

https://dspace.org/

The impact of algorithmic trading on market quality: evidence from the Johannesburg Stock Exchange

Courdent, Aurelie

2021
https://hdl.handle.net/10539/32599

Downloaded from a DSpace repository. https://dspace.org/

THE IMPACT OF ALGORITHMIC TRADING ON MARKET QUALITY: EVIDENCE FROM THE JOHANNESBURG STOCK EXCHANGE

by

AURELIE COURDENT (858765)

MASTER OF COMMERCE IN BUSINESS FINANCE

in the

SCHOOL OF ECONOMICS AND FINANCE

at the

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Supervisor: MR. DAVID MCCLELLAND

Date of submission:

14 December 2020

SCHOOL OF ECONOMIC AND
BUSINESS SCIENCES

Declaration Regarding Plagiarism

I (full names & surname):	Aurelie Courdent
Student number:	858765

Declare the following:

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this assignment is my own, original work. Where someone else's work was used (whether from a printed source, the Internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not copy and paste any information directly from an electronic source (e.g., a web page, electronic journal article or CD ROM) into this document.
4. I did not make use of another student's previous work and submitted it as my own.
5. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his/her own work.

ABSTRACT

High Frequency Trading (HFT) is a trading method that relies on sophisticated algorithms to analyse markets and execute large amounts of orders within milliseconds. In the last two decades, this new technology has gained traction globally and now accounts for the majority of the trading volume on the Johannesburg Stock Exchange. Despite the dominance of HFT, studies on the topic have been scarce outside of the U.S. This study seeks to examine the effects of HFT on market quality in a South African context. Market quality is proxied by two measures, market liquidity and short-term volatility. Firstly, the study makes use of a set of proxies for algorithmic trading (AT), namely average trade size, odd-lot volume ratio and trade-to-order volume ratio. This captures the level of AT activity from intra-day trade and quote data for all JSE-listed stocks from 1 January 2019 to 30 June 2019. Secondly, panel regressions are used to determine the relationship between these proxies and the market quality measures, while controlling for market capitalisation. The panel data estimators indicated a strong positive relationship between average trade size and the liquidity measures. However, they revealed an inverse relationship between the odd-lot volume ratio and the trade-to-order volume ratio. These findings thus suggest that the proxies capture the activity of different types of algorithmic traders (ATs), market-makers in the first case, opportunistic in the second. Finally, the study found a strong positive relationship with short-term volatility. The study concludes that overall, AT has a positive impact on market quality, despite carrying the risk of causing instability in certain markets.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1. BACKGROUND 1
   - *1.1. FACTORS LEADING TO THE ADVENT OF ALGORITHMIC TRADING 2*
   - *1.2. HIGH FREQUENCY TRADING STATISTICS 5*
   - *1.3. THE DEBATE ABOUT THE EFFECTS OF HFT ON MARKET QUALITY 6*
2. PROBLEM STATEMENT 8
3. RESEARCH OBJECTIVES 8
4. IMPORTANCE AND BENEFITS OF THE PROPOSED STUDY 9
5. STUDY LIMITATIONS 9

CHAPTER 2: LITERATURE REVIEW 10

1. INTRODUCTION OF THE LITERATURE REVIEW 10
2. ALGORITHMIC TRADING AND MARKET LIQUIDITY 10
3. ALGORITHMIC TRADING AND PRICE DISCOVERY 13
4. ALGORITHMIC TRADING AND ADVERSE SELECTION COSTS 18
5. ALGORITHMIC TRADING AND SHORT-TERM VOLATILITY 20
6. STOCKS ACTIVELY TRADED BY ALGORITHMIC TRADERS 22
7. CONCLUSION OF THE LITERATURE REVIEW 23

CHAPTER 3: METHODOLOGY 24

1. FACTORS AFFECTING THE SELECTION OF THE RESEARCH METHOD 24
2. SAMPLING AND DATA COLLECTION 25
3. DATA PREPARATION PROCESS 26
4. ALGORITHMIC TRADING PROXIES 27
   - *4.1. AVERAGE TRADE SIZE 28*
   - *4.2. ODD-LOT VOLUME RATIO 29*
   - *4.3. TRADE-TO-ORDER VOLUME RATIO 29*
   - *4.4. CANCEL TO TRADE COUNT RATIO 30*
5. MARKET QUALITY 30
   - *5.1. LIQUIDITY MEASURES 30*

5.2. SHORT-TERM VOLATILITY MEASURES 31
6. PANEL DATA REGRESSIONS 32
6.1. REGRESSION EQUATION 32
6.2. DESCRIPTION OF THE PANELS 33
6.3. IMPLEMENTATION IN R 33
CHAPTER 4: RESULTS 37
1. SUMMARY STATISTICS 37
1.1. DISTINCT NUMBER OF TRADES, ODD-LOTS AND ORDERS 37
1.2. TRADE, ODD-LOT AND ORDER VOLUMES 40
1.3. TRADE VALUE 42
2. GRAPHS DEPICTING THE RELATIONSHIPS BETWEEN KEY VARIABLES 44
2.1. AT PROXIES AND MARKET CAPITALISATION 44
2.2. LIQUIDITY, SHORT-TERM VOLATILITY AND MARKET CAPITALISATION 45
2.3. SIX STOCKS IN DIFFERENT MARKET CAPITALISATION CATEGORIES 47
3. PRELIMINARY RESULTS 51
3.1. STATIONARITY TESTS 51
3.2. THE LAGRANGE MULTIPLIER TEST 51
3.3. THE F TEST FOR INDIVIDUAL AND/OR TIME EFFECTS 51
3.4. THE HAUSMAN TEST 51
4. ALGORITHMIC TRADING ON MARKET LIQUIDITY 52
4.1. AVERAGE TRADE SIZE AND MARKET LIQUIDITY 53
4.2. ODD-LOT VOLUME RATIO AND MARKET LIQUIDITY 54
4.3. TRADE-TO-ORDER VOLUME RATIO AND MARKET LIQUIDITY 54
5. ALGORITHMIC TRADING AND SHORT-TERM VOLATILITY 55
5.1. AVERAGE TRADE SIZE AND SHORT-TERM VOLATILITY 56
5.2. ODD-LOT VOLUME RATIO AND SHORT-TERM VOLATILITY 56
5.3. TRADE-TO-ORDER VOLUME RATIO AND SHORT-TERM VOLATILITY 57
CHAPTER 5: CONCLUSION 58
APPENDIX 67

LIST OF TABLES

Table 1: JSE Colocation Trading Statistics - HFT Volumes January – June 2019. 5
Table 2: JSE Colocation Trading Statistics - HFT trading values January – June 2019. 6
Table 3: Aggregate values for every stock. 27
Table 4: Algorithmic trading proxies. 28
Table 5: Description of the terms in the panel data regression. 32
Table 6: Characteristics of the three panels. 33
Table 7: The regression of the quoted bid-ask spreads on the AT proxies. 52
Table 8: The regression of the effective bid-ask spreads in the AT proxies.. 53
Table 9: The regression of the standard deviations of arithmetic returns on the AT proxies. 55
Table 10: The regression of the standard deviations of logarithm returns on the AT proxies. 56
Table 11: Abbreviations used in this document. 67
Table 12: Automated strategies and potential impact on market quality. 68
Table 13: List of conditional codes. 69
Table 14: List of Bloomberg fields. 69
Table 15: Measures of market quality. 70
Table 16: Related literature. 71
Table 17: Panel A - summary statistics across all stocks per day. 72
Table 18: Panel B - summary statistics across all stocks per day. 73
Table 19: Panel C - summary statistics across all stocks per day. 74
Table 20: Unit root tests. 75
Table 21: Panel A - Lagrange Multiplier Test - (Honda). 76
Table 22: Panel B - Lagrange Multiplier Test - (Honda). 77
Table 23: Panel C - Lagrange Multiplier Test - (Honda). 78
Table 24: Panel A - F test for individual effects. 79
Table 25: Panel B - F test for individual effects. 80
Table 26: Panel C - F test for individual effects. 81
Table 27: Panel A - Hausman Test. 82
Table 28: Panel B - Hausman Test. 83
Table 29: Panel C - Hausman Test. 84
Table 30: Equation: quoted bid-ask spread ~ average trade size + market capitalisation. 85

Table 31: Equation: quoted bid-ask spread ~ odd-lot volume ratio + market capitalisation.....86
Table 32: Equation: quoted bid-ask spread ~ trade-to-order vol. ratio + market capitalisation....88
Table 33: Equation: effective bid-ask spread ~ average trade size + market capitalisation.....89
Table 34: Equation: effective bid-ask spread ~ odd-lot volume ratio + market capitalisation. 91
Table 35: Equation: effective bid-ask spread ~ trade-to-order vol. ratio + market capitalisation. 92
Table 36: Equation: std. dev. (arithmetic) ~ average trade size + market capitalisation.....94
Table 37: Equation: std. dev. (arithmetic) ~ odd-lot volume ratio + market capitalisation. 95
Table 38: Equation: std. dev. (arithmetic) ~ trade-to-order vol. ratio + market capitalisation.....97
Table 39: Equation: std. dev. (logarithm) ~ average trade size + market capitalisation. 98
Table 40: Equation: std. dev. (logarithm) ~ odd-lot volume ratio + market capitalisation. 100
Table 41: Equation: std. dev. (logarithm) ~ trade-to-order vol. ratio + market capitalisation. ...102

LIST OF FIGURES

Figure 1: Graphical representation of the first differences model.....34
Figure 2: Number of trades reported by the JSE and in the sample. 38
Figure 3: Number of orders, trades and odd-lots in the sample. 39
Figure 4: Number of orders, trades and odd-lots for three market capitalisation categories. 40
Figure 5: Trade volumes (in billions of shares) reported by the JSE and in the sample. 41
Figure 6: Trade, odd-lot and order volumes for three market capitalisation categories.....42
Figure 7: Trade value (in billions of Rands) reported by the JSE and in the sample. 43
Figure 8: Trade value (in billions of Rands) for three market capitalisation categories. 43
Figure 9: Average trade size for three market capitalisation categories.....44
Figure 10: Odd-lot and trade-to-order volume ratios for three market capitalisation categories. .45
Figure 11: Liquidity measures for three market capitalisation categories (in bps). 46
Figure 12: Volatility measures for three market capitalisation categories (in bps).....46
Figure 13: Average trade size for six stocks in the market capitalisation categories.....47
Figure 14: Odd-lot volume ratios for six stocks in the market capitalisation categories. 48
Figure 15: Trade-to-order volume ratios for six stocks in the market capitalisation categories. ..49
Figure 16: Quoted spreads for six stocks in the market capitalisation categories (in bps).....50
Figure 17: Standard deviations for six stocks in the market capitalisation categories (in bps).....50

CHAPTER 1: INTRODUCTION

1. BACKGROUND

High-frequency trading (HFT) uses computer algorithms in a low trading cost and a high data frequency environment to capitalise on small profit margins within and across markets as soon as they arise. The attraction of the returns generated by millions of these transactions has resulted in a competitive environment that accounted for more than 50% of the U.S. equities daily volume in 2017 (Meyer, Bullock & Rennison, 2018). From January to September 2019, the Johannesburg Stock Exchange (JSE) reported that high-frequency traders (HFTs) accounted for roughly 49% of the total trading volume on the exchange (JSE Colocation Activity Report). However, despite the ubiquity of these HFTs, there is a lack of consensus as to their impact on markets. This research contributes to the growing empirical studies in the field by measuring the impact of HFT on two aspects of JSE market quality namely, liquidity and short-term volatility. The notion of market quality refers to the degree to which the stock market allocates capital to its most productive use.

HFTs form part of the broader group of algorithmic traders (ATs) (Hendershott, Jones & Menkveld, 2011). In this study, the term "AT" is preferred for a technical reason. Studies with access to private datasets are able to determine whether orders are submitted by HFTs or humans, and use the term "HFT". In contrast, studies reliant on public datasets are unable to make this distinction regarding trader identity and employ the term "AT". Since this study falls into the second category, the term "AT" is used in the chapters following the literature review and in the study's title.

HFT is a novel and complex technology. It is a niche field of finance with many intricacies. One must have an understanding of the fundamental characteristics of HFT to be able to form an assured opinion on whether its consequences are beneficial or hazardous to other market participants. HFTs rely on powerful algorithms to collect and analyse market data and thereafter generate, route and execute orders all in the space of a few milliseconds (Hasbrouck & Saar, 2012). Human traders fail to compete against HFTs at this speed. As a comparison, the blink of a human eye lasts one third of a second (Chakrabarty, Moulton & Wang, 2018). To gain an advantage over competing

traders, HFTs implement low-latency strategies. The term "latency" is commonly defined as the amount of time taken for data to travel across a network from the markets' servers to the traders (Biais & Foucault, 2014). To reduce latency, HFTs rent specialised premises, known as colocation centres, in proximity to the market servers on which matching engines operate. These colocation centres store the hardware on which HFTs function (Zhang, 2010). In South Africa, the JSE launched its first colocation centre in May 2014 (JSE Colocation Statistics, 2019).

HFTs submit thousands of orders daily and earn a small profit on every trade (Zhang & Powell, 2011). They are characterized by very short holding periods which vary from milliseconds to a couple of hours and generally do not carry overnight positions (Zhang, 2010). The algorithms determine all aspects of the orders including the price, size, timing and venue to which the orders are routed (Zhang, 2010). HFT firms invest in smart order routing systems, which instantly identify the best quotes across competing trading venues (Foucault & Menkveld, 2008). This eliminates the risk of human error and emotion affecting the trading process (Chakrabarty, Moulton & Wang, 2018). Most HFT firms trade using their own capital and are therefore classified as proprietary traders, however large investment banks such as Goldman Sachs and Morgan Stanley have their own HFT desks. A smaller portion of HFT firms operate as hedge funds, with an example being Citadel (Biais & Foucault, 2014). Finally, HFTs incur significant expenses such as hardware, employment of qualified developers, subscription to real-time data feeds and colocation fees (Biais, Foucault & Moinas, 2011). Although trading algorithms are costly to develop, Hendershott et al. (2011) showed that the marginal trading costs of HFT firms tend to be low. In addition, HFTs also benefit from large trade volume discounts offered by stock exchanges (Brogaard, 2010).

1.1. FACTORS LEADING TO THE ADVENT OF ALGORITHMIC TRADING

Since the 1990s, intense debate has surrounded the effect of HFT on financial markets. It is apt to begin with the history of HFT in order to substantively engage with the diverse strategies employed. Advancements in technology have played a pivotal role in accessibility to HFT. First, innovative tech companies started building algorithms as a result of exchanges developing electronic trading systems. In 1971, the NASDAQ began trading and became the world's first electronic stock market (McGowan, 2010). Five years later, the NYSE launched the Designated

Order Turnaround (DOT) system, which made it possible for member firms to send their orders electronically to the traders (McGowan, 2010). In the mid-1990s, numerous stock exchanges globally, including the JSE, abandoned the Open Outcry system in favour of electronic trading (Hendershott, Jones & Menkveld, 2011). This system allows HFTs to electronically send their orders to the market within a few milliseconds (Biais & Foucault, 2014).

Second, in 2001, the SEC instructed U.S. exchanges to price stocks in decimals as opposed to fractions (Kim, 2010). Prior to this date, a price quote could move by no less than one sixteenth of a dollar (Chung & Van Ness, 2004). With prices being converted into decimals, the minimum price movement became one cent of a dollar, which lowered the risk of trading over short horizons (Brogaard, 2010). Consequently, this opened the door for the possibility of traders to find even smaller profit margins on which to capitalize.

Third, two years after decimalization, the NYSE started making use of the Auto-quote software to automatically disseminate the best quotes in listed stocks. Prior to this, quotes were manually updated by specialists (Hendershott, Jones & Menkveld, 2011). The new system allowed HFTs to get faster access to information about stock prices and therefore led to a greater use of algorithms. Finally, the world's main stock exchanges began to offer colocation services in the late 2000s, allowing HFT firms to place their servers and trading equipment within the vicinity of market servers (Biais & Foucault, 2014). As a result, HFTs are able to view prices a split second before other traders, who rely on consolidated data (Biais & Foucault, 2014).

In May 2014, the JSE launched its on-site colocation facility, allowing its clients to rent space in proximity to the market's server. HFTs that place their servers at the colocation centre benefit from a latency of about 100 microseconds as opposed to 2400 microseconds if they were located in the general Sandton area (JSE Colocation Statistics, 2019). These enhancements have attracted HFT firms from developed countries. The key market players in the HFT industry are investment banks such as Goldman Sachs and Credit Suisse, hedge funds such as Citadel and Global Electronic Trading Co. (GETCO), as well as proprietary firms like Hudson River Trading (Lewis, 2014).

Along with technological development, the turn of the millennium saw the financial field shaped by the implementation of two major regulations aimed at fostering competition which resulted in an increase in the use of HFT. In 2000 and 2005, the Securities Exchange Commission (SEC) passed Regulation Alternative Trading System (Reg ATS) and Regulation National Market System (Reg NMS), respectively. First, alternative trading platforms are non-exchange trading platforms approved by the SEC (Regulation 300(a) of Reg ATS). Under Reg ATS, they are exempt from registration as a securities exchange and are only required to register as a broker-dealer to compete with public exchanges. This significantly reduced the barriers to entry for ATS and caused the fragmentation of the U.S. stock market. In 2015, eleven stock exchanges and more than 50 ATS competed for order flow in the U.S. (O'Hara, 2015). The surge in the number of trading platforms created advantageous market conditions for HFTs, who capitalised on the price discrepancies between the various platforms (Biais & Foucault, 2014). In line with the global trend of market fragmentation, a series of new stock exchanges have entered the South African market since 2016, namely, ZARX, 4AX, A2X and Equity Express Securities Exchange (2018 Financial Markets Review). This encouraged the growth of HFT in South Africa.

Second, Reg NMS introduced the Securities Information Processor (SIP) to improve public access to market data. This technology collects trade and quote data across registered U.S. exchanges and disseminates this information to all trading venues and ATS. In addition, Rule 611 of Reg NMS, the Order Protection Rule, requires brokers to trade at the best price for their clients' orders. This price is the National Best Bid and Offer (NBBO) which is determined by the SIP. In the event of a broker being unable to execute an order at the NBBO, the regulation stipulates that the order must be cancelled or routed to a different trading venue. Studies have revealed that HFTs exploit a major loophole in Reg NMS. By positioning their servers in proximity to the exchanges' matching engine, they are able to calculate NBBOs at a much faster rate than the SIP and therefore get a more accurate view of the market than slower traders. Finally, the European Parliament and Council promulgated MiFID II (Markets in Financial Instruments Directive) in 2007, which had similar consequences in the EU. To date, there is no South African legislative equivalent. In 2020, the Financial Sector Conduct Authority (FSCA) submitted draft regulations concerning best executions. The draft regulations are expected to be finalised at the beginning of 2021.

1.2. HIGH FREQUENCY TRADING STATISTICS

HFT has become an integral part of the financial markets. Although HFT data is difficult to retrieve, the various sources that published statistics demonstrated the importance of HFT activity. In 2017, HFTs accounted for about 50% of the U.S. equities daily volume (Meyer, Bullock & Rennison, 2018) compared to 61% in 2009 (Sussman, Tabb & Iati, 2009). The explanation for this sudden decrease may be that markets have been less volatile since the global financial crisis, providing fewer profitable opportunities for HFTs. Indeed, several studies have identified a strong positive relationship between HFT and volatility (Brogaard, 2010); (Brogaard, Hendershott & Riordan, 2014). Nonetheless, South African markets seem to be experiencing the opposite trend. Colocation activity represented 49.28% of the total trading volume from January to September 2019 compared to 32.70% from January to September 2015 (JSE Colocation Activity Report). Table 1 reports the HFT volumes as a percentage of the total equity market volumes from January to June 2019 (JSE Colocation Statistics, 2019).

Table 1: JSE Colocation Trading Statistics - HFT Volumes January – June 2019.

January	February	March	April	May	June
35.55%	33.85%	37.56%	34.98%	36.71%	35.05%

Regarding absolute trading volume in the U.S., HFTs traded about 1.6 billion shares per day in 2012 compared to 3.25 billion in 2009 (Chakrabarty, Moulton & Wang, 2018). In addition, HFTs actively trade in countries outside of the U.S. For example, HFTs accounted for 77% of all transactions in the UK in 2011 (Sukumar, 2011). In 2012, they were responsible for roughly 45% of the equity trading volume in the EU, 40% in Japan and 12% in other Asian countries (Goldstein, Kumar & Graves, 2014). Finally, in South Africa, Table 2 presents the HFT trading values as a percentage of the total equity market values from January to June 2019 (JSE Colocation Statistics, 2019).

Table 2: JSE Colocation Trading Statistics - HFT trading values January – June 2019.

January	February	March	April	May	June
39.89%	39.38%	37.11%	37.99%	40.20%	39.11%

In terms of profits, HFTs generated around $1.3 billion in 2014 from trading U.S. stocks compared to $7.2 billion in 2009 (Tabb, 2014). Rosenblatt Securities estimates are slightly lower, with the firm claiming that HFTs made $1.25 billion in 2012 but over $4.9 billion in 2009 (Popper 2012, New York Times).

1.3. THE DEBATE ABOUT THE EFFECTS OF HFT ON MARKET QUALITY

High-frequency trading is at the centre of a rigorous debate. HFT supporters claim that algorithms improve market liquidity, reduce trading costs for other market players and accelerate the price discovery process. Detractors argue that HFTs amplify short-term volatility under stressed market conditions as demonstrated by the May 6, 2010 U.S. Flash Crash. On that day, the Dow Jones Industrial Average (DJIA) fell by approximately 900 points, however recovered the losses to end the day at a similar level to its opening value (Biais & Foucault, 2014).

