differences exist in our sample.

The fee and rebate structure of a trading venue determines the cost of trading and therefore will determine the profitability of trading strategies. One of the ways in which the exchange generates its revenue is by charging fees to those who want to trade on them while, one of the ways in which it attracts market participants and increases its market share, is offering a rebate on the Limit Orders. The difference between the fees charged and rebates paid is the profit made by the exchange. Such pricing model is called "Maker-Taker[4]" and the rationale behind it is that traders who use Market Orders, usually, want to trade as fast as possible and with as little price impact as possible. Hence, such traders will prioritize trading on those exchanges that can offer the most liquidity at the best prices and, presumably, traders are willing to pay for such opportunity. Charging a fee for removing liquidity via a Market Order, also allows the exchange to offer a rebate for traders who send Limit Orders their way. This inflow of Limit Orders allows the exchange to build its liquidity and, as a result, become more desirable to traders who need to trade. Even though the "Maker-Taker" pricing model is the dominant one as is used by the vast majority of trading venues, there exist an interesting alternative: the "Inverted-Pricing" model. As the name suggests, such model uses an inverted pricing system, which means that those who send a Market Order to an "Inverted-Pricing" venue will be paid a rebate while those who submit a Limit Order will be charged a fee. The idea behind such pricing model is that offering a rebate for every Market Order should make that venue considerably more appealing to traders who choose the target venue for their Market Orders. On the other hand, if the ultimate purpose of a Limit Order is execution, a Limit Order submitter might view

[4] The name "Maker-Taker" refers to the fact that the market participant who "makes" the market by submitting his Limit Order receives a rebate, while the one who "takes" from the market with his Market Order will have to pay a fee.

favorably a trading venue that he knows is preferred by Market Orders submitters as such venue would offer him a higher chance of meeting a counterparty. Only a handful of "Inverted-Pricing" venues exist in the US and they have a very small market share compare to the venues that use the regular "Maker-Taker" pricing model.

Table 6 summarizes the fee and rebate structure for the four venues in our sample. During our sample period, January 2011, NASDAQ and EDGE-X are the venues that charge the most for removing liquidity: respectively, \$0.0030 and \$0.0029 per share. On the other hand, BATS-Z is the venue that charges the least for removing liquidity, with only \$0.0025 per share. BATS-Z is also the venue that offers the highest rebate when submitting a limit order, \$0.0024 per share while, the NASDAQ offers the lowest one, paying only \$0.0010 per share. If we now look at EDGE-A, we see that it is a clear example of an "Inverted-Pricing" venue since removing liquidity from this venue results in a rebate of \$0.0002 while posting a new Limit Order requires the payment of a fee of \$0.0030. In essence, out of the three venues with a standard maker-taker fee structure, NASDAQ is the least attractive one while BATS-Z is the most attractive.

It is also interesting to point out how the fee and rebate structure of a venue can change during the lifetime of the trading venue. For example, for EDGE-X we see that in January 2010, it used to charge a fee for posting a hidden order yet, 10 months later, it switches to paying a rebate when doing so. Other, smaller adjustment are also pretty common with the

	EDGE – A	EDGE-X	NASDAQ	BATS – Z
01/01/2010	1. Posting Hidden: -0.0030 2. Posting Visible: -0.0002 3. Removing: 0.0002	1. Posting Hidden: -0.0030 2. Posting Visible: 0.0029 3. Removing: -0.0029	1. Posting Hidden: 0.0010 2. Posting Visible: 0.0020 3. Removing: -0.0030	1. Posting Hidden: 0.0020 2. Posting Visible: 0.0024 3. Removing: -0.0025
09/30/2010	No Change	1. 0.0026 2. 0.0026 3. -0.0030	No Change	No Change
10/5/2010	1. -0.0030 2. -0.00025 3. 0.00015	No Change	No Change	No Change
10/29/2010	No Change	No Change	No Change	1. 0.0020 2. 0.0027 3. -0.0028
3/1/2011	No Change	1. -0.0030 2. 0.0023 3. -0.0030	No Change	No Change
7/1/2011	No Change	No Change	No Change	1. 0.0017 2. 0.0025 3. -0.0029
7/27/2011	1. -0.0030 2. 0.0005 3. -0.0006	No Change	No Change	No Change
9/30/2011	1. -0.0030 2. 0.0004 3. -0.0007	1. 0.0015 2. 0.0023 3. -0.0030	No Change	No Change

Table 6: Fee and Rebate Structure Across Venues. All prices are in USD per share traded. Negative values represent fees charged while positive ones are rebates paid by the exchange. The term "Posting Hidden" refers to the action of submitting a hidden limit order; "Posting Visible" refers to the submission of a visible limit order while "Removing" refers to the action of removing a limit order, hence refers to the action of submitting a market order. Exchanges implement a tiered pricing structure in which better clients receive better pricing. Better pricing usually means lower fees and higher rebates but, occasionally, can also mean that a client will receive a rebate instead of paying a fee. The prices reported in this table represent the lowest tier in the pricing structure for each venue, which is the least competitive one offered to their clients.

NASDAQ being the only exception and not changing at all its fee and rebate structure during the two year period covered in our dataset. One final remark is about the prices stated in Table 6. Every exchange implements a tiered pricing model, where better clients get better pricing. Better pricing, usually, means that the client will pay a lower fee or will receive a higher rebate. Sometime, however, better pricing can mean that a trader will actually receive a rebate instead of paying a fee and vice versa. An example of such case is EDGE-A: based on the information provide in Table 6, it would appear that on 7/27/2011, this venue switched from an “Inverted-Pricing” model to a classic “Maker-Taker” model. However, upon closer analysis, we observe that such pricing is available only for the lowest tier of customers. The better ones, are still treated with the previous fee and rebate structure and receive a rebate when removing liquidity and need to pay a fee when posting it. Different venues use different criteria to rank their clients but, in general, a client that sends a large number of Limit or Market Orders during a trading day is rated higher. Sometime restrictions apply in terms of how long do the Limit Orders need to stay in the book or how aggressively priced they must be in order to qualify for a certain tier but, in general, a client with more orders gets better pricing. The same is true for the case of EDGE-A in our sample: in fact, even though the lowest tier implies a classic “Maker-Taker” fee structure, attaining a better pricing tier is extremely simple on this venue and is possible even with the submission of a very limited number of orders. Hence, it is critical to clarify that the prices provided in Table 6, are those for the lowest tier possible, hence they represent the worst-case scenario for a trader on each of these venues, when he has to pay the highest fee or he receives the lowest rebate.

4.2 Fleeting Liquidity Summary Statistics

We begin now to analyze the fleeting liquidity present in our sample and we do so by means of empirical survival analysis. Table 7 and Figure 5 present the estimated cumulative cancellation and execution probabilities for the stocks in our sample and are generated by pooling together all of the non-marketable limit orders for the eighty-five stocks. Consistent with what was observed in previous studies, a large number of limit orders is cancelled within a very short time period since submission. We observe that 44.2% of all submitted limit orders are cancelled within one second of submission with the percentage increasing up to 52.8% if we consider the cancellations within the first two seconds.

	Cumulative Probability of	
Time	Cancellation	Execution
1 milliseconds	3.7%	0.1%
10	14.8%	0.6%
50	21.7%	1.1%
100	26.3%	1.3%
500	41.1%	2.3%
1 seconds	44.2%	2.9%
2	52.8%	3.8%
10	71.2%	7.7%
1 minutes	88.3%	15.1%
2	92.5%	18.6%
10	97.9%	26.7%
1 hours	99.6%	35.7%

Table 7: Cancellation and Execution Rates of Limit Orders. This table presents the estimated cumulative probabilities of cancelling and executing a limit order within a given time interval after submission. The survival function is estimated using the life-table approach using all non-marketable limit orders for each of the 85 stocks in our sample. Moreover, in the estimation for the cancellation (execution) process, executions (cancellations) are taken to be the censoring events.

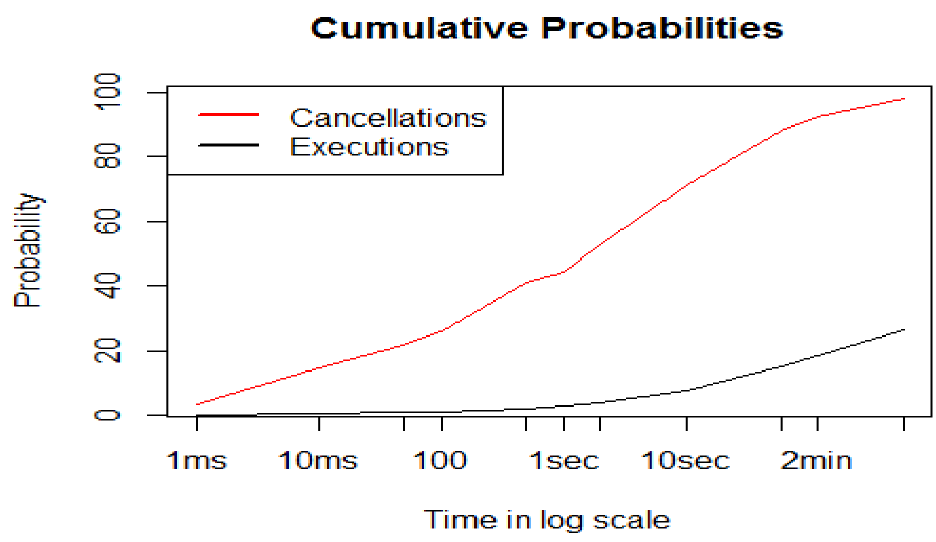

Figure 5: Cancellation and Execution Rates of Limit Orders. This figure presents the estimated cumulative probabilities of cancelling and executing a limit order within a given time after submission. The survival functions are built using all non-marketable limit orders in our sample of eighty-five stocks and is estimated using the life-table approach. Moreover, in the estimation of the cancellation (execution) process, executions (cancellations) are taken as censoring events.

This results show how markets have become considerably faster since the findings of Hasbrouck and Saar (2009). In fact, in that study 29.6% (36.9%) of all limit orders were canceled within one (two) seconds of submission. This suggests that fleeting liquidity has now become even more present in today's equity markets and confirms the importance that market participants place on speed when submitting and cancelling their order flow. The probability of an execution within an hour of submission is 26.7%, which is considerably lower than the 56.8% observed by Hasbrouck and Saar (2009). This indicates that either executions have become rarer than in the past or that execution times have now gotten considerably longer. However, in either case, the cancellation and execution results presented in Table 7 and in Figure 5 support a commonly held belief that, in recent years, markets have experienced an increase in submissions and cancellations but not in executions. Regulators and industry practitioners alike have been pointing out to the fact that limit order book events seem to be driven by other limit order book events rather than by events that alter the true value of the underlying asset. Such behavior, they argue, is against the fundamental role of financial markets that is providing and taking liquidity and does not bode well for the future of equity markets.

Given the increase in cancellations reported in Table 7 and in Figure 5, and in order to select and define as fleeting orders those orders that today resemble the most those studied in Hasbrouck and Saar (2009), we define a limit order as fleeting if it is cancelled within one second of submission rather than two. However, at times we will use the two-second classification as well, in order to make our result more directly comparable to those of Hasbrouck and Saar (2009).

We now study the submission patterns of the two types of limit orders, fleeting and not, and present in Table 8 their distribution across price levels for both one and two seconds limit: the entries in parenthesis represent the results for the two second cut off value, while the others, those for the updated value of one second. we find that 48.8% of all limit orders is cancelled within two seconds of submission while 45.5% is cancelled after more than two seconds. This implies that, overall, 94.3% of all limit orders is cancelled and that only 5.7% is executed. In contrast, Hasbrouck and Saar (2009) report a smaller overall cancellation rate of 92.3% and find that fleeting orders make up only 36.69% of the entire order flow. Moreover, they find that non-fleeting orders make up 56.24% of the total. Our results, once again, point to the fact that today's equity markets are considerably faster than before and are consistent with the results from the empirical survival analysis study reported in Table 5 and in Figure 5. Moreover, they suggest that not only execution rates have become lower than before but also that execution times have now become considerably longer. This confirms the trend of increased market activity, if measured by the number and frequency of submission and cancellation, but not if measured by the number of executions.

	Cancelled Limit Orders	
	Fleeting Orders	Non-Fleeting Orders
Percentage of Orders in Category (relative to all limit orders)	44.2% (48.8%)	50.1% (45.5%)
Relative to same-side BBO at submission		
Price Improving	10.9% (10.8%)	5% (4.8%)
At BBO	53.9% (52.2%)	33.9% (32.5%)
Top 5	27.3% (28.5%)	39.1% (39.9%)
Top 10	4.4% (4.8%)	12.9% (13.3%)
Behind Top 10	3.4% (3.7%)	9.1% (9.5%)
Total	100% (100%)	100% (100%)

Table 8: Submission Patterns of Cancelled Limit Orders. In this table, we distinguish between two types of limit orders: those that are cancelled within one second of submission ("fleeting orders") and those that are cancelled after one second ("non-fleeting orders"). However, in order to allow for a direct comparison of our findings to Hasbrouck and Saar (2009), we provide in parenthesis the results when using the two seconds cut off value. The first line of the table presents the sample average for the proportion of limit orders belonging to each category (out of the totality of all limit orders submitted). In the remaining portion of the table, we describe the submission statistics for orders belonging to each relative price level. Orders that are submitted at prices that are better (higher on the bid side and lower on the ask side) than the best price are called "Price Improving". Those that are submitted at the best price are called "At BBO"; those orders that are submitted between the second best and the fifth best relative price are called "Top 5", while those that are submitted between the sixth and tenth best relative price are called "Top 10". Finally, orders submitted beyond the tenth best relative price are called "Behind Top 10".

Another interesting difference from previous studies is how the fleeting orders distribute across relative prices. As reported in Table 8, only 10.9% of all fleeting orders are submitted at prices that are better, higher for the bid side and lower for the ask side, than the quoted best. This is much lower than the 35.39% observed in Hasbrouck and Saar (2009) and suggests that even if the signaling hypothesis is correct, it can only explain a small portion of the entire fleeting order flow. The remainder of the fleeting orders is split between those submitted at the best price, 53.9%, and those submitted deeper in the limit order book, 35.1%. Compared to the findings in Hasbrouck and Saar (2009), who report, respectively, 34.7% and 29.9% of fleeting orders at the best price or in the deeper portion of the book, it can be noted that today there is considerably more fleeting liquidity at the best price but considerably less at the most aggressive, price improving price levels. It is also interesting to observe that regardless of the cutoff value used to classify a fleeting order, there is little variation in the distribution of orders across relative prices. This suggests that both values can be used to correctly distinguish a fleeting order from a non-fleeting one (hence, a limit order submitted by a fast, impatient trader from the rest of the order flow) and that our choice of updating the cutoff value from two seconds to only one should not affect the rest of the analysis. Overall, the results in Table 8 confirm the extent of the changes in equity market and point to the fact that it is necessary to develop a more general explanation for fleeting liquidity, as the signaling hypothesis can provide only a partial explanation. The results also suggest that a cutoff value of one second can be used to correctly identify, in our data, those orders that are submitted by fast, impatient traders, and support our claim that even with this updated cut-off value we are still able to directly relate our findings to those in Hasbrouck and Saar (2009).

