LOUVAIN
School of Management

Essays on Intraday Order Flow Dynamics and Liquidity Forecasting

Thibaut MOYAERT

Thèse de doctorat 7 | 2014

Université catholique de Louvain
LOUVAIN SCHOOL OF MANAGEMENT RESEARCH INSTITUTE GROUP

LOUVAIN
School of Management

Essays on Intraday Order Flow Dynamics and Liquidity Forecasting

By Thibaut Moyaert

Submitted to
Catholic University of Louvain
in Partial Fulfillment of the
Requirements for the Degree of
Docteur en Sciences Economiques et de Gestion
at Louvain School of management

DOCTORAL COMMITTEE:

Chairman

Prof. Per Agrell CATHOLIC UNIVERSITY OF LOUVAIN, BELGIUM

Supervisor

Prof. Catherine D'Hondt CATHOLIC UNIVERSITY OF LOUVAIN, BELGIUM

Supervisory Panel

Prof. Jonathan Brogaard UNIVERSITY OF WASHINGTON, USA

Prof. Rudy De Winne CATHOLIC UNIVERSITY OF LOUVAIN, BELGIUM

Prof. Sebastien Laurent AIX EN PROVENCE UNIVERSITY, FRANCE

Prof. Mikael Petitjean CATHOLIC UNIVERSITY OF LOUVAIN, BELGIUM

MONS, BELGIUM
JUNE 2014

CONTENTS

[1]This essay is a joint work with Catherine D'Hondt.

[1]This essay is a joint work with Jonathan Brogaard and Ryan Riordan.

[1]This essay is a joint work with Catherine D'Hondt.

[1]This essay is a joint work with Jonathan Brogaard and Ryan Riordan.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This doctoral thesis results from a 3-year research grant led by Prof. Catherine D'Hondt and Mikael Petitjean at the Louvain School of Management. Many people contributed to this research in its current state.

I would like to thank my supervisor, Catherine D'Hondt, for her support, advice and comments that appear through this thesis. Next, I would like to thank the members of my supervisory panel: Jonathan Brogaard, Rudy De Winne, Sebastien Laurent and Mikael Petitjean. Their respective knowledge in various fields has greatly enhanced the quality of this thesis.

I express my gratitude to the institution, especially to the finance team at the Louvain School of Management.

I greatly appreciated the comments and discussions from various people I had the opportunity to meet during the writing of this thesis.

I thank Sandrine Delhaye for the considerable amount of work delivered to ensure a smooth doctoral process.

I thank NYSE Euronext and Nasdaq OMX for providing the data used in this research.

I am grateful to the "Action de recherche Concertées" subsidies from the Communauté française de Belgique granted by the Académie universitaire "Louvain".

CHAPTER

ONE

INTRODUCTION

1.1 Market Microstructure

This doctoral thesis contributes to the field of market microstructure. O'Hara (1995) defines market microstructure as the study of the process and outcomes of exchanging assets under a specific set of rules. It encompasses theoretical, empirical and experimental research on the relationship between market organization and market quality.

It has implications that go well beyond the design of well-functioning stock exchanges, with indirect implications for asset pricing, international finance and corporate finance.

The assessment of market quality is not straightforward as it includes several dimensions such as liquidity, transaction costs, and volatility. Its implications in terms of welfare and the regulation of market structures are also substantial. Agents' behaviors are indeed very sensitive to the information structure shaped by the market organization (Madhavan, 2000).

In this dissertation, we make several contributions in empirical market microstructure by focusing on liquidity issues and market disruptions, that have increased lately according to Hendershott (2011). Aside of providing a novel methodological framework to better manage execution price, we pay specific attention to liquidity dynamics during extreme market states. Our results show that market participants have an incentive not to trade during those periods of time to better manage their execution price. Finally, we document the trading behavior of high frequency traders (HFT) around market disruptions. It is a topical issue for regulators since the advent of HFT is often mentioned as responsible for higher market instability.

Before proceeding to a more detailed overview of our essays, it is important to ground our research in the current market environment. Indeed, in a perfect market, all information should be incorporated in the trading price of an asset and the trading price should reflect the fundamental value of a traded asset resulting from the competition between fully informed market participants. Our essays would be less relevant in such market environment. In practice, market microstructure theory acknowledges that the full efficient allocation of financial resources is in general not reached (Biais et al., 2005). Theoretical and empirical literature outline that trades have both a permanent impact that reflects the informational content of trades and a short term impact that represents transactions costs and market frictions. Hence, market efficiency is more about facilitating and accelerating the convergence of short term prices towards their fundamental value (long term).

The market microstructure literature outlines that there is a link between the market organization and the amplitude of both costs and frictions in a market. While market organization allows to manage those costs and frictions, a perfect market organization does not exist. There are often propagation effects between costs and frictions such that it is impossible to achieve at the same time "transparency, liquidity and immediacy" (Schwartz, 1997). Trade-offs have to be made between those characteristics preferred by market participants, such as high levels of transparency, disclosure, immediacy, liquidity, depth and investors' protection.

Mainly four costs and frictions arise from the literature:

- Professional market operators that hold a sizeable concentration of orders in their hands and continuously monitor the market may have the ability to take advantage of market conditions and private information such as past order flows or information (integrated banking and market operators).

- The prominence of information asymmetry between investors incites liquidity suppliers to widen their spreads to compensate the higher risk of trading against more informed traders (adverse selection) that would result in a loss for them.

  On the one hand, information asymmetry is impacted by the transparency of the market in terms of prices, quantities and identities of market participants. This issue is obviously related to the trading systems design that can be summarized into four main types: order-driven market, quote-driven market, price-taking systems and hybrid market.

  Order-driven market can be either continuous or auction (call) market. Most of the time it combines a closing and an opening auction with a continuous trading session.

This system matches traders' orders through an electronic order book where they confront with each other.

Quote-driven market is a system that relies on a decentralized network of market makers that ensure immediacy with bid-ask quotations to fulfil investors' orders. In such system, market makers have to cope with inventory risk.

Hybrid markets are a mix of order-driven and quote-driven markets. It is today the most prominent trading system and it combines the advantage of both order- and quote-driven systems.

On the other hand, information asymmetry is related to the listing requirements for firms. The higher the disclosure standards, the lower the risk for an uninformed trader to trade with a more informed one.

- Uncertainty surrounding transaction completeness raises transaction costs. Operational risks (such as disruption of the IT structure) or counterparty risk (such as the bankrupt of a broker) also raise transaction costs.

- Implicit and explicit transaction costs also result from trading. Indeed, the final investors will have to pay a spread when trading in a market while the handling of orders by intermediaries and exchanges leads to higher transaction costs for the final investors through the payment of commissions.

1.2 Recent Evolution in Market Microstructure

The evolution of financial markets has been marked by the recent advent of derivatives markets and the spread of continuous electronic trading systems as a standard.

While the principle of financial markets remains unchanged over time, the way trading takes place has witnessed some significant changes in the last decades. The main trigger of those changes were technological advances and new regulations. In this section, we review some major changes that occur lately from trading fragmentation to the emergence of dark pools and high frequency trading.

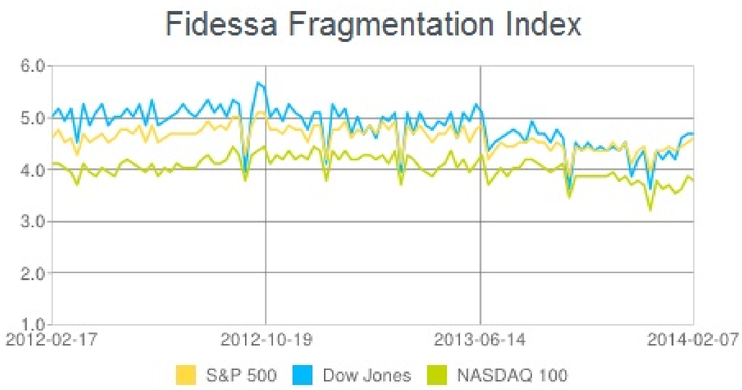
TRADING FRAGMENTATION In the last decade, regulatory changes have favored competition among trading services. The introduction of the Regulation National Market System (Reg NMS in the United States) and the implementation of the Markets in Financial Instruments Directive (MiFID in Europe) put an end to the monopoly of national exchanges and led to the apparition of new trading systems to fulfil trading needs from market participants such as BATS Chi-X.

This competition contributes to lower transactions costs. Indeed, the competition to attract investor's order flows has increased, resulting in a market share loss for incumbent exchanges. Furthermore, the access to electronic trading and more generally financial markets has been improved such that new market participants can easily access it.

Those regulatory changes have significant side effects on market microstructure given they lead to volume fragmentation across multiple trading venues as outlined in Figure

1.1. In turn, we acknoledge more fragmented liquidity, smaller trade sizes and increased technological and supervision costs.

Figure 1.1: Fidessa Fragmentation Index



This figure represents the average number of venues that list the underlying stocks. A value of 1 means the stock is listed on one venue.

DARK POOLS Along with this higher accessibility of electronic financial markets, we observe the emergence of liquidity dark pools. This new type of trading services allows to maintain anonymity and reduces the overall transparency of the market. Indeed, orders and trading activity handled in such systems are not publicly displayed and evade the control of both market participants and regulators. Those new venues increase their market share on a continuous basis [1] and have a significant influence in the current trading context. Those alternative trading systems mainly suit institutional/large traders that can keep their orders hidden from the market. Doing so, they benefit from lower commissions while minimizing the market impact costs of their trades. This comes at the expense of a higher execution uncertainty. The transaction prices in such venues are generally set at a reference markets price. Hence, transactions in liquidity dark pools do not accelerate and smooth price discovery in the main market. This leads to several potential drawbacks of

[1] For a comprehension view on this phenomenon, see http://fragmentation.fidessa.com/

such venues as they get more prominent that include, among others, altering liquidity and price discovery (Degryse et al., 2009).

HIGH FREQUENCY TRADING A recent significant evolution in market microstructure is the emergence of high frequency trading (HFT). HFT has become pervasive in many markets and is a kind of natural consequence of technological advances. Its prominence depends mainly on the possible implementation of low latency trading systems and the presence of fragmented highly liquid markets with small tick sizes as increment. HFT is part of algorithmic trading. A comprehensive definition of what HFT includes is still subject to debate in the literature. According to Castura et al. (2010), HFT encompasses several common characteristics. HFT are professional traders that use high-speed algorithmic strategies. It underlies the use of stock exchanges co-location services and individual data feeds to implement low latency trading systems.

HFT typically exhibits several similar features in their strategy. It implements day trading strategies. It holds no overnight position and closes its position within a single day. Their strategies' time-frames are very short. Finally, HFT is known to implement controversial strategies (Egginton et al., 2013) such as quote stuffing[1] and layering.[2] It leads to the submission of a bunch of orders that are often cancelled before being filled.

Biais et al. (2012) outline that HFT allows to ensure a fair value of assets during the trading session across fragmented markets. Brogaard et al. (2013) empirically show that HFT improves overall market quality and contributes to price discovery. Nevertheless, the

[1]Quote stuffing consists in submitting a large number of orders followed immediately by a cancellation to generate order congestion.

[2]Layering consists in submitting a large number of orders in one side of the book that are not meant to be filled to facilitate the entry on the other side of the book.

externalities of HFT remain a topical issue. Indeed, the flash crash of May, 6th 2010 casts doubt on the soundness of HFT activity. Furthermore, Hendershott (2011) shows that individual stock crashes have increased along with an overall market quality improvement. HFT are often mentioned as the prime suspect for this higher market instability. Indeed, HFT became so prominent that the periodicity in their order submission dynamics from liquidity provision to liquidity taking activities may potentially induce market stability issues. The quality of the HFT liquidity provision is also questioned, most of the submitted HFT orders being cancelled before being filled. Another potential concern on HFT is the adverse selection it creates for slow traders.

All those evolutions show that liquidity is a corner stone of a sound financial market. In such an evolving trading landscape, we contribute to the empirical literature on market microstructure. First, we provide a novel methodology to forecast intraday liquidity that helps market participants reduce the execution price uncertainty surrounding their orders. Second, we focus on the recent increase of market disruptions in the stock market (Hendershott, 2011). In addition to proposing a framework to forecast price jumps, we investigate the behavior of HFT traders around the inception of price jumps. We contribute to a better understanding of those recent evolutions in the financial market. Hence, the implications of our work impact both market participants, market operators and regulators.

1.3 A Three Essays Doctoral Research

1.3.1 Essay 1

In this first essay, we provide a novel methodology to forecast periods of intraday low execution price uncertainty.

The first concern is to define the most suitable liquidity proxy for our purpose. Market microstructure theory outlines that an exhaustive definition of market liquidity typically refers to the four dimensions identified in Harris (2003), i.e. tightness, depth, resiliency and immediacy.

Given its multidimensional feature, liquidity cannot be captured by a single figure. Hence, practitioners must strike a balance between the different liquidity dimensions to select the best proxy according to their needs.

Two broad categories of liquidity proxies coexist in the literature, namely trade-based and order book-based measures, even though they display little correlation. In turn the choice of liquidity proxies may impact our result outcomes such that it is of utmost importance to justify our proxy choice.

As outlined by Irvine et al. (2000), investors can be patient or impatient to trade. On the one hand, impatient investors, such as arbitrageurs, require an immediate execution of their orders. On the other hand, patient investors, such as institutional investors who rebalance or hedge their portfolio, look for occasional high liquidity time period to execute their orders inexpensively.

In this essay, we investigate liquidity from the viewpoint of patient investors. In this framework, a liquid market is a market where the uncertainty surrounding the execution price is low. Hence, we rely on the Amihud ratio that is a liquidity proxy benchmark and measure liquidity through volume price impact.[1] This ex-post liquidity measure neglects the immediacy dimension but allows to account for both potential uncommitted liquidity[2] and market resiliency[3] to assess the level of execution price uncertainty.

Our paper puts forward a novel aggregation methodology of continuous data that is based on a fixed trading volume instead of a time-based interval. This novel methodology introduces a twofold advantage for liquidity forecasting.

First, the Amihud illiquidity ratio underlies a linear relationship between volume and volatility. Kempf and Korn (1997) emphasize that this strong assumption leads to a misleading interpretation of the ratio. Using a fixed trading volume aggregation allows us to bypass this non-linearity issue.

Second, Irvine et al. (2000) address the idea that a market may be at the same time liquid for small orders and illiquid for large orders. Our methodology avoids optimal sampling frequency issues. Furthermore, we consider liquidity as a trade size specific concept given the fixed trading volume of the event chart, which can be adjusted to the order size considered.

[1]The Amihud ratio is defined as follows: $Amihud_t = \frac{|R_t|}{V_t}$, where $|R_t|$ is the absolute return and V_t is the trading volume at time t.

[2]Uncommitted liquidity is hidden liquidity that is not displayed in the order book.

[3]Resiliency is the speed with which prices recover from a random, uninformative shock. Large (2007) and Degryse et al. (2005) outline the particular importance of resiliency in the case of an electronic limit order book, where the price-smoothing agency of a specialist or market maker is absent.

Using time series models on a sample of 81 Euronext stocks included in national indexes, we select the optimal set of explanatory variables through the general-to-specific process presented in Hendry and Krolzig (2001). We then run the Hansen (2005)'s SPA test on two mainstream loss functions to document the performance of our forecasting models against a random walk model, both in- and out-of-sample.

We outline some predictable patterns in the Amihud ratio. All our forecasting models statistically outperform our benchmark random walk model both in and out-of-sample. Consistent with Cont et al. (2014), we show that trade-based and order book-based liquidity measures are informative of forthcoming price impact. To investigate the robustness of our results to order size changes, we carry out the same empirical study on other time frames. We find very consistent forecasting performance across time frames.

Using the ARX-GARCH model, we conclude the empirical part with an illustration of an execution price improvement strategy that statistically outperforms a naive strategy.

On average, the closing of the continuous trading day seems to be the best time for institutionals to hedge or rebalance a portfolio. Indeed, it is characterized by a negative - albeit stock specific- periodicity, a higher forecast ability of the Amihud ratio and a higher trading volume.

1.3.2 Essay 2[1]

In this essay, we investigate whether illiquidity can forecast liquidity. A recent evolution in financial markets is the increase in individual stock price jumps (Hendershott, 2011). While the occurrence of price jumps has numerous implications, the root of those disruptions remains uncertain.

On the one hand, Joulin et al. (2008) and Weber and Rosenov (2005) document that price jumps are liquidity-driven. Indeed, they find that the inception of price jumps lies in a low density of the order book in spite of the sudden surge of trading volume. Boudt and Petitjean (2014) confirm that time-varying liquidity provision is the trigger for about 70% of price jumps in the DJIA index.

On the other hand, Jiang et al. (2010) and Miao et al. (2012) find that most jumps are information-driven using US treasury bond market and stock index futures market.

The question of the predictability of price jumps is a widely unexplored topic. To fill this gap, we provide a framework to forecast such disruptions using the informational content of market microstructure variables on Euronext 100 stocks.

The definition of "a price jump" is here of utmost importance. We consider various detection methods that encompass the straightforward percentile of ex-post return distribution and robust jump tests to time-varying volatility and periodicity in the stock market (Lee and Mykland, 2008). Our dataset is aggregated using a fixed trading volume that allows us to deal with non-trading issues at high frequency aggregation.

[1]This essay is a joint work with Catherine D'Hondt.

Due to higher information asymmetry that translates into a lower liquidity in the early continuous trading session (Cont et al., 2014), we observe most price jumps happen during that time period even after accounting for periodicity in the stock market.

Our forecasting framework includes a logistic model and an extension adjusted to rare events. The optimal set of explanatory variables is set through a general-to-specific model selection process based on the likelihood function (Castle and Hendry, 2011).

We highlight some predictable patterns in price jumps mainly at high frequency aggregation. We find order book measures are informative of forthcoming price jumps, which supports that the trigger for price jumps lies in time-varying liquidity provision. We point out that trading activity during price disruptions is often pretty low. The drivers of this illiquidity is still an open question whether due to order cancellations or a low resiliency of the order book.

We conclude with an illustration of an execution price management strategy for value and passive market participants. We show that there is an incentive not to trade when our model detects a risk of price jumps. We manage to statistically decrease the price execution uncertainty by ruling out intervals with a risk of jump occurrence in our out-of-sample dataset.

1.3.3 Essay 3[1]

The last essay is devoted to the recent emergence of HFT activity. Financial market stability became a topical issue following the 2008 financial crisis and the 2010 flash crash. As mentioned earlier, HFT is often mentioned as a contributor to this higher instability. Preventive actions were already taken by some exchanges to prevent potential detrimental behaviors from HFT.

We shed light on the trading behavior of high frequency traders around idiosyncratic (stock specific) and systematic (market-wide) price jumps. HFT could trigger or emphasize price jumps in various ways. In times of higher instability, HFT firms may exit the market or add to one-sided order imbalance. HFT may also switch from liquidity provision to liquidity taking activity in given market states.

In this essay, we proxy for market instability using the 99% percentile of midquote observations as a threshold.

We use the Nasdaq HFT dataset used in other researches (e.g. Brogaard et al. (2013)). The data divide market participants into two types', HFT and non-HFT (nHFT). The data also disclose which type of participant is taking and providing liquidity for each trade.

We document that HFT are more active during market disruptions. The one-sided order book activity for HFT/nHFT around jumps reveals that they both trade aggressively when in direction of the price jump and passively when against the price jump direction. All in all, we find that neither HFT nor HFT exhibit a significant net position pattern.

[1]This essay is a joint work with Jonathan Brogaard and Ryan Riordan.

Yet, it confirms that HFT accumulate an inventory position in the opposite direction of the jump.

The likelihood of permanent price jumps increases when lagged HFT net position is in the direction of the price jump. It reflects that the higher HFT activity leads to a quicker adjustment of stock prices to information, which in turn may explain the prominence of stock specific price jumps outlined in Golub et al. (2013). At the opposite, transitory price jumps go along with HFT net position against the price jump direction.

To conclude, we evaluate whether HFT firms may have an incentive to try and trigger market disruptions. Their trading profits around price jumps suggest that HFT firms have no obvious incentives to foster the inception of disruptions. Overall, nHFT make profits on their aggressive trading activity while HFT lose money on their passive trading activity during the jump interval. In all, we find no significant profit pattern during price jumps whether for HFT or nHFT.

CHAPTER

TWO

THE INFORMATION CONTENT OF THE AMIHUD RATIO FOR INTRADAY LIQUIDITY FORECASTING

2.1 Introduction

Liquidity is a key concept for market efficiency and financial system stability. The expansion of high frequency trading, the 2010 flash crash and the implementation of new regulations, such as Reg NMS in the United States and MiFID in Europe, made liquidity issues a hot topic drawing the attention of both practitioners and academics.

In the theory on market microstructure, an exhaustive definition of market liquidity typically refers to the four dimensions identified in Harris (2003), i.e. tightness, depth, resiliency and immediacy. The literature brings out two broad categories of liquidity proxies, namely trade-based and order book-based measures. As they show little correlation, a bunch of research focuses on determining which type of measure is the best liquidity proxy. Goyenko et al. (2009), among others, show that the choice of liquidity proxy may significantly impact our results.

Given its multidimensional feature, liquidity cannot be captured by a single figure. Hence, practitioners must strike a balance between the different liquidity dimensions to select the best proxy according to their needs. As outlined by Irvine et al. (2000), investors can be patient or impatient to trade. On the one hand, impatient investors, such as arbitrageurs, require an immediate execution of their orders. On the other hand, patient investors, such as institutional investors who rebalance or hedge their portfolio, look for occasional high liquidity environment to execute their orders inexpensively.

In this paper, we investigate liquidity from the viewpoint of patient investors. In their view, a liquid market is a market where the uncertainty around the execution price is low. In this framework, liquidity is set through volume price impact as measured by the Amihud ratio.[1] This ex-post liquidity measure neglects the immediacy dimension but allows to capture both potential uncommitted liquidity[2] and market resiliency[3] to assess the level of execution price uncertainty.

Our methodology is based on a fixed volume event chart and has a threefold advantage. First, Kempf and Korn (1997) outline a major pitfall in the widely used Amihud illiquidity

[1]The Amihud ratio is defined as follows: $Amihud_t = \frac{|R_t|}{V_t}$, where $|R_t|$ is the absolute return and V_t is the trading volume at time t.

[2]Uncommitted liquidity is hidden liquidity that is not displayed in the order book.

[3]Resiliency is the speed with which prices recover from a random, uninformative shock. Large (2007) and Degryse et al. (2005) outline the particular importance of resiliency in the case of an electronic limit order book, where the price-smoothing agency of a specialist or market maker is absent.

ratio as it implies a linear relationship between volume and volatility, which is typically not the case. Our fixed trading volume aggregation allows us to make trading volume exogenous and bypasses the use of a ratio. Second, Irvine et al. (2000) address the idea that a market may be at the same time liquid for small orders and illiquid for large orders. Our event chart allows us to assess liquidity as a trade size specific concept and avoid optimal sampling frequency issues. Indeed, the fixed trading volume of the event chart can be adjusted to the order size considered by market participants. Third, by setting a fixed trading volume, we turn liquidity into a volatility proxy, which allows us to transpose the extensive literature on volatility forecasting into the liquidity framework.

Using time series models on a sample of Euronext stocks, we document some predictable patterns in volume price impact that statistically outperform a random walk model in- and out-of-sample. Consistent with Cont et al. (2014), we show that trade-based and order book-based liquidity measures are informative of forthcoming price impact. Among the four tested models, the ARX(3)-GARCH(1,1) model stands out from the crowd while the ARMAX model underperforms the other models. Robustness checks confirm the consistency of our results across stocks and different time frames.

We also provide an example of execution price improvement strategy that statistically outperforms a naive strategy. On average, the end of the continuous trading day seems to be the best time for institutionals to hedge or rebalance a portfolio. It is indeed characterized by a slightly -but stock specific- negative periodicity, a higher predictability of the execution price and a higher trading volume.

The paper is structured as follows. Section 2.2 provides an overview of the existing literature related to liquidity proxies and liquidity forecasting. Section 2.3 presents the applied methodology as well as the data sample used in the empirical part. Section 2.4 reports the in-sample and out-of-sample results as well as several robustness checks. Finally, Section 2.6 summarizes the findings and identifies possible extensions.

2.2 Literature Review

The plethora of liquidity proxies in the literature illustrates that liquidity remains an elusive concept. Market liquidity is an unobserved process with multidimensional characteristics, which makes it impossible to measure by a single figure. Practitioners and academics have to strike a balance between the different liquidity dimensions to select the best proxy for their purpose.[1]

In this section, we only present the benchmark of ex-post (trade-based) measures that suits best our research purpose.

The benchmark of ex-post liquidity measures is the Amihud illiquidity ratio, presented in Amihud (2002). This category of proxies is mainly used when market timing is neglected, such as for brokers or hedgers who rather focus on improving execution price. The Amihud ratio measures liquidity through the relationship between traded volume and absolute return:

$$Amihud_t = \frac{|R_t|}{V_t}$$

where $|R_t|$ is the absolute return and V_t is the trading volume at time t.

The shortcomings of this measure, and more broadly of all ex-post measures, is that

[1]For the sake of simplification, two possible biases of liquidity proxies are often neglected in the literature. On the one hand, liquidity is asymmetric as confirmed by Kalay et al. (2004). Buy orders lead to significant higher price impact than sell orders and the price impact is higher for both in market decline. This cannot be accounted for through a unique measure. On the other hand, Hall and Hautsch (2007) show that financial markets are driven by both information and liquidity. Liquidity proxies do not filter price impact as either liquidity driven or information driven.

they cannot be computed in the absence of trading volume. For illiquid stocks, this measure is thus less relevant as it may be lagging because of the absence of trading volume. Furthermore, trade-based measures do no consider liquidity as a trade size specific concept. Irvine et al. (2000) show that a market may be at a point in time liquid for small orders but illiquid for large orders. Eventually, Kempf and Korn (1997) point out the fact that the use of a ratio implies that the relationship between trading volume and absolute return is linear which is typically not the case. This non-linear relationship entails misleading interpretation of the Amihud ratio.

The relationship between liquidity and price movements in the financial market is well documented.

In a quote-driven market environment,[1] Dufour and Engle (2000) extend the Hasbrouck (1991)'s vector autoregressive model for prices and trades to highlight some correlation between trading frequency and price impact. They find that high trading activity is related to both larger quote revisions and stronger positive autocorrelation of trades. Kyle and Obizhaeva (2013) support the idea of an invariant market microstructure across assets when measured in units of business time. In this framework, Bialkowski et al. (2008) outline some predictable patterns in trading volume and show it can be used to better manage transaction cost compared to a VWAP strategy.