The effects of HFT on market quality mainly depend on the type of strategy that is implemented (Hagströmer & Nordén, 2013). These strategies are summarized in Table 12 in the Appendix. HFTs generally follow one of two strategies: market-making or opportunistic (Biais & Foucault, 2014). HFT market-makers send limit orders that specify the maximum (minimum) price at which they are willing to buy (sell) an asset. In the event that the limit order is filled, the HFT makes a profit which corresponds to the bid-ask spread on the trade (Menkveld, 2013). These traders are likely to improve liquidity and price discovery as they constantly monitor markets and update their quotes to prevent their orders from becoming stale (Biais & Foucault, 2014). However, they could also pose a threat to market stability as they are not always required to provide liquidity and may disappear when needed (Zhang, 2010).

Opportunistic strategies include arbitrage, momentum and manipulative strategies (Biais & Foucault, 2014). First, arbitrage strategies entail exploiting short-lived opportunities by purchasing

and selling an asset with different prices in two or more markets at the same time. For instance, HFTs actively trade in the foreign exchange market where triangular arbitrage opportunities (pricing discrepancies among three currency pairs) frequently arise (Biais & Foucault, 2014). HFTs seeking these opportunities are likely to speed up the price discovery process.
Second, momentum strategies are based on short-term, predictable market trends (Biais & Foucault, 2014). Specific types of algorithms thoroughly examine limit order books and generate orders depending on the short-term patterns identified (Brogaard, Hendershott & Riordan, 2014). These traders are likely to amplify short-term volatility.

Third, a subset of HFTs implement manipulative strategies such as electronic front-running, which involves sending multiple small orders inside the bid-ask spread, a tactic defined as "pinging", to detect large hidden orders and trade ahead of other investors (Brogaard, 2010). Biais and Woolley (2011) also describe the practice of quote stuffing, whereby HFTs submit an extremely large amount of orders to the market and thereafter immediately cancel them to disconcert other traders.

Finally, spoofing and layering are deceptive trading strategies designed to artificially increase or reduce stock prices to mislead other market participants (McNamara, 2016). More specifically, spoofing is the act of placing a bid or offer on a stock with the intent to cancel these orders before execution. This creates the false impression of increased demand or supply for the stock (Biais & Woolley, 2011). Layering is a special form of spoofing, which consists in sending a sequence of limit orders to the market with the intent to cancel before execution. These orders are generally small in size, mindful of the risk that some of these orders may be executed before the HFT cancels them (2018 Financial Markets Review).

Most trading algorithms that do not fall within the HFT group are designed to extract subjective information from public news websites or social media platforms to evaluate the prevailing market sentiment (Leinweber, 2009). Unlike HFTs, these algorithms do not operate at extreme speeds due to the complexity involved in interpreting the information. ATs which are not classified as HFTs are therefore likely to improve the process of price discovery but may amplify short-term volatility.

In South Africa, HFT has been left largely unregulated due to the complexity of the strategies followed. To compensate for the lack of regulation, the JSE has created its own set of restrictions. The exchange has built circuit breakers that are designed to cancel or limit the amount of orders executed when stock prices drop too quickly (JSE Colocation Statistics, 2019). In addition, the JSE limits the amount of orders that can be submitted to 300 orders per second, per broker (King & Barnes, 2015).

2. PROBLEM STATEMENT

Academics are divided about the impact of ATs on market quality. While many argue that ATs should be encouraged as they supply ample liquidity and speed up the price discovery process, others are concerned that these traders amplify short-term volatility and impose adverse selection costs on other investors. From January to September 2019, colocation activity accounted for 36.64% of the total order volume and 49.28% of the total trading volume (JSE Colocation Activity Report). However, despite the substantial amount of orders initiated by ATs, these traders are left largely unregulated and could potentially harm retail investors in South Africa (2018 Financial Markets Review). Given the lack of consensus on the role of ATs and the absence of regulation surrounding their activity in South Africa, this study seeks to provide empirical evidence on the effects of ATs on market quality on the JSE.

3. RESEARCH OBJECTIVES

The central aim of this study is to determine the impact of AT on two measures of market quality which are liquidity and short-term volatility. The study has two distinct objectives to achieve this goal. First, a set of AT proxies are developed to identify AT activity from intra-day tick data. Second, the proxies are regressed on two measures of liquidity, namely, quoted and effective bid-ask spreads and on a measure of short-term volatility which is the intra-day correlation in stock returns. This technique will help describe the relationship between AT and the two measures of market quality.

The null hypotheses tested in this study are as follows:

1. AT activity is unrelated to stock market liquidity on the JSE.
2. AT activity is unrelated to short-term volatility on the JSE.

4. IMPORTANCE AND BENEFITS OF THE PROPOSED STUDY

This study contributes to the growing literature on the role of ATs in today's stock markets. Studies on this topic have been limited outside of the U.S. HFTs accounted for about 49.28% of the total trading volume on the exchange from January to September 2019 (JSE Colocation Activity Report). Yet, there has been no empirical analysis to date on the effects of ATs on the JSE. This study aims to fill this research gap by identifying the effects of ATs on the quality of the South African stock market. The findings highlight some of the reasons that would justify regulatory intervention by the Financial Sector Conduct Authority (FSCA) and provide valuable insight for market players who compete against ATs.

5. STUDY LIMITATIONS

This study is subject to two limitations which could be addressed in future research. First, HFTs place a large number of orders in multiple markets at the same time and cover various asset classes (Biais & Foucault, 2014). This study however, is restricted to all the stocks listed on the JSE. This may affect the validity of the empirical results. For example, Menkveld (2013) studied HFT activity on two different exchanges, Euronext and Chi-X. When studying each market separately, the study did not find that the HFTs' aggregate positions mean-reverted although they did across markets. Second, although HFTs represent a subset of ATs (Hendershott, Jones & Menkveld, 2011), the study regards ATs as a homogeneous group of traders. The AT proxies capture the activity of ATs operating at any low and high frequencies.

This study is structured as follows: Chapter 2 presents an overview of academic studies on the effect of ATs on market liquidity, price discovery and short-term volatility in the U.S., Europe and South Africa. Chapter 3 describes the sample selection and methodology followed. Chapter 4 presents the results and provides a detailed interpretation of same. Chapter 5 forms the conclusion.

CHAPTER 2: LITERATURE REVIEW

1. INTRODUCTION OF THE LITERATURE REVIEW

The primary purpose of the stock market is to allocate capital to its most productive use (Fama, 1970). The degree to which it does so is also referred to as market quality, which is proxied by three measures, namely, liquidity, price discovery and efficiency, as well as short-term volatility. This chapter reviews the empirical evidence on the impact of AT on these three measures. The literature review is organised into five sub-sections. First, it surveys prior studies that explore the role of AT on market liquidity. Liquidity is commonly defined as the ease and amount of time needed in order to convert a security into cash (Lippman & McCall, 1986). Second, the review summarizes studies on the impact of AT on the price discovery process and price efficiency. Price discovery is defined as "the speed at which new information is impounded into asset prices" (Biais & Foucault, 2014), p. 6). Price efficiency refers to the degree to which prices reflect public information (Fama, 1970). Third, the review discusses various studies which examine the effects of AT on the adverse selection costs of slower traders. HFTs gain direct access to market information through colocation and therefore benefit from an informational advantage over other traders. This is commonly known as the adverse selection problem (Biais & Foucault, 2014). Fourth, the review presents empirical evidence on AT as a source of market instability. Finally, the review evaluates prior studies which identify the specific types of stocks that are most actively traded by ATs.

2. ALGORITHMIC TRADING AND MARKET LIQUIDITY

In recent years, high-speed trading algorithms have become cheaper, more efficient and reliable. As a result, HFTs have replaced traditional human market-makers. Market-makers post buy and sell orders in securities to provide liquidity on the exchange. They place limit orders in the order book by indicating a maximum (minimum) price they are willing to buy (sell) the asset without any guarantee that these orders will be executed. In the case where limit orders are not filled, they are generally cancelled and resubmitted to ensure that prices reflect market conditions. Most exchanges offer rebates to traders who submit limit orders (Biais & Foucault, 2014). In contrast,

market-takers submit buy (sell) orders that automatically match limit sell (buy) orders. These types of orders remove liquidity from the exchange and have a much higher probability of resulting in an immediate trade (Biais & Foucault, 2014).

Hendershott and Riordan (2009) studied the effect of AT on the liquidity of the 30 largest and most liquid stocks on the Deutsche Boerse. In their study, the authors could directly distinguish the orders initiated by ATs from those initiated by human traders over the period 1 January 2008 to 18 January 2008. The authors regressed two variables, namely an AT trade-initiation variable and a human trade-initiation variable on quoted half-spreads and effective spreads. The AT trade-initiation variable took the value of 1 if the order was initiated by an AT and zero otherwise. The opposite applied if the order was initiated by a human trader. The results indicated that ATs took liquidity away from the market when they initiated market orders but supplied liquidity through limit orders. The demand for liquidity and the supply of liquidity by ATs represented 52% and 50% of the total trading volume respectively. ATs were found to provide liquidity when it was scarce and consumed it when it was abundant.

Hendershott, Jones and Menkveld (2011) corroborates these findings. The authors examined the relationship between AT and liquidity for stocks listed on the NYSE from February 2001 to December 2005. For every month in the sample period, the authors regressed their proxy for AT, which is the number of electronic messages per $100 of trading volume on various measures of liquidity such as quoted and effective bid-ask spreads. The results of their analysis suggested that the increase in AT resulted in narrower spreads for the stocks grouped into the largest-cap quintile. For every unit increase in AT, the spreads narrowed by 0.53 basis points. For small-cap stocks, the study did not find any statistically significant relationship. The study also provided evidence that the average trade size declined over the sample period due to the significant growth in AT activity. These results were robust to controls for share turnover, volatility, inverse of share price and log market cap.

Boehmer, Fong and Wu (2012) also confirmed these results. The authors collected tick data from the Reuters database to analyse the effect of AT on the liquidity of non-U.S. stocks listed on 39 exchanges from 2001 to 2009. They regressed their proxy for AT on three measures of liquidity,

namely, quoted spreads, effective spreads and the Amihud ratio, which is "the absolute value of daily return divided by the contemporaneous daily dollar trading volume" (Boehmer, Fong & Wu , 2012, p. 15). The study showed that the effect of AT on liquidity differed greatly across different categories of stocks. For example, for large-cap and high-priced stocks with low volatility, AT significantly improved liquidity. For small-cap stocks, AT led to higher transaction costs and lowered liquidity levels. For low-priced stocks and stocks that were highly volatile, the authors could not find any significant relationship between the two variables. Finally, on the days when market-making activities were costlier to implement, AT provided less liquidity to the markets.

Hasbrouck and Saar (2012) defined a new measure of low-latency activity to test the impact of HFT on liquidity for NASDAQ stocks. The authors created "strategic runs" which are sequences of order book messages sent in less than one second. The types of messages included submissions, cancellations and executions. Hasbrouck and Saar found that low-latency activity led to tighter quoted and effective spreads as well as a higher level of depth in the limit order book.

Hagströmer and Norden (2013) compared the effects of HFTs who engage in market-making activities to those who implement opportunistic strategies on the quality of the stock market in Sweden. Their sample included all limit order messages about stocks part of the OMXS 30 index during the month of August 2011 and February 2012. The traders who initiated the orders were explicitly identified. The authors relied on the assumption that HFT market-makers posted limit orders at the inside quotes more frequently than other HFTs to capture their activity. The results indicated that within the HFT group, the market-makers were responsible for 71.5% and 62.8% of the total trading volume in August 2011 and February 2012, respectively. Furthermore, they submitted 80.5% and 86.4% of the total number of limit orders over these two months respectively.

Kirilenko, Kyle, Samadi and Tuzun (2017) followed a different approach but reached a similar conclusion. The study examined AT activity on the day of the Flash Crash on the 6th May 2010. In their study, the authors compared the aggressiveness ratios of HFTs with fundamental traders over the period 3 May 2010 to 6 May 2010. They explained that an aggressive trade results from a market order whereas a passive trade rests in the limit order book for some time. The study defined the aggressiveness ratio as the "ratio of aggressive trade executions to total trade

executions" (Kirilenko, Kyle, Samadi & Tuzun, 2017, p. 15). The results indicated that HFTs had a lower aggressiveness ratio (45.68%) than fundamental buyers (64.09%) and sellers (61.13%). These findings suggested that HFTs provided liquidity through limit orders whereas fundamental traders took liquidity away from the market via market orders.

Finally, in contrast to the studies described above, Brogaard (2010) found that HFTs provided much less order book depth than other traders from 2008 to 2009.

In summary, there is a general consensus among academics that HFTs reduce the cost of liquidity provision to the benefit of all market participants. The relationship between HFT activity and liquidity differs depending on certain characteristics specific to stocks. In fact, the relationship is stronger for large-cap and high-priced stocks with low volatility.

3. ALGORITHMIC TRADING AND PRICE DISCOVERY

In efficient markets, stock prices fully reflect all available information (Fama, 1970). AT supporters claim that trading algorithms allow buyers and sellers to trade at efficient prices. They argue that HFTs identify small mispricings and trade away arbitrage opportunities. This theory is consistent with the fact that HFT has become less profitable in the last few years. Arbitrage opportunities may have become scarce as a result of the intense competition among HFT firms. The concept of price efficiency must not be confused with that of price discovery, which refers to the speed at which prices are updated to reflect new information (Biais & Foucault, 2014).

Hendershott and Riordan (2009) measured the contribution of AT on price discovery for 30 stocks in the DAX index on the Deutsche Boerse. The study followed the vector auto-regression framework developed by Hasbrouck (1991a, 1991b). Their results indicated that AT quotes contributed to 51% of the stocks' price discovery as opposed to 39% for human quotes. The remaining 10% resulted from AT and human quotes which were initiated simultaneously. The study further showed that ATs had an average permanent price impact of 0.53 basis points compared to 0.44 basis points for human traders. These findings clearly suggested that the quotes initiated by ATs were more efficient than those initiated by humans. Surprisingly, when comparing

the impulse response functions for ATs and human traders, AT quotes were found to contain more private information than human trades.

Similar findings were provided by Brogaard (2010) who found a strong positive relationship between HFT and three measures of price discovery defined by Hasbrouck (1991a, 1991b, 1995), namely, permanent price impact, aggregate information variance decomposition and information share. Hendershott et al. (2011) confirmed the above findings by showing that an increase in AT resulted in more informative quotes on the NYSE from February 2001 to December 2005.

O'Hara, Yao and Ye (2011) investigated whether odd-lot trades carried any information content. An odd-lot is commonly defined as a trade for less than 100 shares of stock. Prior to the 1990s, the vast majority of odd-lot traders were retail investors who could not afford to buy more than 100 shares of stock at a time. After the mid-1990s, however, odd-lot traders were mainly ATs. Their strategy consisted of dividing large orders into smaller pieces to hide their trading intentions from other market participants. Indeed, in the U.S., odd-lot trades were not report to the consolidated tape before December 2013. The sample analysed by O'Hara, Yao and Ye (2011) consisted of a non-public TAQ dataset for 120 stocks listed on the NASDAQ from January 2008 to December 2009. The results indicated that odd-lot trades had a higher probability of occurring when trades were initiated by a HFT as opposed to a human trader. Over the sample period, the analysis showed that high-priced stocks such as Google, Apple and Amazon were more frequently traded in odd-lots. The authors also studied the weighted price contribution (WPC), which refers to the proportion of the stock's cumulative price change that is attributable to different trade size categories. The study found that the odd-lot trades and round-lot trades (trades of exactly 100 shares) combined contributed 80% of price discovery on the NASDAQ. Odd-lots contributed 30% and round-lots 50%. In contrast, the trades greater than 500 shares contributed 4.1% of price discovery.

Biais, Foucault and Moinas (2011) later confirmed that HFT activity has a positive effect on the informational content of orders and trades. Similarly, Boehmer, Fong and Wu (2012) found that AT generally enhanced informational efficiency across their sample of non-U.S. stocks in 39 markets from 2001 to 2009.

Brogaard, Hendershott and Riordan (2014) tested whether HFTs improved the price discovery process for 120 stocks listed on the NASDAQ and NYSE. Their dataset included all market and limit orders initiated by HFTs over the sample period 2008 to 2009. Their analysis entailed applying a state space model to the trading data to divide price movements into a permanent efficient component and a noise component. The study proved that HFTs played a positive role on price discovery. They traded "in the same direction as future efficient price moves" but "in the opposite direction to transitory pricing errors" on average (Brogaard, Hendershott & Riordan, 2014, p. 2). HFTs achieved this result when they initiated market orders, which take away liquidity from the market. Furthermore, HFTs also improved the price discovery process upon the arrival of new information related to macroeconomic news announcements. They anticipated price movements by analysing the limit order book. These findings are consistent with those of Jiang, Lo and Valente (2014).

Finally, Chaboud, Chiquoine, Hjalmarsson and Vega (2014) examined the effects of AT on the global foreign exchange market. The Electronic Broking Services (EBS) and Reuters provided the authors with trading data covering the euro-dollar, dollar-yen and euro-yen markets from September 2003 to December 2007. The authors were able to explicitly distinguish the identity of the counterparties to each trade within the dataset. They made use of reduced-form and structural VAR systems to prove that ATs reduced the number of triangular arbitrage opportunities that existed among the three currency pairs. AT enhanced the level of price efficiency by acting on quotes posted by human traders and taking liquidity away from the FX market.

Several studies have found evidence that human traders have limited attention and have a tendency to underreact to corporate earnings announcements (Peng, 2005). During times of low investor attention, information is impounded into stock prices at a slower pace than normal. However, although human traders are subject to low-attention constraints, this is not the case for HFTs, who now account for a significant portion of the total trading volume in equities globally (Zhou & Olivari, 2013).

Chakrabarty, Moulton and Wang (2018) examined the role of HFTs in the price discovery process during times of low-human attention. The NASDAQ provided the authors with a dataset which specified all trades initiated by HFTs versus non-HFTs for 120 stocks from 2008 to 2009. The study also made use of 958 corporate earnings announcements from the IBES and Compustat databases. The authors developed six proxies to identify the low-attention earnings announcements in their sample. The types of events classified as low-attention announcements were those made on Fridays and on days with multiple announcements, announcements with slow analyst forecast adjustments, high news distraction, low Electronic Data Gathering Analysis and Retrieval (EDGAR) download volume and Google search volume.

The study confirmed that HFTs facilitated price efficiency in the short-term by accelerating the speed at which prices reflected earnings-related information. This effect did not apply to all of the low-attention events but only to those when low investor attention lead to under-reaction. Furthermore, HFT activity led to smaller long-term post-earnings announcement drifts (PEAD) in prices after the announcements were made. Finally, HFTs improved price efficiency by taking the active sides of the trades and demanding liquidity during times when human traders were less focused on earnings announcements. The results proved to be robust after controlling for firm size and applying various time windows. The study even reported that the effect of HFTs on price efficiency was significantly stronger for smaller stocks.

In contrast, a couple of studies found evidence that ATs have a negative impact on price efficiency and discovery. For example, Zhang (2010) argued that since HFTs do not carry overnight positions, they are indifferent to the financial sustainability of the firms whose stocks they trade in. HFTs make trading decisions based on short-term statistical correlations in stock returns. Zhang examined a sample of U.S. stocks from 1985 to 2009 and found that HFTs tend to drive stock prices away from their fundamental values. The study showed that following the publication of information related to a firm's fundamentals, HFTs caused the market to overreact in the short-term. In the months following the event however, stock prices were reversed. It is worth noting that this study has several limitations. First, it relied on the simplifying assumption that the trading behaviour of individual investors followed that of institutional investors over time, which may not

be the case. Second, Zhang measured HFT at the quarterly level, which seems inadequate since HFTs are day traders and hold stocks over very short periods of time.

Biais and Foucault (2014) made the compelling argument that faster price discovery is not necessarily more desirable. Following an announcement regarding a certain stock, the price updates to reflect the new piece of information and the speed at which it does so whether it is a millisecond or half a minute is inconsequential.

Weller (2018) provided evidence for Biais and Foucault's argument. The author tested whether ATs improved price efficiency at the expense of information acquisition. In the study, the dataset originates from the SEC's Market Information Data Analytics System. It comprises all public messages that were sent to the major U.S. stock exchanges from January 2012 to March 2015. The orders were aggregated by stock, date and exchange. The author estimated several panel data regressions of information acquisition on a set of AT proxies. The dependent variable is the price jump ratio, which captures the extent of information acquisition around quarterly earnings announcements. This ratio is the post-announcement price response divided by the total price movement. A lower ratio suggested that prior to the announcements, ATs were more informed and followed order anticipation strategies. The study found that although ATs improved price efficiency over a few seconds, they strongly discouraged the acquisition of new information by human traders up to a month prior to the earnings announcements. The analysis showed that a one standard deviation increase in AT led to a 8-29% reduction in the extent of information acquisition before the announcements. This inverse relationship was found to be stronger for ATs who followed order anticipation strategies and took liquidity away from the market. The results were robust to various controls for market capitalisation, volatility, bid-ask spreads, analyst coverage, as well as month and stock fixed effects.

To conclude, most of the studies reviewed in this section agree that HFT speeds up the price discovery process and improves price efficiency except from Zhang (2010), who showed that HFTs drive prices away from their fundamentals. However, although HFTs play a beneficial role on the price discovery process in the short-term, this is not obviously the case in the long-term as demonstrated by Biais and Foucault (2014) and Weller (2018).