In conclusion, preliminary findings show that equity markets have become considerably faster in recent years. The activity in the limit order book has increased considerably but the level of executions has not: the overall proportion of limit orders that attain execution has decreased and the execution times have gotten considerably longer. We also find that today fleeting liquidity has become an even more relevant component of the entire order flow and argue that its present submission patterns call for a careful study of its possible causes since previous hypothesis are not able to fully explain this. Finally, the results in Table 5, Figure 5 and Table 8 suggest that in today's equity markets, time is even more compressed and that a cut off value of one second is more appropriate in order to identify the order flow generated by fast, impatient traders.

5 Signaling Hypothesis

We begin our empirical study of fleeting liquidity with a test of the signaling hypothesis proposed by Hasbrouck and Saar (2009) which suggests that the fleeting liquidity submitted inside the spread serves a signaling purpose. In fact, the authors argue that due to market fragmentation there is a co-ordination problem between those traders looking for hidden liquidity inside the spread and those willing to submit it but uncertain about the target venue. They believe that fast, impatient traders may want to signal to other market participants the venue in which they are searching for hidden liquidity to attract more attention to that venue. However, the authors do not test their hypothesis and just state that it is consistent with the observed limit order submission patterns. However, it is unclear whose attention are the fast traders trying to get, as normal traders would not be able to notice a signal based on limit orders that stay in the limit order book for only a few seconds while fast (silicon) traders would not need that much time in order to notice it and react to it.

We provide a test for this hypothesis by investigating the relation between fleeting order submissions and the submission of hidden orders inside the spread. Given that our data does not provide me with the information about the submission of hidden orders, we cannot directly observe whether the signal is effective in attracting them. However, since fast, impatient traders are supposed to be signaling their intention to execute against hidden orders submitted inside the spread, we argue that if the signal were effective in attracting them then we would observe an increased number of hidden executions, which are documented in our dataset by the use of message “T” (see Table 1 for details). Hence, we propose that the signaling hypothesis can be tested by studying the relation between the fleeting orders that are submitted inside the spread and the subsequent execution of hidden liquidity.

To test this hypothesis, we look at whether the present number of hidden executions, which proxies for the unobservable number of hidden orders submitted inside the spread, is affected by the past intensity of the signal, as measured by the submission of fleeting liquidity inside the spread. In order to do so, we use a linear regression approach and, given the lack of consensus about to the time that fast, impatient traders need to react to the signal, we employ an aggregation time of one second. An alternative aggregation time of two seconds, consistent with the alternative cut-off value for the definition of a fleeting order, is used as robustness checks and the results are provided in Section 5.1. Also, in order to make the results over different stocks relatable, the number of hidden executions and that of fleeting order submissions are standardized by, respectively, the number of all, visible and hidden, executions and by the number of all, fleeting and not, limit orders submitted. Moreover, we include in our analysis three control variables. The first is the lag one value of hidden executions, standardized by the number of all, visible and hidden, executions, to control for the well documented persistence in hidden order executions. In fact, when hidden liquidity is discovered, other market participants attempt to make the most of the price improving opportunity and increase the aggressiveness of their order flow. The second is the number of all, visible and hidden, executions standardized by the number of all, fleeting and not, limit order submissions. This controls for overall market activity, which, as it can be easily seen, affects the number of hidden order executions. Finally, we control for the concurrent effect that signaling for hidden liquidity has on finding/executing against it. To that end, we include in our analysis the present number of fleeting orders submitted inside the spread, standardized by the number of all, fleeting and not, limit order submissions.

When investigating the signaling effect of fleeting liquidity on a given venue, we must also include in our analysis the activity on the remaining/competing ones. This is necessary to correctly account for market wide dynamics and for the competition across trading venues. Hence, in our model, we also want to account for the past and concurrent strength of the signal and for overall market activity on each remaining, competing venue. However, in order to reduce the dimensionality of the problem without losing information, we do not include in the model a separate set of parameters for each one of the remaining three venues. Rather, we aggregate executions, hidden and visible, and submissions, fleeting and not, across venues and compute only one set of parameters that captures the state of the competing market. We refer to them as the *Super Book* parameters.

The resulting linear regression equation used in our analysis is of the form

$$\begin{aligned} HiddenExec_{i,t} = {} & \beta_0 + \beta_1^{Own} \times HiddenExec_{i,t-1} + \beta_2^{Own} \times AllExec_{i,t} \\ & + \beta_3^{Own} \times FleetingSubm_{i,t} + \beta_4^{Own} \times FleetingSubm_{i,t-1} \\ & + \beta_5^{Super} \times S.AllExec_{i,t} + \beta_6^{Super} \times S.FleetingSubm_{i,t} \\ & + \beta_7^{Super} \times S.FleetingSubm_{i,t-1} \end{aligned} \tag{1}$$

where $i = \{EDGE - A, EDGE - X, NASDAQ, BATS - Z\}$ while $t - 1$ indicates the lag one value of the variable.

If the signaling hypothesis were correct, for each venue we would expect to observe an increase in hidden order executions as a result of an increase in the lag one submission of fleeting orders inside the spread. Moreover, the control variables should all have a positive effect on hidden executions. The discovery of hidden liquidity, signaled by the execution against a hidden order, induces other market participants to increase the aggressiveness of their order flow (in an attempt to take advantage of the newly discovered price improvement opportunity) hence results in more executions against hidden orders. Higher market activity, measured by number of executions, will result mechanically in more executions against hidden orders as the incoming order flow of market orders executes against them. Finally, the concurrent strength of the signal will once again increase the number of execution against hidden orders, as the fleeting orders submitted inside the spread will execute against the hidden liquidity. For the *Super Book* parameters, we expect that market activity on the competing venues will also have a positive effect on hidden executions on the signaling venue since market activity can spill over across venues. On the other hand, the effect of the lag one strength of the signal should have a negative effect on hidden executions. In fact, if the signaling on other venues increases than the hidden order flow should be diverted to those venues and result in less hidden executions on the remaining one.

In Table 9, we report (we) the averages of the values of the coefficients obtained by running a separate regression for each of the eighty-five stocks on each trading venue and (ii) the number of stocks that have that specific coefficient significant at a 5% level. Several interesting conclusions can be made from the regression results. First, on every venue, an increase in the past intensity of the signal results in an increase in hidden order executions. This suggests that the stronger the signal generated by a fast, impatient trader, the more hidden orders are executed after allowing for a certain reaction time. In light of our decision to use hidden executions as proxy for hidden order submissions (which are not reported in our data), the regression results suggest that an increase in fleeting liquidity inside the spread positively affects the future submission of hidden orders in that same venue, consistent with the signaling hypothesis proposed in Hasbrouck and Saar (2009). Second, the average value for most of the control variables on the Own Venue are consistent with our initial predictions.

		EDGE - A		EDGE - X		NASDAQ		BATS - Z	
		Aver.	Sig. 5%	Aver.	Sig. 5%	Aver.	Sig. 5%	Aver.	Sig. 5%
Own Venue	Lag 1 Hidden Executions	0.057	58	0.039	60	0.066	71	0.077	70
	All Executions	0.139	68	0.145	75	0.255	76	0.302	75
	Fleeting Submissions	-0.002	51	0.046	50	0.133	52	0.062	55
	Lag 1 Fleeting Submissions	0.002	25	0.018	41	0.170	33	0.026	33
Super Book	All Executions	0.013	62	0.013	71	0.016	62	0.058	75
	Fleeting Submissions	0.049	56	0.017	57	0.012	58	0.046	58
	Lag 1 Fleeting Submissions	0.011	15	0.002	29	-0.003	25	0.020	28

Table 9: Regression Analysis of Signaling Hypothesis. This table presents the results for the empirical study of the relation between hidden executions and fleeting liquidity submissions inside the spread. For each of the eighty-five stocks we run the four regressions described in Equation 1, one for each of the four venues available in our dataset. For each venue, we then report the average value of each coefficient across stocks and the number of stocks that have that coefficient significant at a 5% level. The model includes a set of parameters that describe the state of the signaling venue, called the *Own Venue*, and one for the *Super Book*, that represents the competing market. *Super Book* parameters are computed using aggregations of executions (hidden and visible) and submissions (fleeting and non) across venues. For the *Own Venue* parameters, *Lag 1 Hidden Executions* is the number of hidden executions divided by the total number of executions in the previous time interval. *All Executions* is the number of all executions divided by the number of all limit order submissions; *Fleeting Submissions* is the number of fleeting limit orders submitted inside the spread divided by the total number of order submissions. *Lag 1 Fleeting Submissions* is the lagged value of *Fleeting Submissions*. The *Super Book* parameters are defined in the same way as their *Own Book* counterparts but are computed by aggregating executions and submissions across venues. Hence, for example, the Super Book parameters for EDGE-A are computed by aggregating the executions and the submissions on EDGE-X, NASDAQ and BATS-Z.

The only exception is the effect of the concurrent signal on hidden executions on EDGE-A. A possible explanation for this contradicting result is that EDGE-A is the only inverted pricing venue in our sample. As shown in Table 6, EDGE-A is the only venue that actually charges a fee when a hidden order is executed while on every other venue, a hidden order execution generates a rebate. This implies that, holding all else equal, EDGE-A is the least convenient venue for hidden orders, implying that it might not be the most attractive target venue for this type of orders. The negative sign suggests that unless a signal is generated on EDGE-A, little hidden order flow is initially sent to this venue. Hence, posting fleeting orders inside the spread will not result in concurrent executions since there would not be any hidden orders to begin with. Third, overall market activity on the *Super Book*, proxied by the total number of executions, has a positive effect on hidden executions on the signaling venue: this is consistent with our initial assumption that executions will spill over across trading venues. Finally, the result for the lag one strength of the signal on the *Super Book* are definitely surprising. In fact, only for the NASDAQ, when the strength of the signal increases on the competing venues, this results in less hidden order submissions on the NASDAQ. For every other venue, when fleeting submissions increase on the competing market, that results in more hidden executions on the signaling venue as well. In order to better understand why that could be the case, and contradict our initial assumptions, it is important to realize that the NASDAQ is present in the *Super Book* of every trading venue, with the exception of the NASDAQ's. Given the dominant role of the NASDAQ (in terms of executions and order submissions) in our dataset, this result can be explained in the context of competition between exchanges. If the priority for a trader who submits a hidden order is execution, then the NASDAQ could be the first choice when routing their orders since it has the largest proportion of executions (hence, the largest proportion of market orders that serve as potential counter parties) out of the four venues. However, if other market participants signal their willingness to trade against hidden orders on other exchanges, then BATS-Z and EDGE-X become preferable as they offer higher rebates than the NASDAQ. It is important to note that if the NASDAQ is the first choice for a trader submitting a hidden order, this also implies a higher competition to attain a superior queue placement at the top of the hidden order queue, which in turn decreases the probability of execution for the hidden order. This mechanism could explain the opposite coefficients for the past signal on the *Super Book*: when the non-NASDAQ venues signal to attract hidden orders, it results in part of the hidden order flow being diverted away from the NASDAQ, hence reducing future hidden order executions on that venue. On the other hand, when the signaling increases on the NASDAQ, hidden order submitters decide to continue to divert part of their order flow to other venues to try to increase their probability of execution by avoiding placing a hidden order in a poor position in the hidden order queue.

In conclusion, the regression results summarized in Table 9 provide empirical evidence supporting the signaling hypothesis proposed by Hasbrouck and Saar (2009). This suggests that a co-ordination problem between traders could be behind the phenomena of fleeting liquidity inside the spread confirming the authors' initial intuition: to the best of our knowledge, this is the first empirical test for their hypothesis. Moreover, our findings for the control variables on the *Own Book* are in line with our initial assumptions and are consistent with the extant literature, while those for the *Super Book* variables describe the competing nature of the four trading venues and reinforce some of the conclusions drawn from our preliminary findings in Table 2.

5.1 Robustness Check

Our choice of using a one second aggregation level for the empirical analysis of the signaling hypothesis was determined by the decision to use a one second cut off value to define a fleeting order. In fact, if we assume that a fast, impatient trader needs up to one second to analyze and react to changing market conditions, a similar reaction time might be expected when assessing, and reacting to, the intensity of the fleeting liquidity signal. However, previous studies have hypothesized a two second cut off value in order to identify the order flow generated by fast, impatient traders. Hence, as a robustness check, we have re-run the entire empirical analysis using the alternative aggregation time of two seconds and we present our results in Table 10.

		EDGE - A		EDGE - X		NASDAQ		BATS - Z	
		Aver.	Sig. 5%	Aver.	Sig. 5%	Aver.	Sig. 5%	Aver.	Sig. 5%
Own Venue	Lag 1 Hidden Executions	0.064	60	0.049	70	0.065	71	0.081	77
	All Executions	0.107	69	0.162	82	0.298	83	0.365	84
	Fleeting Submissions	-0.005	50	0.051	58	0.035	58	0.049	61
	Lag 1 Fleeting Submissions	0.002	20	0.021	39	0.186	30	0.038	33
Super Book	All Executions	0.017	60	0.017	69	0.009	62	0.071	81
	Fleeting Submissions	0.094	61	0.030	56	0.015	56	0.074	58
	Lag 1 Fleeting Submissions	0.003	25	0.002	22	-0.005	18	0.018	21

Table 10: Regression Analysis of Signaling Hypothesis: Two Second Aggregation Time. This table presents the results for the empirical study of the relation between hidden executions and fleeting liquidity submissions inside the spread. For each of the eighty-five stocks we run the four regressions described in Equation 1, one for each of the four venues available in our dataset. For each venue, we then report the average value of each coefficient across stocks and the number of stocks that have that coefficient significant at a 5% level. The model includes a set of parameters that describe the state of the signaling venue, called the *Own Venue*, and one for the *Super Book*, that represents the competing market. *Super Book* parameters are computed using aggregations of executions (hidden and visible) and submissions (fleeting and non) across venues. For the *Own Venue* parameters, *Lag 1 Hidden Executions* is the number of hidden executions divided by the total number of executions in the previous time interval. *All Executions* is the number of all executions divided by the number of all limit order submissions; *Fleeting Submissions* is the number of fleeting limit orders submitted inside the spread divided by the total number of order submissions. *Lag 1 Fleeting Submissions* is the lagged value of *Fleeting Submissions*. The *Super Book* parameters are defined in the same way as their *Own Book* counterparts but are computed by aggregating executions and submissions across venues. Hence, for example, the Super Book parameters for EDGE-A are computed by aggregating the executions and the submissions on EDGE-X, NASDAQ and BATS-Z.