Next, Engle and Lange (2001) develop a new liquidity proxy, VNET, that captures the depth of the underlying stock corresponding to the net buy and sell volume during a given price movement. Using an event chart where the event is a given price movement, they also forecast the time to complete a fixed price movement relying on an ACD model

[1]A quote driven market is a market that only displays bids and asks of designated market makers and/or specialists for a specific security.

variant. This forecasted time to achieve a given price movement could, then, be used to set up market timing strategies. Next, Tombeur and Wuyts (2011) highlight that order flow imbalance is a better predictor of momentum than order book imbalance in the short run. However, order book imbalance remains informative in the long run which is not the case for order flow imbalance.

Finally, Cont et al. (2014) outline order book-based and trade-based measures are both informative of future price impact. They show that highly correlated market orders lead to momentum while limit orders lead to range-bound market. They also find that order flow imbalance and momentum exhibit a linear relationship with a slope inversely proportional to depth.

2.3 Methodology

2.3.1 Sample

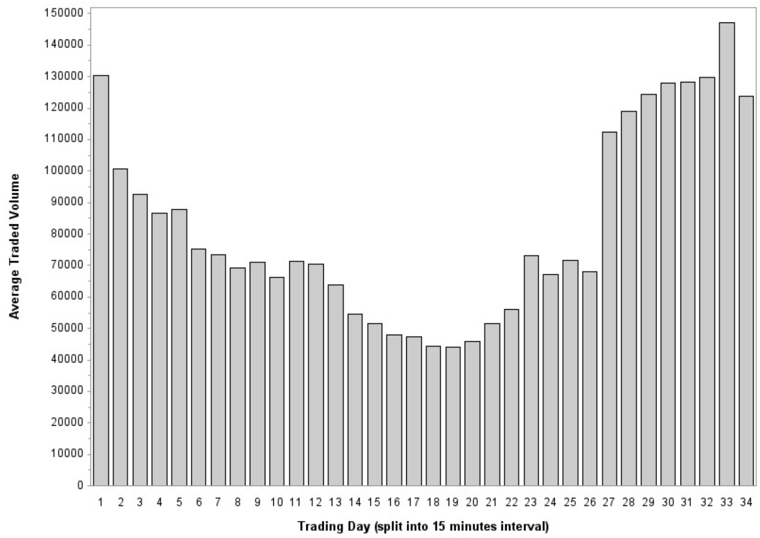
The data pertain to 81 stocks that belong to three national major European NYSE Euronext stock indexes: CAC40, AEX and BEL20. The sample period spans 61 trading days (from February 1, 2006 to April 30, 2006). Relying on trades, orders and quotes, the database has been rebuilt using a replication of the Euronext algorithm. This allows us to use trade and order book data to compute an extensive set of liquidity measures.

We arrange our database in event-based intervals. Hasbrouck (1996) outlines the advantage of event-based charts over mainstream time-based charts for forecasting purposes as they tend to be less noisy. In this paper, an interval event is defined as a given trading volume, which allows us to consider liquidity as a trade size specific concept. Furthermore, it also avoids optimal sampling frequency issue. To ensure consistency across stocks, we set a fixed trading volume for each stock that corresponds to the average 15-minute trading volume observed for the stock over the whole sample period. Indeed, Oomen (2006) shows that sampling in transaction time is superior to calendar time sampling in that it leads to a further reduction of the mean squared error loss function.

Our final data consist in 154,633 intervals split across 81 stocks. For each stock, the observations are split into two equal databases, the first half is for in-sample model fitting and the second half is for out-of-sample model forecasting performance.

We apply some filters to the data in order to ensure the relevance of our results. We

Figure 2.1: Volume Periodicity



Distributional properties of trading volume. The chart represents the average trading volume (in euro) across the 81 stocks. The continuous trading day is discretized in 34 intervals of 15 minutes.

remove the first interval of each continuous trading day to avoid referring to previous day order book data. We also cancel the last observation of the continuous trading day because the event, that is trading volume, on that interval may be uncompleted. We should mention as well that the trigger for a new interval is the order that crosses the fixed volume threshold. The traded volume per interval may thus slightly vary in order to avoid order splitting.

Figure 2.1 reports the distributional properties of the average trading volume across the 81 stocks over our sample period. As expected, the well-known U-shaped volume distribution is observed and our observations are thus outweighed at both the beginning and the end of the continuous trading day.

We detect execution price uncertainty periods through the Amihud ratio. Building on Alizadeh et al. (2002) and De Vilder and Visser (2008) who document the higher efficiency of High-Low compared to absolute returns and squared returns as a volatility proxy, we settle for the High-Low measure. We also consider other measures, such as absolute return and squared returns, that confirm our initial findings[1]. Realized volatility measures were also considered even though it is less relevant for our research purpose given market resiliency could lead to misleading realized volatility measures. Our Amihud ratio (AR) is thus defined as follows:

$$AR_{i,t} = \frac{HL_{i,t}}{V_{i,t}}$$

where $HL_{i,t}$ is the High-Low and $V_{i,t}$ is the traded volume for stock i during interval t.

In the framework of a volume event chart methodology, we set a fixed traded volume such that the measure fines down to the $HL_{i,t}$ measure. This allows us to consider liquidity as related to a given trading size (by adjusting our event chart to the order size considered). Our adjusted Amihud ratio (AAR) measure is defined as:

$$AAR_{i,t} = HL_{i,t}$$

where $HL_{i,t}$ is the High-Low for a fixed traded volume for stock i during interval t.

2.3.2 Explanatory Variables

Thanks to our rich database, we compute an extensive set of measures including Kang and Yeo (2008)'s dispersion and Næs and Skjeltorp (2006)'s slope, among others. For the

[1]Results are available upon request

sake of convenience, this section only reviews the explanatory variables that are used in the final model.

The final model includes an exogenous trade-based measure that is the number of trades. This is a state variable that is thus summed over the interval. In the framework of our event chart, the number of trades ($\#Trades_{i,t}$) for a fixed trading volume represents the average trading size and is defined as follows:

$$\#Trades_{i,t} = \sum_{j=1}^{J} N_{i,j,t}$$

where N is the number of trades for stock i during interval t and J is the number of trades during interval i.

The remaining measures are order book-based measures. They represent a snapshot of the order book at a given point in time and are taken at the end of each interval. Those measures are respectively the depth at the best and the five best bid and ask quotes ($Q1_{i,t}$ and $Q5_{i,t}$), the relative spread ($RS_{i,t}$) and both bid and ask slopes ($BidSlope_{i,t}$ and $AskSlope_{i,t}$) which capture order book imbalance. They are defined as follows:

$$Q1_{i,t} = Q_{B,i,t} + Q_{A,i,t}$$

where $Q_{B,i,t}$ is the committed bid quantities and $Q_{A,i,t}$ is the committed ask quantities display at the best limit prices for stock i at the end of interval t.

$$Q5_{i,t} = \sum_{j=1}^{5} Q_{B,j,i,t} + Q_{A,j,i,t}$$

where $Q_{B,j,i,t}$ is the committed bid quantities and $Q_{A,j,i,t}$ is the committed ask quantities display at the j^{th} best limit prices for stock i at the end of interval t.

$$RS_{i,t} = \frac{P_{A1,i,t} - P_{B1,i,t}}{MQ_{i,t}}$$

where $P_{A1,i,t}$ is the best ask and $P_{B1,i,t}$ is the best bid for stock i at the end of interval t and $MQ_{i,t}$ is the midquote at the end of interval t.

The cost of round trip measure captures both the tightness and the depth dimension of the underlying stock. To calculate it, we take the five best quotes of the order book into account and then assume a perfect elasticity of prices beyond as it is usually done in the literature (Jain et al., 2011). The trade size is lined up with the time frame of the event chart, that is the 15-minute average trading volume on the underlying stock. The measure is computed as follows for stock i at the end of interval t:

$$AbsCRT_{i,t} = \frac{\mid \sum_{k=1}^{5} I_k^{Buy} \left(Midquote_{i,t} - P_k^{Buy} \right) + \sum_{k=1}^{5} I_k^{Sell} \left(P_k^{Sell} - Midquote_{i,t} \right) \mid}{T \times Midquote_{i,t}}$$

where $I_k^{Buy/Sell}$ is the number of shares bought or sold respectively at the k^{th} price limit of the order book and $P_k^{Buy/Sell}$ is the price at the k^{th} quote. T is the total number of shares to be bought or sold, with

$$
\begin{array}{lll}
I_k^h = Q_j^h & \text{if} & T > \sum_{j=1}^{k} Q_j^h, \\
I_k^h = \left(T - \sum_{j=1}^{k-1} Q_j^h\right) & \text{if} & T > \sum_{j=1}^{k-1} Q_j^h \quad and \quad T < \sum_{j=1}^{k} Q_j^h, \\
I_k^h = 0 & & \text{otherwise}
\end{array}
$$

where $h \in [Buy, Sell]$ and Q_j^h is the displayed quantity at the j^{th} price limit of the order book.

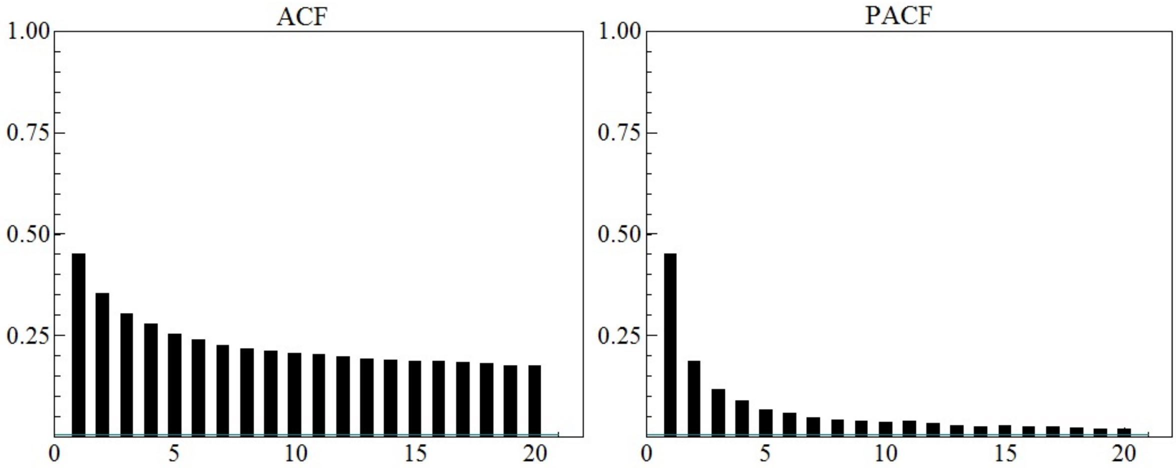
2.3.3 Quantitative Analysis

As a preliminary step, we point out some characteristics of the Adjusted Amihud Ratio (AAR). The main purpose is to outline some stylized features with volatility in the AAR to support our model selection. Hence, we pay specific attention to clustering (distribution skewness and kurtosis), heteroscedasticity and autocorrelation. The stationarity of the measure is tested using the Phillips-Perron test. The AAR displays strongly significative stationarity in our sample.

Figure 2.2 presents the AAR autocorrelogram, which outlines some autocorrelation in the proxy that nevertheless converges rather quickly to zero. We apply the extended sample autocorrelation function, presented in Tsay and Tiao (1984), to test the relevance of an ARMA dynamics. An ARMA (1,1) process fits two out of three stocks of our sample [1].

[1] The ACF suggests the presence of a long memory process that could be considered on a longer dataset

Figure 2.2: Autocorrelation Function of the Adjusted Amihud Ratio



The autocorrelation function (ACF) of the AAR is displayed for 20 lags on the left-hand side chart. The right-hand side chart plots the partial autocorrelation function (PACF) of the AAR up to 20 lags.

As reported in Table 2.1, the AAR displays highly positive excess kurtosis. We carry out ARCH disturbances test on the residuals that reveals a long-lasting and consistent ARCH effect across stocks. This is a strong support for GARCH (1,1) model. The normality test shows that our measure is not normally distributed. The positive skewness suggests that we should consider extensions of the GARCH model that account for asymmetry as well.

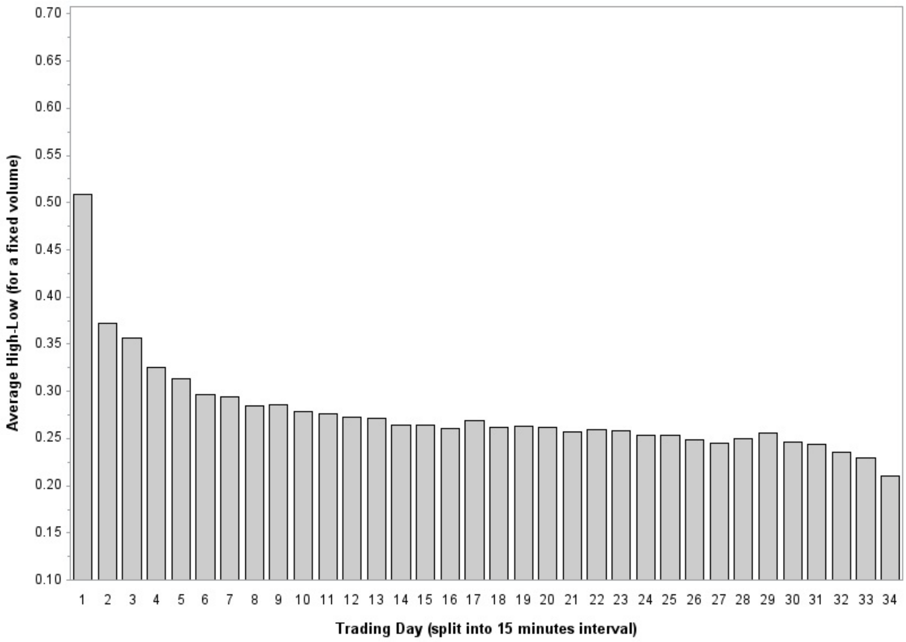
Table 2.1: Descriptive Statistics of the Adjusted Amihud Ratio

Skewness	1.5781	Minimum	0
Excess Kurtosis	6.0783	Maximum	2.7231
Mean	0.2720	Normality test	<0.0001
Standard Deviation	0.1692	ARCH test 1-10	<0.0001
Median	0.2430	AR test 1-2	<0.0001

The table presents key descriptive statistics of the AAR across the 81 stocks for an average 15-minute volume observed for the stock over the whole sample period.

The AAR exhibits periodicity as plotted in Figure 2.3 . At the beginning of the continuous trading day, volume price impact is significantly higher. This is consistent with Cont et al. (2014) who show that there is lower depth in the morning because of information asymmetry. The very end of the continuous trading day displays a slightly lower price impact as well. We should point out that, while trading volume is lower at midday, there is no observed effect on our measure.

Figure 2.3: Adjusted Amihud Ratio Periodicity



The chart plots the average AAR measure across the 81 stocks for an average 15-minute trading volume on the underlying stock over the whole sample period. We rank the observed AAR measures with regard to the time they occur during the trading session. For this purpose, we discretize the continuous trading day in 34 intervals of 15 minutes.

2.3.4 Model Presentation

We use the general-to-specific model selection process, presented in Hendry and Krolzig (2001), to select the optimal set of explanatory variables. This procedure simplifies a general unrestricted model (GUM) to a congruent model. The selection encompasses test for model mis-specifications and diagnostic tests at each simplification step.

We consider three extensions of the model relying on the quantitative analysis[1] of the adjusted Amihud ratio in Section 2.3.3, the models are presented in details hereafter:

- The ARX(3) model

$$Log(AAR_{i,t}) = \beta_{i,0} + \textstyle\sum_{j=1}^{3} \gamma_{i,j} Log(AAR_{i,t-j}) + \beta_{i,1} RS_{i,t-1} + \beta_{i,2} Q1_{i,t-1}$$
$$+\beta_{i,3} Q5_{i,t-1} + \beta_{i,4} AbsCRT_{i,t-1} + \beta_{i,5} \#Trades_{i,t-1}$$
$$+ \textstyle\sum_{j=1}^{2} \alpha_{i,j} SE_{i,j} + \sum_{j=33}^{34} \alpha_{i,j} SE_{i,j} + \varepsilon_{i,t},$$

  where $AAR_{i,t}$ is the High-Low for stock i at interval t, $AAR_{i,t-j}$ is the autoregressive term for stock i, $SE_{i,1}$ and $SE_{i,2}$ are dummies for the two first 15-minute time-based intervals for stock i that capture the beginning of day periodicity in the dependent variable and $SE_{i,33}$ and $SE_{i,34}$ are dummies for the two last 15-minute time-based intervals for stock i. Other explanatory variables were defined previously in section 2.3.2.

- The ARMAX (1,1) model

$$Log(AAR_{i,t}) = \beta_{i,0} + \gamma_{i,1} Log(AAR_{i,t-j}) + \delta_{i,1}\varepsilon_{i,t-1} + \beta_{i,1} RS_{i,t-1} + \beta_{i,2} Q1_{i,t-1}$$
$$+\beta_{i,3} Q5_{i,t-1} + \beta_{i,4} AbsCRT_{i,t-1} + \beta_{i,5} \#Trades_{i,t-1}$$
$$+ \textstyle\sum_{j=1}^{2} \alpha_{i,j} SE_{i,j} + \sum_{j=33}^{34} \alpha_{i,j} SE_{i,j} + \varepsilon_{i,t},$$

  where $AAR_{i,t}$ is the High-Low for stock i at interval t, $AAR_{i,t-j}$ is the autoregressive term for stock i, $\varepsilon_{i,t-1}$ is the moving average term for stock i, $SE_{i,1}$ and $SE_{i,2}$

[1] We also consider component-level forecasting as an alternative modeling. As its forecasting performance is lower than our core methodology, we do not report it in the paper.

are dummies for the two first 15-minute time-based intervals for stock i that capture the beginning of day periodicity in the dependent variable and $SE_{i,33}$ and $SE_{i,34}$ are dummies for the two last 15-minute time-based intervals for stock i. Other explanatory variables were defined previously in section 2.3.2.

- The ARX (3)-GARCH (1,1) model

$$\begin{aligned}
Log(AAR_{i,t}) &= ARX(3)model, \\
\varepsilon_{i,t} &= z_{i,t}\sqrt{h_{i,t}}, \\
z_{i,t} &\sim N(0,1), \\
h_{i,t} &= \omega_{0,i} + \lambda_{i,1}\varepsilon^2_{i,t-1} + \theta_{i,1}h_{i,t-1},
\end{aligned}$$

where $h_{i,t}$ is the conditional variance for stock i at interval t.

- The ARX (3)-EGARCH (1,1) model

$$\begin{aligned}
Log(AAR_{i,t}) &= ARX(3)model, \\
\varepsilon_{i,t} &= z_{i,t}\sqrt{h_{i,t}}, \\
z_{i,t} &\sim N(0,1), \\
log(h_{i,t}) &= \omega_{0,i} + \lambda_{i,1}log(h_{i,t-1}) + \theta_{i,1}\left(\frac{|\varepsilon_{i,t-1}|}{\sqrt{h_{i,t-1}}} - \sqrt{\frac{2}{\pi}}\right) + \zeta_{i,1}\left(\frac{\varepsilon_{i,t-1}}{\sqrt{h_{i,t-1}}}\right),
\end{aligned}$$

where $h_{i,t}$ is the conditional variance for stock i at interval t.

The model calibration is made through maximum likelihood function using Levenberg-Marquart optimization algorithm. The initial values are set to the parameter values of the pooling regression over the whole sample period. As autoregressive models do not impose a strictly positive value for the dependent variables, that is the adjusted Amihud ratio, we take the log-transformation of the dependent variable to ensure positivity of our dependent variable in the empirical part.

As mentioned earlier, we divide the sample period into an estimation period (first half of data) that is used for fitting the model in-sample and an evaluation period (second half of data) used for out-of-sample forecasting. For the out-of-sample data, we perform a rolling regression. Each model is estimated over an initial window that corresponds to the estimation period and a 1-period ahead liquidity forecast is performed. The window is then rolled forward one period to obtain the second forecast and so on and so forth for all the out-of-sample data.

We use two mainstream loss functions to measure the discrepancies of the model compared to the realized value: the Mean Squared Error (MSE) and the Mean Absolute Error (MAE). As a remainder, the MSE loss function gives some extra penalty to models that display large errors as the observed discrepancy is squared. The MSE is defined as:

$$MSE_{i,t} = \frac{1}{N}\sum_{j=1}^{N}\left(\widehat{AAR}_{i,t} - AAR_{i,t}\right)^2$$

where $\widehat{AAR}_{i,t}$ is the estimated High-Low for stock i at interval t and $AAR_{i,t}$ is the realized High-Low for stock i at interval t. The MAE is defined as:

$$MAE_{i,t} = \frac{1}{N}\sum_{j=1}^{N} \mid \widehat{AAR}_{i,t} - AAR_{i,t} \mid$$

where $\widehat{AAR}_{i,t}$ is the estimated High-Low for stock i at interval t and $AAR_{i,t}$ is the realized High-Low for stock i at interval t.

We test the statistical outperformance of our models against a random walk model through the superior predictability ability test (SPA) presented in Hansen (2005). Hansen's

SPA test is less prone to data mining issues than the Diebold-Mariano test. The SPA test addresses the null hypothesis that any alternative forecasting models are better than a benchmark model, here a random walk model.

2.4 Results

2.4.1 In-sample Analysis

We report the fit summary statistics of each model in Table 2.2. For the mean modeling, the ARX model obviously outperforms the ARMAX model with a better Akaike information criteria[1], Schwarz information criteria[2] and adjusted R^2 statistics. The modeling of the conditional variance, through both a GARCH and an EGARCH process, seems to improve the AIC information criteria while the SBC information criteria worsen. The standard deviations across stocks of the fit summary are pretty consistent across models.

Table 2.2: Fit Summary Statistics

	AIC		SBC		Adj. R-square	
	Mean	Std	Mean	Std	Mean	Std
ARX	1355.40	292.94	1418.28	292.29	0.250539	0.057047
ARMAX	1362.26	289.98	1426.40	289.37	0.246155	0.05625
ARX-GARCH	1353.04	292.14	1423.03	291.14	0.252046	0.055865
ARX-EGARCH	1350.89	291.38	1423.42	290.07	0.251585	0.055577

The table reports the mean and standard deviation of key fit statistics across the 81 stocks for the four forecasting models. AIC is the acronym for Akaike Information Criteria. SBC is the acronym for Bayesian Information Criteria or Schwarz criterion. The difference between the SBC and the AIC lies in the penalty term for the number of explanatory variables. SBC penalty term is higher than AIC penalty term.

As diagnostic test for our models, we run the Ljung-Box Q test on the ARX-GARCH model standardized residuals and squared standardized residuals for lags 5, 10, 20 and 50

[1]The AIC is defined as follows: $AIC = -2\ln(ML) + 2k$, where k is the number of parameters in the statistical model and ML is the maximum likelihood function

[2]The SC, also known as BIC or SBC, is defined as follows: $SC = -2\ln(ML) + k\ln(n)$, where k is the number of parameters in the statistical model, n is the sample size and ML is the maximum likelihood function

in Table 2.3. It unveils the remaining presence of some long-memory effects in our model as expected while the heteroscedasticity is fully captured by our model.

Table 2.3: Box-Ljung Test

	Standardized residuals		Squared Standardized Residuals	
	Q-Statistics	P-Values	Q-Statistics	P-values
Lag 5	1.72371	0.1892161	2.02896	0.5664176
Lag 10	9.27432	0.1587278	8.96788	0.3450125
Lag 20	31.1801	0.0127612*	13.7489	0.7453102
Lag 50	81.8871	0.0008887**	53.7309	0.2641563

The table displays the Q statistics ($Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$) where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k and h is the number of lags. The base hypothesis is that there is no serial correlation. Hence, we accept H0 when the p-value is high that is when the Q-statistic is lower than the quantile value of a chi-squared distribution with h freedom degree.

As the ARX model better captures the mean process than the ARMAX model, we only report the ARX model parameter estimates in Table 2.4. The other models parameter estimates are provided in Table 2.11, 2.12 and 2.13 (in Appendix).

Table 2.4: Parameter Estimates of the ARX Model

Variable	Estimate	StdErr	**
AAR_1	0.190860	0.0352	100%
AAR_2	0.082498	0.0338	68%
AAR_3	0.064203	0.0337	46%
$absCRT$	0.619252	0.1648	78%
N	0.002369	0.0008	67%
$Q1$	-0.000010	0.0000	30%
$Q5$	-0.000008	0.0000	84%
RS	0.827505	0.4006	51%
SE_1	0.162219	0.1332	37%
SE_2	0.119509	0.1044	34%
SE_{33}	-0.110880	0.0818	34%
SE_{34}	-0.199313	0.1314	41%

This table reports the parameter estimates of the ARX(3) model. Exogenous variables are those presented in Section 2.3.2. $SE_{i,1}$ and $SE_{i,2}$ are dummies for the two first 15-minute time-based intervals for stock i that capture the beginning of day periodicity in the dependent variable and $SE_{i,33}$ and $SE_{i,34}$ are dummies for the two last 15-minute time-based intervals for stock i. The column ** reports the percentage of stocks for which the parameter is significant at a 95% confidence interval.

The set of explanatory variables consists in both trade-based and order book-based measures. It is consistent with Cont et al. (2014) who show that both measures are informative of future volume price impact. Table 2.4 points out some parameters that are significant across almost all the stocks: the first autoregressive lag, depth at the five best quotes and, to a lesser extent, the absolute cost of round trip. The average trading size and the spread is also informative of future liquidity for about half of the 81 stocks. The first hour of the continuous trading day displays a positive periodicity while the end of the trading day exhibits a negative periodicity. This periodicity is though somewhat stock specific with about 40% of the 81 stocks with a significant periodicity. The economic significance of our explanatory variables outlines that we expect a liquid period when the spread is tight, the average order size is larger, the depth at the best quotes is higher and the cost of round trip trade is low (that is no imbalance in the order book).

The GARCH parameters are reported in Table 2.12 available in Appendix. They all exhibit significant parameter for about half of the sample stocks. The EGARCH model parameters, in Table 2.13 available in the Appendix, also display significant parameter across about half of the stocks, except for the asymmetry parameter "Θ". Only a tiny part of the sample stocks turns to have a significant asymmetry in their conditional variance.

In a second step, we compare the performance of the forecasting models against a random walk model for both loss functions in Table 2.5. All the models outperform the random walk with a 99% confidence level. Nevertheless, the choice of a loss function affects the ranking of the forecasting models. The performance of the GARCH-type models are very close to the ARX model using the MAE loss function. The performance of the ARX model improves when using the MSE loss function.