4. ALGORITHMIC TRADING AND ADVERSE SELECTION COSTS

As the level of HFT activity increases, it becomes easier for slower traders to find counterparties. From this perspective, HFTs improve the trading costs of other traders. However, human traders have an information disadvantage and as a result, systematically lose to HFTs (Biais & Foucault, 2014). This adverse selection problem may discourage other traders from participating (Biais, Foucault & Moinas, 2011). In addition, a subset of HFTs use their low-latency advantage to prey on other investors. More specifically, HFTs use pattern recognition software and send out "ping" orders to detect large impending orders in dark pools. Dark pools are private exchanges in which traders submit orders anonymously. These exchanges are not required to display their order book to the public (Banks, 2014). HFTs front-run slower traders by purchasing (short-selling) and immediately relisting stocks that investors intend to buy (sell and buy back) from non-HFTs at a higher (lower) price (Biais & Woolley, 2011).

Numerous studies have exposed the predatory behaviour of various groups of HFTs. Hirschey (2018) investigated whether HFTs generated profits at the expense of other investors. The study made use of the NASDAQ HFT dataset over the year 2009. The analysis showed that HFTs' aggressive trades often preceded those of slower traders, especially when the stocks traded were mid and small caps. Aggressive trades are defined as "trades where a HFT initiates the transaction by submitting a marketable buy or sell order" (Hirschey, 2018, p. 2). To explain this observation, the study tested the anticipatory trading hypothesis which states that HFTs predict other investors' order flows and as such, are able to trade ahead of them. The study showed that it was easier for HFTs to detect large, hidden orders at times when non-HFTs were impatient and less able to hide their trading intentions from the market. In fact, the relationship between the aggressive trades of HFTs and non-HFTs strengthened near the market open, on high volume and high imbalance days as well as for illiquid stocks. The study singled out certain groups of HFTs that implemented these strategies and were better at predicting future returns than the rest of the HFTs in the sample. Many studies confirmed that HFTs impose higher adverse selection costs on slower traders (Biais & Woolley, 2011; Chaboud, Chiquoine, Hjalmarsson & Vega, 2014).

Hoffman (2013) showed that limit orders that are revised regularly are more likely to be filled. As such, HFTs have an unfair advantage in that they are better equipped to modify their limit orders as new information becomes publicly available. Slower traders will always be worse off because their orders have a smaller probability of execution. The above findings have numerous implications for price efficiency. HFT activity leads to an increase in trading costs for other traders. This may reduce their incentives to acquire new information and spend resources analysing firm fundamentals.

A number of studies however, proved that AT activity in fact leads to a reduction in the trading costs of other market participants. Malinova, Park and Riordan (2013) examined the impact of AT on markets outside of the U.S.. Their dataset was provided by the TMX Group and contained trader-level information about 248 stocks in the S&P/ TSX Composite Index from 1 March 2012 to 30 April 2012. In April 2012, the Investment Industry Regulatory Organization of Canada (IIROC) increased the fees paid by investors based on the number of electronic messages sent to exchanges. As a result, traders that carried out "message-intensive strategies" such as HFTs significantly reduced their daily volumes. In fact, the number of messages dropped by 30%.

Malinova et al. (2013) examined the impact of the sudden decline of HFT activity on the trading costs, measured by effective bid-ask spreads and intra-day returns for two distinct groups of traders: institutional and retail investors. On the one hand, institutional investors who submitted market orders after the event, paid and earned higher effective spreads but their intra-day returns increased. On the other hand, retail investors who traded 54% of their volume with market orders, did not experience any significant increase in their trading costs while their intra-day returns declined. In addition, institutional and retail investors who submitted limit orders following the new regulation, were subject to higher adverse selection costs. Finally, although the intra-day returns of institutions remained constant, those of retail investors substantially declined. In essence, the study provided strong evidence that HFT also benefits retail investors.

In sum, the evidence regarding the impact of AT on the adverse selection costs of slower traders is mixed. Several studies found that HFTs who engage in manipulative strategies increase the

trading costs of other investors. However, Malinova et al. (2013) found that the passing of a new law in Canada aimed at reducing HFT activity resulted in lower trading profits for retail investors.

5. ALGORITHMIC TRADING AND SHORT-TERM VOLATILITY

Academics and investors are concerned with the fact that ATs may amplify short-term volatility. A couple of studies provided evidence that this concern was not necessarily valid. Hendershott and Riordan (2009) examined whether ATs were more likely than human traders to increase the level of volatility for stocks included the DAX index. The authors first identified the times when bid-ask spreads were narrower or wider than average for each stock in the sample. During these periods, they measured the amount of time where ATs and human traders were on the inside of the spread. The results indicated that ATs were on the inside more often than their human counterparts (approximately one hour each day). The study therefore proved that human traders contributed more than ATs to the increase in volatility for these stocks.

Brogaard (2010) confirmed these findings. The author observed that from 2008 to 2009, HFTs traded less with one another than predicted, suggesting that HFT strategies were highly correlated with one another compared to non-HFTs. However, HFTs did not exacerbate volatility. On the contrary, they were found to lessen intra-day volatility. Hagströmer and Norden (2013) compared the effects of HFT on the volatility of stocks in the OMXS 30 index during the month of August 2011, when the market was particularly volatile and February 2012, when the market remained stable. The analysis showed that HFTs, regardless of their strategy (market-making or opportunistic), reduced the levels of intra-day price volatility.

Similarly, Hasbrouck and Saar (2012) examined the impact of low-latency activity on short-term volatility during the month of October 2007, a period in which normal market conditions prevailed and the month of June 2008, a period of extreme market stress. The study showed that during both periods, low-latency activity improved short-term volatility. More specifically, in June 2008, HFTs reduced the volatility of small-cap stocks to a greater extent than that of large-cap stocks.

Finally, Chaboud et al. (2014) argued that since trading algorithms have pre-programmed strategies, these could be highly correlated with one another. In other words, many trading algorithms could react in the same way and amount of time to a market event. To test this possibility, the authors estimated the amount of trades where both counterparties were ATs. The results indicated that ATs traded less with each other than the random matching model predicted. Thus, the authors concluded that the strategies implemented by ATs were highly correlated compared to those followed by human traders. Despite this high level of correlation among AT strategies, the trading algorithms did not cause any excess volatility over the sample period. In fact, ATs actually reduced the degree of autocorrelation of high frequency returns in the exchange rates.

Detractors of HFT argue that trading algorithms are a source of market instability. HFTs that implement similar strategies may react to inaccurate signals at the same time, causing the stock market to temporarily crash as demonstrated by the May 6, 2010 Flash Crash (Biais & Foucault, 2014). Flash crashes occur when stock prices suddenly plummet and shortly thereafter, rise to their previous levels (Biais & Foucault, 2014). Such price movements do not reflect any change in the stocks' fundamentals. Various studies showed that ATs worsen short-term volatility. Zhang (2010) studied quarterly stock turnover data from the CRSP and Thomson Reuters Institutional Holdings databases from 1985 to 2009. The author examined the relationship between HFT and stock price volatility. The study presented evidence of a strong positive relationship between both variables even after controlling for firm fundamental volatility and other drivers of volatility such as the introduction of the automated quote dissemination system by the NYSE in 2003.

Zhang (2010) developed various proxies for fundamental volatility such as earnings surprise and sales growth volatility as well as analyst forecast dispersion, which were all found to be positively related to stock price volatility. Conversely, firm age, institutional holdings, book-to-market ratios were found to be negatively related to volatility. The results of the empirical analysis demonstrated that HFTs increased short-term volatility for large-cap stocks and stocks with high institutional holdings, especially during times of market uncertainty.

Kirilenko et al. (2017) produced a renowned study on this subject. The authors investigated whether HFTs triggered the Flash Crash which occurred on the 6th May 2010. On that day, in 30 minutes the price of the E-mini S&P 500 futures contract dropped by 5.1% before suddenly rising again. The study showed that the Flash Crash occurred as a result of a mutual fund selling 75 000 contracts valued at approximately $4.1 billion during the course of a single day. The authors argued that although HFTs did not trigger the Flash Crash, they aggravated volatility during the event.

Biais, Foucault and Moinas (2011) also claimed that HFT activity leads to an increase in the level of short-term volatility. Similarly, Boehmer, Fong and Wu (2012) found that AT systematically worsened short-term volatility across 39 non-U.S. markets. This finding was especially true for small and low-priced stocks. The authors argued that this relationship could not be explained by the fact that AT led to faster price discovery or tighter spreads. In summary, little consensus exists in the empirical literature regarding the causal effects of AT on short-term volatility.

6. STOCKS ACTIVELY TRADED BY ALGORITHMIC TRADERS

Several studies found that the level of AT activity depends on specific stock characteristics (liquid vs illiquid, large vs small-cap stocks). Hendershott and Riordan (2009) for example, showed that ATs were more likely to initiate trades when bid-ask spreads were narrow. Hagströmer and Norden (2013) corroborates this observation. In addition, Brogaard (2010) observed that HFTs actively trade large-cap stocks, with low market-to-book ratios, larger spreads and more depth. The study compared the NASDAQ HFT dataset with a Compustat dataset from 2008 to 2009. In the HFT dataset, the average market-cap was $17.6 billion compared to $3.37 billion in the other dataset. This difference was found to be statistically significant. Brogaard, Hendershott and Riordan (2014) found similar results. The study showed that from 2008 to 2009, HFTs made higher revenues when they traded large stocks compared to smaller ones. In addition, the study indicated that HFT activity was mainly concentrated in large stocks. This could be explained by the fact that large stocks have narrower spreads and are therefore more attractive to these types of traders.

O'Hara et al. (2011) found an inverse relationship between the number and volume of odd-lot trades and the volatility of stocks from January 2008 to December 2009. As mentioned previously, the majority of odd-lot traders are HFTs. The author explained that although splitting large trades into smaller ones reduces transaction costs, doing so over long time periods increases execution risk. As a result, ATs are less likely to trade volatile stocks. On the contrary, Brogaard (2010) as well as Brogaard, Hendershott and Riordan (2014) observed that the number and volume of trades initiated by HFTs significantly rose on the days where the NASDAQ stocks were the most volatile from 2008 to 2009. Brogaard (2010) found that during the most volatile days, the level of HFT activity remained stable whereas in the least volatile days, it dropped to a lower level than average.

7. CONCLUSION OF THE LITERATURE REVIEW

The academic literature on the topic of AT largely suggests that AT improves market quality. In fact, the general consensus seems to be that AT enhances market liquidity, price discovery and efficiency. In addition, prior studies provided strong evidence that AT imposes adverse selection costs on other market players. Finally, the findings regarding the overall impact of AT on short-term volatility are mixed. This study aims to evaluate the impact of AT on market quality and the next chapter describes the methodology employed to achieve this.

CHAPTER 3: METHODOLOGY

The aim of this study is to examine the effects of AT on two measures of market quality, namely, liquidity and short-term volatility. To achieve this goal, the first step is to construct a set of proxies to measure the level of AT activity for all JSE listed stocks, for each trading day in the sample period. The second step consists in explaining the variations in market quality across the JSE listed shares and trading days. Due to the cross-sectional and time series nature of this enquiry, panel data regression models are used to establish a relationship between AT and market quality. In the regressions, the dependent variables are the two measures of market quality and the explanatory variables are the proxies for AT activity.

This chapter is divided into six sub-sections. The first part presents the reasons for selecting the particular research approach based on prior literature. The second part describes the sample and explains the methods of data collection. The third part details the process of cleaning and consolidating the raw data into a table with daily aggregate values. The fourth part describes the methodology used to compute the AT proxies and the theory behind their selection. The fifth part defines the measures of market quality. Finally, the sixth part describes the technique used to estimate the panel data regressions.

1. FACTORS AFFECTING THE SELECTION OF THE RESEARCH METHOD

Intra-day data on trades and quotes that is publicly available do not specify the identity of traders. As such, it is impossible to pinpoint the exact orders initiated by ATs from public sources. The identification of AT activity therefore poses a major challenge for studies that seek to examine the effects of ATs on market quality. In prior literature, studies followed one of two approaches to measure AT activity. Some studies had access to non-public datasets that directly identified orders initiated by ATs. For example, Brogaard (2010), O'Hara, Yao & Ye (2011), Gai, Yao & Ye (2013), Brogaard et al. (2014), Chakrabarty et al. (2018), Hirschey (2018) all studied the same NASDAQ dataset, which consists of messages for 120 stocks from 2008 to 2009. Hendershott and Riordan (2009) examined trading data provided by the Deutsche Boerse. Kirilenko et al. (2017) analysed all trades in the E-mini contracts from the Computerized Trade Reconstruction (CTR) database.

This approach has several drawbacks. The NASDAQ dataset for example is incomplete as it only includes the order flows of independent proprietary trading firms. The orders of HFTs which form part of larger broker-dealers such as Goldman Sachs or Merrill Lynch are excluded (Brogaard, Hendershott & Riordan, 2014). Furthermore, the NASDAQ dataset does not provide any information on the type of strategy implemented by HFTs. Other studies inferred the portion of AT from publicly available information such as Hendershott et al. (2011) and Boehmer et al. (2012). This study follows a similar approach. There are two major advantages to examining a public dataset. Firstly, there are no barriers to the accessibility of the data thus allowing for potential future replications of the study. Secondly, public datasets from Bloomberg can be customized based on specific parameters such as the exact time period, stock exchanges and trade types so that studies can vary in focus.

2. SAMPLING AND DATA COLLECTION

As mentioned previously, electronic messages at the order level are not publicly available. This study therefore examines a sample of order book changes at the best bid and offer (BBO). This methodology is similar to that of Boehmer, Fong, and Wu (2015), who argue that AT activity in the BBO dataset is likely to be highly correlated to that in the order level dataset.

The dataset contains all the daily BBO quote updates between 09:00 and 17:00 over the period 1 January 2019 to 31 December 2019 on the JSE, with the exception of three days, where the data downloaded from Bloomberg was corrupted (12 to 14 June). In total, the sample period covers 119 trading days. The BBO data is time stamped to the 1/100 of a second and include order submissions, executions and cancellations. For each order submitted a number of order characteristics were recorded. These include, the type of order: “BEST_ASK”, “BEST_BID”, “MID_PRICE”, “TRADE”, the price at which these orders were submitted, the volume and conditional codes. These codes as well as the Bloomberg fields in the BBO dataset are listed in Tables 13 and 14 in the Appendix, respectively.

The sample period was chosen for two reasons. First, although intra-day BBO data can be exported from Bloomberg via the Excel plug-in, the amount that can be downloaded is limited to 90 days historically. The process for collecting data for this study started in February 2019 and at that time, data for 2018 was inaccessible. Second, it is costly to download and store such a large amount of data (16 Gigabytes in total) and a six month period is enough to identify daily patterns.

The data collection process can be divided into four steps. The first step was to identify all of the JSE listed stocks over the sample period. For each trading day, the Bloomberg EQS function shortlisted the universe of securities and the JSE helpdesk also sent the full list of listed stocks over the period, which allowed for a cross check. The second step consisted in downloading the intra-day tick data from Bloomberg for each stock on the list. The third step was to collect the market capitalisation figures for all listed stocks over the period. This information was sent by the JSE helpdesk.

The 50 Excel workbooks contain approximately 400 worksheets, with data for one stock per worksheet. Each worksheet displays information for hundreds of thousands of orders. In order to clean, summarize and run regressions on such a large amount of data, this study made use of scripts written for R and RStudio as well as SQL queries[1].

3. DATA PREPARATION PROCESS

The first R Script created for this study extracts all of the data in the worksheets for each workbook and combines it into one table saved in CSV format. The R script also removed data between 09:00 to 09:15 and 16:45 to 17:00 to avoid incorporating abnormal values associated with market open and close when orders are aggregated into an auction. Brogaard, Hendershott and Riordan (2014) for example excluded data before 09:30 and after 16:00 to address this problem. The CSV files were then imported into a SQL database. For each business day and share code, a SQL query aggregated a set of values. Extreme outliers such as data points with less than five trades were

[1] Please contact author for R scripts and SQL queries.

excluded from the analysis as they resulted in biased estimations of the volatility and liquidity measures. The aggregate values used in the computation of the AT proxies are defined in Table 3.

Table 3: Aggregate values for every stock.

Aggregate values per stock, per day	Description
(1) The total number of trades	Number of distinct orders with the type "TRADE" and conditional code "AT" (automated trade).
(2) The total number of orders	Number of distinct orders.
(3) The total number of odd-lots	Number of distinct orders with the type "TRADE" submitted in quantities smaller than 100 shares.
(4) The order volume	Total volume across all orders.
(5) The trade volume	Total volume across orders with the type "TRADE" and conditional code "AT" (automated trade).
(6) The odd-lot volume	Total volume across orders with the type "TRADE" submitted in quantities smaller than 100 shares.
(7) The trade value	The product of the volume and price for each trade.

The SQL query ignored in the calculations the orders with the type "MID_PRICE" as those orders are the average of the best ask and bid quotes. Furthermore, orders with specific market codes such as "IP" and "OC" which denote irregular trades were excluded from the sample (see Appendix for the list of Conditional Codes). This is consistent with the methodology of Boehmer, Fong & Wu (2012). The three algorithmic trading proxies, namely, average trade size, odd-lot volume ratio and trade-to-order volume ratio, were then estimated based on the daily aggregate values.

4. ALGORITHMIC TRADING PROXIES

In this study, three algorithmic trading proxies are constructed, namely, the average trade size, the odd-lot volume ratio and the trade-to-order volume ratio. Table 4 describes the AT proxies developed in this study.

Table 4: Algorithmic trading proxies.

AT Proxy	Relationship	Description
Average Trade Size	Negative	The trade volume divided by the total number of trades.
Odd-Lot Volume Ratio	Positive	The trade volume executed in quantities smaller than 100 shares divided by the total trade volume.
Trade-to-Order Volume Ratio	Negative	The trade volume divided by the order volume.

Prior literature suggests that AT is characterized by lower average trade sizes, higher odd-lot volume ratios and lower trade to order ratios. This section provides the theoretical reasons for this.

4.1. AVERAGE TRADE SIZE

A subset of HFTs attempt to front-run large institutional investors by placing a series of small orders to detect large hidden orders, a tactic known as "pinging". Similarly, HFTs who apply the technique of layering, submit orders to the market that are small in size to mislead other market participants. In some cases, these orders are executed, despite the HFT intending to cancel them. Several studies have shown that the average trade size fell significantly in the last decade. For example, the average trade size for large cap stocks listed on the NYSE fell from 1,600 shares in 1993 to 200 shares in 2011 (Stoll, 2014). Brogaard (2010) confirmed that HFTs trade in smaller quantities. The author compared the average trade size in two datasets from 2008 to 2009 for stocks listed on the NASDAQ and NYSE. The first dataset consisted of trades initiated by HFTs and the second one of trades initiated by human traders. The average trade size was 208 shares in the HFT dataset compared to 243 in the other dataset. This difference was statistically significant.
In this study, the average trade size is equal to:

$$\text{Average trade size} = \frac{\text{Trade volume}}{\text{Total number of trades}} \quad (1)$$

4.2. ODD-LOT VOLUME RATIO

ATs tend to split large blocks of shares into smaller ones to conceal their trading intentions from other market participants. As a result, they are more likely to trade odd-lots. Odd-lot trades are trades for less than 100 shares. Consistent with this theory, O'Hara et al. (2011) proved that the overwhelming majority of odd-lot traders after the mid-1990s were ATs. In addition, Weller (2018) suggested that HFTs who send out a series of "ping" orders to detect large hidden orders by institutional investors were more likely to trade odd-lots than ATs who followed market-making strategies. In sum, the odd-lot volume ratio is likely to capture the activity of ATs who take liquidity away from the market. In this study, on any given day and for every stock, the odd-lot volume ratio is computed as follows:

$$\text{Odd-lot volume ratio} = \frac{\text{Total volume of odd-lot trades}}{\text{Total trade volume}} \qquad (2)$$

4.3. TRADE-TO-ORDER VOLUME RATIO

On the one hand, ATs take advantage of their immense computing power to transmit thousands of order messages per second to various trading platforms. On the other hand, ATs avoid trading with each other because their strategies are highly correlated. They trade more frequently with human traders who implement different strategies (Chaboud, Chiquoine, Hjalmarsson & Vega , 2014). As a result, although ATs generate vast amounts of orders, only a small portion ultimately are executed. This means that ATs are more likely to have a lower ratio of trades per orders placed than other traders. This theory is consistent with the findings of Brogaard, 2010; Hendershott et al., 2011 and Boehmer et al., 2012).

In addition, ATs who engage in market-making activities frequently modify or cancel their limit orders to prevent their orders from becoming stale. As such, among all ATs, the group of market-makers have the highest probability of having a low trade to order volume ratio. Hagströmer and Norden (2013) corroborates this notion. The daily trade to order volume ratio for a given stock is calculated as follows:

$$\text{Trade-to-order volume ratio} = \frac{\text{Total trade volume}}{\text{Total volume across all orders placed}} \qquad (3)$$

4.4. CANCEL TO TRADE COUNT RATIO

Numerous studies also use the cancel-to-trade count ratio as a proxy for AT activity. The SEC's Concept Release on Equity Market Structure noted that HFTs tend to place a large amount of orders which are cancelled within a short period of time (SEC Concept Release, 2010, p. 3606). HFTs market-makers frequently cancel their limit orders to prevent them from becoming stale. In addition, HFTs who engage in manipulative practices such as quote stuffing or spoofing and layering cancel the vast majority of their orders. The "Autorité des marchés financiers" (AMF) surveillance department stated that in April 2010, the three HFT firms responsible for 39.6% of orders on the CAC-40, the major French stock market index, cancelled 96.5% of these orders (Zhang & Powell, 2011). The SEC made public the aggregate number of orders cancelled per stock per day. This information however, is only available for U.S. stocks. Thus, this study does not include this metric in the regressions when testing the impact of AT on the JSE.