All of the coefficients for the parameters describing the effect of changes on the *Own Venue* and the *Super Book* have the same direction and magnitude as those in the main study reported in Table 9. Moreover, the number of stocks that have those specific coefficients significant at a

5% is, roughly, the same across the two aggregation times. It is interesting to point out that the results for the *Super Book* parameters on the NASDAQ and, in the specific, those for the Lag one of the strength of the signal, are also consistent with those reported in Table 9. This suggest that the dynamics of the interaction between trading venues are robust to the choice of aggregation time and reinforces our conclusions on how the fleeting liquidity submitted inside the spread could serve a signaling purpose.

6 Fragmentation and Fleeting Liquidity

In this section, we elaborate on our main thesis that adverse changes in the placement of the orders submitted by fast, impatient traders are behind the phenomenon of fleeting liquidity and propose a methodology to test it. We argue that fast, impatient traders, which previous studies have associated with fleeting liquidity, represent a subgroup of traders interested in attaining a quick execution. In order to attain such quick execution, their orders need to be placed as close as possible to the top of the Limit Order Book. However, the introduction of Rule 611 from Reg-NMS, discussed in detail in Section 2.3, resulted in the aggregation of all the Limit Orders Books across exchanges, creating one large queue for each side of the market. A consequence of such aggregation is that the actual placement of a Limit Order in the overall queue to execution does not depend anymore solely from its position in the Limit Order Book of the exchange where it was submitted: rather, it will also require accounting for the outstanding depth on all the remaining venues as well. This considerably increases the difficulty of correctly assessing the placement of an outstanding limit order, given that it requires to monitor the evolution of every trading venue in the US National Market System. However, we believe that fast, impatient traders have both the financial and computational resources necessary to actively monitor all trading venues and correctly determine to position of their limit orders. Hence, we argue that fast, impatient traders react to changes that occur on all trading venues, when these result in a worse queue placement for their orders. This suggests that the determinants of fleeting liquidity should be looked for in events that occur on venues other than the one where the orders are submitted.

In the spirit of Ranaldo (2004) and Hasbrouck and Saar (2009), we look at the cancellation determinants of fleeting orders through a logistic regression[5] model. In order to test our hypothesis, we evaluate two nested regression models: in the first model, called the *partial model*, we study the cancellation determinants by using solely variables that describe the change in the state of the limit order book in which the order is submitted. In the second one, called the *full model*, we add a matching set of variables that describe the change in the state of the other trading venues as well. If our assumption is correct and fast, impatient traders do monitor the state of all trading venues, then we would expect a significant improvement in the goodness of fit from the *partial model* to the *full one*.

6.1 Variable Definition

The predictor variables (determinants) used in our study reflect our focus on understanding how fast, impatient traders react to events that alter the placement of their limit orders. Our approach is to follow the changes in the key features of the order book during the lifetime of the order, hence between its submission and ultimate cancellation or execution. To begin with,

[5] For a detailed explanation of the methodology we recommend the excellent book "Applied Logistic Regression" by Hosmer, Lemeshow and Sturdivant (3rd Edition, April 2013).

we will look at changes in the relative price of the order. The relative price of an order identifies the distance of the limit order from the best price. If, for example, a limit order is submitted at the best price, the highest bid or lowest ask, then the relative price is zero. If the limit order is submitted at the second best price, then it will have relative price equal to one, and so forth. Hence, higher values for the relative price indicate orders that are submitted at deeper portions of the limit order book, while smaller ones, orders that are closer to the market. It is important to clarify that the relative price does not tell us how many cents is the order away from the market, rather how many price levels are there between the order and the market. Hence, an order that has relative price equal to one could be several cents away from the market but still be at the next best price after the market, with no queues of orders in between the two. Changes in the relative price of an order quantify the change in the number of queues, or price levels, of limit orders that precede the outstanding order. An increase in relative price suggests that new aggressive queues have appeared in the limit order book with the potential of delaying the execution of our order as more limit orders, at more favorable prices, need to be executed before ours does.

A second variable that we use to measure the change in the placement of an outstanding limit order is the amount of shares that precede it. In fact, for an order to fall behind in the queue to execution, it is not necessary for a new, more aggressive price level to appear: if new orders are submitted in front of ours at pre-existing price levels, that could still delay the execution of the order as more liquidity needs to be cleared before our order can be executed. Hence, an increase in the volume ahead of the order, could suggest a potential delay in execution. In order to account for the differences in stock characteristics, we standardize the change in volume ahead by the total amount of shares submitted on that side of the book at the top five price levels, which allow us to compare correctly the results for different stocks.

Given the nature of our study, which aims at explaining the cancellation determinants of fleeting liquidity, we include a number of control variables that are known to affect the decision to cancel an order. The first control variable is outstanding depth imbalance, as research points to the fact that it is a good signal of future price movements. If the market moves away from the order this implies that the best Bid and Ask are moving farther away from it. This suggests that need to re-price the limit orders as the fundamental value of the assets has now changed compared to submission (with the link between the state of the order book and limit order placement strategies discussed in Cao, Hansch and Wang (2008), Hollifield, Miller and Sandas (2004), Ranaldo (2004) and Parlour (1998)). As in Cao, Hansch and Wang (2009), we measure the depth imbalance at the top five price levels of the limit order book with an own-opposite perspective rather than a buy-sell one. To expand, this measure of depth imbalance will capture the extent of the imbalance between the side on which the order is submitted and the opposite one rather than the imbalance between the buy and sell side. Such approach is necessary given that what is relevant for our analysis is to know whether prices might move away or towards an outstanding limit order rather than, simply, whether they might increase or decrease. This way of measuring the change relative to the opposite side captures the dynamics of limit order books better in our view.

The second control variable is relative spread, defined as the ratio of quoted spread over mid-price, and it proxies for changes in information asymmetry. A number of theoretical and empirical papers (Handa and Schwartz (1996), Handa, Schwartz and Tiwari (2000), Foucault

(1999), Ranaldo (2004), Hasbrouck and Saar (2002) and Cao, Hansch and Wang (2009)) find that relative spread is a significant determinant of order flow and limit order aggressiveness. In fact, if information asymmetry increases on a trading venue, than that means that there is a higher risk of trading against an informed trader. In order to account for this increase in risk, traders need to adjust their limit prices accordingly so that they become indifferent to whether their counter party is informed or not.

Finally, in light of the findings of recent empirical and theoretical models (most recently, van Kervel (2015) or Baldauf and Mollner (2014)) suggesting that cancellations and executions are used by market participants as proxies for the arrival of new information, we look at the number of limit order executions and cancellations that occur during the lifetime of a limit order. In order to account correctly for varying intensity of market activity, we standardize the number of executions and cancellations reported during the lifetime of the order by the average number of executions and cancellations that have occurred during a certain look-back period. In our study, we set the look-back period to be equal to five times the lifetime of the order, unless such period exceeds the duration of regular trading hours, in which case it extends only until the beginning of regular trading on each exchange.

Our methodology is related to the models of limit order cancellation (and execution) suggested by Hasbrouck and Saar (2009), Lo, MacKinlay and Zhang (2002), Ellul, Holden, Jain and Jennings (2007), Ranaldo (2004) or Cao, Hansch and Wang (2008) but our specification differs from the above in two key aspects. First, in these studies the predictors measure the change in the state of the limit order book between the moment of cancellation and the previous limit order book event. This essentially describes the change in the state of the market in the instant prior to the order’s cancellation. However, we believe that it is more appropriate if the predictors relate the state of the book at the moment of cancellation to that at submission as we argue that traders actively monitor each order through its lifetime. In fact, if we want to make a case for a change in relative spread as a cancellation determinant, it is more appropriate to assume that a trader would want to revise his order if the relative spread has changed considerably compared to what it was at submission, when the limit price was set, rather than what it was at the previous limit order book event. Similarly, if a trader is concerned with a potential delay in execution, as a result of an increase in relative price, it is more reasonable to assume that the limit order would be revised if the relative price changes considerably from what it was at submission rather than what it was in the previous limit order book event. We feel that defining the predictors this way has three major advantages: first, it is closer to capturing the decision process of a trader who is more likely to benchmark the present state of the market against that at submission. Second, it allows to make the most of the granularity of the data by measuring the totality of the changes that have occurred in the market during the lifetime of the order rather than just those that have occurred in the instant prior to cancellation. Finally, it reduces the risk of introducing a considerable amount of noise in our analysis by not relating order revision decisions to, potentially, noisy high frequency limit order book events. The only variables that are not defined as differences, between submission and cancellation, are the number of cancellations and executions that occur during the lifetime of the limit order. These two measures, however, by construction compare the extent of market activity during the lifetime of the order to that in a given look-back period hence already capture changing market activity during the life of the limit order.

A second major difference in our approach is the inclusion of multiple venues. To the best of our knowledge, no previous study on order flow dynamics has investigated cancellation determinants that go beyond the venue where the order is submitted. Yet, the regulatory changes brought by Reg-NMS have created interconnected markets in which limit order book events across all trading venues affect the placement of every order. Hence, we expect the dynamics of any given order book to be affected by the events occurring on the competing venues. Such setting requires the active monitoring of all trading venues, which considerably increases the complexity of the problem. Our solution to reducing the dimensionality of the problem is to rebuild, for each limit order, the limit order book in which it is submitted, which we will call the *Own Book*, and the one resulting from the aggregation of the remaining three venues, which we will call the *Super Book*. This means that, for example, given a limit order submitted on the NASDAQ, we rebuild the limit order book for that venue and another one for the order book resulting from the aggregation of all the limit orders submitted on the two EDGEs and on BATS-Z. Such setup allows me to account correctly for the presence of multiple trading venues without assigning a separate set of variables for each venue, which would result in an explosion of the number of variables to be considered. It also allows me to represent correctly the perspective of a trader who can always compare the state of his venue to that of the competing ones. Once the two limit order books are generated, we will use one set of variables to describe the changes in the state of *Own Book* and another one to describe the changes in the *Super Book*.

6.2 *Super Book* Variables Vs *Own Book* Variables

When studying the changes that occur in the state of the *Super Book* during the lifetime of the order, we need to distinguish between two different objectives. On one hand, we need to quantify the change in the placement of the order, which means that we need to study how have the *Super Book*'s volume ahead and relative price changed since submission. This can be done by simply comparing the *Super Book*'s values at cancellation to those at submission and it is the same approach that we use to quantifying the changes in order placement on the *Own Book*. On the other hand, we need to represent the perspective of the trader who actively monitors and compares the limit order book where his order is submitted to the competing market. Hence, it is important that the *Super Book* parameters that describe the changes in depth imbalance and relative spread capture the relation between the two books. To that extent we are interested in knowing how, for example, the relative spread on the *Super Book* relates to that on the *Own Book* and whether this relation has changed during the lifetime of the order. Hence, for those two *Super Book* variables we want to go beyond knowing whether the value has increased or decreased during the lifetime of the order: we want to know whether it has moved in the same direction as that on the *Own Book* and whether the change had the same magnitude. To attain this, we carefully re-define these two *Super Book* variables so to always benchmark the state of the *Super Book* to that of the *Own Book*. In the specific, we still look at the difference between cancellation and submission but instead of using the actual values of the *Super Book*'s depth imbalance and relative spread, we use their relative values, obtained by comparing the state of the *Super Book* to the *Own Book*. Hence, for example, if the *Super Book's* relative spread at submission is 0.0061 while at cancellation is 0.0065, and if the relative spread on the *Own Book* is 0.0062 at submission and 0.0064 at cancellation, then the change in *Own Book*'s relative spread is defined as $0.0064 - 0.0062 = 0.0002$ while that on the *Super Book*'s is $(0.0065 - 0.0064) - (0.0061 - 0.0062) = 0.0002$. Such set up allows me to say that, during the lifetime of the order (we) the relative spread on the *Own Book* has increased and that (ii) the relative spread on the *Super Book* has also increased and that the magnitude of the change is larger than that on the *Own Book*. Thus, we can say that, compared to the *Own Book*, the relative spread on the *Super Book* is higher at cancellation than what it was at submission. In general, for the *Super Book* variables, a positive sign means that that *Super Book* characteristic has increased during the lifetime of the order, when compared to the *Own Book*. Table 11 summarizes the definitions of all the predictors in our logistic regression.

Own Book Variables	
Notation	Variable Definition
ΔRelative Price	$= Relative\ Price_{Cancellation} - Relative\ Price_{Submission}$
ΔVolume Ahead	$= Volume\ Ahead_{Cancellation} - Volume\ Ahead_{Submission}$, where $Volume\ Ahead_i = \frac{Volume}{Top5}$ and where *Volume* is number of shares in front of order, while *Top5* is total number of shares at top 5 price levels on same side of book.
ΔDepthImbalance	$= DepthImbalance_{Cancellation} - DepthImbalance_{Submission}$, where $DepthImbalance_i = \frac{Top5_{Own} - Top5_{Opposite}}{Top5_{Own} + Top5_{Opposite}} * 10$ *and* $Top5_{Own\ (Opposite)}$ *is the aggregate depth at top 5 price levels on the same (opposite) side of the book*
ΔRelativeSpread	$= RelativeSpread_{Cancellation} - RelativeSpread_{Submission}$, where $RelativeSpread_i = 2 * \frac{(\text{Best Ask}_i - \text{Best Bid}_i)}{(\text{Best Ask}_i + \text{Best Bid}_i)}$
Cancellations	Number of cancellations occurring on the own Book during the lifetime of the limit order standardized by the average number of cancellations that have occurred in a look-back period (equal to five times the lifetime of the order or until the beginning of regular trading).
Executions	Number of executions occurring on the own Book during the lifetime of the limit order standardized by the average number of executions that have occurred in a look-back period (equal to five times the lifetime of the order or until the beginning of regular trading).