Table 2.5: Average Loss Functions

MAE	MEAN	STD	SPA	MSE	MEAN	STD	SPA
RW	0.126246	0.129008		RW	0.032581	0.091846	
ARX	0.098754	0.101259	***	ARX	0.020006	0.066981	***
ARMAX	0.099201	0.102070	***	ARMAX	0.020259	0.068989	***
ARX-GARCH	0.098760	0.101306	***	ARX-GARCH	0.020016	0.067690	***
ARX-EGARCH	0.098782	0.101259	***	ARX-EGARCH	0.020011	0.066806	***

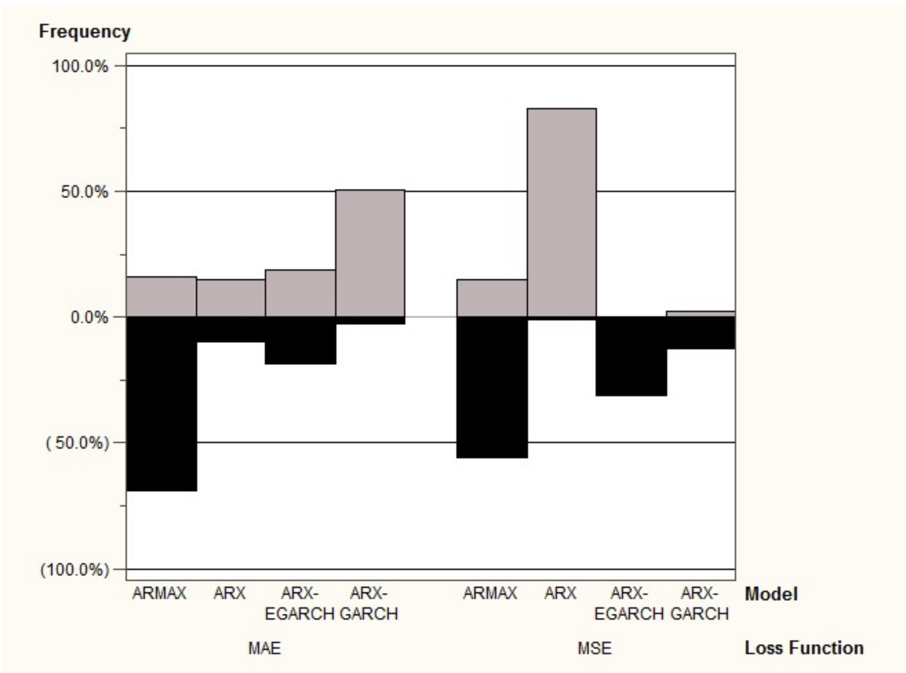
The table exhibits the mean loss function value and the standard deviation across the 81 stocks for the four forecasting models and the random walk model (benchmark). The column SPA reports the statistical outperformance of each forecasting model compared to the random walk model (RW). *** means the model outperforms the benchmark with a 99% confidence level.

The competition summary plotted in Figure 2.4 confirms that the performance of the GARCH-type models is strongly affected by the choice of the loss function. Based on the mean average error, the GARCH-type models seem to outperform the ARX model while their performances strongly deteriorate when the mean squared error is considered at the benefit of the ARX model. This result is in line with our expectations for two reasons.

On the one hand, the loss functions measure discrepancy in the mean process. Hence, we do not expect a strong improvement from the GARCH-type models. On the other hand, the MAE loss function is robust to the presence of outliers, which is not the case of the MSE loss function. The presence of outliers in the data explains the deterioration of the GARCH models performance when using the MSE loss function.

The ARMAX model underperforms the ARX model for the mean modeling across both loss functions but it still strongly outperforms the random walk model.

Figure 2.4: Summary of the Forecast Competition



Summary of the forecast competition between the four forecasting models for the 81 stocks. The figure plots the frequency of minimum (gray) and maximum (black) in-sample loss for the two loss functions considered in the paper. The sum of each frequencies (respectively gray and black) amount to 1 by loss function. For instance, focusing on the MAE loss function, we observe that the ARX-GARCH model displays the lower loss function (grey rectangle) for about 50% of the stocks while the ARMAX model has the higher loss function (black rectangle) for about 70% of the stocks.

2.4.2 Out-of-sample Analysis

The out-of-sample rolling regression results, reported in Table 2.6, are consistent with the in-sample findings. The random walk model strongly underperforms all the forecasting models at a 99% confidence level. Overall, the ranking of the models is unchanged. The ARX model outperforms the ARMAX model. The GARCH-type models are still affected by the choice of the loss function but their out-of-sample performances are better on average than in-sample.

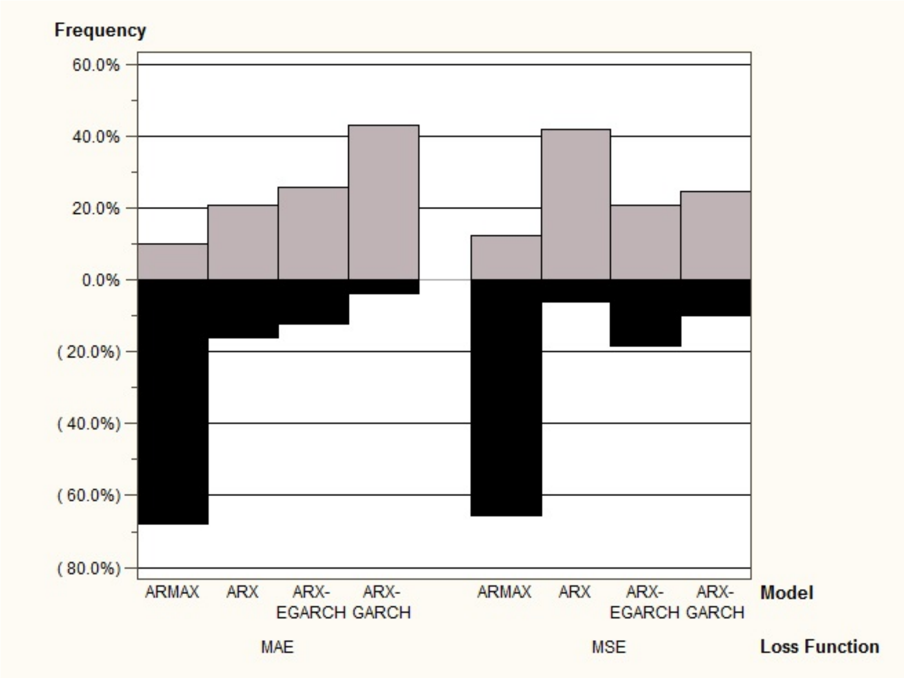
Table 2.6: Average Loss Functions

MAE	MEAN	STD	SPA	MSE	MEAN	STD	SPA
RW	0.119882	0.121063		RW	0.029028	0.073281	
ARX	0.094072	0.094685	***	ARX	0.017815	0.050748	***
ARMAX	0.094603	0.095225	***	ARMAX	0.018017	0.051721	***
ARX-GARCH	0.094019	0.094586	***	ARX-GARCH	0.017786	0.050762	***
ARX-EGARCH	0.094012	0.094679	***	ARX-EGARCH	0.017802	0.050642	***

The table exhibits the mean loss function value and the standard deviation across the 81 stocks for the four forecasting models and the random walk model (benchmark). The column SPA reports the statistical outperformance of the four forecasting models compared to the random walk model (RW), *** means the model outperforms the benchmark with a 99% confidence level.

The same conclusions are drawn from Figure 2.5. The ARX model dominates the ARMAX model for the mean modeling across both loss functions. The forecast competition also confirms the better out-of-sample behavior of the GARCH-type models.

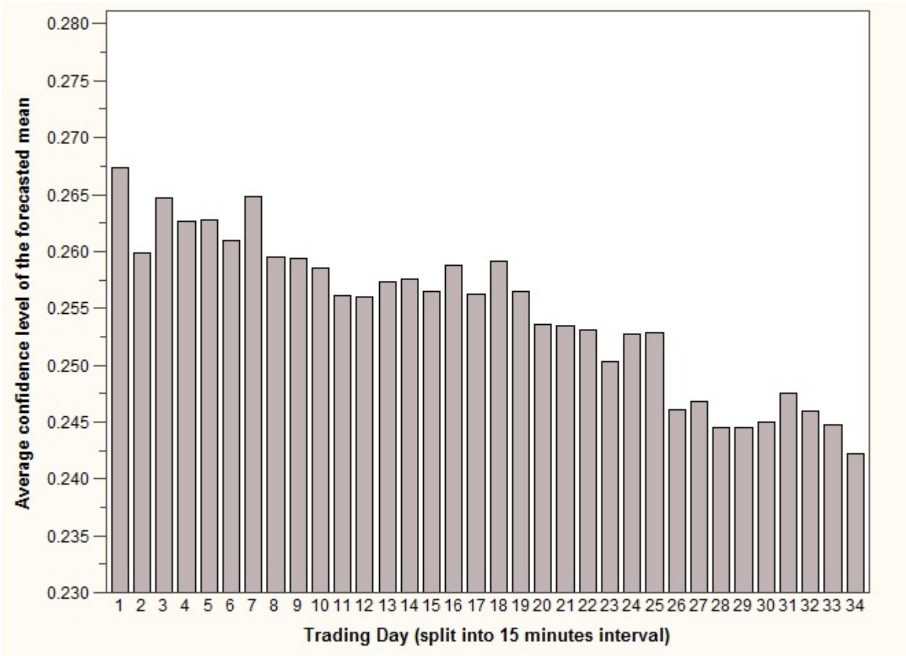
Figure 2.5: Summary of the Forecast Competition



Summary of the forecast competition between the four forecasting models for the 81 stocks. The figure plots the frequency of minimum (gray) and maximum (black) out-of-sample loss for the two loss functions considered in the paper. The sum of each frequencies (respectively gray and black) amount to 1 by loss function. For instance, focusing on the MAE loss function, we observe that the ARX-GARCH model displays the lower loss function (grey rectangle) for about 45% of the stocks while the ARMAX model has the higher loss function (black rectangle) for about 70% of the stocks.

To gauge the uncertainty surrounding our mean forecast, we plot the evolution of the GARCH variance process throughout the trading day in Figure 2.6. The average confidence interval around our liquidity forecast exhibits a downward trend. This indicates that the uncertainty around our forecast decreases along with the trading day. The conditional variance process of the EGARCH model in Figure 2.7, available in Appendix, displays the same pattern.

Figure 2.6: Evolution of the Confidence Interval Throughout the Trading Day for the GARCH Model



The figure plots the evolution of the expected conditional variance surrounding our forecasted mean at an average 15-minute trading volume across the 81 stocks throughout the trading day. In the case of a GARCH process, the squared residual follows an ARMA process. The squared residual is defined as: $e_t^2 = \alpha_0 + (\alpha_1 + \beta_1)e_{t-1}^2 + v_t - \beta_1 v_{t-1}$. We rank our forecasts with regard to the time they occur during the trading session. For this purpose, we discretize the continuous trading day in 34 intervals of 15 minutes.

Finally, we apply the model confidence sets presented in Hansen et al. (2003) to determine a set of forecasting models that stand out from the crowd for a given confidence interval. The results are summarized in Table 2.7. The set of models consists in the ARX model and its extensions ARX-GARCH and ARX-EGARCH models. It confirm the outperformance of the ARX model over the ARMAX model to model the mean. The MAE loss function yields more distinct results which is expected given the lower sensitivity to outliers of this loss function compared to the MSE loss function. On our dataset, the ARX(3)-GARCH(1,1) model seems to better capture the adjusted Amihud ratio dynamics.

Table 2.7: Model Confidence Sets

	MCS p-value	
	MAE	MSE
RW	0.0000	0.0000
ARX	1.0000*	1.0000*
ARMAX	0.0872	0.0345
ARX-GARCH	0.1293**	0.2241*
ARX-EGARCH	0.1965**	0.2241*

The table presents the MCS p-value for the range statistic deviation from the common average as presented in Hansen et al. (2003) on our selection of forecasting models. We carry out the test using 10,000 resamples for a block length of 10. ** means the model is in the model confidence set with a 90% confidence level while * means the model is in the confidence set with a 75% confidence level.

All our results highlight some predictable patterns in the adjusted Amihud ratio against a random walk model. In the absence of more sophisticated strategies of execution price improvement, the end of the continuous trading day seems to be the best time for institutional to hedge or rebalance a portfolio. Indeed, it is characterized by a slightly -but stock specific- negative periodicity, a higher adjusted Amihud ratio predictability and a higher trading volume.

2.4.3 Robustness Check on Other Time Frames

To test the robustness of our out-of-sample results, we consider different time frames that correspond to different trading sizes, specifically the 10-minute average trading volume (Table 2.8) and the 30-minute average trading volume (Table 2.9).

In both time frames, the results support our initial findings. The random walk model underperforms at a 99% confidence level all the forecasting models. The ranking of the model remains unchanged. As mentioned earlier, the loss function choice impacts the performance of the models. Nevertheless, the mean modeling is better captured by the ARX process than the ARMAX process across both loss functions.

The average loss function reduction lies in the same scope as in our initial results. Both time frames display very similar results, which suggests our findings are robust across several trading sizes.

Table 2.8: Average Loss Functions on 10-minute Volume Time Frame

MAE	MEAN	STD	SPA	MSE	MEAN	STD	SPA
RW	0.103615	0.106676		RW	0.022116	0.057721	
ARX	0.082373	0.083122	***	ARX	0.013695	0.048041	***
ARMAX	0.082670	0.083842	***	ARMAX	0.013864	0.050967	***
ARX-GARCH	0.082405	0.082785	***	ARX-GARCH	0.013644	0.044321	***
ARX-EGARCH	0.082427	0.082536	***	ARX-EGARCH	0.013606	0.042007	***

The table exhibits the mean loss function value for a 10-minute volume time frame event chart and the standard deviation across the 81 stocks for the four forecasting models and the random walk model (benchmark). The column SPA reports the statistical outperformance of the four forecasting models compared to the random walk model (RW), *** means the model outperforms the benchmark with a 99% confidence level.

Table 2.9: Average Loss Functions on 30-minute Volume Time Frame

MAE	MEAN	STD	SPA	MSE	MEAN	STD	SPA
RW	0.164799	0.159287		RW	0.052531	0.124639	
ARX	0.124352	0.121167	***	ARX	0.030145	0.076002	***
ARMAX	0.124704	0.123314	***	ARMAX	0.030757	0.095553	***
ARX-GARCH	0.124396	0.120888	***	ARX-GARCH	0.030088	0.075605	***
ARX-EGARCH	0.124486	0.121066	***	ARX-EGARCH	0.030153	0.075964	***

The table exhibits the mean loss function value for a 30-minute volume time frame event chart and the standard deviation across the 81 stocks for the four forecasting models and the random walk model (benchmark). The column SPA reports the statistical outperformance of the four forecasting models compared to the random walk model (RW), *** means the model outperforms the benchmark with a 99% confidence level.

2.5 Illustration

Our paper outlines some predictable patterns in liquidity that statistically outperform a random walk model. For the sake of completeness, we provide an illustration of an execution price improvement strategy.

We show how to significantly reduce the risk surrounding execution prices using the ARX-GARCH model on our sample data. The results are especially relevant for patient investors who hedge or rebalance a position. Indeed, they are mainly concerned with the execution cost of their orders in the market. It is also relevant to use as a circuit breaker for algorithmic trading strategy. This section illustrates how they can use our findings as a circuit breaker to filter entry signal that goes along with high uncertainty of execution prices in the market.

To investigate the reduction in execution price, we rule out our out-of-sample prediction that lies above a given threshold. We set as a threshold for staying out of the market, observations that lie above the third quartile of the in-sample observations. We consider three order sizes from 25, 50 to 100, which correspond to the average 15-minute trading volume multiplied respectively by 25, 50 and 100. We run 1,000 simulations for each investigated strategy. For each scenario, we select a stock at random and compute a naive strategy and a strategy that allows trading when our model forecasts an execution price uncertainty below the given threshold. Table 2.10 reports the median comparison of the simulations as well as the outperform frequency of our strategy over the naive strategy using the Wilcoxon signed rank test. The reduction in execution price uncertainty is pretty consistent across order size and period of the day. It suggests our model performance is broadly unaffected by the order size and the period of the continuous trading day considered. However, the outperformance frequency of our strategy increases along with the order size. For instance, it ranges from 48% to 85% for the median comparison on all the sample.

We carry out our analysis focusing on large and small caps separately, it unveils that our forecasting model performs better on large caps than small caps. It is in line with our expectations since the Amihud ratio is an ex-post liquidity measure that is lagging in an illiquid environment.

In appendix, we report the mean comparison of the simulations in table 2.14 that supports our median comparison analysis.

Our findings emphasize the relevance of execution price management for market participants wishing to better manage their execution price.

Table 2.10: Execution Strategy Comparison: Lower Quartiles Trading Strategy Versus Naive Trading

Panel A : Median comparison ALL	Size 25			Size 50			Size 100		
	strategy	naive		strategy	naive		strategy	naive	
AM	0.186168	0.237942	49.3%***	0.186070	0.236500	68.3%***	0.186829	0.240771	85.9%***
PM	0.184672	0.238284	49.3%***	0.186640	0.240392	69.7%***	0.186018	0.235174	82.7%***
ALL	0.188088	0.239006	48.7%***	0.186373	0.234506	68.9%***	0.187130	0.238947	85.2%***
Panel B: Median comparison LC	Size 25			Size 50			Size 100		
	strategy	naive		strategy	naive		strategy	naive	
AM	0.197444	0.258732	57.6%***	0.199519	0.262478	69.3%***	0.197955	0.267067	88.8%***
PM	0.197304	0.261072	52.8%***	0.197759	0.265987	72.0%***	0.199336	0.263883	87.9%***
ALL	0.194102	0.254084	54.0%***	0.199258	0.264384	69.5%***	0.199295	0.264946	87.3%***
Panel C: Median comparison SC	Size 25			Size 50			Size 100		
	strategy	naive		strategy	naive		strategy	naive	
AM	0.175541	0.236430	44.4%***	0.179158	0.239562	66.9%***	0.183655	0.238635	74.5%***
PM	0.182760	0.238949	43.6%***	0.185014	0.241646	66.6%***	0.184758	0.243410	79.1%***
ALL	0.181384	0.237754	45.1%***	0.184672	0.242234	68.8%***	0.185586	0.242075	79.9%***

The table reports the results for the median comparison between a naive strategy (naive) and an execution price strategy (strategy) that consists in trading when our model forecasts an execution price below the third quartile of the in-sample execution price distribution. For each simulated order execution, the high-low across the trade prices is used as a proxy for price uncertainty. We exhibit the outperformance frequency of our strategy compared to a naive strategy and perform a Wilcoxon signed rank test on the median to assess the statistical outperformance between both strategies. *** means the results are significant at a 99% level. The size refers to the order size that correspond to the average 15-minute trading volume on the underlying stock multiplied respectively by 25, 50 and 100. We consider three periods of time that is, the entire day (ALL), the morning (AM) and the afternoon (PM). Panel A, B and C report the results across stock capitalization respectively, the whole sample (ALL), large caps (LC) and small caps (SC). All the results are based on 1000 simulations. Each simulation mimics order execution on a given stock that is selected at random among the whole sample of stocks.

2.6 Conclusion

In this paper, we investigate the predictability of intraday low execution price uncertainty periods in the stock market. We detect those periods using an adjusted Amihud ratio.

Our forecasting methodology relies on an event chart as advocated by Hasbrouck (1996). This approach has two main advantages. First, Kempf and Korn (1997) outline a major pitfall in the widely used Amihud illiquidity ratio. This ratio implies a linear relationship between the variables, that is volume and volatility, which is typically not the case. Setting a fixed trading volume allows us to bypass the use of a ratio and to avoid optimal sample frequency issues. Second, Irvine et al. (2000) address the idea that a market may be at the same time liquid for small orders and illiquid for large orders. Our event chart allows us to assess liquidity as a trade size specific concept. Indeed, the fixed trading volume of the event chart can be adjusted to the order size considered by market participants.

We select the optimal forecasting model through the general-to-specific process presented in Hendry and Krolzig (2001). We use the Hansen (2005)'s SPA test on two mainstream loss functions to document the performance of our forecasting models against a benchmark, both in- and out-of-sample.

Our results outline some predictable patterns in the adjusted Amihud ratio that statistically outperform a random walk model. The ARX(3)-GARCH(1,1) model stands out from the crowd among the four tested forecasting models. Trade-based and order book-based liquidity measures are informative of forthcoming price impact as confirmed by

Cont et al. (2014). On average, the closing of the continuous trading day seems to be the best time for institutionals to hedge or rebalance a portfolio. Indeed, it is characterized by a slightly -but stock specific- negative periodicity, a higher adjusted Amihud ratio predictability and a higher trading volume.

We carry out robustness checks of our results on other time frames. The results are similar to our initial findings, which suggests our models are robust across several order sizes.

Using the ARX-GARCH model, we provide an illustration of an execution price improvement strategy that statistically outperforms a naive strategy.

This paper may be extended in various directions. First, we can consider market timing issues within the same methodological framework. Second, we can transpose volatility literature extensions into the liquidity framework such as realized kernels and intraday jumps.

2.7 Appendix

Table 2.11: Parameter Estimates of ARMAX Model

Variable	Estimate	StdErr	**
AAR_1	0.389884	0.0629	96%
δ_1	0.196663	0.0679	60%
$absCRT$	0.465964	0.1456	70%
N	0.001612	0.0008	48%
$Q1$	-0.000009	0.0000	33%
$Q5$	-0.000006	0.0000	79%
RS	0.703332	0.4145	43%
SE_1	0.169060	0.1107	44%
SE_2	0.115925	0.0921	27%
SE_{33}	-0.053447	0.0721	9%
SE_{34}	-0.100376	0.1721	18%

This table reports the parameter estimates of the ARMAX model. Exogenous variables are those presented in Section 2.3.2. $SE_{i,1}$ and $SE_{i,2}$ are dummies for the two first 15-minute time-based intervals for stock i that capture the beginning of day periodicity in the dependent variable and $SE_{i,33}$ and $SE_{i,34}$ are dummies for the two last 15-minute time-based intervals for stock i. The column ** reports the percentage of stocks for which the parameter is significant at a 95% confidence interval.

Table 2.12: Parameter Estimates of ARX-GARCH Model

Variable	Estimate	StdErr	**
AAR_1	0.192798	0.0382	99%
AAR_2	0.084436	0.0356	69%
AAR_3	0.064279	0.0349	47%
$absCRT$	0.617198	0.1665	78%
N	0.002363	0.0008	67%
$Q1$	-0.000010	0.0000	31%
$Q5$	-0.000008	0.0000	84%
RS	0.826726	0.4203	48%
SE_1	0.160206	0.1345	37%
SE_2	0.117924	0.1049	29%
SE_{33}	-0.112195	0.0832	37%
SE_{34}	-0.201614	0.1434	39%
$ARCH_0$	0.142051	0.0732	81%
$ARCH_1$	0.059119	0.0346	69%
$GARCH_1$	0.31269	0.2103	59%

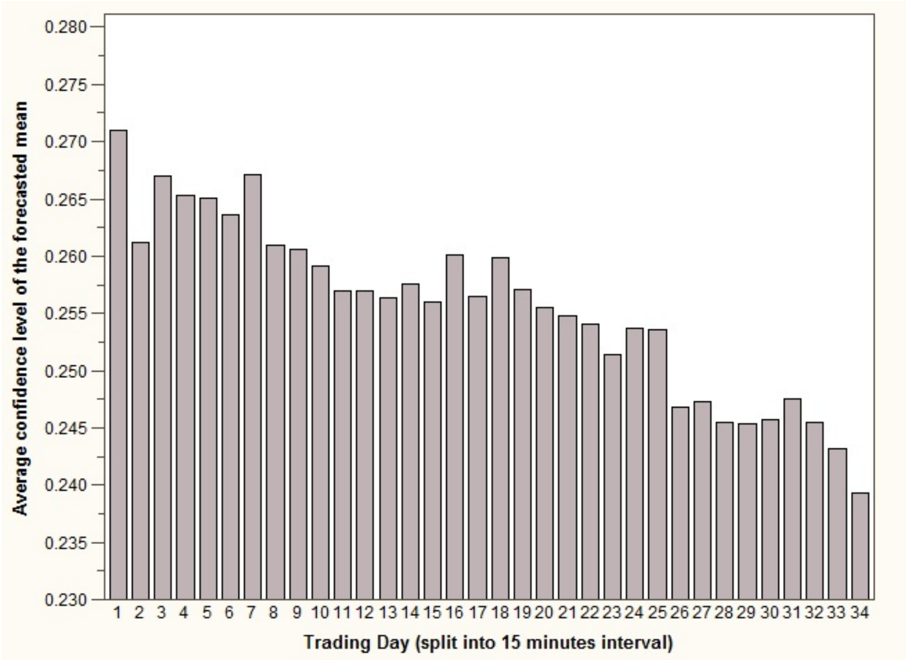
This table reports the parameter estimates of the ARX(3)-GARCH(1,1) model. Exogenous variables are those presented in Section 2.3.2. $SE_{i,1}$ and $SE_{i,2}$ are dummies for the two first 15-minute time-based intervals for stock i that capture the beginning of day periodicity in the dependent variable and $SE_{i,33}$ and $SE_{i,34}$ are dummies for the two last 15-minute time-based intervals for stock i. The column ** reports the percentage of stocks for which the parameter is significant at a 95% confidence interval.

Table 2.13: Parameter Estimates of ARX-EGARCH Model

Variable	Estimate	StdErr	**
AAR_1	0.198697	0.0383	100%
AAR_2	0.082458	0.0347	69%
AAR_3	0.062717	0.0344	44%
$absCRT$	0.617122	0.1642	79%
N	0.002342	0.0008	67%
$Q1$	-0.000009	0.0000	32%
$Q5$	-0.000008	0.0000	86%
RS	0.810575	0.4009	51%
SE_1	0.164549	0.1350	32%
SE_2	0.117052	0.1043	31%
SE_{33}	-0.110822	0.0834	35%
SE_{34}	-0.201128	0.1434	41%
$EARCH_0$	-1.105938	0.3623	74%
$EARCH_1$	0.069241	0.0669	41%
$EGARCH_1$	0.260583	0.1496	47%
$THETA$	-0.733986	1.1800	8%

This table reports the parameter estimates of the ARX(3)-EGARCH(1,1) model. Exogenous variables are those presented in Section 2.3.2. $SE_{i,1}$ and $SE_{i,2}$ are dummies for the two first 15-minute time-based intervals for stock i that capture the beginning of day periodicity in the dependent variable and $SE_{i,33}$ and $SE_{i,34}$ are dummies for the two last 15-minute time-based intervals for stock i. The column ** reports the percentage of stocks for which the parameter is significant at a 95% confidence interval.