5. MARKET QUALITY

In this study, two measures of market quality are regressed on the set of AT proxies described in Chapter four. The following section describes these measures and the theories behind their selection.

5.1. LIQUIDITY MEASURES

Two measures of liquidity are regressed on the AT proxies in the study. The first measure is the daily quoted bid-ask spread, standardized by the quote midpoint as described by Boehmer, Fong & Wu (2012). For a stock on any given day, the quoted spread is the amount by which the median ask price exceeds the median bid price. As such, LIQ_1 is defined as:

$$\mathrm{LIQ_1} = \frac{(\text{median ask} - \text{median bid})}{(\text{median ask} + \text{median bid})/2} \qquad (4)$$

The second measure is the daily effective bid-ask spread LIQ_2 which is the difference between the median trade execution price and the midpoint of the best bid and offer at the time of the trade. This measure is also standardized by the quote midpoint.

$$\text{LIQ}_2 = \frac{\text{median trade execution price} - (\text{median ask} + \text{median bid})/2}{(\text{median ask} + \text{median bid})/2} \qquad (5)$$

To account for the irregularity of intra-day tick data, the liquidity measures are initially computed for every hour in the sample period instead of the full business day. For each share code, the SQL query groups the orders based on the hour during which the orders were initiated and computes median values within these intervals. Thereafter, based on the hourly liquidity measures, the median values for the business day are estimated. Both quoted and effective bid-ask spreads are negatively related to liquidity.

5.2. SHORT-TERM VOLATILITY MEASURES

The volatility measures are the standard deviations of all intra-day arithmetic and logarithm returns. First, for any given share code, the arithmetic return is calculated as follows:

$$\text{Arithmetic return} = \frac{\text{trade price}_t - \text{trade price}_{t-1}}{\text{trade price}_{t-1}} \qquad (6)$$

Second, the logarithm return is the log of the price at which the stock is traded divided by the price at which it was previously traded.

$$\text{Logarithm return} = \log\left(\frac{\text{trade price}_t}{\text{trade price}_{t-1}}\right) \qquad (7)$$

The SQL built-in STDEV function returned the standard deviations based on these returns for a stock on any given day. Equations (11) and (12) below describe the mathematical formula of the STDEV function.

$$\text{Mean} = \frac{\text{Sum of each individual return}}{\text{Total number of items}} \qquad (8)$$

$$\text{Variance} = \frac{((\text{Return}_1 - \text{Mean})^2 + \ldots + (\text{Return}_x - \text{Mean})^2)}{\text{Total number of items} - 1} \quad (9)$$

$$\text{Standard deviation} = \sqrt{\text{Variance}} \quad (10)$$

6. PANEL DATA REGRESSIONS

6.1. REGRESSION EQUATION

The dataset in this study is multi-dimensional in that it comprises a cross-section and a time series element. The cross-sectional data is the daily measures of AT activity and market quality for a set of JSE listed stocks. The sample period spans 119 trading days. As such, panel data regressions may be used to establish relationships between the AT proxies and the measures of liquidity and volatility. Equation 11 describes the panel data regressions estimated in the study and Table 5 defines the terms included in the equation.

$$\text{Regression equation: } \mathcal{Y}_{it} = \alpha_i + X_{it}\beta + controls_{it} \times \gamma + \mathcal{U}_{it} \quad (11)$$

$$\text{Error term: } \mathcal{U}_{it} = \mu_i + \mathcal{V}_{it} \quad (12)$$

Table 5: Description of the terms in the panel data regression.

$\mathcal{Y}$	The two measures of market quality, namely, liquidity (LIQ_1, LIQ_2) and short-term volatility (VOL). The liquidity metrics are the quoted and effective bid-ask spreads and the short-term volatility is the degree to which stock returns are correlated.
i	The individual stocks.
t	The trading day.
α	The intercept which is constant across all stocks and trading days.
X	The AT proxies: average trade size, odd-lot volume ratio and trade-to-order volume ratio.

β	The effect of the AT proxies on the measures of liquidity and volatility ($\beta \neq 0$).
$controls_{it}$	Market capitalisation values, which were collected from the JSE.
γ	The effect of the control variables on the measures of liquidity and volatility.
μ_i	The individual-specific error component which captures any unobserved effects that are different across stocks but fixed in time.
$\mathcal{V}_{it}$	The stochastic regression disturbance, this term varies across stocks and trading days.

6.2. DESCRIPTION OF THE PANELS

Not all the stocks listed on the JSE are traded on a daily basis and as such, the full sample does not include data for every stock on every trading day. Therefore, different versions of the dataset are included in the regression analysis. Table 6 summarizes the characteristics of the three datasets in this study. The first dataset is a balanced panel and only consists of stocks with data for every trading day (Panel A). The second dataset only includes the stocks that have data for at least 100 of the 119 trading days (Panel B), and the third dataset is the full sample (Panel C).

Table 6: Characteristics of the three panels.

	Panel A	Panel B	Panel C
Number of stocks (n)	136	192	389
Minimum number of trading days for which all stocks have data	119	100	1
Number of observations (N)	16,184	22,450	28,654

6.3. IMPLEMENTATION IN R

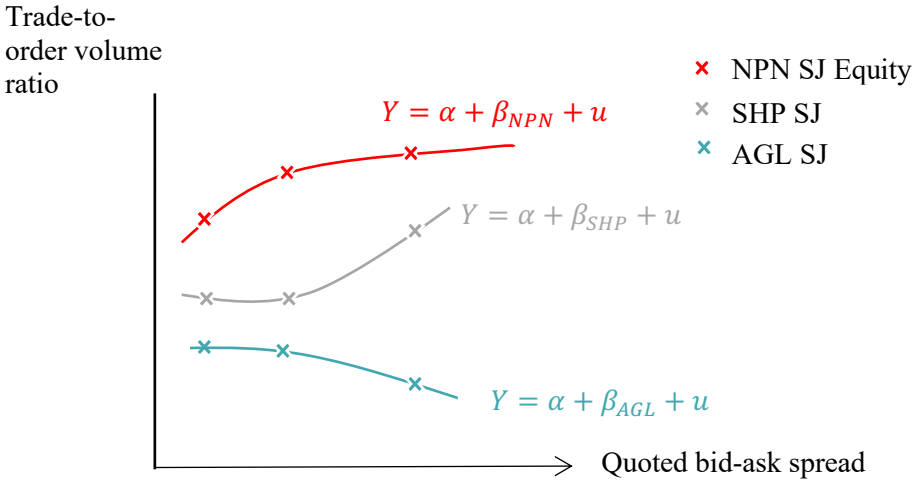
The Linear Models for Panel Data ("plm" package) in R, provides a range of functions that are specifically designed for panel data. First, the panel data has a time series component and as such, needs to be tested for a unit root as the stationarity assumption is implicit in the panel regression models. The linear models for panel data "plm" package in R provides a function "purtest" which

can test unit root hypotheses for panel data. This is important because including non-stationary variables in the regressions may lead to spurious results.

The R package includes a set of estimators for panel data regressions such as the pooled OLS (Ordinary Least Square), first differences, fixed effects, between and the random effects model. In this study, the first differences and fixed effects models are relevant as they rely on the assumption that the sample of stocks is the same over time. It should be noted that in Panel C, no stocks were removed from the sample and thus, there are no survivorship effects during the time period. In addition, these estimators disregard the differences in the average levels of liquidity and volatility between the different stocks because they are due to stock-specific characteristics.

In a first differences model, the first difference of the dependent variable is regressed on the first difference of the AT proxy and on that of the error term $\mathcal{U}_{it}$ (see graph below). As a result, the unobserved heterogeneity term α_i is removed from the equation. This model identifies all the data points that relate to a specific stock over time, draws a line connecting them (a line of best fits) and thereafter estimates the β of that line. This process is repeated for every stock in the sample as shown in Figure 1. Once all of the βs have been estimated, the average is computed.

Figure 1: Graphical representation of the first differences model.



In the fixed effects or "within" model, the time-demeaned dependent variable (liquidity or volatility) is regressed on the time-demeaned AT proxy variable plus time-demeaned error as follows:

$$\text{Fixed effects equation: } \ddot{Y}_{it} - \bar{Y}_t = \beta(X_{it} - \bar{X}_t) + \ddot{U}_{it} \quad (13)$$

This technique removes the unobserved heterogeneity term α_i from the equation. This estimator therefore generates a vector of the values in deviation from the individual means to compute β. The "between" model is similar to the "within" model, with the exception that the model regresses the actual time means as opposed to regressing the deviations from the time means of the variables.

The Pooled OLS estimator and the random effects model are likely to be inadequate in the context of this study. Indeed, the former assumes that the sample of stocks differs over time and treats the data like one broad cross-section. The latter assumes that the regression equation includes enough control variables so that the variation in the dependent variable is fully explained. Since market capitalisation is the only control variable included in the study, it is highly unlikely that this condition will be met.

Numerous tests in R can be performed to confirm that the first difference and fixed effects models are preferable, such as the Lagrange Multiplier (LM) test, the F test for individual and/or time effects and the Hausman test. First, the "plmtest" function in R performs the Lagrange Multiplier (LM) test, which is designed to test the presence of individual and/ or time effects. Individual effects are specific to the individual stocks, but constant over time, whereas time effects are specific to the time period, but constant for all the stocks. If the same sample is observed over time, these effects are generally present. The "plmtest" in R is founded on the equations presented in Honda (1985):

$$y_{it} = x_{it}\beta + \omega_{it} \quad (14)$$

$$\omega_{it} = \mu_i + \lambda_t + \upsilon_{it} \quad (15)$$

In the equations, i stands for the individual observations (1 to N) and t, the time periods (1 to T). Furthermore, x_{it} is an $1 \times k$ non-stochastic vector and β is $k \times 1$ unknown coefficient parameters. μ_i, λ_t and υ_{it} are independently and identically distributed with mean zeroes and variance ${\sigma^2}_{\mu}$,

$\sigma^2{}_\lambda$, $\sigma^2{}_\upsilon$ respectively. μ_i, λ_t and υ_{it} are mutually independent. The equations can be written in matrix form as follows:

$$y = X\beta + u \tag{16}$$

$$E(uu') = \sigma^2{}_\upsilon \mathrm{I_{NT}} + \sigma^2{}_\mu(\mathrm{I_N} \otimes \mathrm{e_T e'_T}) + \sigma^2{}_\lambda(\mathrm{e_N e'_N} \otimes \mathrm{I_T}) \tag{17}$$

Where y and u are $TN \times 1$ and X is $TN \times k$, non-stochastic and of rank k. $\otimes$ is the Kronecker product. The null hypothesis states that both individual and time effects do not exist:

$$\text{Null hypothesis: } \sigma^2{}_\mu = \sigma^2{}_\lambda = 0 \tag{18}$$

In the case where μ_i, λ_t and υ_{it} are normally distributed, the Lagrange multiplier (LM) test statistic can be written as follows:

$$\text{LM stat.} = \left\{\tilde{u}'\left(I_N \otimes (e_T e'_T - I_T)\right)\tilde{u}/\tilde{\sigma}^2{}_\upsilon \sqrt{2L}\right\}^2 + \left\{\tilde{u}'\left((e_N e'_N - I_N) \otimes I_T\right)\tilde{u}/\tilde{\sigma}^2{}_\upsilon \sqrt{2M}\right\}^2 \tag{19}$$

Where $\tilde{u}$ is $TN \times 1$ ordinary least squares residuals, $\tilde{\sigma}^2{}_\upsilon$ is equal to $\tilde{u}'\tilde{u}/TN$, L is equal to $TN(T-1)$ and M, $TN(N-1)$.

Second, the "pFtest" function in R compares the pooling and fixed model for individual and/or time effects. Finally, the "phtest" function in R runs the Hausman test where the null hypothesis is that the random effects model is preferred to the fixed effects model. More specifically, the Hausman test estimates the difference of the vectors of coefficients of two models, in this case the random and fixed effects model. The equation presented in Hausman (1978) is:

$$y = X\beta + \tilde{X}\alpha + \upsilon \tag{20}$$

In this test, the null hypothesis is equal to:

$$\text{Null hypothesis: } \alpha = 0. \tag{21}$$

The results of these preliminary tests are reported in the third part of the next chapter.

CHAPTER 4: RESULTS

This Chapter reports the main findings of this study and provides an interpretation of the results. It is divided into six sub-sections. The first sub-section presents summary statistics in the sample, including the number of trades, odd-lots and orders, trade volumes as well as trade values. The second sub-section provides numerous graphs representing the relationship between the key variables in the sample and for six stocks in different market capitalisation categories. The third sub-section details the results of stationarity tests and other tests that determine the most appropriate panel regression model given the sample. Finally, the fourth, fifth and sixth sub-sections report the results of the panel data regressions that test the relationship between AT and the measures of market quality (liquidity and short-term volatility, respectively).

1. SUMMARY STATISTICS

Over the sample period, the number of JSE listed companies ranged from 205 to 283 with an average of 241. The firms in the sample had a market capitalisation value of R52 billion on average. The largest company listed on the JSE during the period, ANH SJ Equity (Anheuser-Busch InBev SA/NV), had an average market value of R 1,94 trillion.

1.1. DISTINCT NUMBER OF TRADES, ODD-LOTS AND ORDERS

Figures 2, 5 and 7 compare summary statistics published by the JSE to those included in the sample. These statistics include the total number of trades, trade volume and trade value over the first six months of 2019. When reporting these values, the JSE included both sides of the trades (buying and selling). As such, the statistics in Figures 2 and 5 were multiplied by two so as to provide a meaningful comparison. This, however, is not the case for the summary statistics reported in Tables 17 to 19 in the Appendix.

Figure 2 below compares the aggregate number of trades for the equity market, sub-categorised into colocation and non-colocation activity, to those obtained in the sample. The diagram shows that HFTs initiated a significant portion of the trades over the period. In addition, the sample

amounts are lower than the JSE's estimates because the orders initiated in the first and last 15 minutes of each trading day were excluded from the study. The results indicate that about 520,000 trades occurred per day on average compared to roughly 590,000 as reported by the JSE. In every month within the sample, NPN SJ Equity (Naspers Ltd) was the most frequently traded stock, with a monthly average of 396,200 trades. Collectively, these trades represented about 3.83% of the total number of trades during the period.

Figure 2: Number of trades reported by the JSE and in the sample.

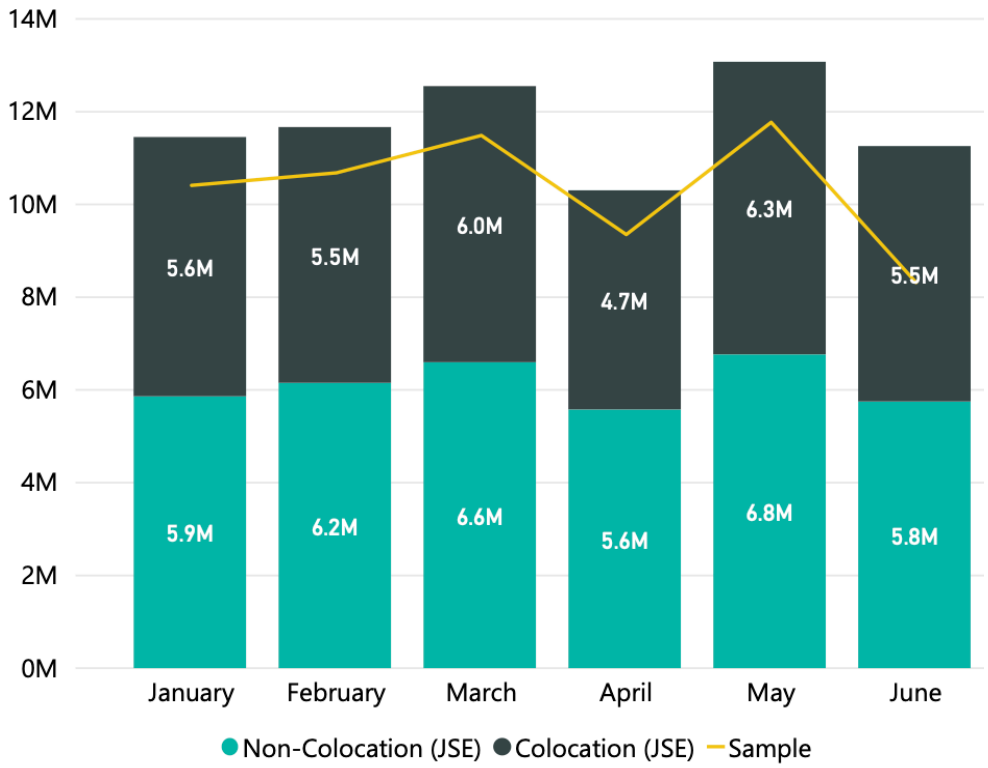


Figure 3 reports the total number of trades, odd-lots and orders in the sample, while Figure 4 groups these values based on their market capitalisation categories (small, medium and large). In this study, trades for less than 100 shares of stock are defined as odd-lot trades. On average, odd-lot trades represent about 35% of the total number of trades during the sample period. Consistent with the above findings, NPN SJ Equity was the single most frequently traded stock in odd-lots, with an average of 332,830 trades per month. The results suggest that 84% of Naspers trades are odd-lots. Over the sample period, the share with the highest number of orders is BTI SJ Equity (British

American Tobacco PLC) with about 12,27 million orders, closely followed by NPN SJ Equity with 12,26 million orders.

Figure 3: Number of orders, trades and odd-lots in the sample.

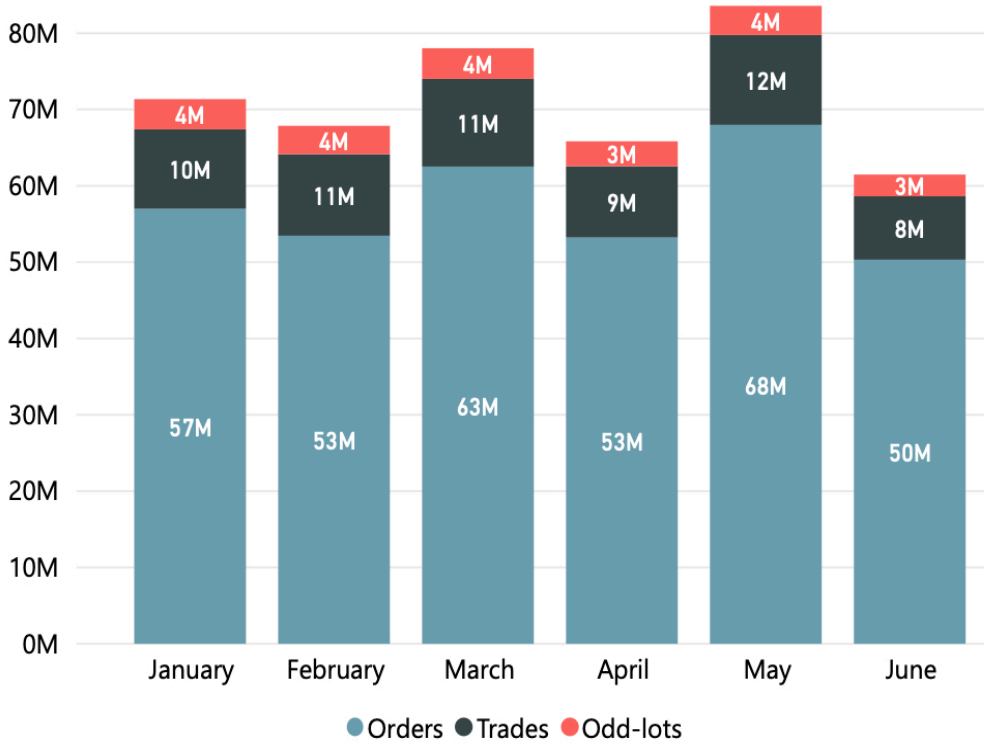
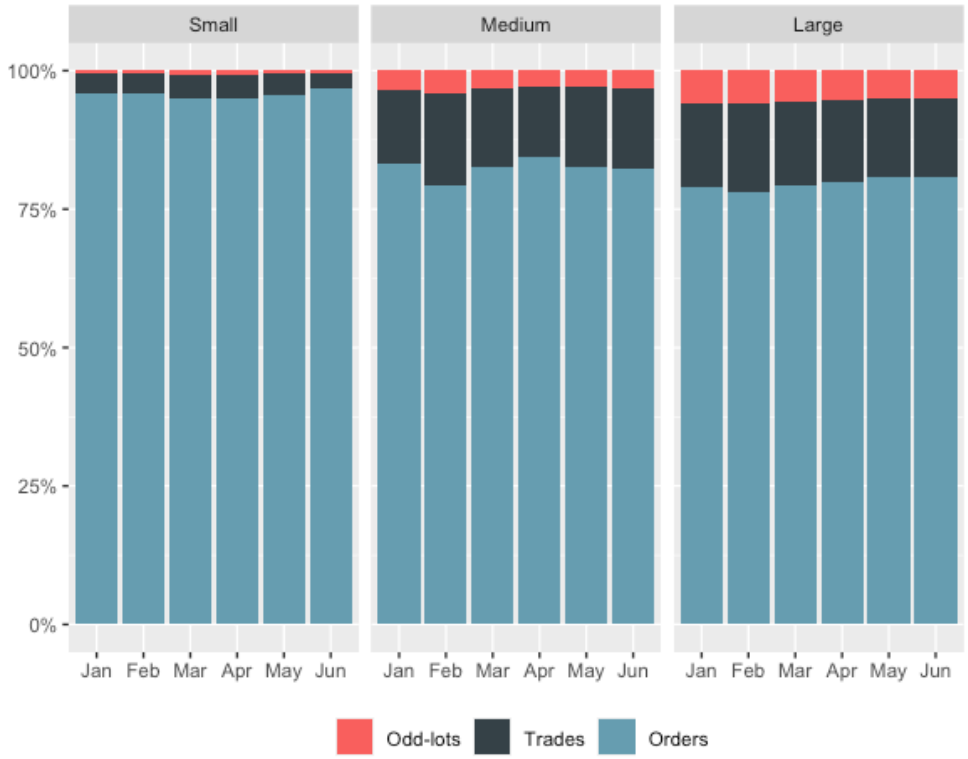


Figure 4: Number of orders, trades and odd-lots for three market capitalisation categories.