Super Book Variables	
Notation	Variable Definition
ΔS.Relative Price	$= S.Relative\ Price_{Cancellation} - S.Relative\ Price_{Submission}$
ΔS.Volume Ahead	$= S.Volume\ Ahead_{Cancellation} - S.Volume\ Ahead_{Submission}$, where $S.Volume\ Ahead_i = \frac{S.Volume}{S.Top5}$ and where *S.Volume* is number of shares in front of order on *Super Book*, while *S.Top5* is total number of shares at top 5 price levels on same side of *Super Book.*
ΔS.DepthImbalance	$= (S.DepthImbalance_{Cancellation} - DepthImbalance_{Cancellation}) - (S.DepthImbalance_{Submission} - DepthImbalance_{Submission})$, where $S.DepthImbalance_i = \frac{S.Top5_{Own} - S.Top5_{Opposite}}{S.Top5_{Own} + S.Top5_{Opposite}} * 10$ *and* $S.Top5_{Own\ (Opposite)}$ *is the aggregate depth at top 5 price levels on the same (opposite) side of the book*
ΔS.RelativeSpread	$= (S.RelativeSpread_{Cancellation} - RelativeSpread_{Cancellation}) - (S.RelativeSpread_{Submission} - RelativeSpread_{Submission})$, where $S.RelativeSpread_i = 2 * \frac{(S.\text{Best Ask}_i - S.\text{Best Bid}_i)}{(S.\text{Best Ask}_i + S.\text{Best Bid}_i)}$
S.Cancellations	Number of cancellations occurring on the Super Book during the lifetime of the limit order standardized by the average number of cancellations that have occurred in a look-back period (equal to five times the lifetime of the order or until the beginning of regular trading).
S.Executions	Number of executions occurring on the Super Book during the lifetime of the limit order standardized by the average number of executions that have occurred in a look-back period (equal to five times the lifetime of the order or until the beginning of regular trading).

Table 11: Definition of Predictors for Study on Cancellation Determinants. We define two matching sets of predictors: one, to characterize the change in the state of the limit order book in which the order is submitted and another one for the change in the state of the competing market, resulting from the aggregation of the remaining three venues and called the *Super Book*. *Δ* stands for the difference in the value of the variable between the moment of cancellation and the moment of submission of the order while the *S.* prefix refers to the variables describing the state of the *Super Book*. *Relative Price* quantifies the number of price levels between the limit order and the market. A value of 0, means that the order is submitted at the most aggressive price, a value of 1 means that the order is at the second best price and there is only one queue of orders in front of it, and so forth. *Volume Ahead* quantifies the number of shares that belong to limit orders that are in front of the order. This includes orders that are at the same price but were submitted at an earlier time, and all orders at better prices. *Depth Imbalance* captures the extent of the imbalance in the outstanding depth between the side of the order and the opposite one. Using an Own/Opposite approach instead of a Bid/Ask one to study depth imbalance allows to represent correctly our intention to use this parameter to understand if traders expect prices to move away or towards the order rather than, simply, if prices will increase or decrease. *Relative Spread* proxies for the level of information asymmetry and is defined as the ratio of quotes spread and mid-price. Finally, *Cancellations* and *Executions* measure changes in market activity, which proxy for the arrival of new information, by looking at the number of cancellations and executions that occur during the lifetime of the order. To account for different levels of market activity across stocks, these two measures are standardized by the number of cancellations and executions that occur during a look back period.

6.3 Data Cleaning

Given our focus on the determinants of cancellation of outstanding limit orders, after rebuilding the limit order books we exclude from the logistic regression analysis all the messages that indicate the submission of a new limit order. This serves a twofold purpose: first, it allows to reduce considerably the sample size, second it mitigates the risk of potential biases due to the violation of the independence assumption of the observations. In fact, logistic regression requires each observation to be independent yet, dependence may occur if the same order has an entry (submission) and an exit (by cancellation or execution) message.

We also exclude from the analysis all orders submitted more than ten price levels away from the best bid or ask of each order book, since there is very limited theoretical framework available to explain the dynamics of orders submitted this far away from the market. Based on the distribution of fleeting order flow across price levels, reported in Table 8, we find such decision affects only a very small portion of orders but should allow to eliminate possible sources of pure noise in the data.

Finally, given the substantial difference between regular and after-hours trading, we only include in our study those limit orders that are submitted during regular trading hours and when markets are not locked or crossed.

In conclusion, the sample used in the logistic regression analysis is comprised of only those messages that describe the execution or the cancellation of those limit orders that are submitted within ten relative price levels from the best price on each order book, during regular trading hours when markets are not locked or crossed.

6.4 Preliminary Analysis

Before investigating the determinants of limit order cancellations, we look at the average values of the predictors for the two possible outcomes: execution and cancellation. In fact, if the average values of the predictors are the same across the two outcomes that would cast some doubt on the ability of our model to discriminate between them. The results are presented in Table 12.

Starting with fleeting orders, we see that there appears to be a significant difference in the average values of the predictors between the two outcomes. On the *Own Book,* fleeting orders that attain execution do so after clearing more volume ahead of them than those that are cancelled. In fact, the average value for the change in the standardized volume ahead for orders that are executed and cancelled are, respectively, -0.196 and -0.018. The negative signs suggest that in both cases there is less volume ahead of the order when it leaves the book than at submission, which must be true in the case of an execution given that there can be no volume ahead of the order as, at that moment, it must be at the top of the queue. On the other hand, if we look at the change in relative price, we see that fleeting orders that are executed do so after reducing their distance from the market while those that are cancelled are actually further away at cancellation than they were at submission. This result manages to reconcile the apparently contradicting finding from the change in volume ahead: fleeting orders that are cancelled might clear some volume ahead of them but still end up falling behind in the execution queue, as prices appear to have moved away from them. These results suggest that a cancellation determinant for a fleeting order could be the lack of progress in its path to execution, consistent with the assumption that fleeting orders are submitted by impatient traders who prioritize a quick execution. A similar conclusion appears to be substantiated by the results for the number of executions and cancellations on the *Own Book*: compared to orders that are executed, those that are cancelled exist during periods of lower (higher) market activity when measured by the number of executions (cancellations) that occur during the lifetime of the order. Given that limit orders can be executed only if they are matched by incoming Market Orders, a less active market, as measured by the number of executions, might suggest to an impatient trader a delay in execution, causing him to cancel his orders.

Similar conclusions hold if we compare the average values for the two outcomes of fleeting orders by looking at the variables that describe the changes in the *Super Book*. However, an interesting remark can be made if we focus on volume ahead: fleeting orders that are cancelled have, an average, more volume ahead of them at cancellation than they had at submission. This is not the case for those that attain execution. This information provides a more complete picture of the changes that occur in the market during the lifetime of the order and might help better understand the decision to cancel a fleeting order. In fact, based on the variables that describe the state of the *Own Book*, it appeared that cancelled fleeing orders cleared a considerable amount of volume ahead of them before cancellation. However, if we now look at the broader picture and look at the changes that have occurred in the *Super Book*, we see that these orders have actually fallen behind in the overall queue since there is now more volume in front of them, even though this volume is on other venues.

A first conclusion that we can make based on the results in Table 12 is that our set of variables can accurately capture the key differences between fleeting orders that are executed and those that are cancelled. Moreover, we also find first evidence to support the claim that only by accounting for the changes in the *Super Book* we are able to explain correctly the dynamics of fleeting liquidity.

	Fleeting Orders		Non-Fleeting Orders	
	Execution	Cancellation	Execution	Cancellation
Own Book				
ΔRelative Price	-0.021	0.102	-0.258	0.242
ΔVolume Ahead	-0.196	-0.018	-0.277	0.002
ΔDepth Imbalance	-0.730	-0.342	-0.500	-0.208
ΔRelative Spread	-0.006	-0.003	-0.006	-0.003
Number of Exec.	0.863	0.166	1.369	0.778
Number of Canc.	2.006	2.285	1.236	1.398
Super Book				
ΔS.Relative Price	-0.019	0.065	-0.275	0.188
ΔS.Volume Ahead	-0.002	0.015	-0.056	0.072
ΔS.Depth Imbalance	0.615	0.277	0.376	0.144
ΔS.Relative Spread	0.007	0.004	0.005	0.003
S.Number of Exec.	1.263	0.603	3.426	3.067
S.Number of Canc.	2.823	2.801	1.340	1.562

Table 12: Average Values of Predictors Across Order Types and Outcomes. In the upper portion, we present the average values for the predictors describing the state of the order book in which the limit order resides (the *Own Book*) while in the lower portion we present those for the book resulting from the aggregation of the three remaining competing venues (the *Super Book*). Δ(S.)Relative Price refers to the change in the *Own* (Super) book's relative price between the moment of cancellation and submission; Δ(S.)Volume Ahead refers to the change in the number of shares preceding the order on the *Own* (Super) book between the moment of cancellation and submission; Δ(S.) Depth Imbalance refers to the (relative) change in the own (Super) book's depth imbalance between the moment of cancellation and submission; Δ(S.) Relative Spread refers to the (relative) change in the own (Super) book's relative spread between the moment of cancellation and submission; (S.) Number of Exec, and (S.) Number of Canc. refer to the average number of executions and cancellations, respectively, that have occurred on the own (Super) book during the lifetime of the order.

Another interesting comparison that can be made from Table 12 is the one between fleeting and non-fleeting orders. For both executions and cancellations there are significant differences between the two order types, confirming previous findings suggesting that a distinction between the two is, in fact, meaningful. If we look at the average change in the Relative Spread and Volume Ahead on both books, we see that fleeting orders, during their lifetime, change their position in the queue by much less than non-fleeting ones. This is suggests that these orders must have been priced more aggressively than their non-fleeting counterparts which is consistent with our previous findings reported in Table 8 and with the impatient nature of the traders who submitted them. Also, if we look at the changes in market activity, we see that fleeting orders appear to exist during periods of lower activity when measured by the number of executions, and heightened activity when measured by the number of cancellations. These differences indicate that fast, impatient traders could employ very different trading strategies from their slower and more patient counterparts, suggesting that the timing of events could be an important differentiator between the two. These results support the case made by Hasbrouck and Saar (2009) that it is possible to distinguishing between the two types of traders based on the order flow they generate and bodes well for our future analysis.

Finally, an interesting observation can be made when comparing the average values of depth imbalance and relative spread on the *Own Book* to those on the *Super Book*. For these two predictors, the averages have similar magnitudes but have opposite signs. A possible explanation for this result could lay in the fact that these two *Super Book* variables are built to benchmark the changes in the *Super Book* against those on the *Own Book*. Values that are close in magnitude but with opposite signs, suggest that the *Super Book* has remained fairly stable during the lifetime of the order. In fact, given a certain change on the *Own Book*, if no change at all were to occur on the *Super Book*, then the *Super Book* variables would have exactly the same values as those on the *Own Book* but with opposite sign. This would have been the case since, relative to the *Own Book*, the *Super Book* would have changed by the same amount but in the opposite direction. Given that in our case the values have opposite signs but are also slightly smaller, this suggests that the *Super Book* has also changed during the lifetime of the order but, albeit, by a considerably smaller degree. This is not surprising given that the *Super Book* is the result of the aggregation of the three remaining trading venues, which should give a much more stable limit order book than the one specific to the venue where the order is submitted.

In conclusion, the results in Table 12 confirm that, in line with the findings of Hasbrouck and Saar (2009), a distinction between fleeting and non-fleeting orders is meaningful. Moreover, the results point to the fact that the timing of the limit order book events could be an important differentiator between order types and suggests that fast and slow traders could be employing very different trading strategies. Finally, the differences between fleeting orders that attain execution and those that are cancelled, indicate that our variables are able to capture the critical differences between the two outcomes and can be used to study the cancellation determinants for these orders.

6.5 Empirical Results

We begin now our study of the cancellation determinants of fleeting liquidity. The data that is analyzed here is in the form of panels with panel level information not expected to vary much over time. Therefore, we follow here the standard econometric approach. First, we construct models that account for the variation over time and the resulting estimates of the model parameters are then related to panel level information. Thus, the time-varying and cross-sectional aspects are labeled as within-variation and between-variation. For each stock, the influence of trading characteristics is related to cancellation probabilities, which we call studying with-in variation and the estimates for each stock are then regressed on stock-level variables such as market capitalization or price which are labeled as studying between variation.

6.5.1 Within Variation

In this section we test the hypothesis that adverse changes in order placement are behind the phenomenon of fleeting liquidity. Recall our reasoning that fast, impatient traders are interested in attaining a quick execution and that in order to attain it, they need to place their orders as close as possible to the top of the Limit Order Book. However, the introduction of Rule 611 from Reg-NMS, resulted in the aggregation of all the Limit Orders Books across exchanges which means that the placement of a Limit Order in the overall queue to execution depends now from the shape of every Limit Order Book. Hence, we argue that fast, impatient traders react to changes that occur on all trading venues, when these result in a worse queue placement for their outstanding orders.

To test this hypothesis, we carry out an analysis of the cancellation determinants of fleeting orders. To be specific, we investigate how limit order book events, such as changes in the relative price, volume ahead, outstanding depth imbalance or relative spread, and changing market condition, proxied by the number of executions and cancellations, affect the probability of a cancellation. In our study, we use two logistic regression models: in the first one, called the *partial model*, we use only the variables that describe the state of the limit order book in which the order is submitted. In the second one, called the *full model*, we add a matching set of variables describing the state of the *Super Book*.

If we denote with π the probability of observing the cancellation of a limit order, the logistic regression equation for the *partial model* is:

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1^{Own} \times \Delta Relative\ Price + \beta_2^{Own} \times \Delta Volume\ Ahead$$
$$+\beta_3^{Own} \times \Delta Depth\ Imbalance + \beta_4^{Own} \times \Delta Relative\ Spread$$
$$+ \beta_5^{Own} \times Number\ of\ Exec. + \beta_6^{Own} \times Number\ of\ Canc. \qquad (2)$$

while the equation for the *full model* builds on the partial one by adding the set of predictors describing the state of the *Super Book*. Hence is given by:

$$log\left(\frac{\pi}{1-\pi}\right) = partial\ model + \beta_1^{Super} \times \Delta S. Relative\ Price + \beta_2^{Super} \times \Delta S. Volume\ Ahead$$
$$+\beta_3^{Super} \times \Delta S. Depth\ Imbalance + \beta_4^{Super} \times \Delta S. Relative\ Spread$$
$$+\beta_5^{Super} \times S. Number\ of\ Exec. + \beta_6^{Super} \times S. Number\ of\ Canc. \qquad (3)$$

If our assumption is correct that fast, impatient traders actively monitor all trading venues in order to correctly determine to position of their limit orders, then we would expect a significant improvement in the goodness of fit of the *full model* over the *partial one*. Moreover, given the choice of predictors, and in light of our objective of determining the effect of changes in queue placement on the cancellation probability of a fleeting order, we can hypothesize the following effect of our predictors on the cancellation probability:

1. An increase in relative price on either the *Own Book* or the *Super Book*, should increase the probability of observing a cancellation since it would signal that the order is falling behind as a result of new, more aggressive orders appearing in the market.
2. An increase in the volume ahead on either the *Own Book* or the *Super Book*, should increase the probability of observing a cancellation. In fact, this would signal that new orders are appearing at more aggressive, but pre-existing, prices levels increasing the outstanding depth that needs to be executed before the order does.