Figure 2.7: Evolution of the Confidence Interval Throughout the Trading Day for the EGARCH Model



The figure plots the evolution of the expected conditional variance surrounding our forecasted mean at an average 15-minute trading volume across the 81 stocks throughout the trading day. In the case of a GARCH process, the squared residual follows an ARMA process. The squared residual is defined as: $e_t^2 = \alpha_0 + (\alpha_1 + \beta_1)e_{t-1}^2 + v_t - \beta_1 v_{t-1}$. We rank our forecasts with regard to the time they occur during the trading session. For this purpose, we discretize the continuous trading day in 34 intervals of 15 minutes.

Table 2.14: Execution Strategy Comparison: Lower Quartiles Trading Strategy Versus Naive Trading

Panel A : Mean comparison ALL	Size 25			Size 50			Size 100		
	strategy	naive		strategy	naive		strategy	naive	
AM	0.205062	0.265216	32.0%***	0.207636	0.267204	46.1%***	0.207841	0.262602	60.3%***
PM	0.202408	0.261766	32.2%***	0.205129	0.261019	44.3%***	0.207450	0.258473	58.1%***
ALL	0.206807	0.263689	31.8%***	0.204578	0.261672	44.2%***	0.206148	0.258223	56.6%***
Panel B: Mean comparison LC	Size 25			Size 50			Size 100		
	strategy	naive		strategy	naive		strategy	naive	
AM	0.207476	0.275498	39.8%***	0.212849	0.285243	55.5%***	0.214034	0.291852	66.6%***
PM	0.210967	0.281496	44.3%***	0.212202	0.288847	57.9%***	0.213972	0.286366	65.5%***
ALL	0.205845	0.271945	40.3%***	0.211237	0.285818	54.0%***	0.211869	0.285853	63.3%***
Panel B: Mean comparison SC	Size 25			Size 50			Size 100		
	strategy	naive		strategy	naive		strategy	naive	
AM	0.198578	0.256946	28.9%***	0.203852	0.270078	41.6%***	0.205295	0.269632	55.1%***
PM	0.202624	0.263342	29.9%***	0.204999	0.272255	44.7%***	0.206482	0.271830	57.0%***
ALL	0.201927	0.262880	29.8%***	0.204422	0.272069	44.2%***	0.207509	0.276848	57.0%***

The table reports the results for the mean comparison between a naive strategy (naive) and an execution price strategy (strategy) that consists in trading when our model forecasts an execution price below the third quartile of the in-sample execution price distribution. For each simulated order execution, the high-low across the trade prices is used as a proxy for price uncertainty. We exhibit the outperformance frequency of our strategy compared to a naive strategy and perform a Wilcoxon signed rank test on the median to assess the statistical outperformance between both strategy. *** means the results are significant at a 99% level. The size refers to the order size that correspond to the average 15-minute trading volume on the underlying stock multiplied respectively by 25, 50 and 100. We consider three periods of time that is, the entire day (ALL), the morning (AM) and the afternoon (PM). Panel A, B and C report the results across stock capitalization respectively, the whole sample (ALL), large caps (LC) and small caps (SC). All the results are based on 1000 simulations. Each simulation mimics order execution on a given stock that is selected at random among the whole sample of stocks.

CHAPTER

THREE

DETECTING AND FORECASTING HIGH FREQUENCY PRICE JUMPS IN THE STOCK MARKET

3.1 Introduction[1]

The distributional properties of financial asset returns have raised the interest of researchers and practitioners for decades. Among the several stylized features outlined in the literature, price jumps take on an utmost importance as highlighted by Press (1967). Indeed, the occurrence of price jumps in the stock market has numerous implications in terms of risk management as shown by Duffie and Pan (2001) or Bakshi and Panayotov (2010), derivative pricing as documented in Bates (2000) or Eraker et al. (2003), and

[1]This essay is a joint work with Catherine D'Hondt.

portfolio allocation with its influence on the optimal strategy as mentioned in Jarrow and Rosenfeld (1984) or Liu et al. (2003).

Although their relevance is well-established, the drivers of price jumps depend on the characteristics of the investigated market.

In the stock market, price jumps are found to be mainly liquidity-driven. Using both NYSE and NASDAQ data, Joulin et al. (2008) and Weber and Rosenov (2005) document that a lower density in the order book is at the source of price jumps, instead of the surge of significant trading volume underlying the arrival of information in the market. Farmer et al. (2004) find similar evidence on the London Stock Exchange. Aside of the impact of time-varying liquidity, Farmer et al. (2004) also emphasize that jump risks seem to be stock-specific and that illiquid stocks tend to suffer from jumps more often. Recently, Boudt and Petitjean (2014) assess that 70% of price jumps are related to liquidity in the DJIA index. Moreover, they investigate the dynamics of the four dimensions of liquidity around price jumps. They find that the effective spread and the number of trades are both informative of forthcoming jumps.

In other financial markets, some recent studies reveal that price jumps are mainly information-driven. For example, Jiang et al. (2010) and Miao et al. (2012) find that most jumps appear at pre-scheduled macroeconomic news in the US treasury bond market and in the stock index futures market.

In this paper, we address the predictability of price jumps in the stock market, which is a widely unexplored topic. In particular, we investigate the informational content of market microstructure variables for price jumps forecasting. For this purpose, we build on the literature devoted to the relationship between liquidity and price changes.

Some forecasting applications focus on the link between liquidity provision and short-run price movements. Hellström and Simonsen (2009) document that the order book, especially at the best quote, is informative of future price changes at the 1 and 2-minute aggregation levels while predictability strongly deteriorates at the 5 and 10-minute time frames. This result suggests very short-term informational content of the order book.

As for imbalances, the literature agrees that trade imbalance seems to be more of an efficient driver for very short-term price moves than order book imbalance (Cont et al., 2014). Nevertheless, Tombeur and Wuyts (2011) show that the informational content of the order book imbalance tends to be long-lasting while the trade imbalance informational content vanishes quickly. They also highlight that the predictability of price movements is higher when information asymmetry is large, that is, at the beginning of the trading day. In addition, Jain et al. (2011) find that the cost of round trip trade is the most informative measure to forecast future volatility, prices and trading frequency.

When dealing with jumps, the definition of what we call 'a jump' is of paramount importance. Christensen et al. (2011) document that price jumps are often confused with volatility jumps. For instance, Lanne (2007) considers jumps in the realized volatility to show that splitting the continuous sample path from the jump component leads to an improvement of the out-of-sample volatility forecasts. In contrast, Zheng et al. (2012) define a jump as a change in the midquote and outline some predictability of inter-trade price jumps using LASSO logistic regressions.

In this paper, we define price jumps as a market disruption, that is an extreme price movement induced by a fixed trading volume. Our methodology relies on an fixed volume sampling scheme as advocated by Hasbrouck (1996), allowing to control for trading volume and deal with non trading issues at high frequency aggregation. Theoretically, jump detection in this framework can be related to the widely used Amihud (2002) illiquidity ratio[2]. This jump definition allows us to focus on liquidity-driven jumps. Indeed, price jumps identified in this sampling scheme go along with illiquidity spike in the Amihud ratio. Our fixed volume sampling scheme bypasses the non-linearity issue between absolute returns and trading volume in the Amihud ratio (Kempf and Korn, 1997).

We detect price jumps using return distribution percentiles as well as robust jump tests to time-varying volatility and periodicity in financial markets. Relying on a logistic model and an extension adjusted to rare events, we first select the optimal set of explanatory variables for the most robust jump test (JM WSD) using a general-to-specific model selection process based on the likelihood function (Castle and Hendry, 2011). We then evaluate the forecasting performance of our models, both in and out-of-sample.

Our results reveal some predictable patterns in price jumps, essentially at high frequency aggregation. While trade imbalance is the key driver behind short-run price movements, the determinant of price jumps seems to lay in gaps within the order book. Indeed, our empirical findings point out to a low density of the order book that tends to last for few intervals while trading activity is pretty low.

Overall, price disruptions happen in majority during the early continuous trading session. This intraday pattern remains significant, albeit in a lower magnitude, after accounting for periodicity in the stock market, which is consistent with a lower liquidity in the

[2]The Amihud ratio is defined as follows: $Amihud_t = \frac{|R_t|}{V_t}$, where $|R_t|$ is the absolute return and V_t is the trading volume at time t.

morning due to information asymmetry as suggested by Cont et al. (2014). Out-of-sample and based on the most restrictive jumps detection method, our relogit model allows to isolate about 13% of the whole database where the majority of price jumps (almost 70%) occur.

To check the robustness of our approach, we change the time frame and the threshold set to discretize the continuous model probabilities into events and non-events. The results are consistent with our initial findings. However, the informational content of the order book is mostly short-lived with a deterioration of the forecasting ability from the average 5-minute trading volume time frame.

Since our empirical findings show that our models perform pretty well in forecasting price jumps, we finally illustrate how market participants could take advantage of such models to improve the implementation of their investment decisions. Using several bootstrap simulations, we find there is an obvious incentive not to trade when our relogit model detects a risk of price jumps, confirming its usefulness for investors mainly concerned with price uncertainty.

The paper proceeds as follows. Section 3.2 details the data and the liquidity proxies that are used in the empirical part. Section 3.3 presents the applied methodology. Section 3.4 documents both in-sample and out-of-sample results as well as several robustness checks. Section 3.5 illustrates how market participants could use our models to reduce uncertainty around execution prices. Section 3.6 sums up the findings.

3.2 Data

3.2.1 Sample

The database used in this paper consists of tick data for the Euronext 100 index constituents. This index comprises the 100 largest and most liquid stocks traded on Euronext,[1] representing 80% of the total market capitalization of the European exchange. The data span 61 trading days, from February 1 to April 30, 2006. We rely on trades, orders and order book data[2] to quantify the liquidity dynamics through an extensive set of proxies, both trade-based and order book-based.

We arrange our database in event-based intervals. Hasbrouck (1996) outlines that event-based aggregation tend to be less noisy than time-based one for forecasting purposes. In this paper, the event is defined as a given trading volume, which allows us to control for trading volume effects and avoid non trading issues at high frequency aggregation. For the sake of consistency, we set a fixed trading volume for each stock that corresponds to the average 2-minute trading volume observed on that stock over the whole sample period. Several studies outline that the 2-minute frequency is a good compromise between microstructure noise issues and the relevance of higher frequency data (Andersen et al., 2010, 2012; Bajgrowicz and Scaillet, 2011; Boudt and Petitjean, 2014).

[1]Euronext is the European regulated cash market of NYSE Euronext group. The latter is the holding company created in 2007 by the combination of NYSE Group, Inc. and Euronext N.V. Some years earlier, Euronext N.V was created to bring together major marketplaces across Europe and become the first pan-European market. NYSE Euronext is today referred to as the European regulated cash market of NYSE Euronext group, with nearly 1 750 firms listed across four countries (Belgium, France, Portugal and The Netherlands).

[2]The database is the same used in DeWinne and D'Hondt (2007). Based on both public and private data, we are able to rebuild the limit order book second by second.

We apply some filters to our data sample to ensure the robustness of our results. We remove the first and the last intervals of each trading day. The first interval is removed because the explanatory variables would refer to the previous day order book data. The last interval is excluded due to inconsistency as the fixed trading volume may not be completed. It is worthwhile to mention as well that the trigger for a new interval is the trade that crosses the fixed volume threshold. The trading volume per interval may thus slightly vary to avoid order splitting.

3.2.2 Liquidity Proxies

In this section, we review the explanatory variables that will be considered in our forecasting models. We measure multidimensional liquidity dynamics by computing an extensive set of both trade-based and order book-based proxies.

Trade-based measures include trading duration, average trading size and trade imbalance, both in number of shares and in monetary volume.

Trading duration ($TimeBar_{i,t}$) is computed as follows:

$$TimeBar_{i,t} = EndTime_{i,t} - StartTime_{i,t}$$

where $EndTime_{i,t}$ is the closing time in seconds and $StartTime_{i,t}$ is the opening time in seconds of the fixed volume interval t for stock i.

Order book-based measures encompass depth both at the best quote and at the five best quotes, the absolute depth imbalance ($AbsImb\_5_{i,t}$), the relative spread ($RS_{i,t}$), Kang and Yeo (2008)'s dispersion ($Dispersion_{i,t}$), Næs and Skjeltorp (2006)'s slope and the absolute cost of round trip trade ($AbsCRT_i$).

The absolute depth imbalance is computed as follows:

$$AbsImb\_5_{i,t} = \frac{|\sum_{k=1}^{5} Q_{B,k,i,t} - Q_{A,k,i,t}|}{\sum_{k=1}^{5} Q_{B,k,i,t} + Q_{A,k,i,t}}$$

where $Q_{B,k,i,t}$ is the displayed bid quantities and $Q_{A,k,i,t}$ is the displayed ask quantities at the k^{th} price limit of the order book for stock i at the end of interval t.

The relative spread is:

$$RS_{i,t} = \frac{P_{A1,i,t} - P_{B1,i,t}}{MQ_{i,t}}$$

where $P_{A1,i,t}$ is the best ask and $P_{B1,i,t}$ is the best bid for stock i at interval t. $MQ_{i,t}$ is the midquote of stock i at the end of interval t.

We measure the dispersion at the five best quotes with the Kang and Yeo (2008)'s measure, which is defined as:

$$Dispersion_{i,t} = \frac{1}{2}\left(\frac{\sum_{j=1}^{n} w_{i,j,t}^{Bid} Dst_{i,j,t}^{Bid}}{\sum_{j=1}^{n} w_{i,j,t}^{Bid}} + \frac{\sum_{j=1}^{n} w_{i,j,t}^{Ask} Dst_{i,j,t}^{Ask}}{\sum_{j=1}^{n} w_{i,j,t}^{Ask}}\right)$$

where $w_{i,j,t}^{Bid/Ask}$ are the bid or ask quantities available at the j^{th} price limit for stock i at the end of interval t normalized by the total depth at the five best quotes, $Dst_{i,j,t}^{Bid} = (Price_{i,j-1,t}^{Bid} - Price_{i,j,t}^{Bid})$, and $Dst_{i,j,t}^{Ask} = (Price_{i,j,t}^{Ask} - Price_{i,j-1,t}^{Ask})$. The midquote is used for the distance of the first best quotes.

The cost of round trip captures both the tightness and the depth dimension of liquidity. To calculate it, we take the five best quotes of the order book into account and then assume a perfect elasticity of prices beyond as it is usually done in the literature (Jain et al., 2011). The trade size is lined up with the time frame of the event chart, that is the 2-minute average trading volume on the underlying stock. The measure is computed as follows for stock i at the end of interval t:

$$AbsCRT_{i,t} = \frac{|\sum_{k=1}^{5} I_k^{Buy}\left(Midquote_{i,t} - P_k^{Buy}\right) + \sum_{k=1}^{5} I_k^{Sell}\left(P_k^{Sell} - Midquote_{i,t}\right)|}{T \times Midquote_{i,t}}$$

where $I_k^{Buy/Sell}$ is the number of shares bought or sold respectively at the k^{th} price limit of the order book and $P_k^{Buy/Sell}$ is the price at the k^{th} quote. T is the total number of shares to be bought or sold, with

$$\begin{array}{lll} I_k^h = Q_j^h & \text{if} & T > \sum_{j=1}^{k} Q_j^h, \\ I_k^h = \left(T - \sum_{j=1}^{k-1} Q_j^h\right) & \text{if} & T > \sum_{j=1}^{k-1} Q_j^h \quad and \quad T < \sum_{j=1}^{k} Q_j^h, \\ I_k^h = 0 & & \text{otherwise} \end{array}$$

where $h \in [Buy, Sell]$ and Q_j^h is the displayed quantity at the j^{th} price limit of the order book.

3.3 Methodology

3.3.1 Price Jump Detection

As mentioned earlier, price jump detection in a volume sampling scheme can be related to illiquidity-driven jumps such as measured by the Amihud ratio (Amihud, 2002). This sampling scheme allows us to solve for the non-linearity issue between trading volume and returns (Kempf and Korn, 1997) in the ratio. We look at two distinct categories of jumps.

First, we consider an ex-post percentile of the return distribution as a threshold for a price jump. In the empirical part, we use the percentile 99.5%. This simple and straightforward method (JM 99.5% hereafter) captures high elasticity of prices to a fixed trading volume. Nevertheless, it assumes no periodicity and a constant volatility, which makes it questionable.

Second, we implement more restrictive jump detection tests that are robust to the violation of those assumptions. For this purpose, we consider the Lee and Mykland (2008)'s jump test (JM LM hereafter) as well as an extension (JM WSD hereafter) used in Boudt et al. (2011). Aside of being typically designed for high frequency data, those non parametric jump tests allow us to account for time-varying volatility and periodicity in the stock market, respectively. Those tests are initially designed for time sampling scheme. Volume sampling scheme has no effect on the mechanism of the Lee and Mykland (2008)'s jump test. Nevertheless, it impacts the computation of the realized bipower variation. Relying on Oomen (2006) who shows that the realized variance computation is less noisy in volume sampling scheme, we expect the local volatility estimate to be more robust than in time sampling scheme, which in turn increases the robustness of the Lee and Mykland (2008)'s jump test.

Lee and Mykland (2008)'s jump test assumes prices follow a continuous time geometric brownian process such as commonly assumed in asset pricing, defined as follows:

$$dS_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t$$

where S_t is the stock price, μ_t is the drift term, σ_t is the spot volatility process and W_t a standard Brownian motion process. κ_t is the jump size while dq_t is a binary variable that takes 1 in the presence of a jump and 0 otherwise. All these terms are defined at time t.

Based on this price model, the authors rest on absolute return standardized by a jump-robust estimate of the local volatility to test the null hypothesis that the return is unaffected by a jump.

The local volatility is computed through the realized bipower variation that is defined as the sum of products of a given number of consecutive absolute returns:

$$RBV_{i,t} = \mu_1^{-2} \sum_{t=2}^{M} |r_{i,t}||r_{i,t-1}|$$

where $RBV_{i,t}$ is the realized bipower variation for stock i at interval t and $\mu_1 = E|\mu| = \sqrt{2/\pi}$. μ follows a standardized Gaussian distribution. $r_{i,t}$ and $r_{i,t-1}$ are the absolute returns for stock i at time t and $t-1$.

Barndorff-Nielsen and Shephard (2004) show that the realized bipower variation converges, when M tends to infinity, to the integrated variance. The Lee and Mykland (2008)'s test assumes that the local volatility is constant over the entire window used to compute it. Given those windows typically expand on days for intraday aggregation sampling, this assumption introduces a significant bias in the jump test. To cope with this issue, Boudt et al. (2011) develop a filtered test and advocate the use of the weighted standard deviation (henceforth JM WSD) as a periodicity filter, which improves the jump detection accuracy. Indeed, spurious jump detection may arise from neglecting intraday volatility periodicity.

The length of the rolling window used to compute the local volatility depends on the sampling frequency. It should account for both, time-varying volatility and jump-robust local volatility estimates. Lee and Mykland (2008) propose to set the length of the rolling window to $\sqrt{252 \times obs}$, where obs is the number of intraday observations. The authors show through simulations that an increase of the window size beyond this threshold brings only marginal contribution while increasing the computational burden.

The jump test statistic ($J_{i,t}$) informs us on the precise timing of a jump as well as its size and is defined as follows:

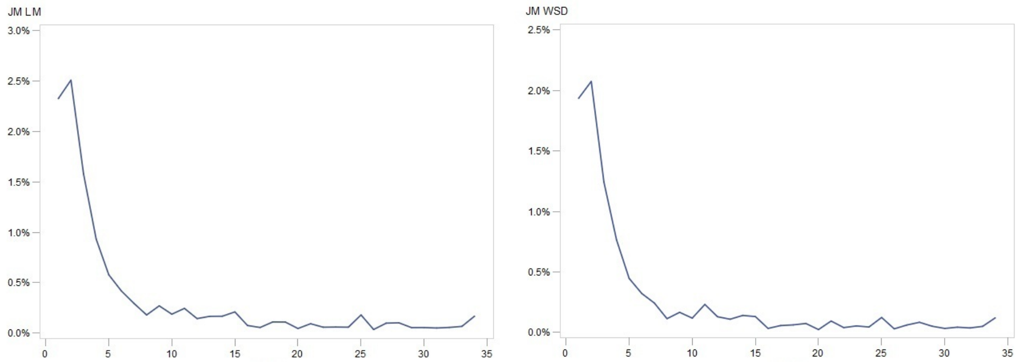
$$J_{i,t} = \frac{|r_{i,t}|}{\xi_{i,t} f_{i,t}}$$

where $r_{i,t}$ is the return for stock i at interval t, $\xi i, t$ is the square root of the realized bipower variation and $f_{i,t}$ is the periodicity estimate for stock i at interval t.

Finally, Lee and Mykland (2008) suggest to set the critical level to reject the null hypothesis of no jump based on a quantile function of a standard Gumbel distribution. In this paper, we reject the null of no jump with a 10% type 1 error probability, such as in Boudt and Petitjean (2014), which is a conservative approach.

As highlighted in Figure 1, the frequency of jumps in our sample is higher at the beginning of the continuous trading day. It still holds when periodicity and time-varying volatility are taken into account, even though the amplitude is somewhat lower. Cont et al. (2014) outline the presence of higher information asymmetry at the beginning of the continuous trading day, which may explain our observations. We also note that the US market opening does not translate into a spike in jump detection on the Euronext stocks under scrutiny.

Figure 3.1: Frequency of Jumps



Both figures represent the frequency of jumps in our sample for the two more restrictive tests considered, that are the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD). We rank price jumps with regard to the calendar time at which they occur during the trading session. For this purpose, we discretize the continuous trading day in 34 intervals of 15 minutes.

As documented by Aït-Sahalia and Jacod (2004), the jump frequency decreases along with the data sampling frequency due to time averaging effects. To disentangle jumps from the continuous diffusion process, high sampling frequency is required. In our sample, the frequency of jumps varies from about 0.4% at the average 2-minute trading volume frequency to 0.03% at the average 15-minute trading volume frequency. However, market microstructure effects may bring noise in our results at very high frequency. Relying on Andersen et al. (2010), Andersen et al. (2012), Bajgrowicz and Scaillet (2011), and Boudt and Petitjean (2014), we settle for the 2-minute frequency, which strikes a balance between the additional informational content of higher frequency data and market microstructure noise.

Table 3.1 presents descriptive statistics for the jumps detected in our sample with both the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD). In the empirical section, we rule out stocks that display less than 20 jumps in-sample to ensure a robust calibration of our models. The jump frequency ranges from 0,3% to 0,37% depending on the detection test. Similarly, the average number of jumps by stock varies from 46 to 56. Our final sample consists in 931,205 intervals across 61 stocks.

Table 3.1: Descriptive Statistics

	JM LM	JM WSD
Count	3,418	2,785
Frequency	0,37%	0,30%
Observations	931,205	
Jump/stock	56	46
Stdev	11	14

The table reports the number of jumps ($Count$) and their frequency ($Frequency$) based on the 2-minute average trading volume. These statistics refer to jumps detected by the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD). They are computed on a total of 931,205 observations. The average number of jumps by stock ($Jump/stock$) as well as its standard deviation ($Stdev$) are also provided.

3.3.2 Model Presentation: Relogit

Forecasting rare events, such as jumps in financial markets, is a challenging issue. King and Zeng (2001) highlight that classical logistic regression tends to be biased towards the dominant binary dependent variable which, in turn, leads to underestimate the probability of rare events. In practice, researchers often settle for increasing the sample in order to collect more events while leaving the rare events biases unresolved. In this paper, we first use a standard logit model and then we address issues raised by rare events through data sampling. For this purpose, we rely on the relogit model developed by King and Zeng (2001). Their adjustments consist in balancing the distribution of the binary dependent variable. The underweight of the majority class implies however two corrections to ensure model efficiency.

First, the rebalancing made on the dependent variable leads to a significant bias in the intercept while the other explanatory variables are almost unaffected. Indeed, binary models intercept depends on the mean of the dependent variable. The mean of a binary variable is the relative frequency of events in the data.

Second, the likelihood function is designed for random sampling. Several methods exist to correct the likelihood function and ensure an accurate calibration of the model. Relying on King and Zeng (2001), we apply the "weighting" method as they document its outperformance, especially when a large sample is available and the functional form is unknown. In this paper, we focus on the logistic distribution. Indeed,the "weighting" method is designed to cope with potential mis-specification in the distribution. Furthermore, in the case of a binary choice model, alternative distributions display very close results as shown in Ruud (1983).

As in the standard logistic regression, the stochastic component for the rare events logistic regression is :

$$Y_i \sim Bernoulli\left(\pi_i\right)$$

where Y_i is the binary dependent variable and takes the value 1 when there is a price jump and 0 otherwise. π_i is the systematic component and represents the conditional probability that a price jump occurs given the beta parameters. It is defined as:

$$P\left(Y_i = 1|\hat{\beta}\right) = \pi_i = \frac{1}{1+\exp\left(-x_i\hat{\beta}\right)}$$

where x_i is the set of explanatory variables, and $\hat{\beta}$ is the maximum-likelihood estimate.

Let the subscript 1 denotes observations for which we observe a price jump and the subscript 0 denotes observations for which we observe no price jump. We use normalized weights that sum to n in order to adjust the model such that:

$$w_1 = \frac{\tau}{\bar{y}}$$

and

$$w_0 = \frac{(1-\tau)}{(1-\bar{y})}$$

where τ is the fraction of price jumps in the population and $\bar{y}$ is the fraction of price jumps in the sample.

Then the vector of weights w_i is:

$$w_i = w_1 Y_i + w_0\left(1 - Y_i\right)$$

and the weighted log-likelihood function is:

$$lnL_w\left(\beta|y)\right) = w_1 \sum_{Y_i=1} ln\left(\pi_i\right) + w_0 \sum_{Y_i=0} ln\left(1-\pi_i\right)$$

where y is the unobserved continuous variable $Y_i^{\star}$.

In most cases, τ is unknown. The convenient way to deal with unknown τ is to specify an upper and a lower bound instead of a point estimate. A robust Bayesian method is then performed to generate a confidence interval for each variable of interest. The model calibration is made through the sandwich package presented in Zeileis (2004).

To select the optimal set of explanatory variables, we use the general-to-specific model selection process on our panel dataset[1], presented in Castle and Hendry (2011). This procedure simplifies a general unrestricted model (GUM) to a congruent model relying on the likelihood function that suits our choice-based sampling design. At each simplification step, the selection encompasses model mis-specification tests as well as diagnostic tests. For the sake of clearness, we only present a single model that is the optimal model for the most robust jump test (JM WSD) and apply it to the less robust jump tests.

Our final forecasting model is:

$$Logit(JM_{i,t}) = \beta_{i,0} + \beta_{i,1}AbsImb\_5_{i,t-1} + \beta_{i,2}AbsCRT_{i,t-1} + \beta_{i,3}Dispersion_{i,t-1} + \beta_{i,4}RS_{i,t-1} + \alpha_i SE_{i,t}$$

where $JM_{i,t}$ is a binary response that takes the value 1 when there is a price jump and 0 otherwise for stock i at interval t, $SE_{i,t}$ is a dummy that captures periodicity of the first continuous trading hour for stock i. The other explanatory variables were defined previously in section 3.2.2 and are all computed at the end of interval t.