1.2. TRADE, ODD-LOT AND ORDER VOLUMES

Figure 5 compares the aggregate trade volumes for the equity market, sub-categorised into colocation and non-colocation activity, to those in the sample. As previously explained, the values in the sample are lower, due to the exclusion of the data detailing the first and last 15 minutes of the trading day, as well as the absence of the corrupted data for three business days in June. Despite the exclusions, the trends of both datasets remain identical.

Figure 5: Trade volumes (in billions of shares) reported by the JSE and in the sample.

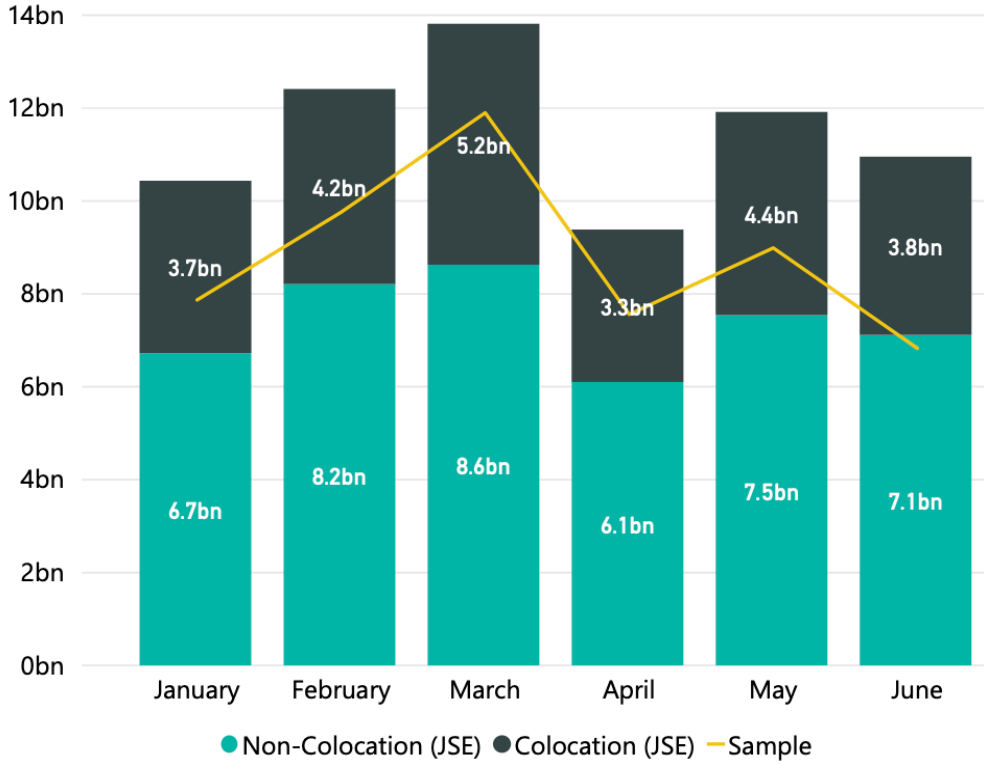
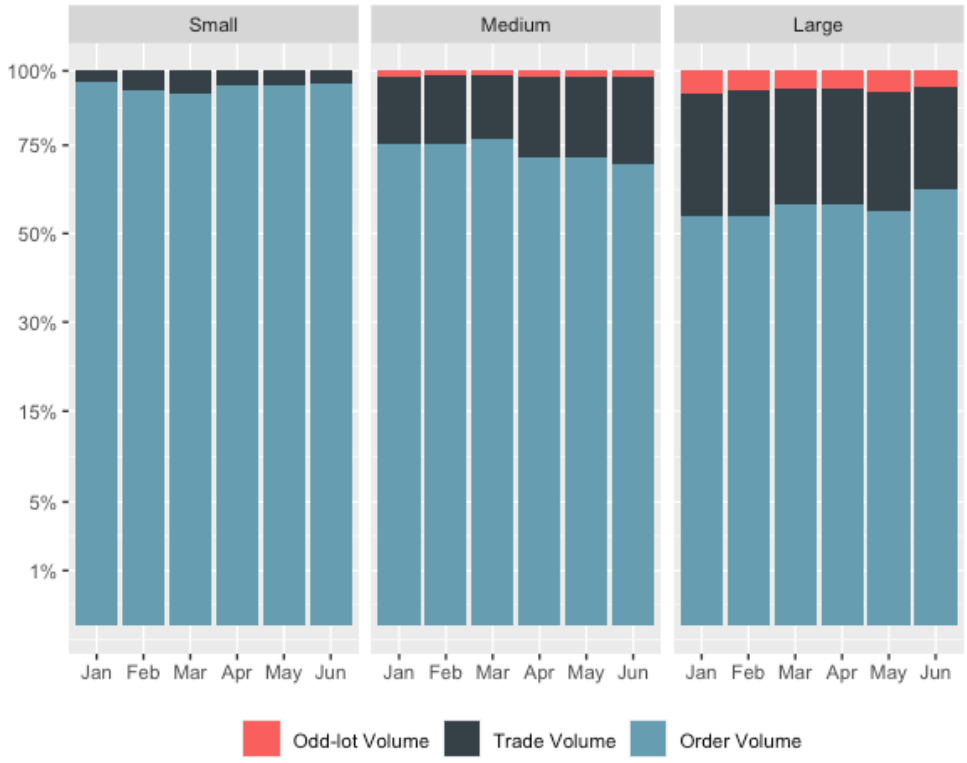


Figure 6 presents the trade, odd-lot and order volumes by market capitalisation category. The diagram indicates that large cap stocks are more frequently traded in odd-lots than medium and small cap stocks. The daily average trade volume is 445 million shares. The mining company AEG SJ Equity (Aveng Ltd) has the highest trade volume over the sample period, with R4 billion shares traded in total. This result is most likely due to the stock's share price, which is relatively low, fluctuating around five cents. The daily average odd-lot trade volume is roughly 8 million shares. NPN SJ Equity is the stock with the highest odd-lot trade volume.

The stock AEG SJ Equity has the highest order volume during the sample period. Regarding average trade size, the median is 993 shares over the sample period. Once again, AEG SJ Equity is the share with the highest median average trade size (183,000 shares). In comparison, the average trade size for NPN SJ Equity is 69 shares.

Figure 6: Trade, odd-lot and order volumes for three market capitalisation categories.



1.3. TRADE VALUE

Figure 7 summarizes both the trade values published by the JSE as well as those included in the study. The pattern follows that of Figure 5. Finally, Figure 8 reports the trade values for the three market capitalisation categories. The daily average trading value is R29 billion. NPN SJ Equity is the stock with the highest trade value, with monthly average of R90 billion. This represents approximately 16% of the total in the sample period.

Figure 7: Trade value (in billions of Rands) reported by the JSE and in the sample.

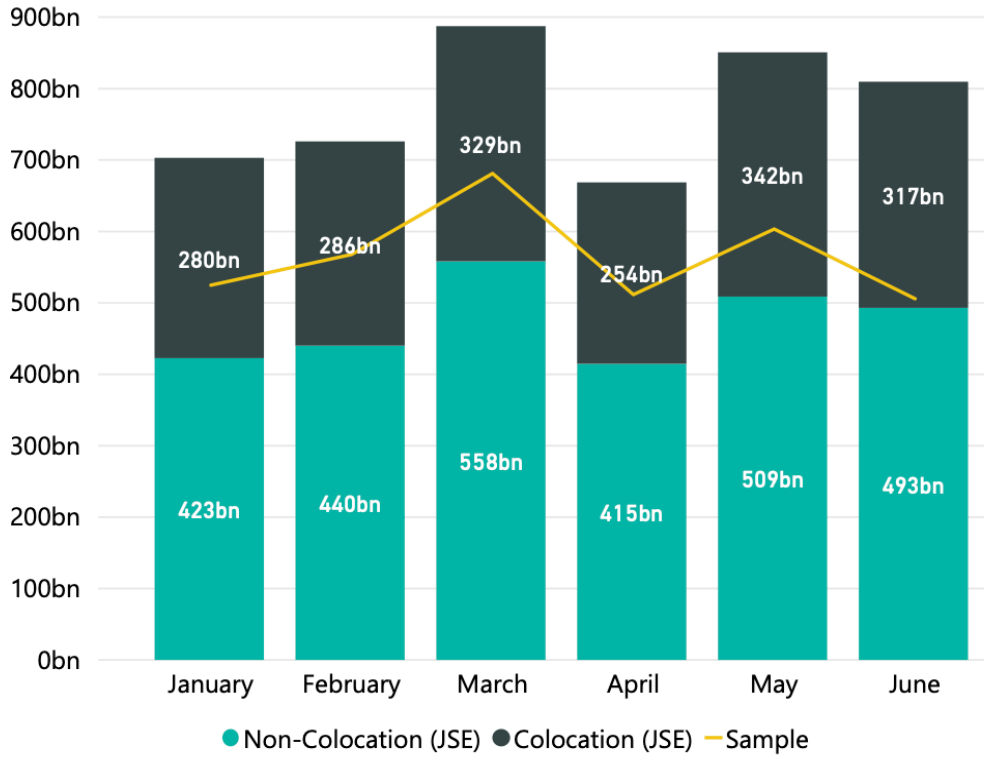
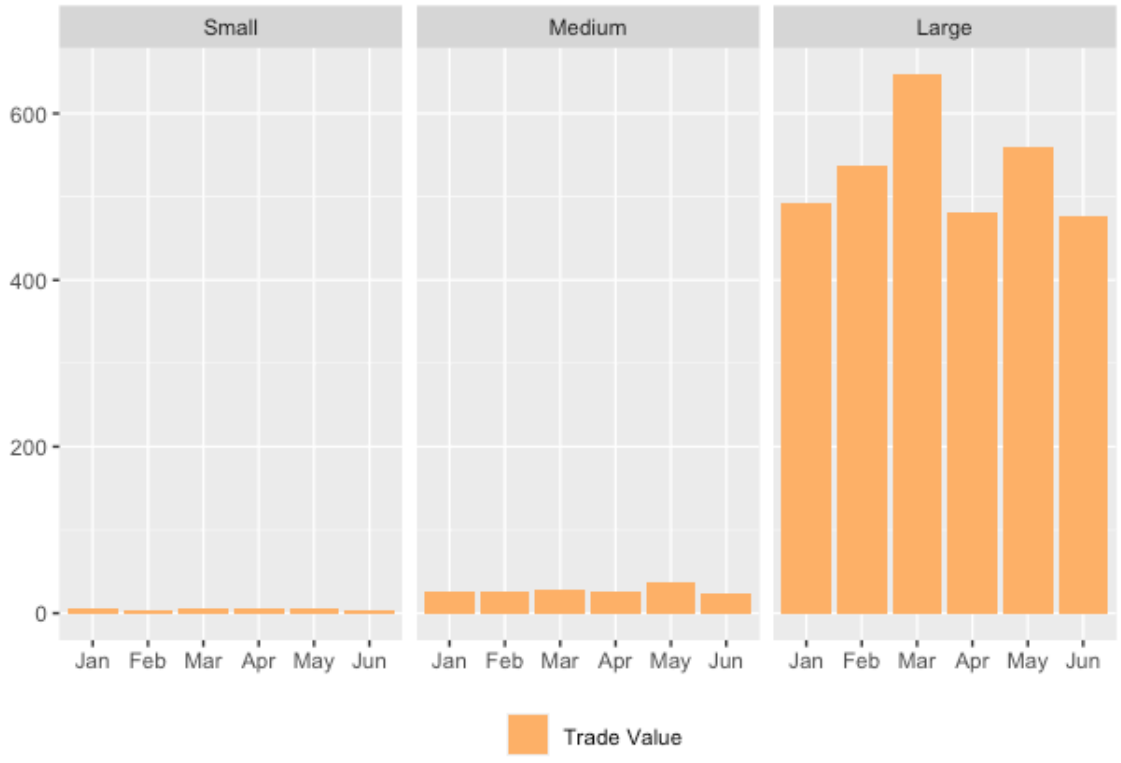


Figure 8: Trade value (in billions of Rands) for three market capitalisation categories.



2. GRAPHS DEPICTING THE RELATIONSHIPS BETWEEN KEY VARIABLES

2.1. AT PROXIES AND MARKET CAPITALISATION

The AT proxies included in the regressions are the average trade size, odd-lot volume ratio, and trade-to-order volume ratio. Figure 9 reports the median average trade size over time for the market capitalisation categories. During the sample period, the average trade size fluctuated around 1,000 shares. Figure 9 clearly shows that this variable is inversely related to the size of the stocks. The finding that smaller stocks are traded in larger quantities is unsurprising and highlights the need to incorporate a control variable for the size of stocks in the panel data regressions.

Figure 9: Average trade size for three market capitalisation categories.

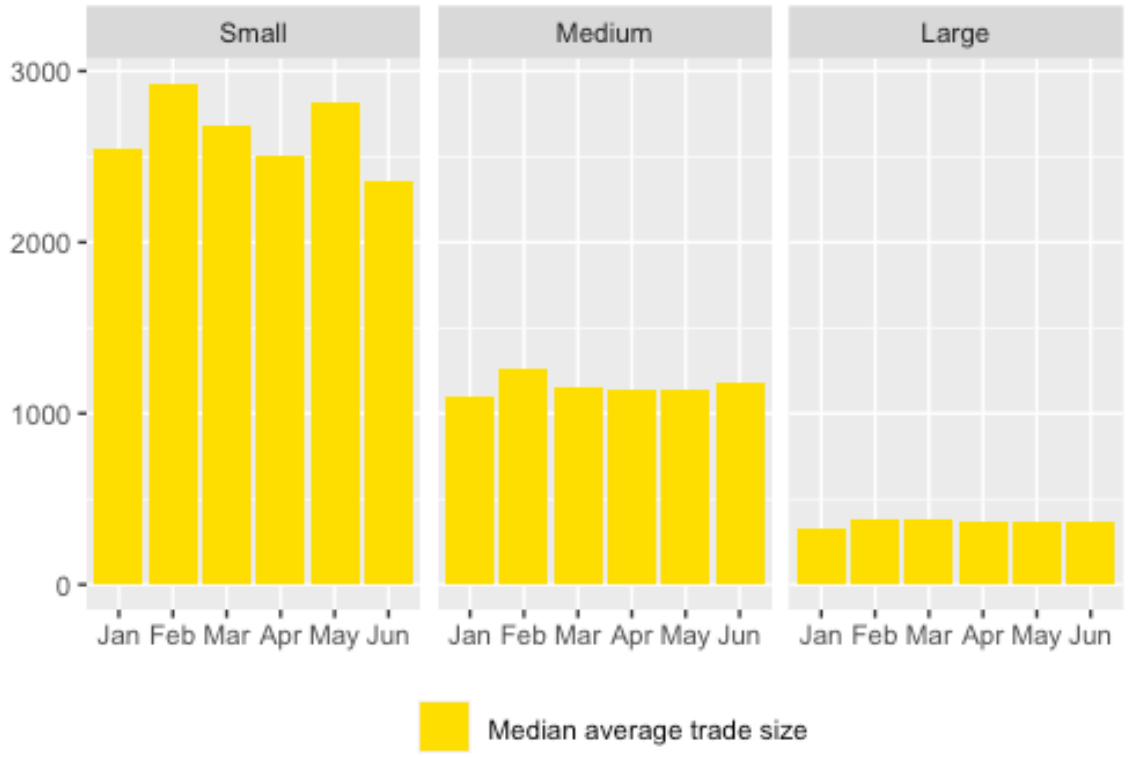
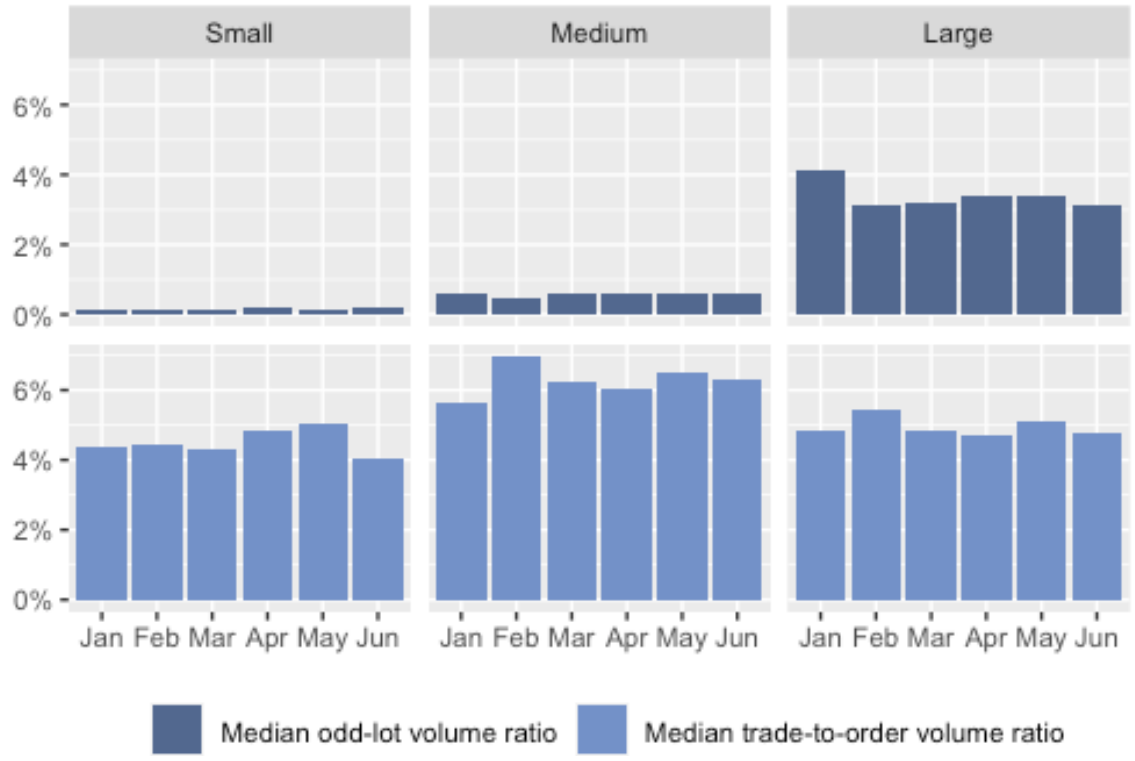


Figure 10 presents the median odd-lot and trade-to-order volume ratios over time for three market capitalisation categories. The diagrams highlight a positive relationship between odd-lot volume ratios and market capitalisation. However, the trade-to-order volume ratio variable does not appear to be correlated with market capitalisation.

Figure 10: Odd-lot and trade-to-order volume ratios for three market capitalisation categories.



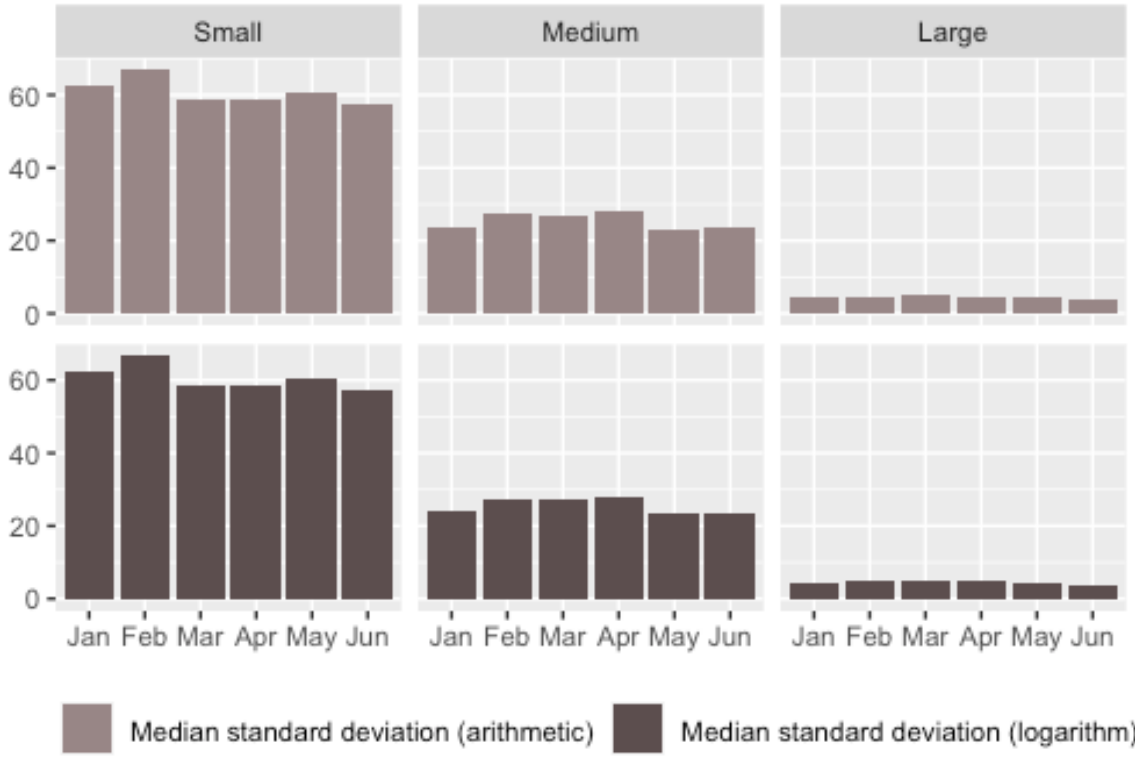
2.2. LIQUIDITY, SHORT-TERM VOLATILITY AND MARKET CAPITALISATION

Figure 11 presents the quoted and effective bid-ask spreads, both standardized by the quote midpoint for three market capitalisation categories. As expected, the diagrams indicate that smaller stocks have large spreads and thus, are less liquid than larger stocks. Finally, Figure 12 compares the median standard deviations of arithmetic and logarithm returns for the three market capitalisation categories. Naturally, small stocks are more volatile than larger stocks. The median standard deviation of daily arithmetic returns during the sample period was 19.85 bps.