3. An increase in depth imbalance on the *Own Book* and on the *Super Book* should increase the probability of observing a cancellation since it could signal future adverse price movements.
4. An increase in relative spread on the *Own Book* should increase the probability of observing a cancellation since traders would have to re-price their orders in order to account for a higher risk of trading against an informed trader. On the other hand, an increase on the *Super Book* should have the opposite effect since it would make the competing venues worse than the one in which the orders resides making it less likely that a fast trader would want to cancel his order in order to resubmit it on another, competing venue.
5. An increase in market activity (proxied by the number of executions and cancellations) on both the *Own Book* and the *Super Book* should increase the cancellation probability since market participants would have to revise the limit prices of their outstanding orders to account for the arrival of new information.

In Table 13, we report the median values of the regression coefficients obtained from running eighty-five individual logistic regressions, one per stock. We also report the number of stocks that have each predictor significant at a 5% level. Several interesting remarks can be made from the results of the *partial model*. First and foremost, we see that for, respectively, 77 and 60 stocks, an increase in the relative price or volume ahead increases the probability of observing a cancellation. This result supports our hypothesis that when the placement of an order deteriorates as a result of the arrival of new orders, than this increases the probability that this order will be cancelled by the fast, impatient trader. Second, we see that an increase in the depth imbalance also increases the likelihood of a cancellation. This is consistent with the extant literature (Cao, Hansch and Wang (2008), Hollifield, Miller and Sandas (2004), Ranaldo (2004) and Parlour (1998)) on the use of depth imbalance as a signal of future price movements and confirms that a fast, impatient trader will cancel his orders if he believes that prices might be moving away from him. Finally, we find that an increase in the relative spread positively affects the probability of a cancellation. This is in line with previous studies (Ellul, Holden, Jain and Jennings (2007) or Ranaldo (2004)) and is consistent with the notion that relative spread is used, by market participants, as proxy for the level of information asymmetry in the order book.

		P(Cancellation | Fleeting Order)			
		Coefficient	Sig. 5%	Coefficient	Sig. 5%
Own Venue Variables	Δ.Relative Price	0.5731	77	0.7899	77
	Δ.Volume Ahead	0.0001	60	0.0001	49
	Δ. Depth Imbalance	0.0001	71	0.0001	63
	Δ. Relative Spread	16.9070	70	5.3477	51
	Number of Exec.	-0.2046	83	-0.2030	81
	Number of Canc.	0.0191	55	0.0215	57
Super-Book Variables	ΔS.Relative Price			0.9463	63
	ΔS.Volume Ahead			0.0001	39
	ΔS.Depth Imbalance			0.0001	42
	ΔS.Relative Spread			-19.2683	51
	S.Number of Exec.			-0.0248	72
	S.Number of Canc.			0.0002	56

Table 13: Logistic Regression Results on Cancellation Determinants. In the first two columns, we report the results for the partial model which uses solely the predictors that define the state of the venue on which the order is submitted; in the next two columns, we present the results for the full model which includes the predictors describing the state of the *Super Book* as well. For each model, we present the median value of the coefficients computed across all eighty-five stocks and the number of stocks that have that predictors significant at a 5% level. Δ(S.)Relative Price refers to the change in the relative price on the Own (Super) Book between the moment of cancellation and submission; Δ(S.)Volume Ahead refers to the change in the volume in front of the order on the Own (Super) Book between the moment of cancellation and submission; Δ(S.) Depth Imbalance refers to the (relative) change in the own (Super) book's depth imbalance between the moment of cancellation and submission; Δ(S.) Relative Spread refers to the (relative) change in the own (Super) book's relative spread between the moment of cancellation and submission; (S.) Number of Exec, and (S.) Number of Canc. refer to the average number of executions and cancellations, respectively, that have occurred on the own (Super) book during the lifetime of the order.

What is surprising, however, are the two coefficients for the effect of changes in the number of executions and cancellations on the probability of cancellation. To be specific, it is surprising that more executions lead to a decrease in the probability of cancellation as significant theoretical and empirical evidence (starting from Easley and O'Hara (1992) and, more recently, van Kervel (2015)) suggests that market participants associate intense trading with a higher risk of information asymmetry. The result is even more puzzling if compared to the fact that a more active market, when measured by the number of cancellations, has the opposite, and expected, effect. A possible explanation for the negative coefficient of executions is that, from the perspective of an impatient trader who is waiting for execution, an increase in the arrival of market orders is actually desirable as it increases the flow of those orders that can be matched against his outstanding order. Hence, an increase in executions can lead an impatient trader to stay put and wait for execution as it can imply a short execution time. Fast traders, on the other hand, can interpret cancellations as a signal that other market participants are revising their limit prices in anticipation of the arrival of new information, especially if such cancellations are made by other fast, impatient traders. A good way of testing this hypothesis would be to make a distinction in our analysis between fleeting and non-fleeting cancellations. In any case, interpreting the opposing coefficients of executions and cancellations on the *Own Book* presents a number of challenges and suggests that it might be necessary to revise the definition of the two proxies in order to disentangle the conflicting results.

The results for the *full model*, obtained by adding the second set of regressors that describes the dynamics of the competing venues, present some interesting findings. First, the addition of the second set of predictors does not affect the sign or magnitude of the predictors from the

Own Book. This is definitely very encouraging given that our argument is that monitoring the *Super Book* allows the fast, impatient traders to assess more accurately the position of an order rather than provide a completely different perspective on Limit Order Book dynamics. Hence, we find encouraging that the dynamics of the *Own Book* are correctly modeled by both the partial and the full model. Second, we see that an increase in relative price or volume ahead on the *Super Book* has the same effect on the probability of cancellation as an increase on the *Own Book*. This is definitely the most interesting result as it provides strong empirical evidence to support our claim that because of the implementation of Reg-NMS, and of Rule 611 in the specific, fast and impatient traders cancel their orders as a result of events that occur on other venues, when such events negatively affect the overall queue placement of their orders. In fact, given that Reg-NMS aggregates all the limit order books across exchanges, the effect of a change in relative price or volume ahead on the *Super Book* has to be the same as a similar change on the *Own Book*. Moreover, this result supports the hypothesis that fast, impatient traders actively monitor all trading venues after order submission since only by doing so they are able to correctly asses the placement of their orders.

Third, the direction of the coefficient for relative spread is in line with our stated hypothesis about the competing nature between venues, which is a result from our initial assumption that fast, impatient traders compare the state of their target venue to that of the rest of the market. In fact, we notice that when a trader observes an increase in relative spread on the competing venues, compared to that on his own venue, he is less likely to cancel his order as his venue is "better" than the competing ones in terms of the level of information asymmetry. This means that there is no advantage in revising and resubmitting the order to another venue as it would be exposed to a higher risk of being picked off by an informed trader.

Fourth, the effect of an increase in executions and cancellations on the *Super Book* is the same one as that on the *Own Book*. Once again, we see that increased market activity, when proxied by the number of executions (cancellations), results in a lower (higher) probability of cancellation. The result for the *Super Book* cancellations is consistent with the idea that market participants perceive them as events that carry new information regardless of where they occur and will adjust their outstanding limit orders accordingly. On the other hand, the fact that executions occurring on the *Super Book* still have a negative effect on the probability of observing a cancellation on another venue, could signify that fast, impatient traders anticipate that the increased market activity on the overall market will, eventually, spill over to their own venue as well. This, in turn, would lead to a higher probability of execution due to the increased market order flow.

Finally, the result for the *Super Book*'s depth imbalance is also consistent with our initial assumptions. In fact, in light of the relation between depth imbalance and future price movements, we expected an increase in the probability of cancellation when the depth imbalance on the same side of the *Super Book* increases, as that could signal the need to chase a running price.

In conclusion, our results for the study of the cancellation determinants are consistent with our initial assumptions. We find evidence that fast, impatient traders actively monitor all trading venues in order to determine correctly the placement of their orders and that they react to events that negatively affect the placement of their orders, regardless of the venue. Moreover, our results suggest that, after order submission, fast, impatient traders actively compare the state of their target venue, to that of the competing market, and evaluate the possibility of resubmitting the order onto a better venue.

We now focus on the goodness of fit of the two models. If our assumption were correct that fast, impatient traders relay on events occurring on all major trading venues in order to assess the correct placement of an order, then we would expect a considerable improvement in the goodness of fit of the model when switching from the partial to the full one.

In the first two columns of Table 14, we report the average value, and the standard deviation, for the goodness of fit of the full model for the eighty-five stocks in our sample. In the following two columns we report the average change, and the standard deviation of the changes, in the goodness of fit when we switch from the partial model to the full one (hence, the change in goodness of fit when we add the second set of regressors describing the state of the *Super Book*). First, the average goodness of fit of the full model indicates that it is able to describe fairly well the probability of observing a fleeting cancellation. However, the rather high values for the standard deviations also suggest that the model fits considerably better some stocks rather than others, indicating that fast, impatient traders could prefer some stocks to others. If we look now at the changes in the goodness of fit resulting from the addition of the second sets of variables we see that, on average, there is a considerable improvement in the fit of the model. Most measures indicate an 18% improvement in predicting the probability of observing a cancellation when we add the information generated on the competing venues. This result, once again, strongly supports our initial research hypothesis that fast, impatient traders, actively monitor all trading venues in order to determine correctly the placement of their orders and that they react to events that negatively affect the placement of their orders, regardless of the venue. However, we feel that such conclusion is warranted not only because of the substantial improvement in the goodness of fit of the model, but also because of the number of stocks in our sample that have the events that occur in the *Super Book* as significant cancellation determinants, as previously shown in Table 13.

	Goodness of Fit (Full model)		Change in Goodness of Fit (Partial => Full)	
	Mean	Std. Dev	Mean	Std. Dev
McFadden	7.9%	7.6%	18%	27.4%
Cox Snell	2.5%	2.9%	17.9%	27.3%
Nagelkerke	9%	8.6%	17.9%	27.3%
McKelvey Zavoina	16.1%	16.7%	39.2%	111%
Effron	3.1%	5.2%	24.8%	62.8%
Corrected AIC	337,927	568,073	-1%	1.4%
Residual Deviance	337,901	568,073	-1%	1.3%

Table 14: Summary of Goodness of Fit of Partial and Full Models. In the first two columns, we present the mean and standard deviation of the distribution of the goodness of fit for the two models for the eighty-five stocks in our sample. In the next two columns, we present the mean and standard deviation of the change in goodness of fit, when switching from the partial to the full mode, for the eighty-five stocks in our sample.

In the final step of our analysis, we test the initial claim, made in section 6, that it is more appropriate to compute the predictors as differences in the state of the limit order book between cancellation and submission rather than as differences between cancellation and the previous limit order book event (as done so far in the literature). In order to test this claim we re-run the entire analysis on the cancellation determinants of fleeting liquidity using the traditional way of computing the predictors. If our claim were correct, we would expect such model to have a considerably worse fit than the initial one. The results in Table 15, report the change in goodness of fit, for both the partial and full model, when switching from the old way of computing the predictors to the new one proposed in our study. We find that for both models there is a considerable improvement when switching from the traditional way to the new one. In fact, with the exception of one measure, all others show a considerable improvement in fit, confirming our initial claim that to correctly capture the perspective of a fast trader, we need to benchmark the changes in the limit order book against the state at submission rather that the state at the previous limit order book event.

Change in Fit from Old to New (our) Variables		
	Mean	
	Partial	Full
McFadden	67%	57%
Cox Snell	66%	57%
Nagelkerke	66%	57%
McKelvey Zavoina	-14%	-14%
Effron	19%	53%
Corrected AIC	-2%	-2%
Residual Deviance	-2%	-2%

Table 15: Effect of definition of predictors on goodness of fit. The table presents the average change in the goodness of fit of both the partial and full models, when the analysis on the cancellation determinants is performed with our novel approach to computing the values of the predictors. We propose that the cancellation determinants should be computed by comparing the state of the limit order book at cancellation to that at submission rather than to the state in the previous limit order book event, as used in the extant literature.

In conclusion, the study on the goodness of fit of the two models provides strong empirical evidence for our initial assumptions. Switching from the partial model to the full one, allows to considerably improve the ability of the model to predict the cancellation of a fleeting order. This suggests that, to understand correctly fleeting liquidity dynamics it is necessary to monitor the changes that occur on all trading venues and interpret them in terms of their effect on the placement of the order. This supports our claim that fast, impatient traders actively monitor all trading venues after order submission to correctly asses the path to execution of their orders. Such approach has become necessary because of the implementation of Rule 611, which extends price priority across all trading venues in the National Market System, de facto, aggregating in one large queue the liquidity submitted across all trading venues. Finally, the results for the change in fit when switching from the old way of computing the predictors to the new one proposed in our study, supports the claim that such approach is necessary to better replicate the perspective of a fast, impatient trader and to limit the amount of noise introduced in the measures.

6.5.2 Between Variation

As the results in Table 13 and Table 14 suggest, there are significant stock level differences in the estimated coefficients. We find that not all stocks are sensitive to all predictors and that not all of them react in the same way to the addition of the second set of regressors describing the state of the *Super Book*. We believe that this variation can be explained, in part, by stock level characteristics. In fact, if our model is supposed to describe the actions of fast, impatient traders we would expect it to work best for those stocks that are preferred by such traders. Anecdotal evidence suggests that fast, impatient traders prefer stocks that are (we) very liquid, (ii) actively traded, (iii) with low risk and (iiii) low price. In fact, the impatient nature of the traders implies that they would prefer to trade in stocks that allow them to trade large quantities in a short period of time. On the other hand, their fast nature, hence their speed advantage over the other market participants, implies that they can successfully engage in trading strategies based on making a small profit on each trade, sometimes as small as the rebate paid by the trading venue, as long as such trade is repeated multiple times. Hence, they would prefer as little uncertainty as possible and they would rather trade in cheap stocks since rebates are paid per number of shares traded and not per dollar value. We will use market capitalization to proxy for the level of liquidity of a stock, turnover as a measure of trading activity, volatility and market beta as measures of risk along with the price level of a stock. The values of these stock characteristics are computed using monthly data from CRSP and COMPUSTAT for years 2009 and 2010.

We will carry out this second stage analysis by looking at the effect of stock level characteristics on the logistic regression coefficients since this should allow me to understand the mediation effect that these characteristics have on the role of the predictors as cancellation determinants. Given that each regression coefficient in Equation 3 has an associated standard error, we need to incorporate this in the analysis and use a weighted linear regression approach. The resulting second level regression is of the form

$$\beta_i^{Own/Super} = \beta_0 + \beta_1 \times log(MCAP_i) + \beta_2 \times log(Price_i) + \beta_3 \times Turnover_i + \beta_4 \times Volatility_i + \beta_5 \times Beta_i \qquad (4)$$

for $i = 1, \dots, 85$ and with weights $w_i = \frac{1}{s_i^2}$.