[1]The model selection is carried out on the relogit model for the most robust jump test and applied to alternative models.

Our sample is then cut into an estimation period (first half of data) which is used for model calibration and an evaluation period (second half of data), which is used for out-of-sample forecasting. In the case of unbalanced sample, the usual 50% threshold to discretize the continuous model probabilities into events and non-events is not optimal. Lemmens and Croux (2006) document this threshold issue in the case of unbalanced samples. The optimal threshold level should strike a balance between the proportion of real jumps detected and the prominence of the model type 1 error. In turn, this threshold lies around the observed event occurrence probability. Hence, we set the threshold level at 1% to assess the forecasting performance of our model. In appendixes, we also consider alternative thresholds: 0,5% and 2%.

3.4 Results

3.4.1 In-sample

We report the in-sample confusion matrixes for the three jump detection methods and for both forecasting models in Table 3.2. The findings reveal that there are some predictable patterns in price jumps. Overall, the logit model seems to detect less often price jumps than the relogit model. The performance of our models are quite robust across the three different jump detection methods considered. As expected, we mention a slightly better behavior of our models when we detect jumps using less restrictive detection methods as a percentile of the ex-post return distribution.

Table 3.2: Confusion Matrices

JM 99.5% Logit			JM 99.5% Relogit		
T:1%	Actual 0	Actual 1	T:1%	Actual 0	Actual 1
Predicted 0	8,956	34	Predicted 0	8,911	15
Predicted 1	996	14	Predicted 1	1,041	33
JM LM Logit			**JM LM Relogit**		
T:1%	Actual 0	Actual 1	T:1%	Actual 0	Actual 1
Predicted 0	9,158	23	Predicted 0	8,876	14
Predicted 1	807	12	Predicted 1	1,089	21
JM WSD Logit			**JM WSD Relogit**		
T:1%	Actual 0	Actual 1	T:1%	Actual 0	Actual 1
Predicted 0	9,227	16	Predicted 0	8,745	9
Predicted 1	746	11	Predicted 1	1,228	18

The table exhibits the confusion matrices for our logit and relogit models. Results across the 61 stocks are provided for the three jump detection methods, that are the percentile 99.5% return distribution (JM 99.5%), the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD). For both logit and relogit models, we use a 1% threshold to split events and non-events from the returned model probabilities. Each matrix displays on the one hand the actual split between events and non-events, and on the other hand, the predicted events and non-events. The results are rebased at 10,000 for an easier interpretation.

To confirm the relevance of our models, we test the joint significance of the explanatory variables using the Wald and the Score tests. Results are reported in Table 3.3, in Panel A for the logit model and in Panel B for the relogit model. A small p-value rejects the null hypothesis that all the slope parameters of the model are jointly zero. This hypothesis is strongly rejected by the two tests for both models and for all the jump detection methods.

Table 3.3: Testing Global Null Hypothesis

Panel A: Logit	JM P99.5%		JM LM		JM WSD	
Test	ChiSq	Pr>ChiSq	ChiSq	Pr>ChiSq	ChiSq	Pr>ChiSq
Score	696.99	<.0001	647.25	<.0001	694.81	<.0001
Wald	62.78	<.0001	76.74	<.0001	69.60	<.0001
Panel B: Relogit	**JM P99.5%**		**JM LM**		**JM WSD**	
Test	ChiSq	Pr>ChiSq	ChiSq	Pr>ChiSq	ChiSq	Pr>ChiSq
Score	378.84	<.0001	490.33	<.0001	330.69	<.0001
Wald	160.15	<.0001	115.32	<.0001	106.74	<.0001

The table displays two tests (Score and Wald) for global beta zero which means the model does not capture any effect of the jump dynamics. Panel A refers to the logit model and Panel B to the relogit model. The figures are average of the stock-by-stock tests for the three jump detection methods, that are the percentile 99.5% return distribution (JM 99.5%), the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD).

For the sake of brevity, we only report in Table 3.4 the parameter estimates for the most restrictive jump detection method, i.e. JM WSD. Similar results for the two other jump detection methods are provided in Tables 3.9 and 3.10 available in appendix.

Table 3.4: Parameter Estimates of JM WSD Logit and Relogit Models

Panel A: Logit				Panel B: Relogit			
Variable	Estimate	StdErr	**	Variable	Estimate	StdErr	**
AbsImb_5	1.6067	0.9631	60.66%	AbsImb_5	2.1679	1.3290	45.90%
AbsCRT	0.0001	0.0000	37.70%	AbsCRT	0.0121	0.0045	59.02%
Dispersion	21.8421	19.6116	47.54%	Dispersion	39.0013	20.9430	67.21%
RS	5.8891	5.2365	52.46%	RS	3.5906	1.5018	70.49%
SE	2.9273	0.8598	91.80%	SA_morning	2.4776	0.8738	85.24%

The table reports the cross-sectional average parameter estimates for the most restrictive jump detection method, i.e. the weighted standard deviation jump detection test (JM WSD). Panel A refers to the logit model and Panel B to the relogit model. The exogenous variables are those presented in Section 3.2.2, that are the absolute depth imbalance at the five best quotes (AbsImb_5), the absolute cost of round trip (AbsCRT), the dispersion (Dispersion) and the relative spread (RS). The last variable (SE) is the dummy for periodicity of the first hour of the trading day. The column ** reports the percentage of stocks for which the parameter estimate is significant at a 95% confidence interval.

Consistent with Farmer et al. (2004); Joulin et al. (2008); Weber and Rosenov (2005), Table 3.4 contains only order book-based liquidity measures for both forecasting models. Those authors conclude that jumps are due to liquidity gaps in the order book, instead of the sudden surge of trading volume in the underlying stock. In our sample, the average duration of intervals where jumps occur is 200 seconds. This means that trading volume during jumps is below the average trading volume that corresponds to 2-minute (120-second) in our time frame and suggests the trading frequency is not especially high when jumps arise. In addition, our set of explanatory variables confirms that the order book is informative beyond the best bid and ask quotes. However, the most significant liquidity proxies vary somewhat across both forecasting models. Overall, the parameter coefficients are higher using the relogit model, except for the relative spread and the first hour periodicity dummy. The latter reveals for both models positive periodicity at the beginning of the continuous trading day for a very large majority of stocks. This is a support for Cont et al. (2014), who outline the lower liquidity at early stage of the continuous trading session due to information asymmetry.

3.4.2 Out-of-sample

The confusion matrices in Table 3.5 summarize the forecasting performance of our models out-of-sample. In line with the in-sample findings, we observe that the predictive ability is robust across jump tests.

Focusing on the most restrictive jump test (JM WSD) and the relogit model, we observe that it allows us to capture 87% of the non-events ($8717/(8717 + 1251)$) and 66% of the events ($21/(21 + 11)$). Basically, our model allows to isolate about 13 % of the database ($(1,251 + 21)/10,000$) where the majority (almost 70%) of price jumps occur. However, it is worthwhile to notice that the forecasting model has difficulties to pinpoint the exact interval where a jump occurs. In particular, it detects price jumps during about 12 intervals in a row around the actual occurrence of a price jump.[1] It corroborates that we capture mainly liquidity-driven jumps instead of information-driven jumps. The origin of this illiquidity remains an open question, it may be due to whether order cancellations or a low resiliency of the order book.

[1] As we rely on order book data to detect price jumps, we could infer the jump direction from our data. In this paper, we rather focus on execution price such that this aspect is not investigated here.

Table 3.5: Confusion Matrices

JM 99.5% Logit			JM 99.5% Relogit		
T:1%	Actual 0	Actual 1	T:1%	Actual 0	Actual 1
Predicted 0	8,863	33	Predicted 0	8,822	21
Predicted 1	1,086	18	Predicted 1	1,127	30
JM LM Logit			**JM LM Relogit**		
T:1%	Actual 0	Actual 1	T:1%	Actual 0	Actual 1
Predicted 0	9,063	23	Predicted 0	8,740	14
Predicted 1	898	16	Predicted 1	1,221	25
JM WSD Logit			**JM WSD Relogit**		
T:1%	Actual 0	Actual 1	T:1%	Actual 0	Actual 1
Predicted 0	9,145	24	Predicted 0	8,717	11
Predicted 1	823	8	Predicted 1	1,251	21

The table exhibits the confusion matrices for our logit and relogit models. Results across the 61 stocks are provided for the three jump detection methods, that are the percentile 99.5% return distribution (JM 99.5%), the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD). For both logit and relogit models, we use a 1% threshold to split events and non-events from the returned model probabilities. Each matrix displays on the one hand the actual split between events and non-events, and on the other hand, the predicted events and non-events. The results are rebased at 10,000 for an easier interpretation.

3.4.3 Robustness Checks

Focusing on the relogit model that exhibits the best forecasting power, we conduct some variants of our initial analyzes to check the robustness of the results. First, we apply the same methodology on a higher time frame, that is the average 5-minute trading volume on the underlying stock. For this purpose, we consider only the percentile-based jump detection method (JM 99.5%) since the jump frequency of more restrictive jump tests strongly decreases along with the time frame (Aït-Sahalia and Jacod, 2004). Second, we change the threshold level that is set to discretize the continuous model probabilities into events and non-events. Since the threshold should lie around the observed event frequency, we test 0.5% and 2% threshold levels. These tests are performed using the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD).

Higher Time Frame

As outlined in Table 3.6, the results tend to deteriorate on this higher time frame. Based on the percentile jump detection method, we still manage to isolate about half of the price jumps occurring at the average 5-minute trading volume time frame. Liquidity dynamics remains informative of price jumps when considering higher time frame, although its predictive ability is less strong and straightforward than at a higher frequency trading volume. The lower significance of our results on a higher time frame is consistent with the short-lived relationship between liquidity dynamics and price jumps in the stock market.

Table 3.6: Confusion Matrix for Average 5-minute Trading Volume

JM 99.5% Relogit		
T:1%	Actual 0	Actual 1
Predicted 0	8,792	22
Predicted 1	1,159	27

The table exhibits the confusion matrix for the relogit model. Results across the 61 stocks are provided for the the percentile 99.5% return distribution (JM 99.5%). We use a 1% threshold to split events and non-events from the returned model probabilities. The matrix displays on the one hand the actual split between events and non-events, and on the other hand, the predicted events and non-events. The results are rebased at 10,000 for an easier interpretation.

Changing the Threshold

For concision, we focus on the two most restrictive jump tests in this section.[1] Table 3.7 highlights that the results are consistent with our initial findings. As expected, the predictive power of our models is sensitive to the threshold level used to discretize the continuous probabilities into events and non-events. The optimal threshold level should lie around the observed event frequency in our dataset. We observe that a lower threshold leads to an improvement of events detection while it deteriorates the model accuracy for non-events detection.

[1] Results for the percentile-based jump test are reported in Table 3.11 of the appendix.

Table 3.7: Confusion Matrices for Different Thresholds

JM LM Relogit			JM WSD Relogit		
T:0.5%	Actual 0	Actual 1	T:2%	Actual 0	Actual 1
Predicted 0	9,423	21	Predicted 0	9,582	19
Predicted 1	538	18	Predicted 1	386	13
JM LM Relogit			**JM WSD Relogit**		
T:2%	Actual 0	Actual 1	T:0.5%	Actual 0	Actual 1
Predicted 0	8,035	10	Predicted 0	7,959	8
Predicted 1	1,926	29	Predicted 1	2,009	24

The table exhibits the confusion matrices for our relogit model. Results across the 61 stocks are provided for the two jump detection methods, that are the Lee and Mykland (2008)'s jump test (JM LM) and its extension that accounts for periodicity (JM WSD). For both logit and relogit models, we vary the threshold from 2% to 0.5% to split events and non-events from the returned model probabilities. Each matrix displays on the one hand the actual split between events and non-events, and on the other hand, the predicted events and non-events. The results are rebased at 10,000 for an easier interpretation.

3.5 Illustration

Our empirical findings show that our models perform pretty well in forecasting price jumps. In this section, we illustrate how market participants could take advantage of such models to improve the implementation of their investment decisions.

Using predictable patterns in price jumps could help reduce uncertainty around execution prices. This objective is especially relevant for value and passive investors, who are mainly concerned with preservation of asset value. For such investors, not trading when a price jump is likely to occur should be a simple and elementary trading strategy. Our models could be used to predict the occurrence of price jumps and then implement this execution strategy.

To check whether such an approach leads to a significant reduction in execution price uncertainty, we focus on our relogit model using the most restrictive jump detection test (JM WSD) and run several bootstrap simulations on our out-of-sample dataset. They will allow us to compare the trading strategy mentioned above with a naive one (i.e. trading without considering the risk of price jumps). For this purpose, we consider three different order sizes that are equivalent to the average 2-minute trading volume for the stock multiplied by 30, 50 and 100 respectively. Basically, for each size, we first simulate order execution resulting from a strategy that prevents from trading when the model predicts a price jump. We then simulate order execution randomly to mimic a naive strategy. For each simulated execution, we compute the high-low across the trade prices as a proxy for price uncertainty. We finally conduct both mean and median comparisons using Wilcoxon signed rank tests for the high-low between both strategies. To take into account the time-of-day effect, we also distinguish different periods in the trading day (morning, afternoon, and entire day respectively). For each order size and trading period under scrutiny, we run 1000 simulations for each execution strategy. Each simulation mimics order execution on a given stock that is selected at random among the whole sample of stocks.

Results are reported in Table 3.8. We can observe in both panels that a large percentage of order executions lead to a significant lower price uncertainty when the risk of price jumps is considered. For example, in Panel A, this percentage ranges from 67% for the smallest order size to more than 81% for the largest orders. This reveals that the price uncertainty reduction increases with order size. The findings do not really vary when we focus on different periods in the trading day.

To check for a potential liquidity effect across stocks, we replicate our analysis on the most and the least liquid stocks respectively. Both are defined with respect to market capitalization quartiles. Tables 3.12 and 3.13 available in appendix exhibit the results for lower quartile stocks and upper quartile stocks respectively. Overall, the results remain similar except for the least liquid stocks which do not reveal a clear positive relationship between price uncertainty reduction and order size.

All our findings show that there is an obvious incentive not to trade when our model detects a risk of price jumps, confirming its usefulness for investors mainly concerned with price uncertainty.

Table 3.8: Execution Strategy Comparison: Avoiding Price Jumps versus Naive Trading

Panel A : Median comparison	WSD_T1_30			WSD_T1_50			WSD_T1_100		
	price jumps	naive		price jumps	naive		price jumps	naive	
AM	0.000581	0.000661	71.5%***	0.000633	0.000667	74.6%***	0.000605	0.000663	80.8%***
PM	0.000637	0.000659	67.4%***	0.000637	0.000644	73.7%***	0.000605	0.000655	74.6%***
ALL	0.000642	0.000665	68.5%***	0.000627	0.000647	73.3%***	0.000652	0.000664	78.9%***
Panel B: Mean comparison	WSD_T1_30			WSD_T1_50			WSD_T1_100		
	price jumps	naive		price jumps	naive		price jumps	naive	
AM	0.000758	0.000848	78.6%***	0.000752	0.000854	78.4%***	0.000786	0.000860	86.0%***
PM	0.000794	0.000834	71.4%***	0.000784	0.000831	75.0%***	0.000750	0.000833	82.5%***
ALL	0.000794	0.000842	71.9%***	0.000767	0.000846	78.8%***	0.000770	0.000845	83.1%***

The table reports the results for the comparison of high-low across trade prices for each trading strategy. For each simulated order execution, the high-low across the trade prices is used as a proxy for price uncertainty. We then perform Wilcoxon signed rank tests for both the median (Panel A) and the mean (Panel B). WSD_T1_30 refers to an order size equivalent to the average 2-minute trading volume for the stock multiplied by 30. WSD_T1_50 and WSD_T1_100 stand for the same reference volume multiplied by 50 and 100 respectively. Different periods in the trading day are considered: morning (AM), afternoon (PM), and the entire day (ALL). $pricejumps$ refers to the median (Panel A) or average (Panel B) high-low computed across the execution prices simulated for the strategy avoiding to trade when there is a risk of price jump. $naive$ refers to the median (Panel A) or average (Panel B) high-low computed across the execution prices simulated for the naive strategy. The third column exhibits the percentage of simulations that lead to a significant lower high-low for the strategy based on price jumps. *** means the results are significant at a 99% level. All the results are based on 1000 simulations. Each simulation mimics order execution on a given stock that is selected at random among the whole sample of stocks.

3.6 Conclusion

In this paper, we investigate the predictability of price jumps on the Euronext 100 stocks using trades, orders and order book data. We apply three jump detection methods from a given percentile of the return distribution that implies constant volatility to more restrictive jump tests that account for time-varying volatility and periodicity in financial markets. Our methodology relies on an event chart, which allows us to control for trading volume effects and to deal with non trading issues at high frequency aggregation. Using a logistic model and an extension adjusted to rare events, we first select the optimal set of explanatory variables with a general-to-specific model selection process. We then evaluate the forecasting performance of our models, both in and out-of-sample.

Our results outline some predictable patterns in price jumps mainly at high frequency aggregation. Our empirical results point out to a low density of the order book that tends to last for few intervals while trading activity is pretty low during price jumps. Although these findings suggest that jumps are caused by temporary illiquidity, what drives this illiquidity is still an open question, whether due to order cancellations or a low resiliency of the order book.

Overall, price disruptions happen in majority during the early continuous trading session. This intraday pattern remains significant, albeit in a lower magnitude, after accounting for periodicity in the stock market, which is consistent with a lower liquidity in the morning due to information asymmetry (Cont et al., 2014). Out-of-sample and using the most restrictive jumps test, our relogit model allows to isolate 13% of the whole database where the majority (66%) of price jumps occur.

We also assess the robustness of our approach by changing the time frame and the threshold set to discretize the continuous model probabilities into events and non-events. The results are consistent with our initial findings. The informational content of the order book is however mostly short-lived with a deterioration of the forecasting ability from the average 5-minute trading volume time frame.

Since our empirical findings show that our models perform pretty well in forecasting price jumps, we illustrate how market participants could take advantage of them to reduce uncertainty around execution prices. This objective is especially relevant for value and passive investors, who are mainly concerned with preservation of asset value. Using several bootstrap simulations for different order sizes and periods of the day, we find that a large percentage of order executions lead to a significant lower price uncertainty when the risk of price jumps is taken into account. There is an obvious incentive not to trade when our model detects a risk of price jumps, confirming its usefulness for investors mainly concerned with price uncertainty.

3.7 Appendix

Table 3.9: Parameter Estimates for the JM 99.5% Relogit Model

Variable	Estimate	StdErr	**
AbsImb_5	1.7267	1.1674	47.54%
AbsCRT	0.0206	0.0063	52.46%
Dispersion	42.1583	9.9585	77.05%
RS	3.2564	1.3474	62.29%
SE	3.7338	0.4167	96.72%

The table reports the cross-sectional average parameter estimates for the relogit model using the percentile-based jump detection method (JM 99.5%). The exogenous variables are those presented in Section 3.2.2, that are the absolute depth imbalance at the five best quotes (AbsImb_5), the absolute cost of round trip (AbsCRT), the dispersion (Dispersion) and the relative spread (RS). The last variable (SE) is the dummy for periodicity of the first hour of the trading day. The column ** reports the percentage of stocks for which the parameter estimate is significant at a 95% confidence interval.

Table 3.10: Parameter Estimates of JM LM Relogit Model

Variable	Estimate	StdErr	**
AbsImb_5	2.3438	1.2344	62.29%
AbsCRT	0.0062	0.0021	68.85%
Dispersion	48.3460	19.3173	86.88%
RS	2.7169	1.4905	63.93%
SE	2.5458	0.7167	72.13%

The table reports the parameter estimates for the relogit model using the Lee and Mykland (2008)'s jump test (JM LM). The exogenous variables are those presented in Section 3.2.2, that are the absolute depth imbalance at the five best quotes (AbsImb_5), the absolute cost of round trip (AbsCRT), the dispersion (Dispersion) and the relative spread (RS). The last variable (SE) is the dummy for periodicity of the first hour of the trading day. The column ** reports the percentage of stocks for which the parameter estimate is significant at a 95% confidence interval.

Table 3.11: Confusion Matrices for Different Thresholds

JM 99.5% Relogit		
T:0.5%	Actual 0	Actual 1
Predicted 0	9,347	33
Predicted 1	604	16

JM 99.5% Relogit		
T:2%	Actual 0	Actual 1
Predicted 0	7,365	12
Predicted 1	2,586	37

The table exhibits the confusion matrices for our relogit model. Results across the 61 stocks are provided for the percentile-based jump detection method (JM 99.5%). We vary the threshold from 2% to 0.5% to split events and non-events from the returned model probabilities. Each matrix displays on the one hand the actual split between events and non-events, and on the other hand, the predicted events and non-events. The results are rebased at 10,000 for an easier interpretation.

Table 3.12: Execution Strategy Comparison on Lower Quartile Stocks

Panel A : Median comparison	WSD_T1_30			WSD_T1_50			WSD_T1_100		
	price jumps	naive		price jumps	naive		price jumps	naive	
AM	0.000686	0.000711	74.7%***	0.000693	0.000709	75.8%***	0.000687	0.000724	77.6%***
PM	0.000689	0.000703	73.7%***	0.000669	0.000683	74.9%***	0.000672	0.000694	73.3%***
ALL	0.000668	0.000697	75.2%***	0.000672	0.000687	76.1%***	0.000691	0.000719	76.2%***
Panel B: Mean comparison	**WSD_T1_30**			**WSD_T1_50**			**WSD_T1_100**		
	price jumps	naive		price jumps	naive		price jumps	naive	
AM	0.000751	0.000964	79.1%***	0.000748	0.000957	83.9%***	0.000739	0.000969	84.7%***
PM	0.000725	0.000938	80.1%***	0.000788	0.000936	79.3%***	0.000756	0.000935	83.0%***
ALL	0.000751	0.000955	79.9%***	0.000729	0.000942	82.8%***	0.000737	0.000961	82.5%***

The table reports the results for the comparison of high-low across trade prices for each trading strategy. For each simulated order execution, the high-low across the trade prices is used as a proxy for price uncertainty. We then perform Wilcoxon signed rank tests for both the median (Panel A) and the mean (Panel B). WSD_T1_30 refers to an order size equivalent to the average 2-minute trading volume for the stock multiplied by 30. WSD_T1_50 and WSD_T1_100 stand for the same reference volume multiplied by 50 and 100 respectively. Different periods in the trading day are considered: morning (AM), afternoon (PM), and the entire day (ALL). $pricejumps$ refers to the median (Panel A) or average (Panel B) high-low computed across the execution prices simulated for the strategy avoiding to trade when there is a risk of price jump. $naive$ refers to the median (Panel A) or average (Panel B) high-low computed across the execution prices simulated for the naive strategy. The third column exhibits the percentage of simulations that lead to a significant lower high-low for the strategy based on price jumps. *** means the results are significant at a 99% level. All the results are based on 1000 simulations. Each simulation mimics order execution on a given stock that is selected at random among the least liquid stocks.

Table 3.13: Execution Strategy Comparison on Upper Quartile Stocks

Panel A : Median comparison	WSD_T1_30			WSD_T1_50			WSD_T1_100		
	price jumps	naive		price jumps	naive		price jumps	naive	
AM	0.000563	0.000584	71.7%***	0.000559	0.000586	75.1%***	0.000551	0.000590	81.9%***
PM	0.000568	0.000576	67.9%***	0.000557	0.000579	74.7%***	0.000536	0.000571	80.2%***
ALL	0.000566	0.000572	66.3%***	0.000566	0.000578	73.2%***	0.000542	0.000578	80.7%***
Panel B: Mean comparison	**WSD_T1_30**			**WSD_T1_50**			**WSD_T1_100**		
	price jumps	naive		price jumps	naive		price jumps	naive	
AM	0.000697	0.000814	74.6%***	0.000735	0.000829	80.8%***	0.000708	0.000828	89.3%***
PM	0.000716	0.000820	71.9%***	0.000724	0.000795	82.0%***	0.000726	0.000798	87.2%***
ALL	0.000708	0.000803	74.5%***	0.000706	0.000807	79.2%***	0.000716	0.000805	86.8%***

The table reports the results for the comparison of high-low across trade prices for each trading strategy. For each simulated order execution, the high-low across the trade prices is used as a proxy for price uncertainty. We then perform Wilcoxon signed rank tests for both the median (Panel A) and the mean (Panel B). WSD_T1_30 refers to an order size equivalent to the average 2-minute trading volume for the stock multiplied by 30. WSD_T1_50 and WSD_T1_100 stand for the same reference volume multiplied by 50 and 100 respectively. Different periods in the trading day are considered: morning (AM), afternoon (PM), and the entire day (ALL). $pricejumps$ refers to the median (Panel A) or average (Panel B) high-low computed across the execution prices simulated for the strategy avoiding to trade when there is a risk of price jump. $naive$ refers to the median (Panel A) or average (Panel B) high-low computed across the execution prices simulated for the naive strategy. The third column exhibits the percentage of simulations that lead to a significant lower high-low for the strategy based on price jumps. *** means the results are significant at a 99% level. All the results are based on 1000 simulations. Each simulation mimics order execution on a given stock that is selected at random among the most liquid stocks.

CHAPTER

FOUR

HIGH FREQUENCY TRADING AND MARKET STABILITY

4.1 Introduction[1]

In the aftermath of the 2008 financial crisis and the May 2010 flash crash the stability of financial markets has been debated. Market disruptions have numerous implications in terms of risk management (Duffie and Pan, 2001), derivative pricing (Bates, 2000; Eraker et al., 2003) and portfolio allocation with its influence on the optimal strategy (Jarrow and Rosenfeld, 1984; Liu et al., 2003). While overall market quality has improved lately (Castura et al., 2010), individual stock mini-crashes are prevalent (Golub et al.,

[1]This essay is a joint work with Jonathan Brogaard and Ryan Riordan.

2013). Those short-lived crashes could originate from several sources as outlined by Hendershott (2011): HFT activity, market structure changes, trading fragmentation and/or the disappearance of designed market makers.

In this paper, we analyze the relationship between HFT activity and high frequency market disruptions. We detect price jumps through the 99.99% percentile of ex-post observations. We made two cutoffs: idiosyncratic jumps and co-jumps; permanent and transitory jumps. Co-jumps are jumps that happen simultaneously in several individual stocks while idiosyncratic jumps are isolated individual stock jumps. A transitory jump is defined as a jump that reverses within thirty seconds of its inception while a permanent jump does not reverse in the same time period.