Figure 11: Liquidity measures for three market capitalisation categories (in bps).



Figure 12: Volatility measures for three market capitalisation categories (in bps).



2.3. SIX STOCKS IN DIFFERENT MARKET CAPITALISATION CATEGORIES

Figures 13 to 17 present the AT proxies, liquidity and volatility measures for six stocks with extreme values in each market capitalisation category. Figure 13 shows the monthly median average trade sizes for the six stocks. As described previously, stocks associated with lower average trade sizes are more likely to have been traded by ATs. This is particularly the case of opportunistic ATs, who break their orders into smaller pieces to hide their trading intentions from the market. The results suggest that, within the large caps category, ATs are more likely to trade NPN SJ Equity than SHP SJ Equity. Similarly, they are more likely to trade GLD SJ Equity than AIL SJ Equity in the medium caps and ADI SJ Equity than REB SJ Equity in the small caps.

Figure 13: Average trade size for six stocks in the market capitalisation categories.

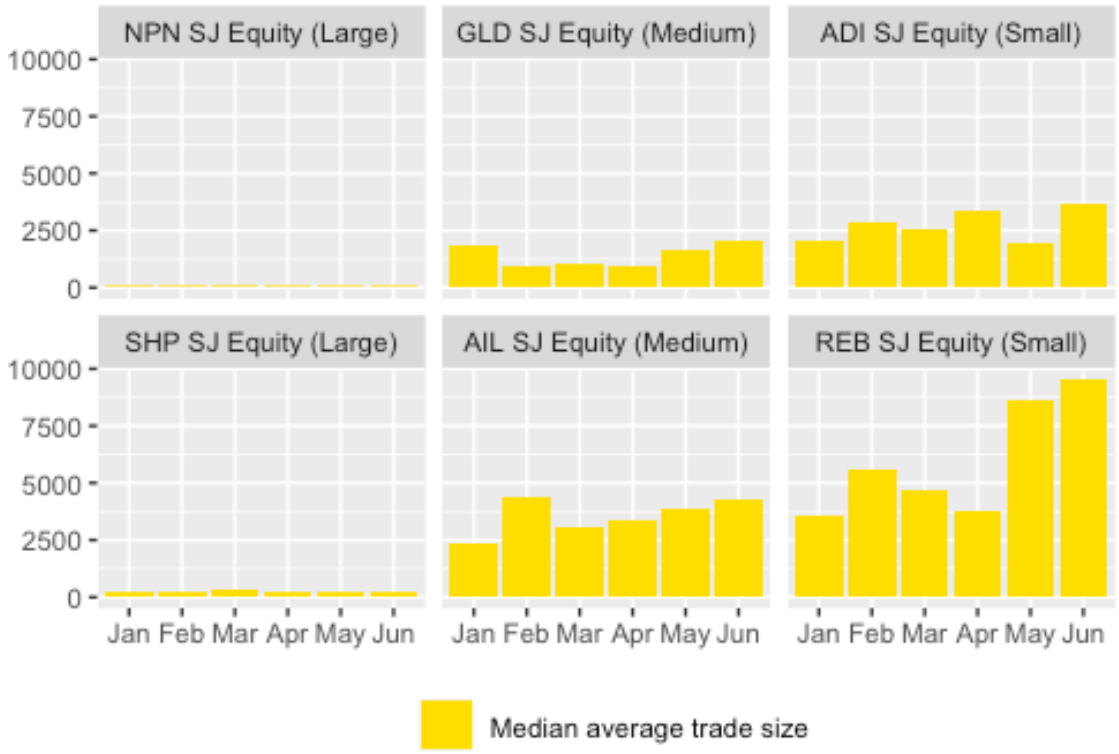


Figure 14 presents the monthly median odd-lot trade volume ratios for the stocks. The results suggest a high level of correlation between the AT proxy and market capitalisation. Consistent with previous findings, within the three market cap categories, ATs are more likely to trade NPN SJ Equity than SHP SJ Equity, GLD SJ Equity than AIL SJ Equity and ADI SJ Equity than REB

SJ Equity. Notably, NPN SJ Equity had a particularly high median odd-lot trade volume ratio of 50.43% over the sample period.

Figure 14: Odd-lot volume ratios for six stocks in the market capitalisation categories.

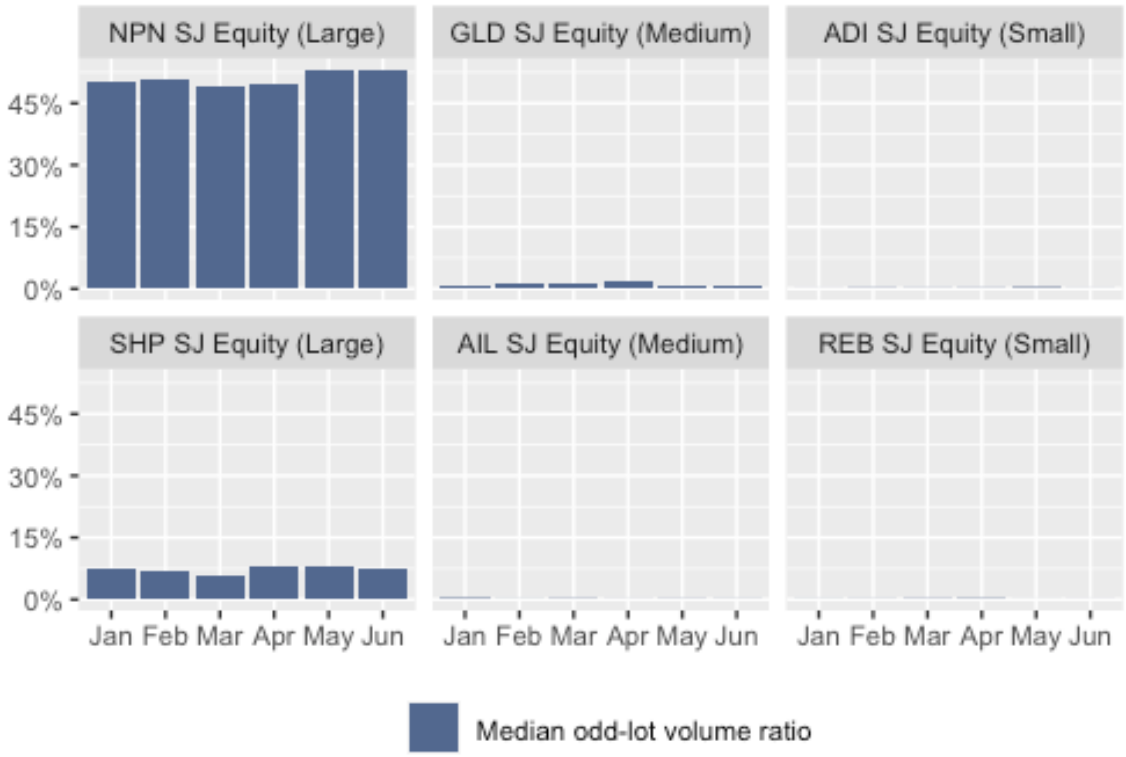
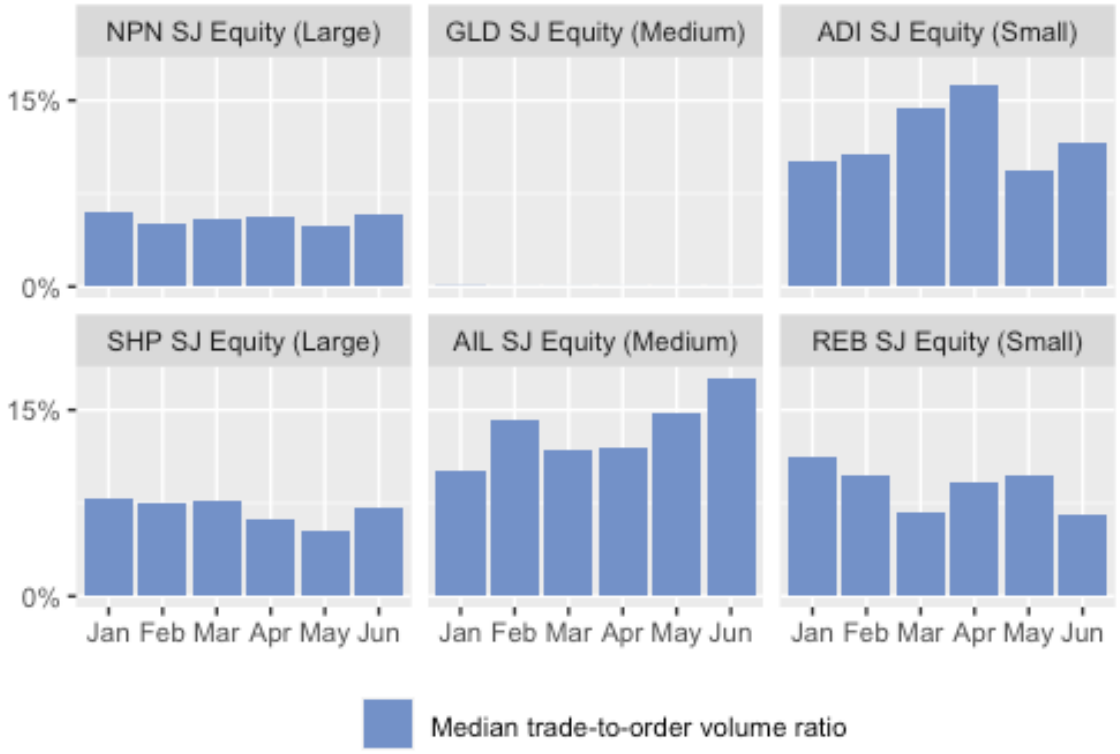


Figure 15 reports the median trade-to-order volume ratios for the stocks. HFT market-makers submit a large number of orders, which are constantly modified and often cancelled. Stocks traded by ATs are therefore likely to have low trade-to-order volume ratios. Consistent with previous findings, the results suggest that NPN SJ Equity and GLD SJ Equity are prone to AT activity. In contrast to the previous results, REB SJ Equity has lower ratios than ADI SJ Equity in the small caps category. Possibly, the AT proxies capture AT activity that relate to different trading strategies. In fact, ATs who follow market-making strategies are likely to initiate large amounts of orders, whereas opportunistic ATs are more likely to trade in small quantities. To conclude, overall Figure 13 to 15 suggest that the stocks in the top three diagrams are more likely to be traded by ATs. These stocks have lower average trade sizes, higher odd-lot volume ratios and lower trade-to-order volume ratios.

Figure 15: Trade-to-order volume ratios for six stocks in the market capitalisation categories.



Next, Figures 16 and 17 display the liquidity and short-term volatility measures for the stocks respectively. Figure 16 indicates that large cap-stocks have tighter bid-ask spreads than small caps. More importantly, the three stocks prone to AT activity, in the top diagrams, are also the most liquid in their respective market capitalisation category. As such, AT activity seems to enhance the liquidity of the stocks. The panel data regressions in this study formally test this relationship. Finally, Figure 17 suggests that large cap stocks are less volatile than small cap stocks. Once again, the stocks with a high expected level of AT activity are the least volatile in their market capitalisation category. The relationship between AT and short-term volatility is formally tested in the panel data regressions.

Figure 16: Quoted spreads for six stocks in the market capitalisation categories (in bps).

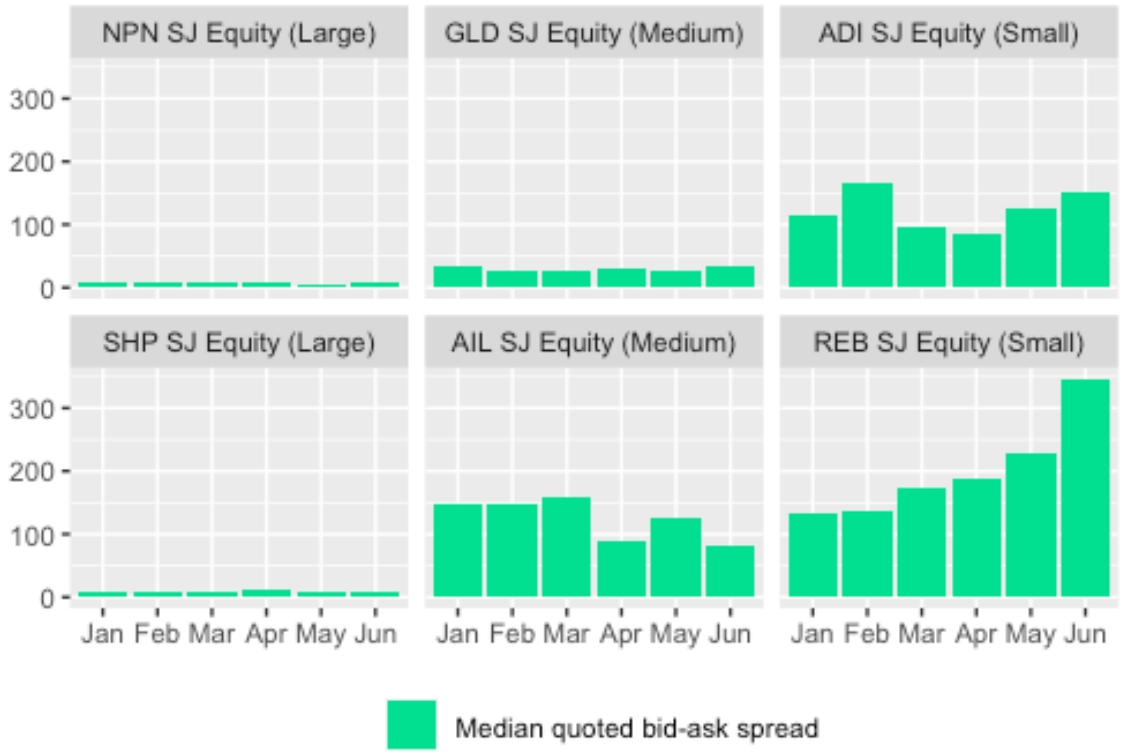
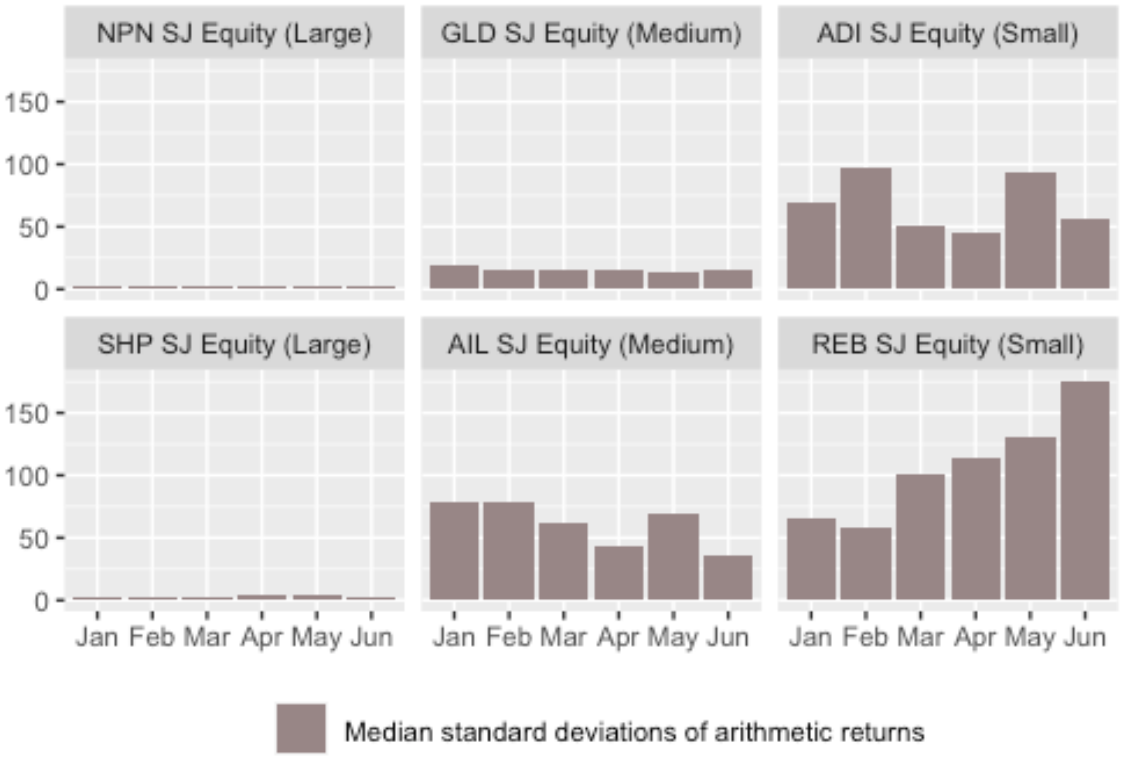


Figure 17: Standard deviations for six stocks in the market capitalisation categories (in bps).



3. PRELIMINARY RESULTS

3.1. STATIONARITY TESTS

When testing the balanced panel (Panel A), the Levin-Lin-Chu test rejected the null hypothesis of a unit root in favour of the alternative i.e. the data is stationary. Similarly, the Maddala-Wu unit root test rejected the unit root null for Panels B and C. The full results are reported in Table 20 in the Appendix. The panel data models rely on the assumption that the data is stationarity to provide accurate results. These tests prove that this condition has been met.

3.2. THE LAGRANGE MULTIPLIER TEST

The Lagrange Multiplier (LM) test is performed to determine the panel data model that is the most appropriate given the sample. The results of this test are presented in Tables 21 to 23 in the Appendix. As expected, for all the panels, the P-values in all equations indicate that there are significant effects present, thereby rejecting the null hypothesis. Hence, the fixed effects model is more appropriate than the pooled OLS model. In the tables, the degree of freedom (df1) is the number of individual stocks minus one and (df2) is the total number of observations minus one.

3.3. THE F TEST FOR INDIVIDUAL AND/OR TIME EFFECTS

The F test for individual and/or time effects compares the pooling and fixed model for individual and/or time effects. The results of this test are presented in Tables 24 to 26 in the Appendix. Once again, the results provide more support for the Fixed Effects model compared to the Pooled OLS model. The results also recommend using the first difference and "between models" for the panel data regressions.

3.4. THE HAUSMAN TEST

The Hausman test compares the random effects and the fixed effects models. The null hypothesis is that the random effects model is preferred to the fixed effects model. If the difference between the P-value and the chi square is less than 0.05, then the fixed model estimator is recommended. The results in Tables 27 to 29 for the three panels in the Appendix, indicate that one of the models

is inconsistent and that the fixed effects model is preferable to the random effects model. To conclude, the results for all panels support the use of the fixed, between and first differences models to estimate the panel data regressions.

4. ALGORITHMIC TRADING ON MARKET LIQUIDITY

Tables 7 and 8 below summarize the panel regression results on the relation between AT and the median quoted and effective bid-ask spreads, respectively. Both liquidity measures are standardized by the quote midpoint.

Table 7: The regression of the quoted bid-ask spreads on the AT proxies.

Panel	**Estimator**	**Average trade size**		**Odd-lot volume ratio**		**Trade-to-order volume ratio**	
A	Fixed effects	(+)	***	Insignificant		(-)	**
	Between	(+)	***	Insignificant		Insignificant	
	First differences	Insignificant		Insignificant		(-)	*
B	Fixed effects	(+)	***	(+)	*	(-)	***
	Between	(+)	***	Insignificant		Insignificant	
	First differences	Insignificant		(+)	*	(-)	***
C	Fixed effects	(+)	***	(+)	**	(-)	***
	Between	(+)	***	(-)	*	(+)	***
	First differences	Insignificant		Insignificant		(-)	***

Refer to Tables 30 to 32 in the Appendix for full results.

Table 8: The regression of the effective bid-ask spreads in the AT proxies..

Panel	Estimator	Average trade size		Odd-lot volume ratio	Trade-to-order volume ratio	
A	Fixed effects	(+)	***	Insignificant	Insignificant	
	Between	(+)	***	Insignificant	(-)	*
	First differences	(-)	**	Insignificant	Insignificant	
B	Fixed effects	(+)	**	Insignificant	Insignificant	
	Between	(+)	***	Insignificant	Insignificant	
	First differences	(-)	*	Insignificant	(-)	*
C	Fixed effects	Insignificant		Insignificant	Insignificant	
	Between	(+)	**	Insignificant	Insignificant	
	First differences	Insignificant		Insignificant	(-)	*

Refer to Tables 33 to 35 in the Appendix for full results.