If our assumptions about the type of stocks preferred by fast, impatient traders are indeed correct, we expect to find that stock level characteristics are significant determinants of all regression coefficients and that they can explain a substantial portion of their variation.

From the results in Table 16 we can see that stock level characteristics are, indeed, significant for a number of regressors. Market capitalization is a significant, and mostly positive, determinant of almost every regression coefficient. The only exceptions are Number of Cancellations on both the *Own* and the *Super Book* and the Volume Ahead on the *Own Book*. What this means is that for stocks with high market capitalization, whether a fast, impatient trader decides to cancel an order or not, will mostly depend from changes in Relative Price, Depth Imbalance and Relative Spread on the *Own Book* and in Volume Ahead, Depth Imbalance and Number of Executions on the *Super Book*. On the other hand, Stock Price is also a significant, but mostly negative, determinant of almost every coefficient, with the exception of Number of Executions and Number of Cancellations on the *Super Book*. In this case, the results suggest that for low priced stocks, the decision to cancel a fleeting order will be driven by changes in Relative Price, Volume Ahead, Depth Imbalance, Relative Spread and Number of Cancellations on the *Own Book* and by changes in Volume Ahead on the *Super Book*. Turnover is a significant determinant of Relative Price, Depth Imbalance, Relative Spread and Number of Cancellations on the *Own Book* but is not significant for any of the *Super Book* coefficients. Surprisingly, market beta is significant for only Depth Imbalance, on both books, and for the Number of Executions on the *Super Book*, while Volatility is significant for Relative Price and Number of Cancellations on the *Own Book* and Depth Imbalance and Number of Executions on the *Super Book*.

If we now look at the extent of the variation of each coefficient explained by stock characteristics we see that it varies considerably ranging from -1.6% of the Number of Cancellations on the *Super Book* to a surprising 96.8% for the Depth Imbalance on the *Own Book*. Overall, the adjusted R-squares are pretty high for most of the coefficients and confirm that stock level characteristics can explain a good portion of the variations observed in the study of the cancellation determinants reported in Table 13 and in Table 14.

In conclusion, the results for the second stage analysis reported in Table 16 provide empirical evidence to support the claim that stock level characteristics can explain some of the in-sample variation observed in the study of fleeting liquidity determinants.

		log(MCAP)	Turnover	Volatility	beta	log(Price)	$Adj\text{-}R^2$
Own Venue Variables	Δ.Relative Price	0.004*	0.03*	-0.06268*	-0.005	-0.018*	56.3%
	Δ.Volume Ahead	0.00001	0.00001	-0.00001	-0.00001	-0.00001*	18.7%
	Δ.Depth Imbalance	0.00001*	0.00007*	-0.00002	-0.0001*	-0.00001*	96.8%
	Δ.Relative Spread	8.164*	45.025*	-90.576	16.458	-18.32*	26.0%
	Number of Exec.	-0.014*	-0.025	0.011	0.006	0.02*	22.5%
	Number of Canc.	0.00005	0.0009*	-0.005*	0.0005	-0.0005*	31.0%
Super-Book Variables	ΔS.Relative Price	-0.006*	-0.011	0.004	-0.0006	0.011*	56.7%
	ΔS.Volume Ahead	0.00001*	0.00001	0.00001	0.00001	-0.00001*	58.8%
	ΔS.Depth Imbalance	0.00001*	-0.00001	0.00002*	-0.0001*	0.000002*	74.6%
	ΔS.Relative Spread	-7.622*	-8.935	27.794	-8.296	8.34*	41.4%
	S.Number of Exec.	0.0009*	-0.003	0.006*	0.004*	0.0005	36.5%
	S.Number of Canc.	-0.00001	-0.0001	0.0004	-0.0001	0.00001	-1.6%

Table 16: Effect of Stock Characteristics on Logistic Regression Coefficients. The table reports the results of the second stage regression analysis, used to investigate the assumption that stock level characteristics can explain the variation observed in the study of the cancellation determinants of fleeting liquidity. Given that each regression coefficient from the study of the cancellation determinants has an associated standard error, we need to incorporate this in the second stage analysis and use a weighted linear regression approach. Equation 4 is the resulting second level regression. The values for the stock characteristics are computed using monthly data from CRSP and COMPUSTAT for years 2009 and 2010. Market capitalization proxies for the level of liquidity of a stock, turnover is a measure of trading activity, while volatility and market beta are measures of risk. The star indicates significance at 5%.

Our findings suggest that stock characteristics determine what limit order book changes are most relevant for fast, impatient traders and point to the fact that not all market changes will be relevant cancellation determinants for all stocks.

7 Overbooking and Order Clustering

In this second part of our study, we examine how fast, impatient traders have learned to make the most of recent technological developments and how they leverage their speed advantage. In fact, a number of recent theoretical studies, most notably van Kervel (2015) and Baldauf and Mollner (2015), posit that if a sub group of traders enjoy a systematic speed advantage over the other market participants, this allows them to anticipate the order flow of the slower traders.

We argue that, if that is the case, this should allow them to leverage the availability of multiple trading venues and engage in an “overbooking” trading strategy, based on the simultaneous submission of multiple limit orders across different exchanges with the objective of executing only one of them. In fact, such trading strategy would allow them to increase their probability of execution, reduce execution time, and could be effectively implemented by them thanks to their speed advantage, which allows them to quickly cancel the remaining orders, avoiding over execution.

In order to test this hypothesis, we first develop a procedure to identify those orders submitted simultaneously by the same trader, and we call them *clustered orders.* Such first step is necessary since our data does not provide me with any information about the identity of the order submitter, making it impossible to match directly individual orders to specific traders. Then, we carry out an analysis of the nature and composition of these *clusters* of limit orders and we compare their performance to remaining orders. Finally, we investigate when such trading strategy is most effective and we test the notion that they belong to the same *cluster.*

7.1 Cluster Construction

One of the key contributions of this study is the identification of a new type of trading strategy, based on the ability of certain traders to engage in multi-venue trading by leveraging their speed and technological advantage. In their theoretical studies, van Kervel (2015) and Baldauf and Mollner (2015) suggest that fast traders can exploit their ability to anticipate the other market participants by submitting Market Orders against stale quotes. In fact, as soon as new price-altering information arrives in the market, fast traders can react to it before anybody else does and execute against orders that belong to the slower market participants who have not yet re-priced them. Following the same spirit and the empirical evidence suggesting that traders use, in their trading strategies, a mix of market and limit orders, we posit that some fast, impatient traders can benefit from submitting simultaneously limit orders across exchanges with the objective of executing only one. In fact, the simultaneous submission on multiple venues allows them to attempt execution on several trading venues rather than on only one, which should allow them to reduce execution time and increasing execution probability. On the other hand, their ability to anticipate the other market participants should allow them to cancel the outstanding orders as soon as the desired execution of one is attained, reducing their risk of over execution.

In order to identify those limit orders that are part of a multi-venue trading strategy employed by a fast trader, we define two limit orders as submitted simultaneously if they verify the following conditions:

- They are submitted on the same side of the book.
- They have the same price.
- They have the same size.
- They are submitted on different venues.
- They are submitted at most 50 milliseconds apart.
- They are cancelled or executed at most 50 milliseconds apart.

These conditions reflect the critical assumption that the simultaneous submission of multiple limit orders on multiple venues is aimed at increasing the probability of execution, or decreasing the execution time, of one of the orders rather than all of them. In this sense, our approach should not be confused with the practice of splitting one large parent order into multiple smaller child orders, which has at its core the objective of executing all of the orders rather than only one. Under this assumption, it is obvious that the trader who submits such a *cluster* of orders is indifferent between which of the orders attains execution, hence all of the orders in the same *cluster* must be identical. This assumption explains why we impose the condition that orders that belong to the same *cluster* must be on the same side of the book, have the same price and the same size. Given that the orders are submitted across exchanges, we cannot have, inside the same cluster, multiple orders on the same venue, as that does not serve the purpose of the cluster. Finally, the last two conditions are set in order to enforce the "simultaneous" aspect of order submission: if a trader submits a cluster of orders, it is possible that due to network latencies and delays in the matching engine, these orders will not appear in the limit order books at exactly the same time. Hence, we allow for a delay of up to 50 milliseconds, which is consistent with an upper bound reported for latencies in those years. Similarly, since the purpose of the cluster is the execution of one order, rather than all of them, if a cluster is no longer needed all of the orders will be simultaneously cancelled. Once again, attaining immediate, simultaneous cancellation of all orders might not be possible due to delays and network latencies hence, we allow for a 50 milliseconds delay in the report of these events.

7.2 Cluster Analysis

7.2.1 Introductory Findings

We run the clustering procedure on the eighty-five stocks in our sample and carry out an introductory analysis on cluster size and composition. As reported in Table 17, we find over nineteen million clusters of limit orders in our sample. The most common cluster (69.3%) is the one made of two limit orders (hence, of two limit orders submitted on two distinct venues) while clusters made up of orders submitted on all four venues are rather uncommon (6.2%). This could be explained by the fact that in our sample of four exchanges, one of them, EDGE-A, is an inverted pricing venue in which a trader is charged a fee for posting a limit order rather than given a rebate. This implies that EDGE-A could be seen as a venue of last resort for a limit order trader who would prefer to submit his order on any of the other three trading venues. It is also interesting to note that the most common cluster is that of size two rather than three. This suggests that simply submitting a limit order on each venue might not be cost effective and indicates that there must be an additional cost associated with submitting a cluster of limit orders rather than a just a single one.

Size of Cluster	2	3	4	Total
Number of Clusters	13,634,367	4,827,785	1,225,165	19,687,317
Percentage	69.3	24.5	6.2	100

Table 17 Cluster Size S*tatistics.* In the top row of the table, we report the number of clusters by size, while in the bottom one the percentage that cluster size represents. The size of the cluster is the number of limit orders that make up the cluster: given that each limit order in a cluster must be submitted at a different exchange, the size of the cluster also tells how many exchanges are used in a cluster.

A careful analysis of cluster's composition, reported in Table 18, shows how the most common cluster (31.3%) is the one made up of one limit order submitted on BATS-Z ("Z") and another one on the NASDAQ (“Q”). This is not surprising since these two exchanges are, respectively, the most convenient one, fee and rebate structure wise, and the one with highest market share. A rather distant second (17.2%) is the cluster made up of three limit orders sent to the NASDAQ (“Q”), BATS-Z (“Z”) and EDGE-X (“K”). Then, two clusters of size two follow: one with orders sent to EDGE-X (“K”) and BATS-Z (“Z”), and the other with orders sent to EDGE-X (“K”) and the NASDAQ (“Z”), which make up, respectively, 14.2% and 10.6% of the total. Not surprisingly, very few clusters include limit orders send to EDGE-A (“J”): the most common cluster that includes this venue is the one made up of limit orders submitted to all four exchanges (6.2%) while all other possible venue combinations that include EDGE-A are considerably less frequently used. The results in Table 18 confirm that EDGE-A, given its fee and rebate structure that makes it the only “Inverted-Pricing” venue in our sample, might be considered as a venue of last resort, since it is not a frequent target for clusters of limit orders.

JKQ	JQZ	JKZ	JK	JZ	JQ	JKQZ	KQ	KZ	KQZ	QZ
1.6%	2.7%	3.0%	3.8%	3.8%	5.6%	6.2%	10.6%	14.2%	17.2%	31.3%

Table 18: Venue Composition of Clusters. The table presents the percentage of clusters submitted to each possible combination of exchanges. The NASDAQ is denoted by “Q”, BATS-Z by “Z”, EDGE-X by “K” and EDGE-A by “J”. For example, a value of 6.2% for venue composition “QZKJ” means that only 6.2% of clusters is made up of 4 limit orders submitted on each of the 4 possible exchanges.

We further investigate cluster composition by looking at a breakdown by venue. In the top row of Table 19, we see how the NASDAQ appears in 31.7% of all clusters and BATS-Z in 33.1% of them. On the other hand, EDGE-A is the least common venue appearing in only 11.3% of clusters. It is also interesting to look at what proportion of the entire limit order flow on each exchange belongs to some cluster. In the bottom row of Table 19, we see that EDGE-X is the venue with the highest proportion of clustered limit orders since 26.3% of all the submitted limit orders belong to some cluster. On the other hand, only 13.1% of all limit orders submitted to the NASDAQ are clustered while on average, across all venues, 20.2% of all limit orders in our sample belong to a cluster. The low proportion of clustered limit orders on the NASDAQ can be explained by the fact that the NASDAQ is the venue with the largest market share, both in terms of limit order submission and execution. Hence, even though the NASDAQ is one of the most common venues when submitting a cluster of limit orders, given the large volume of limit orders submitted there, the overall proportion of clustered orders is still fairly small. These results suggest that clusters of limit orders are a significant component of the entire order flow, consistent with the empirical evidence pointing to an extensive presence of algorithmic traders in today's equity market, and underlines the need to better understand its dynamics.

Venue	J	K	Q	Z	Total
Proportion	11.3%	23.9%	31.7%	33.1%	100.0%
Proportion of Limit Orders in Clusters	20.0%	33.7%	13.1%	26.3%	20.2%

Table 19: Cluster Composition by Venue. In the upper portion of the table, we report the proportion of clusters that include a limit order submitted to each specific venue. In the lower portion, we report the proportion of limit orders on each venue that belong to a cluster. Finally, in the last column we report the proportion of clustered limit orders across all exchanges. The NASDAQ is denoted by "Q", BATS-Z by "Z", EDGE-X by "K" and EDGE-A by "J".

It is also important to better understand whether any significant differences between the limit orders that belong to a cluster and their non-clustered counterparts exist. In fact, it is reasonable to assume that, just like for the case of fleeting and non-fleeting orders, clusters are used by traders who have very different trading strategies and objectives than the classic patient liquidity supplier, hence they should also be significantly different from the rest of the limit order flow.

We investigate the submission patterns for the two order types and present the results in Table 20. We see that, on average, clusters are priced more aggressively than the other orders. In fact, 72.4% of clusters are submitted at prices that are better than, or at, the Best Bid or Ask compared to 63.7% for non-clustered orders. This difference in order aggressiveness becomes even more evident if we look at the break down by relative price: the proportion of clustered orders that are price improving is 16.2% while that for the rest of the orders is only 9.1%. This higher aggressiveness of clustered orders over their non-clustered counterparts is not surprising considering that the entire purpose of submitting a cluster is to increase the probability of execution and reduce execution time. Hence, submitting a cluster far away from the market would have a limited effect on the probability of execution while it would still involve a cost in order to be implemented. It is interesting to point out that the results in Table 20 are consistent with those in Table 8 that compared fleeting orders to non-fleeting ones. In both cases, we see that the order flow generated by fast, impatient traders is considerably more aggressive than that of the order market participants and, in both cases, it is consistent with the actions of traders who prioritizes a quick execution over everything else. Overall, the results in Table 20 confirm that clustered limit orders are more similar in nature to fleeting orders rather than non-fleeting ones and that they could be used by traders with similar trading strategies.