The mechanism of price jumps are not well understood. Farmer et al. (2004) outline that large price fluctuations in a short period of time are driven by time-varying liquidity supply. On the other hand, Jiang et al. (2010) and Miao et al. (2012) find that most jumps appear at pre-scheduled macroeconomic news in the US treasury bond market and in the stock index futures market. High frequency trading (HFT) activities are often accused of triggering or enhancing these events. In response to these concerns exchanges and regulators have already taken action. The Euronext stock exchange now imposes a cancelation fee to avoid the implementation of some HFT strategies. Several regulators are discussing the implementation of minimum liquidity provision by HFT firms.

Price jumps could be affected by HFT in several ways. In times of heightened market instability, when price jumps are more likely to occur, HFT firms may exit the market or could add to one-sided order imbalance, which could ignite or enhance a jump in prices. If so, the observation that HFT firms provide liquidity on average may hide that in times of market instability they switch from liquidity provision to liquidity taking.

We use the Nasdaq HFT dataset used in other researches (e.g. Brogaard et al. (2013)). The data divide market participants into two types, HFT and non-HFT (nHFT). The data also disclose which type of participants is taking and providing liquidity for each trade.

We find HFT do not cease their trading activity during or around price jumps. Moreover the increase in HFT activity is on their passive trading activity. Then, we investigate HFT/nHFT net position around jumps. HFT and nHFT trade on average aggressively when in direction of the price jump and passively when against the price jump direction. In net, we find that neither HFT nor nHFT exhibit a significant net position.

The likelihood of a permanent price jump to occur is higher when lagged HFT net position is in the direction of the price jump. At the opposite, transitory price jumps go along with HFT net position against the price jump direction. It reflects that HFT activity can be related to a quicker adjustment of prices to information which in turn may explain the prevalence of stock specific price jumps. The remainder of the paper proceeds as follows. Section 2 provides a review of the existing literature. Section 3 describes the dataset used in this paper. Section 4 presents the applied methodology. Sections 5 to 7 report the empirical results. Section 8 concludes.

4.2 Literature Review

High frequency trading (henceforth HFT) is one of the latest major development in financial markets. The investigation of its externalities in the market is of utmost importance given its prominence. Indeed, several papers (Zhang, 2010; Brogaard, 2011) estimate that HFT accounts for about 70% of trading volume in the U.S. capital market as from 2009.

A comprehensive definition of what HFT activities include remains elusive. According to Castura et al. (2010), it encompasses professional market participants that present some characteristics: high-speed algorithmic trading, the use of exchange co-location services along with individual data feeds, very short investment horizon and the submission of a large number of orders during the continuous trading session that are often cancelled shortly after submission.

The existing literature outlines overall that HFT activity improves market quality. Indeed, the rising of HFT activity went along with a reduction of the spread, a liquidity improvement and a reduction of intraday volatility (Castura et al., 2010; Angel et al., 2010; Hasbrouck and Saar, 2011). Hanson (2011) and Menkveld (2012) even describe HFT as the new market makers in U.S. financial markets. Indeed, HFT acts mostly as liquidity providers and engages in price reversal strategies (Brogaard, 2011).

Lately HFT activity was under the spotlight following the liquidity-induced flash crash on May 6th, 2010 that casts doubt on the soundness of HFT activity and its externalities on market stability and price efficiency.

Golub et al. (2013) document that along with a market quality improvement, individual stock mini-crashes are prominent as well. Hendershott (2011) puts forward several potential origins of those liquidity-driven crashes such as HFT activity, market structure changes, trading fragmentation and/or the disappearance of designed market makers.

In this paper, we document the relationship between HFT and market stability. We isolate period of instability by looking at high frequency price jumps and investigate the behavior of HFT from a microstructure viewpoint during those periods. Our focus straddles two literatures. First, we briefly review the literature on the link between price movements and liquidity provision. Second, we summarize the expanding literature on HFT and market stability.

Several papers document the relationship between order book imbalances and price movements. Chordia and Subrahmanyam (2004) report a positive correlation between daily order book imbalance, stock returns and volatility. Chordia et al. (2008) also highlight that return predictability is lower when the spread is tight. Cao et al. (2009) investigate the informational content of the order book and show it fosters price discovery in the market.

On higher frequency aggregation sampling, Harris and Panchapagesan (2005) confirm a relationship between the limit order book and future price movements. Hellström and Simonsen (2009) point out that the information content of the order book is short-lived. Indeed, they find some predictability at a 1 and 2-minute aggregation sampling on the Stockholm stock exchange while it vanishes quickly on lower frequency sampling.

The relationship between market disruptions and liquidity provision is not straightforward and depends on the market.

The consensus is that the jump frequency observed in the stock market is not fully explained by news, whether macroeconomic or firm specific. Indeed, Farmer et al. (2004) show that the trigger of large price fluctuations on the London Stock Exchange is time-varying liquidity supply. Hence, jump risks is stock specific. Illiquid stocks tend to suffer from jumps more often than liquid ones. Consistent results are found on US market data. Joulin et al. (2008) and Weber and Rosenov (2005) document that the root of price jumps is a lower density of the order book in spite of the surge of significant trading volume underlying the arrival of information. Lately, Boudt and Petitjean (2014) estimate that around 70% of jumps are liquidity-related on the DJIA index. They emphasize that the effective spread and the number of trades is informative of forthcoming jumps.

Recent papers (Jiang et al., 2010; Miao et al., 2012) on the US treasury bond market and in the stock index futures market contrast those results. They show that the majority of price jumps appear at pre-scheduled macroeconomic news.

Market instability is by definition a rarely occurring event which makes it a challenging issue to investigate. The behavior of HFT during those periods of instability remains broadly unexplored. The flash crash on May 6, 2010 is often given as an example of HFT deteriorating market stability. However, it relies on the strong assumption that the market would behave in a given way in the absence of HFT. Furthermore, the potential impact of a trader is related to its position. In this view, non-HFT who could attempt to buy/sell a large position quickly are more of a threat to market stability than HFT who typically limit their positions in a market.

Brogaard et al. (2013) show that HFT activity is positively correlated to public information, market-wide movements and limit order book imbalances. This higher HFT activity doesn't seem to prevent from market overreaction. Indeed, Kirilenko et al. (2011) highlight that while the flash crash was originated by a bad execution of a large order initiated by a non-HFT trader, HFT exhibits abnormal behaviors and exacerbates market volatility that day.

Several papers outline a tight relationship between high frequency activity and stock specific volatility (Brogaard, 2012; Zhang, 2010; Kirilenko et al., 2011). This higher stock price volatility could be the results of the interaction between HFT and fundamental traders (Zhang, 2010).

Theoretically, Biais et al. (2012) develop a framework where they show that HFT increases adverse selection costs for non-HFT. Jovanovic and Menkveld (2011) found similar results. This additional adverse selection cost comes from the higher speed of information processing by HFT. Foucault et al. (2013) show that this speed advantage leads to a higher fraction of trading volume that is made by informed traders, it increases trading volume, decreases liquidity, induces price changes that are more correlated with fundamental value movements, and reduces informed order flow autocorrelations.

Bernales (2013) outlines that HFT traders are more profitable in high volatile periods (volatility shocks), which suggests they may have an incentive to manipulate market volatility. In a theoretical framework, Goettler et al. (2009) found that the limit order market acts as a "volatility multiplier" in that prices are more volatile than the fundamental value of the asset. This is all the more true when the fundamental volatility of the asset is higher or when there is information asymmetry across traders.

In this framework, market stability could be affected by HFT controversial strategies that could indirectly generate volatility in the market. Hendershott et al. (2011) outline that over 90% of the orders submitted by HFT are either cancelled or modified (cancelled and resubmitted) before being filled. Lately, Gai et al. (2012) show that HFT increases the order cancellation/execution ratio, which supports the significant implementation of quote stuffing[1], and layering[2] strategies in the market. Finally, Egginton et al. (2013) show that quote stuffing is ubiquitous in the US stock market. They find that stocks that experience quote stuffing display lower liquidity, higher trading costs, and higher short term volatility. Lately some regulators consider imposing a cancellation fee to prevent HFT potential detrimental externalities on the stock market. There is also talks to impose obligations on HFT to provide a minimum amount of liquidity and prevent in such a way monetary drying up of liquidity, a role that was ensured previously by designed market makers.

On the other hand, market quality improves mainly as from 2006 and is tough to relate directly to the emergence of HFT activity (Castura et al., 2010). Indeed, almost at the same time, the market structure acknowledges major changes both in the U.S. and in Europe with respectively the implementation of RegNMS and MiFID. Furthermore, Golub et al. (2013) document that individual stock mini-crashes are prominent along with overall market quality improvement. Hendershott et al. (2011) puts forward several potential origins of those liquidity-driven crashes; among them high frequency trading activity.

Some papers report a tight relationship between high frequency activity and stock specific volatility (Kirilenko et al., 2011; Zhang, 2010; Brogaard, 2012). Indeed, HFT

[1]Quote stuffing consists in submitting a large number of orders followed immediately by a cancellation to generate order congestion.

[2]Layering consists in submitting a large number of orders in one side of the book that are not meant to be filled to facilitate the entry on the other side of the book.

generates the majority of order flow while it displays periodicity in order submission and a high rate of order cancellations and modifications. Significant changes in their market activity from liquidity providers to liquidity takers suggest HFT may emphasize volatility/price movements and cause overreaction in the market.

4.3 Data

The database consists in 40 large market capitalization stocks that are listed half-and-half on NASDAQ and the New York Stock Exchange (NYSE).[1] It is the same data used in other academic studies including Brogaard et al. (2013) and O'Hara et al. (2013).

The data span two years from January 1, 2008 to December 31, 2009. NASDAQ categorizes market participants as a high-frequency trading firm or non-high-frequency trading (nHFT) firm, which allows us to identify investor types. A limitation of using market participant identifiers (MPIDs) is that NASDAQ is unable to disentangle HFT activity by large integrated firms that also engage in low-frequency trading strategies. Our data cover trading activity on the NASDAQ trading venue, other trading venues activity is thus not accounted for in this paper.

The dataset identifies 26 HFT firms that act as independent HFT proprietary trading firms.[2] The dataset includes whether the buyer or seller initiated the trade and identifies the type of trader on both sides of the trade.

We supplement the NASDAQ HFT dataset with the National Best Bid and Offer (NBBO) from TAQ. The NBBO measures the best prices prevailing across all markets to focus on market-wide price discovery.

[1]NASDAQ OMX provides us the HFT database to academics under a non-disclosure agreement. Our data cover stocks such as Apple and GE.

[2]Some HFT firms were consulted by NASDAQ in the decision to make data available. No HFT firm played any role in which firms were identified as HFT and no firms that NASDAQ considers HFT are excluded. While these 26 firms represent a significant amount of trading activity and according to NASDAQ fit the characteristics of HFT, determining the representativeness of these firms regarding total HFT activity is not possible.

We remove trades that occur before 9:30 and after 16:00 as well as trades that take place during the opening and closing auction to focus on the stock market continuous trading hours. The filtered database consists in 41,342,013 10-second intervals.

4.4 Methodology

We arrange the database into 10-second intervals to detect market disruptions.[1] We consider the 99.99% percentile of midquote observations as a threshold for a price jump. This methodology assumes static volatility. In turn, our methodology focuses on volatility spike that is extreme midquote changes in a 10-second interval.

An alternative methodology to detect jumps is to use a jump test robust to time-varying volatility. In this framework, Lee and Mykland (2012) propose a jump test that account for microstructure noise. This methodology fits better a market disruption definition, that is an extreme midquote change in a given market volatility context. This definition is not broached in the current paper.

There is a choice between using midpoint quote price movements or transaction price movements. Relying on Boudt and Petitjean (2014) that show most price jumps originate from a low density of the order book in the stock market, we focus on midquote jumps. [2]

Price jumps occur for various reasons. They may be a consequence of new information or they may be the result of liquidity price pressures. The former we expect to result in a permanent shift in prices, while the latter prices should drift back toward their pre-jump level. Permanent price changes are a sign of price discovery and thought to be beneficial, while transitory price jumps are a sign of market instability and should be reduced. A transitory jump is defined as a jump that fully reverses within 30 seconds of its inception

[1]We repeat the analysis with 1-minute, and 5-minute intervals to control for the robustness of our results. Those results are available upon request.

[2]We repeat the analysis on transaction jumps as a robustness check. Those results are available upon request.

while a permanent jump does not fully reverse in the short run.

Another possible cutoff to disentangle informative/noise jumps is to match the release of public information either micro- or macro-news to jump occurrence. We opt for the price path methodology as it better reflects the detrimental or beneficial effect of a price jumps. Indeed, we may have jumps happening around news release that are noise and may not lead to a lasting price change.

An alternative measure of an information event is to model the efficient price. We do so using the Hasbrouck (1991) model. An informative jump is one that is within two standard deviations of the efficient price. By definition, larger jumps have more chance to be noise jumps.

We are also interested in jumps that vary based on their pervasiveness. We separate price jumps into idiosyncratic and systematic. Systematic jumps, referred to as co-jumps, are jumps that occur in several stocks within a small period of time. Specifically we define a jump to be a co-jump if at least 10% of our sample stocks experience a jump within the same minute. By contrast, idiosyncratic jumps are isolated individual stock jumps.

This cutoff permit us to distinguish between individual stock instability and market-wide instability.

In the analysis we report two 10-second intervals before and after the price jump. If several jumps occur in this event window, it may lead to contagion effects in our results. A bit more than 11% of our jumps occur in the same event window as another jump which allows us to state that contagion effects could only play a minor effect on our event window statistics.

Table 4.1 reports descriptive statistics of some microstructure variables during and around jumps.

Panel A reports descriptive statistics for all the jump events while Panel B and C highlight respectively the permanent and transitory jumps.

Using the 99.99% percentile methodology and a 10-second window, we isolate 3,431 jumps. Most of those jumps are permanent (2,669) with only 762 being transitory. By definition the percentile definition is roughly evenly distributed among stocks, small differences arise from no trading intervals.

Price dynamics unsurprisingly unveil a spike during the jump interval. Transitory jumps exhibit a wider scope than permanent jumps on average. The t+1 return shows the interval following jump inception is a pullback on average for transitory jumps while permanent jumps acknowledge no pullback. The trading volume both in shares and in US dollar is higher during the jump interval. Again transitory jumps go along with more trading activity than permanent jumps both during and around the jump interval. As expected, the net overall volume (in shares and in US dollar) is in the price jump direction. Net overall volume reverses after the jump inception for transitory price disruptions while it only moderates for permanent price jumps. Overall, we find that transitory price disruptions happen in a lower liquidity market context than permanent price disruptions. Indeed, the spread is wider and the depth lower for transitory jumps compared to permanent ones.

Table 4.1: Descriptive Statistics

Panel A: All	Mean (t-2)	Mean (t-1)	Mean (t)	Mean (t+1)	Mean (t+2)	Median (t)	Std. Dev. (t)	25 percentile	75 percentile
High Low	0.01222	0.01335	0.02349	0.01286	0.01160	0.01223	0.05579	0.00940	0.01663
Return	-0.00060	0.00035	0.01422	-0.00092	0.00060	0.00840	0.04264	0.00543	0.01207
Squared Return	0.00146	0.00153	0.00202	0.00152	0.00145	0.00007	0.01691	0.00003	0.00015
$ Volume	1488249	1557947	2017609	1498472	1448238	435693	7032470	158389	1409150
Share Volume	30545	32900	52979	33960	30386	13234	156977	4445	42120
Share OIB	-762.05	2499.67	16542.87	1577.15	1855.09	2600.00	71832.75	0.00	12259.00
$ OIB	-56553	84710	606914	264516	225261	90091	3986265	0	406961
Spread	0.00278	0.00260	0.00243	0.00236	0.00207	0.00077	0.00687	0.00035	0.00217
Depth	19.99	19.95	19.33	17.17	17.96	3.00	91.70	1.00	9.00
Number of Jumps	3431								
Panel B: Permanent Jumps	**Mean (t-2)**	**Mean (t-1)**	**Mean (t)**	**Mean (t+1)**	**Mean (t+2)**	**Median (t)**	**Std. Dev. (t)**	**25 percentile**	**75 percentile**
High Low	0.01003	0.01050	0.01955	0.00881	0.00799	0.01200	0.04397	0.00931	0.01629
Return	-0.00125	-0.00111	0.01320	0.00001	0.00081	0.00866	0.03671	0.00602	0.01216
Squared Return	0.00107	0.00119	0.00152	0.00069	0.00068	0.00008	0.01482	0.00004	0.00015
$ Volume	1438794	1428740	1955523	1441427	1430814	470616	6042691	178052	1429984
Share Volume	28619	31291	52665	32638	29657	14500	149577	5283	43254
Share OIB	-922.26	1732.99	17766.86	3260.77	3623.52	3294.00	72147.59	200.00	14019.00
$ OIB	-145530	17443	576065	272708	375728	109938	2940286	7906	457268
Spread	0.00251	0.00233	0.00202	0.00211	0.00185	0.00075	0.00621	0.00037	0.00191
Depth	21.02	21.57	20.86	18.14	19.34	3.00	101.40	1.00	9.00
Number of Jumps	2669								
Panel C: Transitory Jumps	**Mean (t-2)**	**Mean (t-1)**	**Mean (t)**	**Mean (t+1)**	**Mean (t+2)**	**Median (t)**	**Std. Dev. (t)**	**25 percentile**	**75 percentile**
High Low	0.02056	0.02537	0.03729	0.02737	0.02433	0.01309	0.08372	0.00963	0.01806
Return	0.00186	0.00652	0.01778	-0.00424	-0.00016	0.00689	0.05879	0.00270	0.01163
Squared Return	0.00295	0.00299	0.00377	0.00447	0.00418	0.00005	0.02268	0.00001	0.00014
$ Volume	1676297	2103813	2235071	1702750	1509675	312926	9738149	100244	1301479
Share Volume	37316	38549	54076	38598	32941	8766	180627	2379	34737
Share OIB	-199.07	5191.65	12255.71	-4328.78	-4337.89	1128.50	70598.45	-400.00	6780.00
$ OIB	-281779	516281	714966	-235179	-305298	44983	6426243	-16542	250317
Spread	0.00374	0.00356	0.00387	0.00322	0.00286	0.00090	0.00868	0.00029	0.00415
Depth	16.35	14.22	13.94	13.74	13.09	3.00	42.29	1.00	9.00
Number of Jumps	762								

Panel A reports the descriptive statistics for all jumps, Panel B focuses on permanent jumps while Panel C is transitory jumps. The considered variables are the High Low (High minus Low), return (relative close to close, positive if in the jump direction and vice versa), squared return, $ volume (in US dollar), share volume (in shares), share OIB (net position in shares, positive if in the jump direction and vice versa), $ OIB (net position in US dollar, positive if in the jump direction and vice versa), spread (relative bid-ask spread) and depth (depth in the jump direction at the best quote). We display the mean of the variables during and around jump occurrence as well as the median, standard deviation and the percentile 25% and 75%. The last row reports the number of jumps in total and for permanent/transitory cutoff.

Table 4.2 reports the Pearson correlation coefficients for all the variables of interest in this paper.

Net position is positively correlated with HFT and nHFT demand while negatively correlated with HFT and nHFT supply. It suggests that imbalance in trading activity comes mainly from aggressive trading for both HFT and nHFT. HFT net position demand/supply and nHFT net position demand/supply are positively related. At the opposite, net HFT demand is higher when there is less nHFT net position supply and vice versa.

Table 4.2: Pearson Correlation Coefficients

	High Low	Return	Squared Return	$ Volume	Share Volume	Share OIB	$ OIB	Spread	Depth	OIB HFT Demand	OIB HFT Supply	OIB nHFT Demand	OIB nHFT Supply
High Low	1.00	0.79	0.73	-0.01	0.00	0.00	0.00	0.00	-0.03	0.00	0.00	0.00	0.00
Return	0.79	1.00	0.94	-0.01	0.00	0.00	0.00	0.00	-0.03	0.00	0.00	0.00	0.00
Squared Return	0.73	0.94	1.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02	-0.01	0.01	-0.01	0.01
$ Volume	-0.01	-0.01	-0.01	1.00	0.59	0.49	0.32	0.00	0.17	0.29	-0.37	0.45	-0.44
Share Volume	0.00	0.00	-0.01	0.59	1.00	0.28	0.53	-0.01	0.04	0.16	-0.19	0.26	-0.28
Share OIB	0.00	0.00	-0.01	0.49	0.28	1.00	0.65	0.00	0.07	0.63	-0.76	0.89	-0.90
$ OIB	0.00	0.00	-0.01	0.32	0.53	0.65	1.00	0.00	0.01	0.41	-0.43	0.58	-0.63
Spread	0.00	0.00	-0.01	0.00	-0.01	0.00	0.00	1.00	-0.01	0.00	0.00	0.00	0.00
Depth	-0.03	-0.03	-0.02	0.17	0.04	0.07	0.01	-0.01	1.00	0.06	-0.08	0.05	-0.04
OIB HFT Demand	0.00	0.00	-0.01	0.29	0.16	0.63	0.41	0.00	0.06	1.00	-0.46	0.20	-0.59
OIB HFT Supply	0.00	0.00	0.01	-0.37	-0.19	-0.76	-0.43	0.00	-0.08	-0.46	1.00	-0.68	0.40
OIB nHFT Demand	0.00	0.00	-0.01	0.45	0.26	0.89	0.58	0.00	0.05	0.20	-0.68	1.00	-0.79
OIB nHFT Supply	0.00	0.00	0.01	-0.44	-0.28	-0.90	-0.63	0.00	-0.04	-0.59	0.40	-0.79	1.00

The Table displays the Pearson correlation coefficients between our considered variables. The variables are the High Low (High minus Low), return (relative close to close, positive if in the jump direction and vice versa), squared return, $ volume (in US dollar), share volume (in shares), share OIB (net position in shares, positive if in the jump direction and vice versa), $ OIB (net position in US dollar, positive if in the jump direction and vice versa), spread (relative bid-ask spread) and depth (depth in the jump direction at the best quote).

4.5 Trading Behavior around Jumps

4.5.1 Summary

We find HFT do not cease their activity during or around price jumps. Instead HFT activity significantly increases in such market condition. Looking more into details, we show that it is passive HFT activity that spike while aggressive activity remains broadly unchanged.

To investigate the one-sided of the order book activity, we report HFT/nHFT net position around jumps. HFT and nHFT trade on average aggressively when in direction of the price jump and passively when against the price jump direction. In all, we find that neither HFT nor nHFT exhibit a significant net position pattern during price jumps. The permanent/transitory cutoff supports our initial finding.

The likelihood of a permanent price jump to occur is higher when lagged HFT net position is in the direction of the price jump. At the opposite, transitory price jumps go along with HFT net position against the price jump direction. It reflects that HFT activity can be related to a quicker adjustment of prices to information which in turn may explain the prominence of stock specific price jumps outlined in Golub et al. (2013).

The size of the jump is mostly due to market condition. Volatile market environment (low depth and wide spread) as well as the sudden surge of trading volume are positively correlated to the jump size.

To evaluate whether HFT firms may have an incentive to try and trigger market disruptions, we evaluate their trading profits around price jumps. The results suggest that HFT firms have no obvious incentives to foster the inception of disruptions. Overall, nHFT make profit on their aggressive trading activity while HFT lose money on their passive trading activity during the jump interval. In all, we find no significant profit patterns during price jumps whether for HFT or nHFT.

4.5.2 Trading Volume around Jumps

A first concern that is often mentioned when market acknowledges unstable period is that HFT may withdraw the market. In this first table we investigate HFT activity in shares in the market and consider three cutoffs: All trading activity, Aggressive trading activity and passive trading activity.

Table 4.3 outlines that HFT do not cease or even decrease its activity during and around jumps. Indeed, HFT activity is 18% higher during jump interval compared to non-jump interval. This higher HFT trading activity is also true 10-second prior and after the jump interval with an increase HFT activity of 30%. nHFT display similar patterns with a 50+% increase in trading activity during jump occurrence. This higher activity is also true around the jump occurrence but to a lesser extent both before and after jumps inception. It confirms that jumps happen amid a strong trading activity context for all market participants whether HFT or not.

Table 4.3: Trading Volume around Jumps

	t-2		t-1		T		t+1		t+2	
HFT All	90.5%	***	127.0%	***	118.0%	***	130.6%	***	110.2%	***
	(5.13)		(5.36)		(6.18)		(7.1)		(6.74)	
HFT Demand	66.8%	***	104.5%	***	85.3%	***	107.2%	***	89.8%	***
	(4.38)		(5.57)		(5.47)		(6.83)		(6.43)	
HFT Supply	107.0%	***	138.6%	***	134.5%	***	147.1%	***	126.8%	***
	(5.13)		(4.83)		(5.81)		(6.64)		(6.44)	
	t-2		t-1		T		t+1		t+2	
nHFT All	113.6%	***	146.7%	***	158.5%	***	142.2%	***	115.0%	***
	(5.22)		(4.81)		(7.004)		(7.3)		(6.9)	
nHFT Demand	135.3%	***	163.2%	***	186.7%	***	159.7%	***	130.2%	***
	(5.26)		(4.38)		(6.8)		(7.01)		(6.69)	
nHFT Supply	110.3%	***	142.2%	***	159.7%	***	131.7%	***	102.5%	***
	(4.62)		(4.41)		(6.63)		(6.69)		(6.45)	

The table reports the ratio of the average HFT (nHFT) trading volume (in shares) during and around jump intervals over the average HFT (nHFT) trading volume (in shares) during non-jump intervals. HFT (nHFT) All sums up aggressive and passive HFT (nHFT) volume, HFT (nHFT) Supply is passive HFT (nHFT) volume and HFT (nHFT) Demand is aggressive HFT (nHFT) volume. The Table shows relative trading activity as well as its t-value against 100%. ***, **, * mean respectively that the mean trading activity is significant at 99%, 95% and 90%.

Looking more closely at the HFT activity breakdown, we find that the spike of HFT trading activity originates from a spike in their passive trading activity (+35%) while HFT aggressive activity lies lower (-15%) than in non-jump market conditions. It suggests that HFT increase their liquidity provision in times of market instability. nHFT display a spike in both their liquidity taking and providing activity. nHFT demand spike (+87%) is more than nHFT supply spike (+60%).

4.5.3 Net Positions around Jumps

Table 4.4 depicts HFT and nHFT net position during and around price jumps for the idiosyncratic/co-jumps cutoff. All in all, HFT and nHFT exhibit similar behavior. They trade aggressively in the direction of the jump and passively against the direction of the jump. In net, HFT are trading on average in the jump direction. Yet this behavior is not significant.

It is also worthwhile to mention that HFT tend to become significantly aggressive the interval prior to a jump inception while nHFT are net liquidity provider. It is an indication that HFT may be more informed about future price evolution than nHFT.