4.1. AVERAGE TRADE SIZE AND MARKET LIQUIDITY

Table 7 indicates that both the fixed effects and "between" estimators established a positive and statistically significant relationship between the average trade size and the median quoted bid-ask spreads. The first differences estimator generated insignificant results. It must be emphasized that the results of the fixed effects estimator will be more accurate than the other two as this model accounts for both individual and time effects. As described previously, popular AT strategies entail breaking down large orders into smaller ones to prevent their trading intentions from becoming known to other investors. In theory, the average trade size is therefore inversely related to the level of AT activity. Since the bid-ask spreads are inversely related to market liquidity, AT activity is positively related to liquidity. This finding is consistent with prior literature such as Hendershott and Riordan (2009), Hendershott, Jones and Menkveld (2011), Kirilenko, Kyle, Samadi and Tuzun (2017), Boehmer, Fong and Wu (2012), Hagströmer and Norden (2013), Hasbrouck and Saar (2012).

Similarly, in Table 8, both the fixed effects and "between" estimators indicate a strong positive correlation between the average trade size and median effective bid-ask spreads in Panels A and B. For Panel C, the "between" estimator confirms the previous finding whereas the fixed effects

estimator presents insignificant results. Regarding the first differences model, the relationship is found to be negative and statistically significant for Panels A and B and insignificant in Panel C.

4.2. ODD-LOT VOLUME RATIO AND MARKET LIQUIDITY

Table 7 indicates that none of the tests revealed significant relationships between the odd-lot volume ratios and the median quoted spreads for Panel A. For Panel B, the fixed effects and first differences models highlight a positive and statistically significant relationship. The "between" estimator produces insignificant results. For Panel C, although both fixed effects and "between" models confirm a statistically significant relationship, the former indicates a positive correlation whereas the latter a negative correlation. Table 8 shows that when replacing the liquidity measure with the effective spreads, none of the models present significant results for all three Panels. As noted earlier, ATs have a tendency to slice large orders into odd-lots and are therefore characterised by a high odd-lot volume ratio. In theory, the expectation is that the spreads are inversely related to the odd-lot volume ratio. The evidence in this study is inconclusive on this question.

4.3. TRADE-TO-ORDER VOLUME RATIO AND MARKET LIQUIDITY

Finally, Table 7 shows that the fixed effects and first differences estimators reveal a strong inverse correlation between the trade-to-order volume ratios and median quoted bid-ask spreads. The "between" estimator produces insignificant results for Panels A and B, but a strong positive relationship for Panel C. In Table 8, the "between" model specifies a negative and statistically significant relationship between the AT proxy and the effective bid-ask spread in Panel A. The first differences model also highlights a similar relationship for Panels B and C. All other results in the table are insignificant. Theoretically, the trade-to-order ratio is inversely related to AT activity and the bid-ask spreads are inversely related to market liquidity. Consequently, the results suggest that AT activity is associated with a reduction in the level of market liquidity. This is in contradiction to the previous findings. Brogaard (2010) reached the same conclusion and showed that in his sample, on average, HFTs provide less liquidity than non-HFTs. Brogaard (2010) constructed a different liquidity measure to the one in this study, namely, the depth of the order book.

In sum, the panel regression results differ depending on the AT proxy incorporated in the equations. The inclusion of the average trade size variable led to the expected conclusion that AT activity is positively associated with market liquidity, consistent with the vast majority of prior literature on the topic such as Hendershott and Riordan (2009), Hendershott, Jones and Menkveld (2011), Kirilenko, Kyle, Samadi and Tuzun (2017), Boehmer, Fong and Wu (2012), Hagströmer and Norden (2013), Hasbrouck and Saar (2012). However, the inclusion of the odd-lot volume ratio and trade-to-order volume ratio suggest the opposite conclusion, consistent with Brogaard (2010). This contradiction may be explained by the fact that the AT proxies capture the activity of ATs engaged in different strategies. The average trade size proxy may capture the activity of market-makers whereas the other two proxies may represent the activity of opportunistic traders. Furthermore, the results were relatively similar for both measures of liquidity: the quoted and effective bid-ask spreads.

5. ALGORITHMIC TRADING AND SHORT-TERM VOLATILITY

Tables 9 and 10 report the panel regression results on the relation between AT and the standard deviation of arithmetic and logarithm returns, respectively.

Table 9: The regression of the standard deviations of arithmetic returns on the AT proxies.

Panel	**Estimator**	**Average trade size**		**Odd-lot volume ratio**	**Trade-to-order volume ratio**	
A	Fixed effects	(-)	***	Insignificant	(-)	*
	Between	(+)	***	Insignificant	Insignificant	
	First differences	(-)	***	Insignificant	(-)	*
B	Fixed effects	(-)	***	Insignificant	(-)	***
	Between	(+)	***	Insignificant	Insignificant	
	First differences	(-)	***	Insignificant	(-)	***
C	Fixed effects	(-)	**	Insignificant	(-)	***
	Between	(+)	***	Insignificant	Insignificant	
	First differences	(-)	***	Insignificant	(-)	***

Refer to Tables 36 to 38 in the Appendix for full results.

Table 10: The regression of the standard deviations of logarithm returns on the AT proxies.

Panel	Estimator	Average trade size		Odd-lot volume ratio		Trade-to-order volume ratio	
A	Fixed effects	(-)	***	Insignificant		(-)	***
	Between	(+)	***	Insignificant		Insignificant	
	First differences	(-)	***	(+)	*	(-)	**
B	Fixed effects	(-)	***	Insignificant		(-)	***
	Between	(+)	***	Insignificant		Insignificant	
	First differences	(-)	***	(+)	*	(-)	***
C	Fixed effects	(-)	**	Insignificant		(-)	***
	Between	(+)	***	Insignificant		(-)	***
	First differences	(-)	***	Insignificant		Insignificant	

Refer to Tables 39 to 41 in the Appendix for full results.

5.1. AVERAGE TRADE SIZE AND SHORT-TERM VOLATILITY

Tables 9 and 10 reveal that both the fixed effects and first differences models indicate a significant inverse relationship for all three Panels, whereas surprisingly, the results of the "between model" lead to the opposite conclusion, a strong positive relationship. The "between" estimator ignores the time variation in the data, whereas the fixed effects estimator accounts for both individual effects and time effects. As such, in the context of this study, the results of the fixed effects model are probably more accurate than those of the "between" model. In sum, it can be concluded that AT activity is positively associated with short-term volatility.

5.2. ODD-LOT VOLUME RATIO AND SHORT-TERM VOLATILITY

Regarding the results of the regression of the standard deviation of arithmetic returns on the odd-lot volume ratio, Table 9 shows that none of the tests found any significant relationship. In Table 10 however, the first differences estimator specified a strong positive correlation between the AT

proxy and the standard deviation of logarithm returns for Panels A and B, meaning that AT is associated with an increase in short-term volatility, in accordance with the previous finding.

5.3. TRADE-TO-ORDER VOLUME RATIO AND SHORT-TERM VOLATILITY

Table 9 reveals that the trade-to-order volume ratio exhibits a strong negative correlation with the standard deviation of arithmetic returns. The "between" model however, could not establish any significant relationship. Table 10 presents similar results. The fixed effects model concludes a strong negative relationship between the AT proxy and the standard deviation of logarithm returns for all three Panels. The "between" model confirms this for Panel C and the first differences for Panels A and B. As mentioned previously, ATs are more likely to have a lower trade-to-order ratio than other traders because they submit a large amount of orders to the market but only a small portion of these orders are filled and result in trades. As such, the results suggest that a high level of AT activity is associated with an increase in short-term volatility. This finding is consistent with Zhang (2010), Biais, Foucault and Moinas (2011), Kirilenko, Kyle, Samadi and Tuzun (2017) and Boehmer, Fong and Wu (2012) as outlined in Table 16.

To conclude, the results presented in this chapter suggest that AT activity is positively related to short-term volatility. Indeed, both the average trade size and trade-to-order volume ratio variables, which are inversely related to AT activity, are also inversely related to the volatility measures. This evidence is agreement with Zhang (2010), Biais, Foucault and Moinas (2011), Kirilenko, Kyle, Samadi and Tuzun (2017) and Boehmer, Fong and Wu (2012). The speed and extent to which AT can react to an event that affects the market may result in an increase in short-term volatility. This may not be an undesirable outcome, as long as short-term volatility is a reflection of newly available information and not the product of irrational trading behaviour such as during the May 2010 flash crash (Kirilenko et al., 2017).

CHAPTER 5: CONCLUSION

The colocation statistics for August 2020 published by the JSE revealed that HFTs were responsible for more than 70% of the total trade volume for the month. Notwithstanding the dominance of HFT in the stock market, academic studies on the topic are scarce. This study aims to address this research gap by investigating the relationship between AT and two measures of market quality, namely, liquidity and short-term volatility.

The research methodology can be broadly divided into two phases. The first phase consists in developing a set of proxies to capture the level of AT activity for all JSE listed stocks from 1 January to 30 June 2019. The three AT proxies are the average trade size, odd-lot volume ratio and trade-to-order volume ratio. These metrics are calculated for each stock and each trading day from a raw sample of intra-day trade and quote data. The second phase involves explaining the variations in market quality across the stocks during the sample period. The data has a cross-sectional and a time series component and therefore, panel regression models are used to determine the relationship between AT and market quality. In the regressions, the dependent variables are the two measures of market quality, namely liquidity and short-term volatility and the explanatory variables are the AT proxies.

The proxies selected in this study have been widely accepted in prior literature as suitable measures of AT activity. The first two proxies are the average trade size and odd-lot volume ratio. The majority of ATs split large orders into smaller ones to conceal their trading intentions from other market participants. Another subset of ATs place a series of small orders to detect large hidden orders submitted by institutional investors or to simply mislead other ATs. As such, low average trade sizes and high odd-lot volume ratios are associated with AT activity. The third proxy is the trade-to-order volume ratio. ATs tend to modify of cancel many of their orders before these get filled. Consequently, ATs have lower trade-to-order volume ratios than human traders.

Panel regressions were estimated to test the relation between AT and market liquidity, which is measured by the daily median quoted and effective bid-ask spreads. The findings differed depending on the AT proxy selected. The fixed effects estimator indicated a positive and

statistically significant relationship between average trade size and the bid-ask spreads after controlling for market capitalisation. This is consistent with the view that ATs increase market liquidity as argued in prior literature. However, the study reached the opposite conclusion when incorporating the odd-lot volume ratio and trade-to-order volume ratio variables in the equations. It is possible that the proxies capture the activity of different types of ATs, market-makers in the former case, and opportunistic traders in the latter.

Furthermore, the study showed that AT is positively correlated with short-term volatility after controlling for market capitalisation. Both the average trade size and trade-to-order volume ratio, which are inversely related to AT, are also inversely related to the measures of short-term volatility. This is consistent with prior literature. One caveat is that a high level of short-term volatility does not necessarily imply poor market quality, it could simply be the result of ATs reacting to newly available information and updating prices accordingly.

With reference to the debate regarding the net effect of AT and the merits of its overall influence on the markets, this study has led to the conclusion that AT is indeed beneficial. The value of this study is persuasive rather than definitive as it is among the first study of its kind in South Africa. As such, certainty as to the impact of AT on the JSE can only be gauged with further analysis. While not free of valid criticisms and potential to cause harm, AT is the inevitable consequence of technological innovation and market fragmentation. In light of its benefits, caution must be adopted in mitigating its potential hazards, especially in relation to volatility.

REFERENCE LIST

Banks, E. (2014). *Dark Pools: Off-Exchange Liquidity in an Era of High Frequency, Program, and Algorithmic Trading.* New York, NY: Springer.

Biais, B., & Foucault, T. (2014). HFT and Market Quality. *Bankers, Markets & Investors, 128*(1), 5-19.

Biais, B., & Woolley, P. (2011). High Frequency Trading. *Working paper*. Toulouse School of Economics and London School of Economics.

Biais, B., Foucault, T., & Moinas, S. (2011). Equilibrium High Frequency Trading. *In Proceedings from the fifth annual Paul Woolley Centre conference.* London School of Economics.

Boehmer, E., Fong, K., & Wu, J. (2012). International Evidence on Algorithmic Trading. *Working Paper*. EDHEC Business School.

Bongaerts, D., & Van Achter, M. (2016). High-Frequency Trading and Market Stability. *Available at SSRN 2698702*.

Brogaard, J. (2010). High Frequency Trading and its Impact on Market Quality. *Working Paper. Northwestern University.*

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-Frequency Trading and Price Discovery. *The Review of Financial Studies, 27*(8), 2267-2306.

Chaboud, A., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. *The Journal of Finance, 69*(5), 2045-2084.

Chakrabarty, B., Moulton, P., & Wang, X. (2018). Attention Effects in a High-Frequency World. Cornell University. Available at SSRN 2634621.

Chung, K., Van Ness, B., & Van Ness, R. (2004). Trading Costs and Quote Clustering on the NYSE and NASDAQ After Decimalization. *Journal of Financial Research, 27* (3), 309-328.

Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EC.

Fama. E., (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*(2), 383-417.

Foucault, T., & Menkveld, A. (2008). Competition for Order Flow and Smart Order Routing Systems. *Journal of Finance, 63*(1), 119-158.

Goldstein, M., Kumar, P., & Graves, F. (2014). Computerized and High-Frequency Trading. *Financial Review, 49*(2), 177-202.

Hagströmer, B., & Nordén, L. (2013). The Diversity of High-Frequency Traders. *Journal of Financial Markets, 16*(4), 741-770.

Hasbrouck, J., & Saar G. (2012). Low-Latency Trading. *Johnson School Research Paper Series, 35* - 2010, AFA 2012 Chicago Meetings Paper. Available at: SSRN 1695460.

Hasbrouck, J. (1991a). Measuring the Information Content of Stock Trades. *Journal of Finance, 46*(1), 179-207.

Hasbrouck, J. (1991b). The Summary Informativeness of Stock Trades: An Econometric Analysis. *The Review of Financial Studies, 4*(3), 571-595.

Hasbrouck, J. (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. *Journal of Finance, 50*(4), 1175-1199.

Hausman, J.A. (1978). Specification Tests in Econometrics. *Econometrica: Journal of the Econometric Society, 1251*-1271.

Hendershott, T., & Riordan, R. (2009). Algorithmic Trading and Information. *Working Paper.* University of California, Berkeley.

Hendershott, T., Jones, C., & Menkveld, A. (2011). Does Algorithmic Trading Improve Liquidity? *Journal of Finance, 66*(1), 1-33.

Hirschey, N. (2018). Do High-Frequency Traders Anticipate Buying and Selling Pressure? *Working Paper. Available at SSRN 2238516*.

Hoffman, P. (2013). A Dynamic Limit Order Market With Fast And Slow Traders. *Journal of Financial Economics, 113(1), 156-169*.

Honda, Y. (1985). Testing the error components model with non-normal disturbances. *The Review of Economic Studies, 52(4), 681-690*.

Jiang, G., Lo, I., & Valente G. (2014). High-Frequency Trading Around Macroeconomic News Announcements: Evidence from the US Treasury Market. *Working Paper.* Bank of Canada.

Johnson, B. (2010). *Algorithmic Trading & DMA: An Introduction to Direct Access Trading Strategies* (Vol. 200). London: 4Myeloma Press.

JSE Colocation Statistics (2019), Retrieved from: https://www.jse.co.za/services/technologies/jse-colocation

Kim, K. (2010). *Electronic and Algorithmic Trading Technology: the Complete Guide.* Boston, Massachusetts: Academic Press.

King, B., & Barnes, T. (2014, December, 18) High Frequency Trading. *Norton Rose Fulbright.* Retrieved from: https://www.financialinstitutionslegalsnapshot.com/2014/12/high-frequency-trading/

Kirilenko, A., Kyle, A., Samadi, M., & Tuzun, T. (2017). The Flash Crash: High-Frequency Trading in an Electronic Market . *Journal of Finance, 72*(3), 967-998.

Leinweber, D. (2009). *Nerds on Wall Street: Math, Machines and Wired Markets.* Hoboken, New Jersey: John Wiley and Sons.

Lewis, M. (2014). *Flash Boys: A Wall Street Revolt.* New York, NY: Norton & Company.

Lippman, S., & McCall, J. (1986). An Operational Measure of Liquidity. *The American Economic Review, 76*(1), 43-55.

Malinova, K., Park, A., & Riordan, R. (2013). Do Retail Traders Suffer from High Frequency Traders ? *Working Paper.* University of Toronto. *Available at SSRN 2183806*.

McGowan, M. (2010). The Rise of Computerized High Frequency Trading: Use and Controversy. *Duke Law & Technology Review, 9*(1), 1-25.

McNamara, S. (2016). The Law and Ethics of High-Frequency Trading. *Minnesota Journal of Law, Science & Technology, 17*(1), 71.

Menkveld, A. (2013). High Frequency Trading and the New Market Makers. *Journal of financial Markets, 16*(4), 712-740.

Meulbroek, L. (1992). An Empirical Analysis of Illegal Insider Trading. *Journal of Finance, 47*(5), 1661-1699.

Meyer, G., Bullock, N., & Rennison, J. (2018, January 1) How high-frequency trading hit a speed bump. *The Financial Times.* Retrieved from: https://www.ft.com/content/d81f96ea-d43c-11e7-a303-9060cb1e5f44

Morse, D. (1981). Price and Trading Volume Reaction Surrounding Earnings Announcements: A Closer Examination. *Journal of Accounting Research*, *19*(2), 374-383.

O'Hara, M. (2015). High Frequency Market Microstructure. *Journal of Financial Economics, 116*(2), 257-270.

O'Hara, M., Yao, C., & Ye, M. (2011). What's Not There: The Odd-Lot Bias in Market Data. *Working paper*. Cornell University and University of Illinois.

Peng, L. (2005). Learning with Information Capacity Constraints. *Journal of Financial and Quantitative Analysis, 40*(2), 307-329.

Popper, N. (2012, October 14) High Speed Trading No Longer Hurtling Forward. *The New York Times.* Retrieved from: https://www.nytimes.com/2012/10/15/business/with-profits-dropping-high-speed-trading-cools-down.html

Regulation ATS, Exchange Act Release No. 34-40760 (2000).

Regulation NMS, Exchange Act Release No. 34-51808, 70 Fed. Reg. 37,496, 37,532 n.300 (2005).

Securities Exchange Act Release No. 34-61358, 75 FR 3594, 3606 (2010).

Stoll, H. R. (2014). High Speed Equities Trading: 1993 - 2012. *Asia-Pacific Journal of Financial Studies*, *43(6)*, 767-797.

Sukumar, N. (2011, January 24) High-Frequency Trading is 77% of U.K. Market, Tabb Group Says. *Bloomberg.* Retrieved from: https://www.bloomberg.com/news/articles/2011-01-24/high-frequency-trading-is-77-of-u-k-market-tabb-group-says

Sussman, A., Tabb, L., & Iati, R. (2009, September 2) US Equity High Frequency Trading: Strategies, Sizing and Market Structure. *Tabb Group.* Retrieved from: https://research.tabbgroup.com/report/v07-023-us-equity-high-frequency-trading-strategies-sizing-and-market-structure

Tabb, L. (2014, March 31). No, Michael Lewis, the US Equities Market is not Rigged. *Tabb Forum.* Retrieved November 11, 2019, from http://tabbforum.com/opinions/no-michael-lewis-the-us-equities-market-is-not-rigged?print_preview=true&single= true.)

Warpenaar, K. (2019). The Impact of the 4th Industrial Revolution on the South African Financial Services Market. Centre of Excellence in Financial Services. Retrieved from: https://www.coefs.org.za/.

Weller, B. (2018). Efficient Prices at any Cost: Does Algorithmic Trading Deter Information Acquisition? *Review of Financial Studies, 31*(6), 2184 - 2226.

Zhang, F. (2010). High-Frequency Trading, Stock Volatility, and Price Discovery. *Working Paper*. Yale University. *Available at SSRN 1691679.*

Zhang, F., & Powell, S. (2011). The Impact of High-Frequency Trading on Markets. *CFA Magazine, 22*(2), 10-11.

Zhou, W., & Olivari, N. (2013, August 23) Exclusive : EBS Take New Step to Rein in High Frequency Traders. *Reuters.*

Retrieved from: https://www.reuters.com/article/us-markets-forex-hft/exclusive-ebs-take-new-step-to-rein-in-high-frequency-traders-idUSBRE97M0YJ20130823

2018 Financial Markets Review by the National Treasury, the South African Reserve Bank & the FSCA.
Retrieved from: http://www.treasury.gov.za/publications/other/2018_FMR_07.pdf

APPENDIX

Table 11: Abbreviations used in this document.

Abbreviation	**Meaning**
AT	Algorithmic trading
DJIA	Dow Jones Industrial Average
FSCA	Financial Sector Conduct Authority
GETCO	Global Electronic Trading Co
HFT	High-frequency trading
IIROC	Investment Industry Regulatory Organization of Canada
JSE	Johannesburg Stock Exchange
MiFID	Markets in Financial Instruments Directive
NASDAQ	National Association of Securities Dealers Automated Quotations
NBBO	National Best Bid and Offer
NYSE	New York Stock Exchange
Reg ATS	Regulation Alternative Trading Systems
Reg NMS	Regulation National Market System
SEC	Securities Exchange Commission
SIP	Securities Information Processor

Table 12: Automated strategies and potential impact on market quality.