	Price Improving	*At best Bid or Ask*	*Between 2nd and 5th*	*Between 6th and 10th*	*Deeper*
Cluster	16.2%	56.2%	22.7%	3.7%	1.2%
Non-Cluster	9.1%	54.6%	26.6%	5.4%	4.3%

Table 20: Order Aggressiveness Statistics. Each row of the table presents the submission statistics for the two types of limit orders: clustered and non-clustered. Each column identifies a different portion of the limit order book, based on the relative price. "Price Improving" refers to limit orders that, when submitted, set a new best Bid or Ask, hence improve the previous best price. "At best Bid or Ask" refers to orders submitted at the current best price. "Between 2nd and 5th" refers to orders submitted between the 2nd and 4th best price in the book. "Between 6th and 10th" refers to orders submitted between the 6th and 10th most aggressive price on each side. "Deeper" identifies all orders submitted further away than the 10th price level from the market.

In conclusion, the results of this introductory analysis allow to make a number of interesting remarks. First, the analysis of cluster size and composition suggests that submitting a cluster of limit orders might be considerably more expensive than a simple single-order submission and that submitting an order on each of the trading venues might not be cost effective. Second, it also suggests that EDGE-A is seen as a venue of last resort, given that its fee and rebate structure is set to charge a fee when submitting an order rather than paying a rebate. On the other hand, the results for the study of order aggressiveness find that clusters of limit orders are considerably more aggressive than their non-clustered counterparts, suggesting that the two might be used by traders with very different trading strategies and objectives. Moreover, our results further indicate that clusters of limit orders are more similar in nature to fleeting orders, confirming our initial assumption that the overbooking trading strategy can be effectively implemented only by this sub group of traders as a result of their speed advantage over the other market participants.

7.2.2 Performance of Clustered Orders

In this part of our analysis we investigate the benefits of an overbooking trading strategy over a simple single-venue submission. In fact, even though the speed advantage of fast, impatient traders reduces considerably their risk of over execution they still need to account for a higher cost associated with such strategy as suggested by the results in Table 17, Table 18 and Table 19. The higher cost associated with this trading strategy can be explained by the fact that submitting multiple orders on multiple exchanges implies the need to monitor multiple venues, which certainly leads to the need for additional computational and financial resources. Moreover, even though over execution risk is certainly reduced it cannot be entirely eliminated. In order to investigate the performance of clustered orders, we look at the execution times and probabilities of these orders and at the quality of these executions, measured by the realized spread as in Sofianos and Yousefi (2010). This analysis should allow me to explore two possible advantages derived from the implementation of a multi-venue trading strategy: higher probability of execution or the possibility of making the most of short-term, favorable price movements, thanks to a lower execution time.

In the first step of our analysis we look at the probability of execution across order types and present our results in Table 21. We find that, on average, 8.3% of all clusters lead to an execution while only 4.5% of orders do. This higher probability of execution for clusters of limit orders could be the result of the more aggressive nature of these orders, as suggested in Table 20. Hence, to investigate this possibility, we break down the execution probabilities by relative price. Limit orders submitted at prices better than the best price have the same probability of execution: 16.1% for both clusters and non-clustered orders. This is not surprising and it simply indicates that an aggressive limit order will have a higher probability of execution in virtue of its position at the top of the queue. This also confirms a very intuitive result that higher order aggressiveness leads to increased probability of execution. An important difference between the two order types is in the probability of execution of orders that are submitted at the best price: clusters are far more likely to be executed than normal orders, with a probability of execution of 9.5% compared to only 5.2%. This result suggests that a cluster of limit orders is an effective tool to increase the probability of execution without having to compete, and hence increase, the relative price of the order. In fact, increasing the aggressiveness of an order result in a worse price for the trader who submits it, reducing the potential profit generated by each trade. Hence, a trader who submits a limit order would like to attain an execution at the least aggressive price possible and the results in Table 21 suggest that a cluster of limit orders is an effective tool in order to do so.

The execution probabilities for orders submitted beyond the best price are very low across both order types.

	Price Improving	*At best Bid or Ask*	*Between 2nd and 5th*	*Between 6th and 10th*	*Deeper*	Across all Price Levels
Cluster	16.1%	9.5%	1.0%	0.7%	1.3%	8.3%
Non- Cluster	16.1%	5.2%	0.6%	<0.1%	<0.1%	4.5%

Table 21: Execution Probabilities of Limit Orders. Each row of the table presents the execution probabilities for the two types of limit orders: clustered and non-clustered. Each column identifies a different portion of the limit order book, based on the relative price. "Price Improving" refers to limit orders that, when submitted, set a new best Bid or Ask, hence improve the previous best price. "At best Bid or Ask" identifies orders submitted at the current best price. "Between 2nd and 5th" refers to orders submitted between the 2nd and 4th best price in the book. "Between 6th and 10th" refers to orders submitted between the 6th and 10th most aggressive price on each side. "Deeper" identifies all orders submitted further away than the 10th price level from the market. Finally," Across All Price Levels" present the overall execution probability across all price levels.

A second potential advantage of an overbooking trading strategy based on the submission of a cluster of limit orders, could be the possibility of making the most of short-lived and short-term favorable price movements. If that were the case, clusters of limit orders would have to have shorter execution times than their non-clustered counterparts, and would have to be submitted prior to favorable price movements. To test these hypotheses, we look at the average execution time across order types and we compare their quality of executions, measured by the realized spread.

We start by looking at the average execution times across order type and, as reported in Table 22, we find that non-clustered limit orders have rather long execution times with, on average, over 5 minutes and 15 seconds before attaining execution. On the other hand, clusters have an average execution time of only 2 minutes, 19 seconds and 610 milliseconds, which suggests that traders who use clusters of limit orders are able to reduce by more than half their execution time. This confirms a second key advantage of a trading strategy based on the simultaneous submission of multiple limit orders across exchanges: faster executions.

	Average Execution Time (milliseconds)
Clusters	139,610
Non-Clustered Orders	315,690

Table 22: Average Execution Time Across Order Type. Each row of the table presents the execution times for the two types of limit orders: clustered and non-clustered. Execution times are measured in milliseconds and quantify the time between order submission and execution.

We now compare the quality of the executions across order type by looking at the realized spread. Our approach is in line with the work of Sofianos and Yousefi (2010) and Battalio, Corwin and Jennings (2016) and is based on the assumption that the quality of an execution can be measured by looking at the price movement following the execution. In fact, the higher the price increase (drop) after a buy (sell) limit order is executed the better for the order submitter since it means that the he was able to buy (sell) his shares right before a price increase (drop). The opposite is true if a buy (sell) limit order is executed right before a price drop (increase).

For every execution we compute the realized spread by comparing the mid-price five minutes after the execution to what it was at the moment of the execution: in the specific, for an executed buy limit order, we compute the difference between the future mid-price and the execution mid-price, while for a sell limit order, we multiply such difference by -1. Such definition allows to standardize the measure across the two sides of the book, and state that larger values for the realized spread imply higher quality of execution, regardless of side.

The results for the realized spread in Table 23 find no significant difference in the quality of the executions between the two order types, with clusters of limit orders having slightly higher average realized spreads: 251 compared to 249. This suggests that even though clusters of orders allow to decrease the execution time, this doesn't translate in better executions.

	Average Realized Spread
Cluster	251
Non-Cluster	249

Table 23: Average Realized Spread Across Order Types. Each row of the table presents the average realized spread for the two types of limit orders: clustered and non-clustered. Realized spreads are computed by comparing the mid-price five minutes after the execution to that at the moment of the execution: in the specific, for an executed buy limit order, the realized spread is the difference between the future mid-price and the execution mid-price, while for a sell limit order, we multiply such difference by -1. This definition allows to standardize the measure across the two sides of the book, and say that larger values, imply higher quality of executions regardless of side.

In conclusion, our analysis of the performance of clusters of limit orders allows to identify two key advantages of an overbooking strategy: increased probability of execution and considerable reduction of execution time. These advantages are particularly relevant for traders who aim to attain a quick execution, such as the fast, impatient traders who we argue are behind clustered orders, and explain why traders engage in this type of strategy given its higher cost and risk. Our analysis, however, does not find any evidence suggesting that the ability to attain a quicker execution can be used to make the most of short-term favorable price movements. In fact, the realized spread of clustered executions is not significantly different from that of normal orders, which implies that the two order types attain executions of similar quality.

7.2.3 In-Sample Variation of Cluster Performance

In the last part of the empirical analysis of clustered orders, we look at whether the overbooking trading strategy is equally effective for all stocks in our sample. In light of our findings on fleeting liquidity, and keeping in mind that clusters of limit orders are also submitted by the same type of traders who prioritize a quick execution, we expect overbooking to be particularly effective for those stocks in which there is a high level of competition for a favorable queue position. In fact, as we have seen in Section 6, if a fast, impatient trader wants to attain a quick execution he needs to secure a favorable queue placement for his orders. If, however, he wants to trade a very desirable stock, he might face a very high degree of competition for such spot, and attaining it might prove either too hard or very time consuming. It is in such case that the ability of a trader to engage in a multi-venue trading strategy based on the simultaneous submission of a cluster of limit orders might prove especially helpful in allowing him to increase his probability of execution and bypassing the competition for the top of queue placement.

We also expect overbooking to be particularly effective in improving execution probability for stocks that have a significant extent of execution fragmentation. In fact, if a stock has the majority of trades occurring on one trading venue only, submitting multiple limit orders across multiple exchanges might not bring any significant improvements in execution probability. On the other hand, if a stock has executions that occur in large numbers on multiple exchanges, submitting a cluster of orders might considerably increase the probability of finding a matching incoming market order.

To test our hypothesis on how execution fragmentation and competition for superior queue placement affect the effectiveness of the overbooking trading strategy, we carry out a regression analysis on the determinants of the difference in the execution probability between clustered and non-clustered limit orders. In order to test our hypothesis that clusters of limit orders are particularly helpful when trading stocks with high competition for a favorable queue placement we include, as independent variable, the proportion of fleeting orders out of total order flow. In fact, for a given stock, a higher proportion of fleeting liquidity indicates that there are more fast, impatient traders trying to attain a quick execution, hence a larger number of traders fighting for a favorable queue placement. On the other hand, in order to test the role of execution fragmentation, we build a fragmentation measure that quantifies the level of fragmentation of executions across exchanges. The measure is a Herfindahl index in which high values indicate a stock with high concentration of executions around one specific exchange, implying that most of the executions occur on one venue only. The control variables that we use in this study are in line with those used in Section 6, and represent stock level characteristics that account for the in-sample difference reported in Table 2. In the specific, we look at how the average market cap, turnover (computed as the ratio of traded stocks against all outstanding), closing price and market beta of the stock affect the change in execution probability. The values of these independent variables are obtain by looking at historical monthly data for 2009 and 2010 and are collected from CRSP and COMPUSTAT.

The resulting regression model will be of the form

$$\begin{aligned} Diff.Exec.Prob_i = {} & \beta_0 + \beta_{MCAP} \times Market\ Cap_i + \beta_{TO} \times Turn\ Over_i \\ & + \beta_{Price} \times \log(Price_i) + \beta_{beta} \times Market\ Beta_i \\ & + \beta_{Fragm.Index} \times Fragmentation\ Index_i \\ & + \beta_{Fleeting} \times Fleeting_i + \varepsilon_i \end{aligned} \qquad (5)$$

for *we = 1, …, 85.*

The regression results are presented in Table 24 and provide some interesting insights. First, we see that among all control variables only price is significant: in the specific, it is a significant and negative determinant of the change in execution probability. This means that the lower the price of a stock, the higher the benefit of an overbooking trading strategy. There is no clear answers as to why that would be the case even after controlling for fragmentation and proportion of fleeting orders and it is also somewhat surprising that none of the other control variables is significant. Second, we see that the proportion of fleeting order flow is a significant and positive determinant of change in execution probability. This means that stocks with a high proportion of fleeting liquidity are those that benefit the most from the submission of a cluster of limit orders since the difference in execution probability between clusters and non-clustered orders increases. Given that the proportion of fleeting liquidity proxies for the extent of competition for superior queue placement, this result confirms our assumption that overbooking works best for stocks in which there is considerable competition for attaining an advantageous queue placement. Finally, the significant and negative coefficient for the fragmentation index is consistent with our assumption that clustering limit orders is especially useful for those stocks that have very fragmented executions. In fact, the closer the coefficient is to 1, the less fragmented are the executions.

	Estimated Coefficients	t-statistic
Market Cap	0.0000	-1.2820
Turnover	-0.0033	-0.2320
Price	-0.0002*	-2.5520
Market Beta	-0.0100	-1.3760
Fragmentation Index	-0.1018*	-2.1100
Fleeting Orders	0.0704*	2.2980

Table 24: Cluster Performance Regression Results. The table present the regression results for the study of the determinants of the difference in execution probability between clustered and non-flustered orders. In the specific, for each stock we look at how changes in the proportion of fleeting liquidity, that proxies for the level of competition for superior queue placement, and in the level of execution fragmentation, affect the improvement in execution probability obtained when switching from a simple single-venue trading strategy to overbooking. The control variables used in this study are in line with those used in Section 6 for the study of the cancellation determinants of fleeting liquidity and represent stock level characteristics that account for the in-sample difference reported in Table 2. In the specific we look at how the average market cap, average stock turnover (computed as the ratio of traded stocks against all outstanding), average closing price and market beta affect the change in execution probability between the two order types. The values of the control variables are obtain by looking at historical monthly data for 2009 and 2010, collected from CRSP and COMPUSTAT. (*) denotes those coefficients significant at 5% level.

Hence, a negative coefficient implies that when execution fragmentation increase hence, when the fragmentation coefficient goes to zero, the benefit of submitting a cluster of orders over a single one increases. In conclusion, our analysis of the in-sample variation of cluster performance, confirms our assumptions about which stocks benefit the most from a trading strategy based on the simultaneous submission of multiple orders across multiple trading venues. In fact, we find empirical evidence that a cluster of limit orders is able to attain a significantly higher execution probability than normal orders when used for stocks with high competition for superior queue placement and with highly fragmented executions. These results are, once again, consistent with the notion that fast, impatient traders who prioritize quick execution are behind this type of trading strategy and that when obtaining a favorable queue placement is not possible, or just too time consuming, switching to overbooking can be particularly effective.