Focusing on the co-jumps/idiosyncratic cutoff, HFT exhibit very few significant trading behavior during co-jumps occurrence, we only find them to be aggressive in the co-jumps direction. By contrast, nHFT trade aggressively in the jump direction and passively against the jump direction during and prior to co-jumps occurrence. In net, nHFT activity is on average opposed to the jump direction but not significantly.

As for idiosyncratic jumps, HFT display some significant trading behaviors that suggest they may be more informed. Indeed, prior to the jump occurrence, HFT trade in the jump direction firstly passively and then aggressively while nHFT exhibit significant trading pattern only after the jump occurrence, they trade aggressively in the jump direction.

The idiosyncratic/co-jumps cutoff draws the same conclusion even though HFT net position is more pronounced for idiosyncratic jumps than for co-jumps.

Table 4.4: Net Positions around Jumps

Panel A: All	t-2		t-1		T		t+1		t+2	
HFT All	152.10		276.48		107.85		-3.61		32.80	
	(1.18)		(1.58)		(0.88)		(0.03)		(0.32)	
HFT Demand	-29.64		460.27	***	571.16	***	-43.63		-10.77	
	(0.27)		(3.6)		(5.93)		(0.41)		(1.2)	
HFT Supply	181.74	**	-183.79		-463.30	***	40.02		43.57	
	(2.36)		(1.18)		(5.53)		(0.52)		(0.59)	
nHFT All	-152.10		-276.48		-107.85		3.61		-32.80	
	(1.18)		(1.58)		(0.88)		(0.03)		(0.32)	
nHFT Demand	-606.21	**	564.97		1949.33	***	473.58	**	284.00	*
	(2.1)		(1.17)		(5.96)		(2.17)		(1.81)	
nHFT Supply	454.10		-841.45	**	-2057.19	***	-469.97	**	-316.80	**
	(1.59)		(2)		(6.67)		(1.98)		(2.07)	
Panel B: Cojumps	**t-2**		**t-1**		**T**		**t+1**		**t+2**	
HFT All	-1021.29		418.78		299.64		384.12		265.14	
	(1.28)		(0.67)		(0.49)		(0.69)		(0.52)	
HFT Demand	-1092.91		683.71		823.88	*	132.91		216.19	
	(1.41)		(1.25)		(1.93)		(0.33)		(0.62)	
HFT Supply	71.62		-264.93		-524.24		251.21		48.95	
	(0.24)		(0.63)		(1.32)		(0.53)		(0.14)	
nHFT All	1021.29		-418.78		-299.64		-384.12		-265.14	
	(1.28)		(0.67)		(0.49)		(0.69)		(0.52)	
nHFT Demand	-1977.60		3174.35	**	6084.38	**	646.81		741.73	
	(1.37)		(1.96)		(2.17)		(0.96)		(0.96)	
nHFT Supply	2998.89		-3593.12	**	-6384.02	**	-1030.93		-1006.87	
	(1.61)		(2.25)		(2.46)		(1.45)		(1.27)	
Panel C: Idiosyncratic jumps	**t-2**		**t-1**		**T**		**t+1**		**t+2**	
HFT All	281.48	**	260.86		87.69		-46.09		7.21	
	(2.51)		(1.43)		(0.73)		(0.42)		(0.07)	
HFT Demand	87.60		435.74	***	544.58	***	-62.98		-35.76	
	(1.04)		(3.39)		(5.65)		(0.58)		(0.41)	
HFT Supply	193.88	**	-174.88		-456.90	***	16.89		42.98	
	(2.46)		(1.05)		(5.53)		(0.25)		(0.58)	
nHFT All	-281.48	**	-260.86		-87.69		46.09		-7.21	
	(2.51)		(1.43)		(0.73)		(0.42)		(0.07)	
nHFT Demand	-455.00		278.47		1514.56	***	454.60	**	233.58	
	(1.63)		(0.55)		(7.3)		(1.97)		(1.54)	
nHFT Supply	173.52		-539.33		-1602.25	***	-408.51		-240.80	*
	(0.72)		(1.24)		(7.89)		(1.62)		(1.65)	

Panel A, B and C report respectively all jumps, co-jumps and idiosyncratic midquote jumps cutoffs. The tables display net trading volume for HFT and non-HFT (nHFT). Trading activity statistics are All (Demand and Supply), Demand (Aggressive trading) and Supply (Passive trading). The variables are positive if in the jump direction and vice versa. The Table shows mean trading activity as well as its t-value against zero. ***, **, * mean respectively that the mean trading activity is significant at 99%, 95% and 90%.

All our results suggest neither HFT nor nHFT have a significant net position during the jump interval. It is also worthwhile to outline that the scope of the imbalance is much more sizeable for nHFT than for HFT which supports that HFT does not hold a significant inventory during the continuous trading day.

4.5.4 Permanent and Transitory Jumps

Table 4.5 reports HFT and nHFT net position during and around permanent and transitory jumps. This cutoff is especially relevant to gauge the impact of HFT in the stock market. Indeed, transitory jumps are uninformed price movements (noise) that should be dampened by market participants while permanent jumps are informative price movements. In turn, transitory price jumps are detrimental price movements while permanent price jumps are seen as the incorporation of information into stock prices.

Overall, it supports our initial finding in that HFT and nHFT exhibit similar behaviors on average with aggressive trading in the jump direction and passive trading in the opposite jump direction. In net, HFT tend to exhibit net position in the permanent jump direction yet not significant while transitory jumps go along with HFT net position against the jump direction. nHFT net passive and aggressive activity are more pronounced than HFT net activity. This is another support for nHFT triggering price jumps. It is also worth to mention that net position is more prominent for permanent jumps than transitory ones.

Table 4.6 is a robustness check of Table 4.5 using another permanent/transitory cutoff definition. In this table, we use the Hasbrouck (1991) methodology to estimate an efficient price confidence interval. Jumps that stay within this efficient price confidence interval are said to be permanent, otherwise they are transitory. We find very consistent results between the two methodologies. It confirms that HFT net position is on average in the permanent jump direction and against the transitory jump direction. It is a support for HFT dampening transitory jumps and smoothing the incorporation of information into permanent jumps.

Table 4.5: Permanent and Transitory Jumps Based on Price Path

Panel A: Permanent Jumps	t-2		t-1		T		t+1		t+2	
HFT All	356.80	*	417.38		206.68		49.49		53.71	
	(1.87)		(1.46)		(0.94)		(0.27)		(0.3)	
HFT Demand	56.53		985.80	***	820.71	***	-118.05		-55.57	
	(0.39)		(4.33)		(4.8)		(0.68)		(0.38)	
HFT Supply	300.27	**	-568.42	**	-614.03	***	167.54		109.28	
	(2.24)		(2.25)		(4.22)		(1.23)		(0.85)	
nHFT All	-356.80	*	-417.38		-206.68		-49.49		-53.71	
	(1.87)		(1.46)		(0.94)		(0.27)		(0.3)	
nHFT Demand	-965.93	*	1459.08	*	3163.70	***	495.70		402.38	
	(1.95)		(1.75)		(5.22)		(1.33)		(1.51)	
nHFT Supply	609.13		-1876.46	**	-3370.38	***	-545.19		-456.09	*
	(1.368)		(2.49)		(5.87)		(1.41)		(1.77)	
Panel B: Transitory Jumps	t-2		t-1		T		t+1		t+2	
HFT All	-87.63		112.35		-1.07		-65.79		8.35	
	(0.52)		(0.62)		(0.012)		(0.56)		(0.11)	
HFT Demand	-130.56		-151.89	**	296.11	***	43.52		41.65	
	(0.81)		(1.96)		(4.05)		(0.41)		(0.56)	
HFT Supply	42.93		264.24		-297.18	***	-109.32	**	-33.30	
	(0.75)		(1.62)		(4.09)		(2)		(0.59)	
nHFT All	87.63		-112.35		1.07		65.79		-8.35	
	(0.52)		(0.62)		(0.012)		(0.56)		(0.11)	
nHFT Demand	-184.91		-476.55		610.91	***	447.68	**	145.50	
	(0.77)		(1.23)		(3.87)		(2.42)		(1.05)	
nHFT Supply	272.54		364.20		-609.85	***	-381.88		-153.85	
	(0.8)		(1.47)		(4.55)		(1.53)		(1.1)	

Panel A and B report respectively permanent jumps and transitory midquote jumps cutoffs. The tables display net trading volume for HFT and non-HFT (nHFT). Trading activity statistics are All (Demand and Supply), Demand (Aggressive trading) and Supply (Passive trading). The variables are positive if in the jump direction and vice versa. The Table shows mean trading activity as well as its t-value against zero. ***, **, * mean respectively that the mean trading activity is significant at 99%, 95% and 90%.

Table 4.6: Permanent and Transitory Jumps Based on Hasbrouck (1991)

Panel A: Permanent Jumps	t-2	t-1		T		t+1		t+2
HFT All	114.10	250.24		340.10		108.48		49.34
	(0.62)	(0.26)		(1.075)		(1.45)		(0.22)
HFT Demand	116.85	532.73	***	961.58	***	-133.79		-41.28
	(0.72)	(3.77)		(4.85)		(1.15)		(0.415)
HFT Supply	-2.75	-282.50	***	-621.48	***	242.27	**	90.62
	(0.13)	(3.61)		(5.81)		(1.97)		(0.99)
nHFT All	-114.10	-250.24		-340.10		-108.48		-49.34
	(0.62)	(0.26)		(1.075)		(1.45)		(0.22)
nHFT Demand	282.43	347.84	*	3171.00	***	841.80	**	452.14
	(0.27)	(1.68)		(5.13)		(2.15)		(1.315)
nHFT Supply	-396.53	-598.08	*	-3511.10	***	-950.28	**	-501.48
	(0.49)	(1.77)		(6.16)		(2.23)		(1.51)
Panel B: Transitory Jumps	**t-2**	**t-1**		**T**		**t+1**		**t+2**
HFT All	-24.67	-92.97		-165		-21.17		63.45
	(0.5)	(1.6)		(1.22)		(0.64)		(1.41)
HFT Demand	-45.85	-197.28	**	362.18	***	76.27		20.26
	(0.97)	(2.12)		(2.65)		(1.57)		(0.45)
HFT Supply	21.18	104.31		-527.18	***	-97.44	**	43.2
	(0.44)	(1.24)		(3.57)		(2.33)		(0.99)
nHFT All	24.67	92.97		165		21.17		-63.45
	(0.5)	(1.6)		(1.22)		(0.64)		(1.41)
nHFT Demand	18.6	-78.36		969.88	***	305.6	***	133.61
	(0.216)	(1.03)		(2.906)		(2.78)		(1.56)
nHFT Supply	6.07	171.33		-804.88	***	-284.43	*	-197.06
	(0.08)	(1.54)		(2.945)		(1.67)		(1.09)

Panel A and B report respectively permanent jumps and transitory jumps cutoffs using Hasbrouck pricing error model. The tables display net trading volume for HFT and non-HFT (nHFT). Trading activity statistics are All (Demand and Supply), Demand (Aggressive trading) and Supply (Passive trading). The variables are positive if in the jump direction and vice versa. The Table shows mean trading activity as well as its t-value against zero. ***, **, * mean respectively that the mean trading activity is significant at 99%, 95% and 90%.

4.6 Price Jumps Attributes

The likelihood of a permanent price jump to occur is higher when lagged HFT net position is in the direction of the price jump. At the opposite, transitory price jumps go along with HFT net position against the price jump direction. It reflects that HFT activity can be related to a quicker adjustment of prices to information which in turn may explain the prominence of stock specific price jumps outlined in Golub et al. (2013).

The size of the jump is mostly due to market condition. Volatile market environment (low depth and wide spread) as well as the sudden surge of trading volume are positively correlated to the jump size.

To evaluate whether HFT firms may have an incentive to try and trigger market disruptions, we evaluate their trading profits around price jumps. The results suggest that HFT firms have no obvious incentives to foster the inception of disruptions. Overall, nHFT make profits on their aggressive trading activity while HFT lose money on their passive trading activity during the jump interval. In all, we find no significant profit patterns during price jumps whether for HFT or nHFT.

4.6.1 Price Jumps Determinants

In Table 4.7, we assess the drivers of price jump occurrence through the following probit regressions:

$$JM_{i,t} = \beta_{i,0} + \beta_{i,1}HL_{i,t-1} + \beta_{i,2}Ret_{i,t-1} + \beta_{i,3}Ret^2_{i,t-1} + \beta_{i,4}vol_{i,t-1} + \beta_{i,5}Svol_{i,t-1}$$
$$+\beta_{i,6}Svol_{i,t} + \beta_{i,7}Spread_{i,t-1} + \beta_{i,8}Depth_{i,t-1} + \beta_{i,9}HFTnet_{i,t} + \beta_{i,10}HFTnet_{i,t-1},$$

$$JM_{i,t} = \beta_{i,0} + \beta_{i,1}HL_{i,t-1} + \beta_{i,2}Ret_{i,t-1} + \beta_{i,3}Ret^2_{i,t-1} + \beta_{i,4}vol_{i,t-1} + \beta_{i,5}Svol_{i,t-1}$$
$$+\beta_{i,6}Svol_{i,t} + \beta_{i,7}Spread_{i,t-1} + \beta_{i,8}Depth_{i,t-1}$$
$$+\beta_{i,9}HFTnetag_{i,t} + \beta_{i,10}nHFTnetag_{i,t} + \beta_{i,11}HFTnetag_{i,t-1} + \beta_{i,12}nHFTnetag_{i,t-1},$$

$$JM_{i,t} = \beta_{i,0} + \beta_{i,1}HL_{i,t-1} + \beta_{i,2}Ret_{i,t-1} + \beta_{i,3}Ret^2_{i,t-1} + \beta_{i,4}vol_{i,t-1} + \beta_{i,5}Svol_{i,t-1}$$
$$+\beta_{i,6}Svol_{i,t} + \beta_{i,7}Spread_{i,t-1} + \beta_{i,8}Depth_{i,t-1}$$
$$+\beta_{i,9}HFTnetpas_{i,t} + \beta_{i,10}nHFTnetpas_{i,t} + \beta_{i,11}HFTnetpas_{i,t-1} + \beta_{i,12}nHFTnetpas_{i,t-1},$$

where $JM_{i,t}$ is a binary that take 1 when a jump occurs and 0 otherwise for stock i at interval t. HL_{t-1} is the lagged High-Low, Ret_{t-1} is the lagged return, Ret^2_{t-1} is the lagged squared return, vol_{t-1} is the lagged share volume, $Svol_{t-1}$ is the lagged dollar volume, $Svol_t$ is the dollar volume, $Spread_{t-1}$ is the lagged bid-ask spread, $Depth_{t-1}$ is the lagged best quote, $HFTnet_t$ is the net HFT position, $HFTnet_{t-1}$ is the lagged net HFT position, $HFTnetag_t$ is the net demand HFT position, $nHFTnetag_t$ is the net demand nHFT position, $HFTnetpas_t$ is the net supply HFT position and $nHFTnetpas_t$ is the net supply nHFT position.

Price jumps tend to happen in a low liquidity context as outlined in Table 4.7. Indeed, we find that a wide spread and a lower depth increase the likelihood of price jump occurrence.

To the same extend, we outline that higher lagged trading volume both in shares and in US dollar goes along with more jumps. Price dynamics also outline a more volatile market environment. Lagged high low prominence increases the probability of price jumps while lagged squared returns are negatively correlated to jump occurrence. This price dynamic measure could be expected in the opposite sign. It unveils that lagged extreme price movements decrease the probability of a price jump to happen as the measure gives more weight to large return. Finally, the lagged return sign switches between permanent and transitory jumps. It indicates that permanent jumps tend to happen when lagged return is already in the jump direction which suggests a trending market. At the opposite, transitory jumps tend to develop in sideways market.

Looking at the permanent/transitory cutoff, we show that lagged HFT net position in the jump direction increase the probability of a permanent jump to occur. During permanent price jumps, nHFT are aggressive against the jump direction while HFT net passive position is in the jump direction during and prior the jump occurrence. At the opposite, HFT net position is against the jump direction during transitory jumps. Prior to the jump, we outline that HFT net aggressive position is significant and in the jump direction while nHFT provide liquidity in the jump direction.

Table 4.7: Price Jumps Determinants

Panel A: Permanent Jumps	Scaling	Coef.		ME	Coef.		ME	Coef.		ME
High Low (t-1)	1.E+03	2623.80	***	0.50	2622.90	***	0.50	2622.60	***	0.50
		(192.69)			(192.56)			(192.52)		
Return (t-1)	1.E+03	3254	***	0.62	3287	***	0.63	3220	***	0.61
		(12.81)			(12.95)			(12.24)		
Squared Return (t-1)	1.E+03	-23699	***	-4.53	-23880	***	-4.57	-23483	***	-4.48
		(21.42)			(21.5)			(20.56)		
Share Volume (t-1)	1.E+08	247.00	***	0.05	248.70	***	0.05	256.40	***	0.05
		(535.45)			(517.82)			(548.39)		
$ Volume (t-1)	1.E+09	28.15	***	0.01	27.95	***	0.01	26.37	***	0.01
		(46.14)			(47.28)			(39.28)		
$ Volume (t)	1.E+09	0.32		0.00	1.70		0.00	2.01		0.00
		(0.01)			(0.17)			(0.22)		
Spread (t-1)	1.E+01	210.00	***	0.04	210.20	***	0.04	210.00	***	0.04
		(182.32)			(183.15)			(182.21)		
Depth (t-1)	1.E+06	-1228.80	***	-0.23	-1224.90	***	-0.23	-1230.10	***	-0.23
		(154.73)			(153.69)			(155.23)		
OIB HFT (t)	1.E+08	82.80		0.02						
		(1.44)								
OIB HFT Demand (t)	1.E+08				-12.06		0.00			
					(0.02)					
OIB HFT Supply (t)	1.E+08							112.40	*	0.02
								(1.87)		
OIB nHFT Demand (t)	1.E+08				-52.74	***	-0.01			
					(4.69)					
OIB nHFT Supply (t)	1.E+08							35.02	**	0.01
								(2.17)		
OIB HFT (t-1)	1.E+08	103.00	***	0.02						
		(6.52)								
OIB HFT Demand (t-1)	1.E+08				29.63		0.01			
					(0.47)					
OIB HFT Supply (t-1)	1.E+08							142.60	***	0.03
								(8.29)		
OIB nHFT Demand (t-1)	1.E+08				-11.45		0.00			
					(0.38)					
OIB nHFT Supply (t-1)	1.E+08							-27.83	*	-0.01
								(1.87)		
Stock Fixed Effects		Yes			Yes			Yes		
N		2669			2669			2669		
Pseudo R^2		4.89%			4.88%			4.91%		

The table is a probit regression with stock fixed effects and clustered standard errors. The dependent variable takes 1 when a jump occurs and 0 otherwise. The independent variables are defined such as in Table 1. We report the results for the permanent / transitory cutoff. The Table shows the parameter estimates as well as their wald chi-square. ***, **, * mean respectively that the parameter is significant at 99%, 95% and 90%. The marginal effect (ME) of each explanatory variable is also reported.

Panel B: Transitory Jumps	Scaling	Coef.		ME	Coef.		ME	Coef.		ME
High Low (t-1)	1.E+03	2551.80	***	0.03	2548.90	***	0.03	2554.10	***	0.03
		(26.38)			(26.32)			(26.5)		
Return (t-1)	1.E+03	-37711	***	-0.49	-37134	***	-0.48	-36132	***	-0.48
		(42.04)			(45.1)			(46.56)		
Squared Return (t-1)	1.E+03	-434769	***	-5.67	-426111	***	-5.52	-408942	***	-5.39
		(18.5)			(20.16)			(21.04)		
Share Volume (t-1)	1.E+08	106.80	***	0.00	112.80	***	0.00	118.60	***	0.00
		(34.79)			(32.25)			(35.97)		
$ Volume (t-1)	1.E+09	30.66	***	0.00	28.36	***	0.00	27.79	***	0.00
		(26.55)			(18.95)			(20.26)		
$ Volume (t)	1.E+09	10.36	***	0.00	11.56	***	0.00	12.11	***	0.00
		(3.81)			(3.48)			(4.77)		
Spread (t-1)	1.E+01	218.20	***	0.00	217.90	***	0.00	217.90	***	0.00
		(62.81)			(62.52)			(62.37)		
Depth (t-1)	1.E+06	-775.10	***	-0.01	-818.50	***	-0.01	-821.80	***	-0.01
		(10.26)			(10.95)			(10.88)		
OIB HFT (t)	1.E+08	-232.70	***	0.00						
		(5.34)								
OIB HFT Demand (t)	1.E+08				-113.80		0.00			
					(0.76)					
OIB HFT Supply (t)	1.E+08							-101.90		0.00
								(0.52)		
OIB nHFT Demand (t)	1.E+08				21.73		0.00			
					(0.24)					
OIB nHFT Supply (t)	1.E+08							26.57		0.00
								(0.82)		
OIB HFT (t-1)	1.E+08	-15.44		0.00						
		(0.02)								
OIB HFT Demand (t-1)	1.E+08				-257.90	***	0.00			
					(6.87)					
OIB HFT Supply (t-1)	1.E+08							275.90	***	0.00
								(6.81)		
OIB nHFT Demand (t-1)	1.E+08				5.25		0.00			
					(0.02)					
OIB nHFT Supply (t-1)	1.E+08							-17.23		0.00
								(0.23)		
Stock Fixed Effects		Yes			Yes			Yes		
N		762			762			762		
Pseudo R^2		6.33%			6.39%			6.43%		

The table is a probit regression with stock fixed effects and clustered standard errors. The dependent variable takes 1 when a jump occurs and 0 otherwise. The independent variables are defined such as in Table 1. We report the results for the permanent / transitory cutoff. The Table shows the parameter estimates as well as their wald chi-square. ***, **, * mean respectively that the parameter is significant at 99%, 95% and 90%. The marginal effect (ME) of each explanatory variable is also reported.

4.6.2 Price Jumps Size Determinants

We assess the drivers of price jump size through the following equations:

$$HL_{i,t} = \beta_{i,0} + \beta_{i,1}HL_{i,t-1} + \beta_{i,2}Ret_{i,t-1} + \beta_{i,3}Ret^2_{i,t-1} + \beta_{i,4}vol_{i,t-1} + \beta_{i,5}Svol_{i,t-1}$$
$$+\beta_{i,6}Svol_{i,t} + \beta_{i,7}Spread_{i,t-1} + \beta_{i,8}Depth_{i,t-1} + \beta_{i,9}HFTnet_{i,t} + \beta_{i,10}HFTnet_{i,t-1} + \varepsilon_{i,t},$$

$$HL_{i,t} = \beta_{i,0} + \beta_{i,1}HL_{i,t-1} + \beta_{i,2}Ret_{i,t-1} + \beta_{i,3}Ret^2_{i,t-1} + \beta_{i,4}vol_{i,t-1} + \beta_{i,5}Svol_{i,t-1}$$
$$+\beta_{i,6}Svol_{i,t} + \beta_{i,7}Spread_{i,t-1} + \beta_{i,8}Depth_{i,t-1}$$
$$+\beta_{i,9}HFTnetag_{i,t} + \beta_{i,10}nHFTnetag_{i,t} + \beta_{i,11}HFTnetag_{i,t-1} + \beta_{i,12}nHFTnetag_{i,t-1} + \varepsilon_{i,t},$$

$$HL_{i,t} = \beta_{i,0} + \beta_{i,1}HL_{i,t-1} + \beta_{i,2}Ret_{i,t-1} + \beta_{i,3}Ret^2_{i,t-1} + \beta_{i,4}vol_{i,t-1} + \beta_{i,5}Svol_{i,t-1}$$
$$+\beta_{i,6}Svol_{i,t} + \beta_{i,7}Spread_{i,t-1} + \beta_{i,8}Depth_{i,t-1}$$
$$+\beta_{i,9}HFTnetpas_{i,t} + \beta_{i,10}nHFTnetpas_{i,t} + \beta_{i,11}HFTnetpas_{i,t-1} + \beta_{i,12}nHFTnetpas_{i,t-1} + \varepsilon_{i,t},$$

where $HL_{i,t}$ is the High-Low during a jump interval for stock i at interval t. HL_{t-1} is the lagged High-Low, Ret_{t-1} is the lagged return, Ret^2_{t-1} is the lagged squared return, vol_{t-1} is the lagged share volume, $Svol_{t-1}$ is the lagged dollar volume, $Svol_t$ is the dollar volume, $Spread_{t-1}$ is the lagged bid-ask spread, $Depth_{t-1}$ is the lagged best quote, $HFTnet_t$ is the net HFT position, $HFTnet_{t-1}$ is the lagged net HFT position, $HFTnetag_t$ is the net demand HFT position, $nHFTnetag_t$ is the net demand nHFT position, $HFTnetpas_t$ is the net supply HFT position and $nHFTnetpas_t$ is the net supply nHFT position.

Table 4.8 points out to the positive relation between the size of permanent jumps and lagged price dynamics measure (High Low and Squared return). A sudden surge of dollar trading volume (significantly lower prior to jump and significantly higher during the price jump) is another driver of permanent jumps size. A lower depth at the best quote is also informative for a larger jump size. HFT net position shows that the size of permanent jump is larger when HFT trade aggressively in the jump direction during its inception.

At the opposite the size of transitory jumps seem to be related to dollar trading volume during the jump interval. The transitory jump size is negatively related to lagged return. It suggest that transitory jumps go along with sideways market. HFT and nHFT trading behaviors have no significant effects on the size of transitory jumps.