7.3 Testing Cluster Existence

The clustering procedure that we use to determine whether two limit orders are submitted simultaneously is necessary since the data does not provide any identification information about who submits the limit orders. Hence, it is not possible for me to match directly individual orders to specific traders. In this section, we test our assumption that orders that belong to the same cluster are, indeed, part of the same simultaneous submission. In order to do so, we use the fact that clusters of limit orders are submitted to attain the execution of one of the orders rather than all of them. If that is the case, we should find evidence that when the first order in a cluster attains execution, then the others are cancelled.

The testing procedure is based on running a logistic regression on the cancellation determinants of the limit orders that belong to a cluster. The variable of interest is a dummy variable that signals whether there has been an execution in a cluster: in the specific, it will be set to one if the first order in the cluster is executed and zero otherwise. If our clustering procedure and assumptions about the purpose of a cluster are correct, we expect the dummy variable to be a positive and significant determinant of clustered order cancellations. This would mean that if the first order in a cluster attains execution, then this increases the probability of observing a cancellation for the remaining orders in that cluster. Given that we believe that clusters of limit orders are submitted by fast, impatient traders, we use the same two sets of control variables from our study of the cancellation determinants of fleeting liquidity: one set will control for the changes in the state of the limit order book in which the order is submitted, while the other for those occurring in the *Super Book*[6]. Consistent with our results from that study, reported in Table 13, we expect that an increase in depth imbalance on either the *Own Book* or the *Super Book* will increase the probability of cancellation since such market change signals that prices might be about to move away from the cluster of orders. In light of the findings on the cancellation determinants of fleeting liquidity, we expect that an increase in market activity, measured by the number of executions, on either the *Own Book* or the *Super Book* will decrease the probability of a cancellation. In fact, from the perspective of an impatient trader who is waiting for execution, an increase in the arrival of market orders is actually desirable as it increases the flow of those orders that can be matched against his

[6] For a more detailed explanation on the definition and purpose of the four control variables, please see Section 6.1 and Section 6.2.

outstanding order. Hence, an increase in executions can lead an impatient trader to stay put and wait for execution. On the other hand, an increase in market activity when measured by the number of cancellations will increase the probability of observing an order cancellation since market participants use such measure as a proxy for the arrival of new information. One key difference between the expected results for this study and those from the fleeting liquidity cancellation determinants should be in the effect of changes in the relative spread on the *Super Book*. In fact, in the previous study, we argued that fast, impatient traders actively monitor all trading venues after order submission, comparing the state of their venue to that of the competing market. This implied that when the competing venues are worse than the one in which the order is submitted than, fast, impatient traders would not have any incentive to resubmit their orders to those other exchanges. Hence, we found that an increase in relative spread on the *Super Book* decreased the cancellation probability of a fleeting order. In this case, however, given that we deal with a cluster of limit orders submitted on multiple exchanges, an increase in relative spread on the *Super Book* implies that even though the relative spread is increasing on a different venue, that is still a venue in which one of the remaining orders from the cluster resides. Hence, this means that the increase in information asymmetry occurring on the *Super Book* is still going to affect one of the orders from the cluster and we should expect that such increase will have the same effect on the probability of a cancellation as if the increase occurred on the venue were the order is submitted. Hence, we expect an increase in relative spread on the *Own Book* and the *Super Book* to increase the probability of a cancellation.

In order to study the cancellation determinants of clustered orders, we run one logistic regression per stock in our sample, which results in a total of eighty-five regressions. The final regression equation is of the form

$$
\begin{aligned}
log\left(\frac{\pi}{1-\pi}\right) = \beta_0 &+ \beta_1^{Own} \times \Delta Depth\ Imbalance + \beta_2^{Own} \times \Delta Relative\ Spread \\
&+\beta_3^{Own} \times Number\ of\ Exec. + \beta_4^{Own} \times Number\ of\ Canc. \\
&+\beta_5^{Super} \times \Delta S. Depth\ Imbalance + \beta_6^{Super} \times \Delta S. Relative\ Spread \\
&+\beta_7^{Super} \times S. Number\ of\ Exec. + \beta_8^{Super} \times S. Number\ of\ Canc. \\
&+\beta_9 \times Execution\ Dummy \qquad (6)
\end{aligned}
$$

for *we = 1, …, 85.*

In Table 25, we report the results of the regression analysis and present the median values of the regression coefficients obtained when running the eighty-five individual logistic regressions and the number of stocks that have each predictor significant at a 5% level. Two important remarks can be made: first, the control variables are significant for the majority of the stocks, with numbers ranging from thirty-seven up to sixty-one. Moreover, the directions of the coefficients are consistent with our initial assumptions about the effect that changes in the state of the *Own Book* and *Super Book* will have on the cancellation probability of a cluster of limit orders. Second, the Execution Dummy is positive and significant for almost every stock. This means that after the execution of the first order in the cluster, the probability of cancellation for the remaining orders increases. This result validates our clustering procedure since it provides empirical evidence that the orders that belong to the same cluster depend of each others. Moreover, it also confirms our initial assumption that the objective of this trading strategy, based on the simultaneous submission of multiple orders across different exchanges is the execution of one order, rather than all of them. In fact, after the first order in the cluster is executed, the probability of a cancellation of the remaining orders in the cluster increases considerably.

		Probability of Cluster Cancellation	
		Coefficient	Sig. 5%
Own Venues Variables	Δ. Depth Imbalance	0.0001	56
	Δ. Relative Spread	75.60443	42
	Number of Exec.	-0.06677	59
	Number of Canc.	0.02026	48
Super-Book Variables	ΔS.Depth Imbalance	0.00003	49
	ΔS.Relative Spread	18.22075	37
	S.Number of Exec.	-0.00707	61
	S.Number of Canc.	0.00072	41
	Execution Dummy	2.01042	72

Table 25: Cluster Existence Logistic Regression. In this table, we report the results for study on the cancellation determinants of clustered orders, necessary to test the existence of a cluster. In the first column, we present the median value of the coefficients computed across all eighty-five stocks while in the second one we state the number of stocks that have that predictors significant at a 5% level. Execution Dummy is the variable of interest and is a dummy variable that signals whether there has been an execution in a cluster: in the specific, it is set to one if the first order in the cluster is executed and zero otherwise. Δ(S.) Depth Imbalance refers to the (relative) change in the own (Super) book's depth imbalance between the moment of cancellation and submission; Δ(S.) Relative Spread refers to the (relative) change in the own (Super) book's relative spread between the moment of cancellation and submission; (S.) Number of Exec, and (S.) Number of Canc. refer to the average number of executions and cancellations, respectively, that have occurred on the own (Super) book during the lifetime of the order.

8 Conclusions

In our study, we address the effect of recent regulatory and technological changes on trading dynamics, which have not only transformed how the equity market functions but also how market participants interact with it and between each other's. We build on previous research that finds that algorithmic traders can be characterized as fast, impatient traders who prioritize a quick execution and we study their trading patterns. We investigate how the implementation of Reg-NMS has affected the behavior of fast, impatient traders and explore whether the order flow generated by such traders is consistent with the puzzling phenomenon of fleeting liquidity. We then investigate whether fast, impatient traders are able to leverage their speed advantage and turn market fragmentation in their favor by implementing a trading strategy based on the simultaneous submission of multiple orders across exchanges.

We begin our analysis by testing the signaling hypothesis proposed in Hasbrouck and Saar (2009) to explain the fleeting liquidity submitted inside the spread. We find evidence of a positive relation between the submission of fleeting liquidity inside the spread and the execution of hidden orders, supporting their claim the fleeting liquidity submitted inside the spread can serve a signaling purpose to attract the attention of traders to a specific venue.

We then test our assumption that the implementation of Reg-NSM, combined with the impatient nature of algorithmic traders, can explain the dynamics of fleeting liquidity submitted inside the book, rather than only that submitted inside the spread. We find evidence that fast, impatient traders actively monitor all trading venues in order to correctly asses the position of their orders and that they react to market events on every venue if they negatively affect the placement of their orders.

Finally, following a number of recent theoretical studies, we investigate how fast, impatient traders exploit their ability to anticipate the order flow of other, slower market participants. We find evidence that they are able to successfully engage in an overbooking trading strategy based on the simultaneous submission of limit orders across different exchange. Such strategy allows them to increase their probability of execution, reduce execution time and is particularly helpful when used for stocks that have a high degree of competition for a favorable queue placement.

References

Amihud, Yakov, Beni Lauterbach, and Haim Mendelson. "The value of trading consolidation: Evidence from the exercise of warrants." *Journal of Financial and Quantitative Analysis* 38, no. 04 (2003): 829-846.

Baldauf, Markus, and Joshua Mollner. "High-frequency trading and market performance." (2015).

Barclay, Michael J., and Terrence Hendershott. "Liquidity externalities and adverse selection: Evidence from trading after hours." *The Journal of Finance* 59, no. 2 (2004): 681-710.

Battalio, Robert H. "Third Market Broker-Dealers: Cost Competitors or Cream Skimmers?." *The Journal of Finance* 52, no. 1 (1997): 341-352.

Battalio, Robert, Shane A. Corwin, and Robert Jennings. "Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality." *The Journal of Finance* 71, no. 5 (2016): 2193-2238.

Bennett, Paul, and Li Wei. "Market structure, fragmentation, and market quality." *Journal of Financial Markets* 9, no. 1 (2006): 49-78.

Biais, Bruno. "Price formation and equilibrium liquidity in fragmented and centralized markets." *The Journal of Finance* 48, no. 1 (1993): 157-185.

Bloomfield, Robert, Maureen O’hara, and Gideon Saar. "The “make or take” decision in an electronic market: Evidence on the evolution of liquidity." *Journal of Financial Economics* 75, no. 1 (2005): 165-199.

Boehmer, Beatrice, and Ekkehart Boehmer. "Trading your neighbor’s ETFs: Competition or fragmentation?." *Journal of Banking & Finance* 27, no. 9 (2003): 1667-1703.

Cao, Charles, Oliver Hansch, and Xiaoxin Wang. "Order placement strategies in a pure limit order book market." *Journal of Financial Research* 31, no. 2 (2008): 113-140.

Cao, Charles, Oliver Hansch, and Xiaoxin Wang. "The information content of an open limit-order book." *Journal of futures markets* 29, no. 1 (2009): 16-41.

Cebiroglu, Gökhan, and Ulrich Horst. "Optimal order exposure and the market impact of limit orders." (2013).

Chowdhry, Bhagwan, and Vikram Nanda. "Multimarket trading and market liquidity." *Review of financial studies* 4, no. 3 (1991): 483-511.

Chowdhry, Bhagwan, and Vikram Nanda. "Multimarket trading and market liquidity." *Review of financial studies* 4, no. 3 (1991): 483-511.

Cont, Rama, and Arseniy Kukanov. "Optimal order placement in limit order markets." (2013).

Ellul, Andrew, Craig W. Holden, Pankaj Jain, and Robert Jennings. "Order dynamics: Recent evidence from the NYSE." *Journal of Empirical Finance* 14, no. 5 (2007): 636-661.

Fong, Kingsley YL, Peter L. Swan, and Ananth Madhavan. "Why do markets fragment? A panel-data analysis of off-exchange trading." (2001).

Foucault, Thierry, and Albert J. Menkveld. "Competition for order flow and smart order routing systems." *The Journal of Finance* 63, no. 1 (2008): 119-158.

Gajewski, Jean-François, and Carole Gresse. "Centralised order books versus hybrid order books: A paired comparison of trading costs on NSC (Euronext Paris) and SETS (London Stock Exchange)." *Journal of Banking & Finance* 31, no. 9 (2007): 2906-2924.

Hasbrouck, Joel, and Gideon Saar. "Technology and liquidity provision: The blurring of traditional definitions." *Journal of financial Markets* 12, no. 2 (2009): 143-172.

Handa, Puneet, and Robert A. Schwartz. "How best to supply liquidity to a securities market." *The Journal of Portfolio Management* 22, no. 2 (1996): 44-51.

Handa, Puneet, Robert A. Schwartz, and Ashish Tiwari. "Not held orders: evidence on the value of order timing in an equity market." (2000).

Hendershott, Terrence, and Haim Mendelson. "Crossing networks and dealer markets: competition and performance." *The Journal of Finance* 55, no. 5 (2000): 2071-2115.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. "Does algorithmic trading improve liquidity?." *The Journal of Finance* 66, no. 1 (2011): 1-33.

Hoffmann, Peter. "A dynamic limit order market with fast and slow traders." *Journal of Financial Economics* 113, no. 1 (2014): 156-169.

Hollifield, Burton, Robert A. Miller, and Patrik Sandås. "Empirical analysis of limit order markets." *The Review of Economic Studies* 71, no. 4 (2004): 1027-1063.

Kaniel, Ron, and Hong Liu. "So what orders do informed traders use?." *The Journal of Business* 79, no. 4 (2006): 1867-1913.

Lo, Andrew W., A. Craig MacKinlay, and June Zhang. "Econometric models of limit-order executions." *Journal of Financial Economics* 65, no. 1 (2002): 31-71.

Mendelson, Haim. "Consolidation, fragmentation, and market performance." *Journal of Financial and Quantitative Analysis* 22, no. 02 (1987): 189-207.

O’Hara, Maureen. "High frequency market microstructure." *Journal of Financial Economics* 116, no. 2 (2015): 257-270.

O'Hara, Maureen, and Mao Ye. "Is market fragmentation harming market quality?." *Journal of Financial Economics* 100, no. 3 (2011): 459-474.

Pagano, Marco. "Trading volume and asset liquidity." *The Quarterly Journal of Economics* 104, no. 2 (1989): 255-274.

Parlour, Christine A. "Price dynamics in limit order markets." *Review of Financial Studies* 11, no. 4 (1998): 789-816.

Ranaldo, Angelo. "Order aggressiveness in limit order book markets." *Journal of Financial Markets* 7, no. 1 (2004): 53-74.

Sofianos, George and Ali Yousefi, “Smart routing: Good fills, bad fills and venue toxicity”. *Goldman Sachs Equity Execution Strategies Street Smart* 40 (2000): 1-9.

Stoll, Hans R. "Market fragmentation." *Financial Analysts Journal* (2001): 16-20.

Van Kervel, Vincent. "Competition for order flow with fast and slow traders." *Review of Financial Studies* (2015): hhv023.

Vita

Name of the author: Krzysztof Herman

Place of Birth: Warsaw, Poland

Undergraduate School Attended: Universita di Bologna, Bologna, Italy

Degrees Awarded: B.Sc Mathematics