Table 4.8: Price Jumps Size Determinants

Panel A: Permanent Jumps	Scaling	Coef.		Coef.		Coef.	
High Low (t-1)	1.E+03	249.43	***	253.80	***	253.84	***
		(8.35)		(8.49)		(8.47)	
Return (t-1)	1.E+03	9.20		17.61		18.59	
		(0.59)		(1.05)		(1.12)	
Squared Return (t-1)	1.E+03	277.34	***	250.16	***	245.40	***
		(4.08)		(3.54)		(3.49)	
Share Volume (t-1)	1.E+08	0.40		0.29		0.25	
		(1.51)		(1.07)		(0.91)	
$ Volume (t-1)	1.E+09	-0.45	***	-0.44	***	-0.42	**
		(2.73)		(2.57)		(2.47)	
$ Volume (t)	1.E+09	0.47	***	0.46	***	0.46	***
		(3.29)		(3.1)		(3.1)	
Spread (t-1)	1.E+01	-0.27		-0.27		-0.27	
		(0.32)		(0.32)		(0.32)	
Depth (t-1)	1.E+06	-2.73	***	-2.91	***	-2.71	***
		(2.81)		(2.97)		(2.83)	
OIB HFT (t)	1.E+08	0.58					
		(0.64)					
OIB HFT Demand (t)	1.E+08			1.94	*		
				(1.7)			
OIB HFT Supply (t)	1.E+08					-0.67	
						(0.56)	
OIB nHFT Demand (t)	1.E+08			0.40			
				(0.67)			
OIB nHFT Supply (t)	1.E+08					-0.41	
						(0.53)	
OIB HFT (t-1)	1.E+08	0.79					
		(1.49)					
OIB HFT Demand (t-1)	1.E+08			0.10			
				(0.15)			
OIB HFT Supply (t-1)	1.E+08					0.37	
						(0.6)	
OIB nHFT Demand (t-1)	1.E+08			-0.55	***		
				(3.17)			
OIB nHFT Supply (t-1)	1.E+08					0.54	**
						(2.08)	
Stock Fixed Effects		Yes		Yes		Yes	
N		2669		2669		2669	
Adjusted R^2		76.03%		76.17%		76.12%	

The table is a panel regression with stock fixed effects and clustered standard errors on jump intervals. The dependent variable is the High Low during the 10-second jump interval. The independent variables are defined such as in Table 1. We report the results for the permanent / transitory cutoff. The Table shows the parameter estimates as well as their t-stat. ***, **, * mean respectively that the parameter is significant at 99%, 95% and 90%.

Panel B: Transitory Jumps	Scaling	Coef.		Coef.		Coef.	
High Low (t-1)	1.E+03	248.64		261.09		255.90	
		(1.22)		(1.18)		(1.16)	
Return (t-1)	1.E+03	-404.83	***	-414.54	**	-413.21	**
		(2.73)		(2.41)		(2.5)	
Squared Return (t-1)	1.E+03	7142.05		5900.05		6542.70	
		(0.71)		(0.51)		(0.6)	
Share Volume (t-1)	1.E+08	0.13		-0.22		-0.33	
		(0.11)		(0.13)		(0.19)	
$ Volume (t-1)	1.E+09	-0.34		-0.25		-0.21	
		(0.73)		(0.44)		(0.46)	
$ Volume (t)	1.E+09	1.02	**	0.96	**	0.88	*
		(2.16)		(2.08)		(1.9)	
Spread (t-1)	1.E+01	-1.20		-1.21		-1.22	
		(1.19)		(1.21)		(1.21)	
Depth (t-1)	1.E+06	-2.75		-1.44		-2.73	
		(0.87)		(0.41)		(0.8)	
OIB HFT (t)	1.E+08	-0.45					
		(0.13)					
OIB HFT Demand (t)	1.E+08			0.26			
				(0.11)			
OIB HFT Supply (t)	1.E+08					-20.68	
						(0.99)	
OIB nHFT Demand (t)	1.E+08			2.88			
				(0.77)			
OIB nHFT Supply (t)	1.E+08					-0.24	
						(0.08)	
OIB HFT (t-1)	1.E+08	-0.13					
		(0.06)					
OIB HFT Demand (t-1)	1.E+08			-1.37			
				(0.28)			
OIB HFT Supply (t-1)	1.E+08					-8.60	
						(1.12)	
OIB nHFT Demand (t-1)	1.E+08			-0.03			
				(0.02)			
OIB nHFT Supply (t-1)	1.E+08					1.86	
						(0.76)	
Stock Fixed Effects		Yes		Yes		Yes	
N		762		762		762	
Adjusted R^2		81.84%		81.94%		82.47%	

The table is a panel regression with stock fixed effects and clustered standard errors on jump intervals. The dependent variable is the High Low during the 10-second jump interval. The independent variables are defined such as in Table 1. We report the results for the permanent / transitory cutoff. The Table shows the parameter estimates as well as their t-stat. ***, **, * mean respectively that the parameter is significant at 99%, 95% and 90%.

4.7 Profitability

To more accurately discern HFT firms' interest and anticipation of price jumps, we analyze HFT profits during market dislocations. Profit over a given interval is computed assuming no inventory at the beginning and at the end of the interval. The profit is thus the realized profit corrected for inventory position (valued at the midquote).

Table 4.9 shows HFT and nHFT profits around jumps as well as profits for the permanent/transitory cutoff. All in all, we find that HFT tend to incur losses on average during market disruptions while nHFT make profits on average. Nevertheless, few significant patterns appear in a 10-second interval. Overall nHFT make profits on their aggressive trading activity during jumps while HFT lose money on their liquidity supply activity. At a 90% confidence interval, we confirm this finding on all trading activity with HFT losing money at the expense of nHFT during price jumps.

We cannot associate permanent price jumps with any significant profit pattern for either HFT or nHFT. Still on average, we find that HFT exhibit a negative PnL at the opposite of nHFT.

Finally, HFT profits are statistically insignificantly different from zero for transitory jumps. The splitting of aggressive and passive trading activity unveils that HFT incur a loss on their liquidity provision while nHFT make profit on their aggressive trading activity at a 90% confidence interval. This behavior is consistent with HFT firms acting as market markers. It also show that HFT have no incentives to try and trigger price jumps in the stock market.

A last interesting point to outline is the fact that HFT do not make significant profits on their aggressive trading during price jumps. It is an indication that their aggressive trading activity during price jumps is not about market timing but rather inventory management which further supports the idea of HFT acting as a market maker at that time.

Table 4.9: Profitability around Jumps

Panel A: jumps	t-2	t-1		T		t+1	t+2
HFT All	-1371.69	-1390.30		-1655.59	*	-826.69	112.33
	(0.66)	(1.49)		(1.9)		(0.96)	(0.16)
HFT Demand	896.04	214.72		-81.14		-320.39	153.62
	(0.73)	(0.36)		(0.77)		(0.98)	(0.05)
HFT Supply	-2267.72	-1605.02	*	-1574.45	**	-506.30	-41.29
	(1.53)	(1.66)		(2.19)		(0.82)	(0.18)
nHFT All	1371.69	1390.30		1655.59	*	826.69	-112.33
	(0.66)	(1.49)		(1.9)		(0.96)	(0.16)
nHFT Demand	2585.96	2314.44		2505.53	***	1372.90	337.23
	(1.56)	(0.85)		(2.83)		(0.96)	(0.31)
nHFT Supply	-1214.28	-924.14		-849.94		-546.21	-449.56
	(0.87)	(0.46)		(0.27)		(0.3)	(0.07)
Panel B: Permanent jumps	**t-2**	**t-1**		**T**		**t+1**	**t+2**
HFT All	-3310.44	-637.16		-988.84		-981.63	405.15
	(1.18)	(0.8)		(1.27)		(1.53)	(0.74)
HFT Demand	-506.50	339.68		50.23		148.28	201.22
	(0.79)	(0.42)		(0.7)		(0.74)	(0.57)
HFT Supply	-2803.94	-976.84		-1049.07	*	-1129.91	203.94
	(1.36)	(1.21)		(1.76)		(1.53)	(0.23)
nHFT All	3310.44	637.16		988.84		981.63	-405.15
	(1.18)	(0.8)		(1.27)		(1.53)	(0.74)
nHFT Demand	3166.69	1384.00		1944.50		1604.75	239.98
	(1.41)	(1.02)		(1.49)		(1.17)	(0.69)
nHFT Supply	215.75	-746.84		-945.66		-623.12	-645.14
	(0.59)	(0.2177)		(1.06)		(0.81)	(0.43)
Panel C: Transitory jumps	**t-2**	**t-1**		**T**		**t+1**	**t+2**
HFT All	1708.18	-10787.94		-2477.69		-576.28	-376.99
	(0.56)	(1.4)		(1.46)		(0.29)	(0.23)
HFT Demand	3212.81	-287.19		-351.17		-155.17	50.33
	(1.32)	(0.11)		(1.17)		(0.94)	(0.36)
HFT Supply	-1504.63	-10500.74	*	-2126.52	*	-421.11	-427.32
	(1.59)	(1.76)		(1.92)		(0.64)	(0.29)
nHFT All	-1708.18	10787.94		2477.69		576.28	376.99
	(0.56)	(1.4)		(1.46)		(0.29)	(0.23)
nHFT Demand	1779.14	11476.50		3079.39	**	893.09	510.37
	(1.51)	(1.63)		(2.16)		(0.93)	(0.57)
nHFT Supply	-3487.32	-688.57		-601.70		-316.81	-133.38
	(0.68)	(0.77)		(0.34)		(0.13)	(0.01)

The Table exhibits the mean US dollar HFT profit around all jumps, permanent and transitory jumps. The profit is computed as the realized profit during the interval plus the flattening of the net position valued at the mid-quote at the end of the 10-second interval. The amount reported are the means and their t-value against zero. ***, **, * mean respectively that the profit is significant at 99%, 95% and 90%.

4.8 Conclusion

In this paper, we take advantage of a NASDAQ HFT dataset that identifies investor types (HFT and non-HFT) to investigate the relationship between HFT activity and market disruptions. Market disruptions are detected using a straightforward 99.99% percentile. We cutoff idiosyncratic jumps (isolated individual stock jumps) and co-jumps (jumps that occur simultaneously in several individual stocks).

We find HFT are more active during market disruptions. It suggests that HFT process information quicker than nHFT and act as the main liquidity providers in time of higher information asymmetry.

We investigate the one-sided order book activity for HFT/nHFT around jumps. On average, they trade aggressively when in direction of the price jump and passively when against the price jump direction. In all, we find that neither HFT nor HFT exhibit a significant net position pattern. Yet, it confirms that HFT are providing more liquidity against the jump direction.

The likelihood of permanent price jumps increases when lagged HFT net position is in the direction of the price jump. It reflects that the higher HFT activity leads to a quicker adjustment of stock prices to information, which in turn may explain the prominence of stock specific price jumps outlined in Golub et al. (2013). At the opposite, transitory price jumps go along with HFT net position against the price jump direction.

The size of the jump is mostly due to market condition. Volatile market environment (low depth and wide spread) as well as the sudden surge of trading volume are positively correlated to the jump size.

To evaluate whether HFT firms may have an incentive to try and trigger market disruptions, we evaluate their trading profits around price jumps. The results suggest that HFT firms have no obvious incentives to foster the inception of disruptions. Overall, nHFT make profits on their aggressive trading activity while HFT lose money on their passive trading activity during the jump interval. In all, we find no significant profit pattern during price jumps whether for HFT or nHFT.

CHAPTER

FIVE

CONCLUSION

As it shows through this dissertation, market microstructure and market efficiency are closely related. In the wake of recent financial crashes, financial markets' quality and stability became even more of a topical issue.

Indeed, regulators took actions in order to make markets more efficient in the last decade. Those new regulations profoundly impacted the microstructure of financial markets. The end of national stock exchanges monopoly reduced transaction costs but went along with more fragmented liquidity. Then, multiple trading venues were the breeding-ground for high frequency trading activity that grews up at an impressive pace in the last decade. Along with those more competitive and transparent trading venues, liquidity dark pools with complete opacity on trading activity emerged. Those radical changes in markets' organization bring a lot of uncertainties and questions in the side effects of those evolutions. Indeed, there is very few work done on the pros and cons of such changes in markets' organization and some concerns arise mainly after the flash crash in 2010. That day, the US market index Dow Jones Industrial average crashes about 9% and then recovers within minutes. This market organization in the making raises mainly concerns regarding market stability.

Given the recent evolution of financial markets, the prominence of information asymmetry and adverse selection costs have increased for non-HFT traders, which make the assessment of execution price uncertainty a relevant issue to address in such context.

In those essays, we put forward some predictability of intraday liquidity. Transaction costs and execution price uncertainty are of paramount importance in today financial markets given it represents a significant chunk of trading performance (Fama and Blume, 1966) and it is closely related to market efficiency and stability. It is also related to market quality.

Chapter 2 provides a novel methodology to forecast periods of low intraday execution price uncertainty that is a widely unexplored topic.

We focus on patient traders who care more about market resiliency than immediacy. In this framework, the widely used Amihud ratio (Amihud, 2002) suits best our research purpose.

Our methodology consists in a fixed trading volume aggregation of continuous market data instead of the classical time-based aggregation. Aside of being less noisy for forecasting purpose (Hasbrouck, 1996), it confers several methodological advantages over the existing literature.

Using classical time series modelling, we find some predictable patterns in the Amihud ratio that statistically outperform a random walk benchmark model. We confirm the relevance of our findings by implementing a simple execution price improvement strategy that leads to a significant reduction of price execution uncertainty.

On average, we outline that the best time to trade for uninformed traders is towards the end of the continuous trading day. This time period displays a negative periodicity, a higher predictability of the Amihud ratio and more trading volume.

Chapter 3 investigates whether illiquidity can forecast liquidity. Indeed, we detect and forecast price jumps in the stock market using trades, orders and order book data on Euronext 100 stocks.

The aggregation of our data based on a fixed trading volume allows us to control for non-trading issues at high frequency aggregation. The jump detection methods considered range from the straightforward percentile of ex-post observations to the Lee and Mykland (2008)'s jump test robust to time-varying volatility and periodicity in the stock market.

Relying on a logistic model and an extension adjusted to rare events such as price jumps, we demonstrate some predictable patterns at high frequency aggregation (average 2-minute trading volume). We confirm that jumps are mainly due to liquidity provision gaps in the order book while trading activity remains often pretty low.

The fair out-of-sample behavior of our forecasting models leads us to implement an execution prices improvement strategy for the sake of illustration. We show that uninformed market participants have an incentive not to trade when our model predicts a risk of price jumps in the stock market. Indeed, it allows to significantly decrease the uncertainty surrounding the execution prices for market participants.

Chapter 4 aims at giving more insights on the inception of price jumps in the stock market. The recent changes in market organization lead to an overall improvement of market quality but in the meantime, Golub et al. (2013) outline that individual stocks short-lived crashes are prominent. The prime suspect of this higher instability in the stock market is high frequency trading (HFT) activity. While some regulators already took action such as imposing a cancelation fee to avoid some potential detrimental behaviors from HFT, there is a lack of objective empirical findings on the subject. In this essay, we contribute to this research area by investigating the relationship between high frequency trading (HFT) activity and idiosyncratic/systematic price jumps in the stock market.

We detect price jumps relying on a percentile of midquote observations. On a 2-year Nasdaq HFT dataset, we implement a lead/lag analysis around the occurrence of price jumps to grasp the incidence of HFT activity during those time periods.

We document that HFT are more active during market disruptions. More specifically, we confirm that HFT accumulate an inventory position in the opposite direction of the jump. The likelihood of permanent price jumps occurrence increases when lagged HFT net position is in the direction of the price jump. It reflects that the higher HFT activity leads to a quicker adjustment of stock prices to information, which in turn may explain the prominence of stock specific price jumps outlined in Golub et al. (2013). At the opposite, transitory price jumps go along with HFT net position against the price jump direction.

HFT profitability around price jumps suggests that HFT firms have no obvious incentives to try and trigger the inception of disruptions. Overall, nHFT make profits on their aggressive trading activity while HFT lose money on their passive trading activity during the jump interval. In all, we find no significant profit pattern during price jumps whether for HFT or nHFT.

Our essays deliver relevant insights for both market participants and regulators.

Active market participants can take advantage of our findings to filter entry signals that present a high uncertainty of their execution prices or as a circuit breaker within an algorithmic trading system.

Patient market participants benefit from our results by having more insight on the best time to trade in order to preserve asset value. In this view, we point out the pros of the end of the continuous trading session for less sophisticated and uninformed traders to handle their orders.

We also provide empirical results on the externalities of HFT in the stock market. Indeed, while regulators already took action to tackle potential detrimental behaviors from HFT, there is very few objective research on the externalities and soundness of HFT activity in the market.

Finally, it extends academic knowledge on market microstructure. Our novel methodology answers some gaps of the existing liquidity forecasting literature. Furthermore, it contributes to a better understanding of market disruptions that have numerous implications in modern finance theory.

To conclude, it is worth to outline some potential limitations of our work. As in all empirical work, we can question the generalization of our results outside of our sample data. In the same vein, the HFT identification, the impact of multi-lateral trading facilities and the ageing Euronext dataset may impact our results. This work can be extended in several directions, the relationship outlined between liquidity and volatility opens an avenue for further research. Alternative sampling schemes unveil some significant advantages over the usual time sampling scheme. Finally, the third essay is still a work in progress.

BIBLIOGRAPHY

Aït-Sahalia, Y. and J. Jacod (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*, 487–528.

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *Journal of Finance 57(3)*, 1047–1091.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*, 31–56.

Andersen, T., T. Bollerslev, P. Frederiksen, and M. Nielsen (2010). Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics 25*, 233–261.

Andersen, T., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*, 75–93.

Angel, J., L. Harris, and C. Spatt (2010). Equity trading in the 21st century. *Working paper*.

Bajgrowicz, P. and O. Scaillet (2011). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Swiss Finance Institute*, 11–36.

Bakshi, G. and G. Panayotov (2010). First-passage probability, jump models, and intra-horizon risk. *Journal of Financial Economics 1*(95), 20–40.

Barndorff-Nielsen, O. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*(1), 1–48.

Bates, D. (2000). Post'87 crash fears in S&P 500 futures option market. *Journal of Econometrics 94*, 181–238.

Bernales, A. (2013). How fast can you trade? hft in dynamic limit order markets. *Working paper*.

Biais, B., T. Foucault, and S. Moinas (2012). Equilibrium high frequency trading. *Working paper*.

Biais, B., L. Glosten, and C. Spatt (2005). Market microstructure: a survey of microfondations, empirical results and policy implication. *Journal of Financial Markets 8*(2), 217–264.

Bialkowski, J., S. Darolles, and G. Le Fol (2008). Improving VWAP strategies: A dynamical volume approach. *Journal of Banking and Finance 32*, 1709–1722.

Boudt, K., C. Croux, and S. Laurent (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance 18*(2), 353–367.

Boudt, K. and M. Petitjean (2014). Intraday liquidity dynamics of DJIA stocks around price jumps. *Journal of Financial Markets (Forthcoming) 17*, 121–149.

Brogaard, J. A. (2011). High frequency trading and its impact on market liquidity. *Working Paper*.

Brogaard, J. A. (2012). High frequency trading and volatility. *Working Paper*.

Brogaard, J. A., T. Hendershott, and R. Riordan (2013). High frequency trading and price discovery. *Working Paper*.

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

Castle, J. and D. Hendry (2011). Automatic selection for non-linear models. *System Identification, Environmental Modelling and Control (Forthcoming)*.

Castura, J., R. Litzenberger, and R. Gorelick (2010). Market efficiency and microstructure evolution in the U.S. equity markets: A high frequency perspective. *Working paper-RGM advisors*.

Chordia, T., R. Roll, and A. Subrahmanyam (2008). Liquidity and market efficiency. *Journal of Financial Economics 87*, 249–268.

Chordia, T. and A. Subrahmanyam (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics 72*, 485–518.

Christensen, K., R. Oomen, and P. M. (2011). Fact or friction: Jumps at ultra high frequency. *Working paper*.

Cont, R., A. Kukanov, and S. Stoikov (2014). The price impact of order book events. *Journal of Financial Econometrics 1*(12), 47–88.

De Vilder, R. and M. Visser (2008). Ranking and combining volatility proxies for GARCH and stochastic volatility models. *MPRA paper*.

Degryse, H., F. Jong, M. Ravenswaaij, and G. Wuyts (2005). Aggressive orders and the resiliency of a limit order market. *Review of Finance 9*(2), 201–242.

Degryse, H., M. Van Acher, and G. Wuyts (2009). Dark liquidity pools: Issues in competition, design and regulation of crossing networks. *Facing New Regulatory Frameworks in Securities Trading in Europe, Intersentia Press*.

DeWinne, R. and C. D'Hondt (2007). Hide-and-seek in the market: Placing and detecting hidden orders. *Review of Finance 11*, 663–692.

Duffie, D. and J. Pan (2001). Analytical value-at-risk with jumps and credit risk. *Finance and Stochastics* (5), 155–180.

Dufour, A. and R. Engle (2000). Time and the price impact of a trade. *Journal of Finance 55(6)*, 2467–2498.

Egginton, J., B. Van Ness, and R. Van Ness (2013). Quote stuffing. *Working paper*.

Engle, R. and J. Lange (2001). Predicting VNET: A model of the dynamics of market depth. *Journal of Financial Markets 4*, 113–142.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance 58*, 1269–1300.

Fama, E. and M. Blume (1966). Filter rules and stock-market trading. *Journal of Business 39*(1), 226–241.

Farmer, J. D., L. Gillemot, F. Lillo, S. Mike, and A. Sen (2004). What really causes large price changes? *Quantitative Finance 4*(4), 383–397.

Foucault, T., J. Hombert, and I. Rosu (2013). News trading and speed. *Working paper*.

Gai, J., C. Yao, and M. Ye (2012). The externalities of high-frequency trading. *Working paper*.

Goettler, R., C. Parlour, and U. Rajan (2009). Informed traders and limit order markets. *Journal of Financial Economics 93*(1), 67–87.

Golub, A., J. Keane, and S. Poon (2013). High frequency trading and mini flash crashes. *Working paper*.

Goyenko, R., C. Holden, and C. Trzcinka (2009). Do liquidity measures measure liquidity. *Journal of Financial Economics 92*(2), 153–181.

Hall, A. and N. Hautsch (2007). Modeling the buy and sell intensity in a limit order book market. *Journal of Financial Markets 10*, 249–286.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics 23*(4), 365–380.

Hansen, P. R., A. Lunde, and J. M. Nason (2003). Model confidence sets for forecasting models. *Working paper*.

Hanson, T. A. (2011). The effects of high frequency traders in a simulated market. *Working paper*.

Harris, L. (2003). Trading and exchanges: Market microstructure for practitioners. *Financial Management Association survey and synthesis series. Oxford University Press*.

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

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance 46*, 179–207.

Hasbrouck, J. (1996). Modeling microstructure time series. *Handbook of Statistics 14*, 647–692.

Hasbrouck, J. and G. Saar (2011). Low-latency trading. *Journal of Financial Markets (Forthcoming)*.

Hellström, J. and O. Simonsen (2009). Does the open limit order book reveal information about short-run stock price movements? *Working paper*.

Hendershott, T. (2011). High frequency trading and price efficiency. *The future of computer trading in financial markets-Driver review*.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *The Journal of Finance 66*(1), 1–33.

Hendry, D. and H.-M. Krolzig (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control 25*, 831–866.

Irvine, P., G. Benston, and E. Kandel (2000). Liquidity beyond the inside spread: Measuring and using information in the limit order book. *Unpublished Working paper*.

Jain, P., P. Jain, and T. McInish (2011). The predictive power of limit order book for future volatility, trade price, and speed of trading. *Working paper*.

Jarrow, R. and E. Rosenfeld (1984). Jump risks and the intertemporal capital asset pricing model. *Journal of Business 3*(57), 337–351.

Jiang, J., I. Lo, and A. Verdelhan (2010). Information shocks, liquidity shocks, jumps, and price discovery: Evidence from the U.S. treasury market. *Journal of Financial and Quantitative Analysis 46*(2), 527–551.

Joulin, A., A. Lefevre, D. Grunberg, and J.-P. Bouchaud (2008). Stock price jumps: News and volume play a minor role. *Technical Report*.

Jovanovic, B. and A. Menkveld (2011). Middlemen in limit-order markets. *Working paper*.

Kalay, A., O. Sade, and A. Wohl (2004). Measuring stock illiquidity: An investigation of the demand and supply schedules at the TASE. *Journal of Financial Economics 74*(3), 461–486.

Kang, W. and W. Yeo (2008). Liquidity beyond the best quote: A study of the NYSE limit order book. *Working Paper Series*.

Kempf, A. and O. Korn (1997). Market depth and order size. *Journal of Financial Market 2*(1), 29–48.

King, G. and L. Zeng (2001). Logistic regression in rare events data. *Political Analysis 9*(2), 137–163.

Kirilenko, A., P. Kyle, M. Samadi, and T. Tuzun (2011). The flash crash: The impact of high frequency trading on an electronic market. *CFTD working paper*.

Kyle, A. S. and A. A. Obizhaeva (2013). Market microstructure invariance: Theory and empirical tests. *Working paper*.

Lanne, M. (2007). Forecasting realized exchange rate volatility by decomposition. *International Journal of Forecasting 23*, 307–320.

Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets 10*(1), 1–25.

Lee, S. S. and P. A. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies 21*(6), 2535–2563.

Lee, S. S. and P. A. Mykland (2012). Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics* (168), 396–406.

Lemmens, A. and C. Croux (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 276–286.

Liu, J., F. Longstaff, and J. Pan (2003). Dynamic asset allocation with event risk. *Journal of Finance 1*(58), 231–259.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets 3*(3), 205–258.

Menkveld, A. (2012). High frequency trading and the new-market makers. *Working paper*.

Miao, H., S. Ramchander, and J. Zumwalt (2012). Information driven price jumps and trading strategy: Evidence from stock index futures. *Working paper*.

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

O'Hara, M. (1995). Market microstructure theory. *Oxford; Blackwell*.

O'Hara, M., C. Yao, and M. Ye (2013). What's not there: the odd-lot bias in TAQ data. *Journal of Finance - Forthcoming*.

Oomen, R. (2006). Properties of bias corrected realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics 24*(2), 219–237.

Press, S. (1967). A compound events model for security prices. *Journal of Business 3*(40), 317–335.

Ruud, P. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution. *Econometrica 61*, 123–138.

Schwartz, E. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance 52*(3), 923–973.

Tombeur, G. and G. Wuyts (2011). Does the information in the limit order book help to predict returns? *Working paper*.

Tsay, R. S. and G. Tiao (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association 79*(385), 84–96.

Weber, P. and B. Rosenov (2005). Large stock price changes: Volume or liquidity? *Quantitative Finance 6*(1), 7–14.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software 10*(10), 1 – 17.

Zhang, X. (2010). High-frequency trading, stock volatility, and price discovery. *Working paper*.

Zheng, B., E. Moulines, and F. Abergel (2012). Price jump prediction in limit order book. *Working paper*.

Louvain School of Management
Doctoral Thesis

Essays on Intraday Order Flow Dynamics and Liquidity Forecasting

Thibaut MOYAERT

This dissertation consists in three essays that investigate liquidity provision dynamics up to high frequency sampling scheme. The latest evolution of financial markets towards both more trading fragmentation and the emergence of high frequency traders make liquidity provision a topical issue. The first essay provides a novel methodology to forecast liquidity as a trade size specific concept using time series modeling. The second essay focuses on the prominence of market disruptions in the stock market. It investigates what are the most informative microstructure variables with respect to market disruptions' occurrence. Finally, the third essay documents the trading behavior of high frequency traders (HFT) around market disruptions. Indeed, HFT are often referred to as threatening financial markets' stability. Altogether, those essays hold implications for both market participants and regulators.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
www.uclouvain.be