Traders	**Strategies**	**Description**	**HFT and Market Quality (Hypotheses)**
HFTs	Market-making	HFT market-makers send limit orders that specify the maximum (minimum) price at which they are willing to buy (sell) an asset. Their profit is the bid-ask spread on the trade.	HFTs improve liquidity and price discovery. They pose a threat to market stability.
HFTs	Opportunistic (Arbitrage)	HFTs exploit short-lived opportunities by purchasing and selling an asset with different prices in two or more markets at the same time.	HFTs speed up the price discovery process.
HFTs	Opportunistic (Momentum)	HFTs identify patterns in limit order books and chase short-term trends.	HFTs amplify volatility.
HFTs	Opportunistic (Manipulative)	HFTs mislead other market participants (electronic front-running, quote stuffing, spoofing, layering)/	HFTs impose higher trading costs on other market players.
Non-HFTs	Market Sentiment	ATs extract subjective information from websites and social media platforms.	ATs speed up the price discovery process but amplify volatility.

Table 13: List of conditional codes.

AC	Trade at closing price
AT	Automatic Trade
AU	Intra-day Auction
BL	Book Build
BT	Block Trade
C	Bargain Trade
CA	Closing Auction
CF	Corporate Finance Trade
CP	Converted Price
IP	Indicative Price
L	Late Reported Trade
LC	Delayed Cancellation (off-book)
LT	Late Trade
MID	MID
O	Overnight
OA	Opening Auction
OC	Official Close
OP	Off Order Book Principal Trade
PC	Delayed Cancellation (on-book)
PD	Previous Day Trade
PF	Portfolio Trade
SLT	Synthetic Leg Trade
UC	Uncrossing Trade-Official Open for Sets Stocks
VA	Volatility Auction
VT	EOD Volume Auction Trade
XO	Crossed Order Trade

Table 14: List of Bloomberg fields.

(1)	Dates
(2)	Type
(3)	Price
(4)	Size
(5)	Cond Code
(6)	QRM
(7)	ExchCode
(8)	BrkrCodes
(9)	RPSCodes
(10)	RPTParty
(11)	RPTContra
(12)	ActionCodes
(13)	IndicatorCodes
(14)	TradeTime
(15)	Yield
(16)	Spread
(17)	BICMICCodes
(18)	UpfrontPrice

Table 15: Measures of market quality.

Measures of market quality	**Metric**	**Relationship**	**Description**
Liquidity	Quoted bid-ask spread	Negative	The amount by which the median ask price exceeds the median bid price for a stock, standardized by the quote midpoint.
	Effective bid-ask spread	Negative	The difference between the median transaction price and the median mid-quote price, standardized by the quote midpoint.
Short-term volatility	Standard deviation (arithmetic)	Positive	The standard deviation of the arithmetic returns, which are calculated by taking the difference between the price at which the stock was traded and the price at which it was previously traded divided by the price at which it was previously traded.
	Standard deviation (logarithm)	Positive	The standard deviation of the logarithm returns, which are the log of the price at which the stock is traded divided by the price at which it was previously traded.

Table 16: Related literature.

This table summarizes the empirical findings of prior studies on the impact of AT on various measures of market quality.

Impact	Liquidity	Price discovery	Volatility
Positive	Hendershott and Riordan (2009) Hendershott, Jones and Menkveld (2011) Kirilenko, Kyle, Samadi and Tuzun (2017) Boehmer, Fong and Wu (2012) Hagströmer and Norden (2013) Hasbrouck and Saar (2012)	Hendershott and Riordan (2009) Brogaard (2010) Hendershott, Jones and Menkveld (2011) O'Hara (2011) Biais, Foucault and Moinas (2011) Boehmer, Fong and Wu (2012) Brogaard, Hendershott and Riordan (2014) Jiang, Lo and Valente (2014) Chaboud, Chiquoine, Hjalmarsson and Vega (2014) Chakrabarty, Moulton and Wang (2018)	Brogaard (2010) Hendershott and Riordan (2009) Hagströmer and Norden (2013) Hasbrouck and Saar (2012) Chaboud, Chiquoine, Hjalmarsson and Vega (2014)
Negative	Brogaard (2010)	Zhang (2010) Biais and Foucault (2014) Weller (2018)	Zhang (2010) Biais, Foucault and Moinas (2011) Kirilenko, Kyle, Samadi and Tuzun (2017) Boehmer, Fong and Wu (2012)

Table 17: Panel A - summary statistics across all stocks per day.

Panel A: n = 136, T = 119, N = 16,184

Date – min: 2019-01-02, max: 2019-06-28, median: 2019-03-27

Variable	Mean	Std. deviation	Min	1st Quarter	Median	3rd Quarter	Max
Trades	1,742	2,111	6	195	845	2,747	30,834
Orders	16,131	20,867	18	1,431	7,980	24,626	345,150
Odd lot trades	571	902	0	30	169	726	12,179
Order volume	606,862,361	7,370,929,238	20,553	4,274,919	11,871,575	27,449,779	308,234,000,000
Trade volume	1,415,676	7,065,733	354	180,352	528,319	1,305,246	561,652,563
Odd lot trade volume	25,876	41,401	0	1,158	6,803	33,290	745,005
Trade value (Rands)	79,360,870	143,653,067	9,628	4,506,646	25,211,125	96,296,537	3,851,060,000
Standard deviation (arithmetic)	0.003204	0.019041	0.000000	0.000422	0.000751	0.002024	0.535987
Standard deviation (logarithm)	0.003105	0.017643	0.000000	0.000422	0.000751	0.002023	0.398650
Quoted bid ask spread*	0.005081	0.025079	-0.400000	0.000940	0.001820	0.004413	0.400000
Effective bid ask spread*	0.000810	0.012163	-0.200000	-0.000210	0.000000	0.000300	0.200000
Average trade size	4,582	86,256	16	252	624	1,583	9,579,936
Odd lot volume ratio	0.060074	0.112953	0.000000	0.003040	0.013918	0.064340	0.940955
Trade-to-order volume ratio	0.059704	0.050123	0.000007	0.031396	0.052175	0.073567	0.838612
Market capitalisation	62,684 M	148,470 M	384 M	7,661 M	22,395 M	51,858 M	1,481,590 M

*Note that the quoted and effective bid-ask spreads are standardized by the quote midpoints.

Table 18: Panel B - summary statistics across all stocks per day.

Panel B: n = 192, T = 100-119, N = 22,450

Date – min: 2019-01-02, max: 2019-06-28, median: 2019-03-27

Variable	**Mean**	**Std. deviation**	**Min**	**1st Quarter**	**Median**	**3rd Quarter**	**Max**
Trades	1,359	2,075	6	48	343	1,981	30,834
Orders	13,523	22,078	15	543	3,254	18,238	464,362
Odd lot trades	475	1,020	0	8	63	451	18,655
Order volume	458,409,863	6,270,879,994	14,092	2,199,695	9,179,951	24,202,464	308,234,000,000
Trade volume	1,101,385	6,048,383	49	73,505	333,561	986,791	561,652,563
Odd lot trade volume	21,310	44,900	0	300	2,423	19,576	745,005
Trade value (Rands)	73,491,680	224,511,980	285	1361,172	9,389,515	66,779,607	10,960,500,000
Standard deviation (arithmetic)	0.004153	0.016662	0.000000	0.000495	0.001265	0.003726	0.535987
Standard deviation (logarithm)	0.004076	0.015499	0.000000	0.000495	0.001265	0.003726	0.398650
Quoted bid ask spread*	0.007615	0.023861	-0.400000	0.001135	0.003001	0.007860	0.963995
Effective bid ask spread*	0.000638	0.012547	-0.200000	-0.000310	0.000012	0.000516	0.916168
Average Trade Size	4,988	76,071	7	293	794	2,176	9,579,936
Odd lot volume ratio	0.054573	0.112432	0.000000	0.001983	0.010064	0.050621	1.000000
Trade-to-order volume ratio	0.063209	0.064032	0.000001	0.025336	0.051251	0.078627	0.990223
Market capitalisation	66,196 M	214,652 M	107 M	3,782 M	10,847 M	40,964 M	2,193,600 M

Quoted bid ask spread – 8 NAs

Effective bid ask spread – 8 NAs

Table 19: Panel C - summary statistics across all stocks per day.

Panel C: n = 389, T = 1-119, N = 28,654

Date – min: 2019-01-02, max: 2019-06-28, median: 2019-03-27

Variable	Mean	Std. deviation	Min	1st Quarter	Median	3rd Quarter	Max
Trades	1,082	1,935	6	18	129	1,300	30,834
Orders	10,939	20,385	12	191	1,393	12,639	464,362
Odd lot trades	377	927	0	3	21	273	18,655
Order volume	363,043,438	5,553,698,846	600	1,202,049	6,575,182	20,899,613	308,234,000,000
Trade volume	923,192	5,421,481	25	40,753	217,847	783,217	561,652,563
Odd lot trade volume	16,893	40,811	0	116	770	11,155	745,005
Trade value (Rands)	59,231,813	208,429,521	30	461,738	3,801,136	43,062,545	10,960,500,000
Standard deviation (arithmetic)	0.007372	0.026510	0.000000	0.000567	0.001985	0.005984	1.501295
Standard deviation (logarithm)	0.007181	0.021922	0.000000	0.000567	0.001985	0.005986	0.880827
Quoted bid ask spread*	0.015152	0.042053	-0.400000	0.001426	0.004874	0.013120	1.612903
Effective bid ask spread*	0.000995	0.018853	-0.806362	-0.000443	0.000027	0.001002	0.916168
Average trade size	6,881	85,525	4	341	993	3,027	9,579,936
Odd lot volume ratio	0.049939	0.115814	0.000000	0.001068	0.007241	0.039691	1
Trade to order volume ratio	0.071289	0.077972	0.000000	0.023563	0.053055	0.089621	0.990222
Market capitalisation	52,448 M	191,815 M	839,035	1,515 M	6,668 M	31,603 M	2,193,600 M

Quoted bid ask spread – 104 NAs

Effective bid ask spread – 123 NAs

Table 20: Unit root tests.

Panels	Unit-Root Test	Average trade size	Odd-lot volume ratio	Trade-to-order volume ratio
A (Balanced)	Levin-Lin-Chu	z = -135.49, p-value < 2.2e-16 alternative hypothesis: stationarity	z = -140.65, p-value < 2.2e-16 alternative hypothesis: stationarity	z = -124.38, p-value < 2.2e-16 alternative hypothesis: stationarity
B	Maddala-Wu	chisq = 16168, df = 384, p-value < 2.2e-16 alternative hypothesis: stationarity	chisq = 15842, df = 384, p-value < 2.2e-16 alternative hypothesis: stationarity	chisq = 14500, df = 384, p-value < 2.2e-16 alternative hypothesis: stationarity
C	Maddala-Wu	chisq = 21080, df = 576, p-value < 2.2e-16 alternative hypothesis: stationarity	chisq = 20050, df = 576, p-value < 2.2e-16 alternative hypothesis: stationarity	chisq = 18053, df = 576, p-value < 2.2e-16 alternative hypothesis: stationarity

Additional note: the shares with less than 20 observations over the sample period had to be removed when running the unit root test on Panel C.

Table 21: Panel A - Lagrange Multiplier Test - (Honda).

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	normal = 374.51, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 492.33, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 490.52, p-value < 2.2e-16 alternative hypothesis: significant effects
Effective bid-ask spread, standardised by the quote midpoint	normal = 344.16, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 446.6, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 440.63, p-value < 2.2e-16 alternative hypothesis: significant effects
Standard deviation (arithmetic)	normal = 629.67, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 755.89, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 751.52, p-value < 2.2e-16 alternative hypothesis: significant effects
Standard deviation (logarithm)	normal = 694.08, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 825.09, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 820.95, p-value < 2.2e-16 alternative hypothesis: significant effects

Table 22: Panel B - Lagrange Multiplier Test - (Honda).

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	normal = 432.01, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 540.28, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 542.44, p-value < 2.2e-16 alternative hypothesis: significant effects
Effective bid-ask spread, standardised by the quote midpoint	normal = 296.21, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 378.92, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 375.85, p-value < 2.2e-16 alternative hypothesis: significant effects
Standard deviation (arithmetic)	normal = 735.09, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 869.86, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 870.76, p-value < 2.2e-16 alternative hypothesis: significant effects
Standard deviation (logarithm)	normal = 805.12, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 944.62, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 945.87, p-value < 2.2e-16 alternative hypothesis: significant effects

Table 23: Panel C - Lagrange Multiplier Test - (Honda).

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	normal = 269.84, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 303.48, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 297.84, p-value < 2.2e-16 alternative hypothesis: significant effects
Effective bid-ask spread, standardised by the quote midpoint	normal = 143.68, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 167.75, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 166.69, p-value < 2.2e-16 alternative hypothesis: significant effects
Standard deviation (arithmetic)	normal = 340.6, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 392.07, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 392.44, p-value < 2.2e-16 alternative hypothesis: significant effects
Standard deviation (logarithm)	normal = 476.57, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 542.05, p-value < 2.2e-16 alternative hypothesis: significant effects	normal = 542.54, p-value < 2.2e-16 alternative hypothesis: significant effects

Table 24: Panel A - F test for individual effects.

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	F = 97.78, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects	F = 122.93, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects	F = 122.53, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects
Effective bid-ask spread, standardised by the quote midpoint	F = 85.296, p-value < 2.2e-16 df1 = 135, df2 = 16,046 alternative hypothesis: significant effects	F = 101.96, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects	F = 100.94, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects
Standard deviation (arithmetic)	F = 367.68, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects	F = 411.24, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects	F = 408.93, p-value < 2.2e-16 df1 = 135, df2 = 16,046, alternative hypothesis: significant effects
Standard deviation (logarithm)	F = 589.23, p-value < 2.2e-16 df1 = 135, df2 = 16,046,	F = 654, p-value < 2.2e-16 df1 = 135, df2 = 16,046,	F = 650.48, p-value < 2.2e-16 df1 = 135, df2 = 16,046,

	alternative hypothesis: significant effects	alternative hypothesis: significant effects	alternative hypothesis: significant effects

Table 25: Panel B - F test for individual effects.

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	F = 88.263, p-value < 2.2e-16 df1 = 191, df2 = 22,248, alternative hypothesis: significant effects	F = 106.01, p-value < 2.2e-16 df1 = 191, df2 = 22,248, alternative hypothesis: significant effects	F = 106.73, p-value < 2.2e-16 df1 = 191, df2 = 22,248, alternative hypothesis: significant effects
Effective bid-ask spread, standardised by the quote midpoint	F = 49.907, p-value < 2.2e-16 df1 = 191, df2 = 22,248, alternative hypothesis: significant effects	F = 58.306, p-value < 2.2e-16 df1 = 191, df2 = 22,248, alternative hypothesis: significant effects	F = 57.851, p-value < 2.2e-16 df1 = 191, df2 = 22,248, alternative hypothesis: significant effects
Standard deviation (arithmetic)	F = 322.83, p-value < 2.2e-16 df1 = 191, df2 = 22,256, alternative hypothesis: significant effects	F = 357.62, p-value < 2.2e-16 df1 = 191, df2 = 22,256, alternative hypothesis: significant effects	F = 358.85, p-value < 2.2e-16 df1 = 191, df2 = 22,256, alternative hypothesis: significant effects
Standard deviation (logarithm)	F = 477.71, p-value < 2.2e-16	F = 526.05, p-value < 2.2e-16	F = 528.53, p-value < 2.2e-16

	df1 = 191, df2 = 22,256, alternative hypothesis: significant effects	df1 = 191, df2 = 22,256, alternative hypothesis: significant effects	df1 = 191, df2 = 22,256, alternative hypothesis: significant effects

Table 26: Panel C - F test for individual effects.

Independent **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	F = 93.773, p-value < 2.2e-16 df1 = 386, df2 = 28,161, alternative hypothesis: significant effects	F = 97.593, p-value < 2.2e-16 df1 = 386, df2 = 28,161, alternative hypothesis: significant effects	F = 97.15, p-value < 2.2e-16 df1 = 386, df2 = 28,161, alternative hypothesis: significant effects
Effective bid-ask spread, standardised by the quote midpoint	F = 19.304, p-value < 2.2e-16 df1 = 386, df2 = 28,142, alternative hypothesis: significant effects	F = 20.273, p-value < 2.2e-16 df1 = 386, df2 = 28,142, alternative hypothesis: significant effects	F = 20.222, p-value < 2.2e-16 df1 = 386, df2 = 28,142, alternative hypothesis: significant effects
Standard deviation (arithmetic)	F = 119.12, p-value < 2.2e-16 df1 = 388, df2 = 28,263, alternative hypothesis: significant effects	F = 123.77, p-value < 2.2e-16 df1 = 388, df2 = 28,263, alternative hypothesis: significant effects	F = 124.23, p-value < 2.2e-16 df1 = 388, df2 = 28,263, alternative hypothesis: significant effects

Standard deviation (logarithm)	F = 130.11, p-value < 2.2e-16 df1 = 388, df2 = 28,263, alternative hypothesis: significant effects	F = 136.7, p-value < 2.2e-16 df1 = 388, df2 = 28,263, alternative hypothesis: significant effects	F = 137.39, p-value < 2.2e-16 df1 = 388, df2 = 28,263, alternative hypothesis: significant effects

Table 27: Panel A - Hausman Test.

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	chisq = 832.47, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 1.3304, df = 2, p-value = 0.5142 alternative hypothesis: one model is inconsistent	chisq = 2.2444, df = 2, p-value = 0.3256 alternative hypothesis: one model is inconsistent
Effective bid-ask spread, standardised by the quote midpoint	chisq = 3,636.6, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 0.45727, df = 2, p-value = 0.7956 alternative hypothesis: one model is inconsistent	chisq = 5.4609, df = 2, p-value = 0.06519 alternative hypothesis: one model is inconsistent
Standard deviation (arithmetic)	chisq = 552.1, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 1.3766, df = 2, p-value = 0.5024 alternative hypothesis: one model is inconsistent	chisq = 4.4202, df = 2, p-value = 0.1097 alternative hypothesis: one model is inconsistent

Standard deviation (logarithm)	chisq = 372.6, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 1.1326, df = 2, p-value = 0.5676 alternative hypothesis: one model is inconsistent	chisq = 3.8916, df = 2, p-value = 0.1429 alternative hypothesis: one model is inconsistent

Table 28: Panel B - Hausman Test.

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	chisq = 256.21, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 3.4207, df = 2, p-value = 0.1808 alternative hypothesis: one model is inconsistent	chisq = 2.6903, df = 2, p-value = 0.2605 alternative hypothesis: one model is inconsistent
Effective bid-ask spread, standardised by the quote midpoint	chisq = 1150.9, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 0.90431, df = 2, p-value = 0.6363 alternative hypothesis: one model is inconsistent	chisq = 2.9699, df = 2, p-value = 0.2265 alternative hypothesis: one model is inconsistent
Standard deviation (arithmetic)	chisq = 226.46, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 2.7266, df = 2, p-value = 0.2558 alternative hypothesis: one model is inconsistent	chisq = 2.0509, df = 2, p-value = 0.3586 alternative hypothesis: one model is inconsistent

Standard deviation (logarithm)	chisq = 148.49, df = 2, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent	chisq = 2.0409, df = 2, p-value = 0.3604 alternative hypothesis: one model is inconsistent	chisq = 1.0142, df = 2, p-value = 0.6022 alternative hypothesis: one model is inconsistent

Table 29: Panel C - Hausman Test.

Independent / **Dependent**	Average trade size, market capitalisation	Odd-lot volume ratio, market capitalisation	Trade-to-order volume ratio, market capitalisation
Quoted bid-ask spread, standardised by the quote midpoint	chisq = 22.394, df = 2, p-value = 1.371e-05 alternative hypothesis: one model is inconsistent	chisq = 9.4127, df = 2, p-value = 0.009038 alternative hypothesis: one model is inconsistent	chisq = 1.278, df = 2, p-value = 0.5278 alternative hypothesis: one model is inconsistent
Effective bid-ask spread, standardised by the quote midpoint	chisq = 43.76, df = 2, p-value = 3.145e-10 alternative hypothesis: one model is inconsistent	chisq = 0.42096, df = 2, p-value = 0.8102 alternative hypothesis: one model is inconsistent	chisq = 52.323, df = 2, p-value = 4.348e-12 alternative hypothesis: one model is inconsistent
Standard deviation (arithmetic)	chisq = 9.7616, df = 2, p-value = 0.007591 alternative hypothesis: one model is inconsistent	chisq = 11.079, df = 2, p-value = 0.003928 alternative hypothesis: one model is inconsistent	chisq = 13.834, df = 2, p-value = 0.0009906 alternative hypothesis: one model is inconsistent

Standard deviation (logarithm)	chisq = 2.6313, df = 2, p-value = 0.2683 alternative hypothesis: one model is inconsistent	chisq = 4.2694, df = 2, p-value = 0.1183 alternative hypothesis: one model is inconsistent	chisq = 9.3618, df = 2, p-value = 0.009271 alternative hypothesis: one model is inconsistent

Table 30: Equation: quoted bid-ask spread ~ average trade size + market capitalisation.

Panel A

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	0.00	0.00 – 0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	0.00	-0.00 – 0.00	0.387
Market capitalisation	-0.00	-0.02 – 0.02	0.941	-0.01	-0.01 – -0.00	**0.006**	-0.01	-0.12 – 0.11	0.921
(Intercept)				0.00	0.00 – 0.00	**<0.001**	0.00	-0.00 – 0.00	0.950
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.014 / 0.005			0.968 / 0.968			0.000 / -0.000		

Panel B

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	0.00	0.00 – 0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	0.00	-0.00 – 0.00	0.309
Market capitalisation	-0.00	-0.01 – 0.01	0.870	-0.01	-0.01 – -0.00	**0.012**	-0.00	-0.07 – 0.07	0.950

(Intercept)			0.01	0.00 – 0.01	**<0.001**	-0.00	-0.00 – 0.00	0.941
Observations	22450		192			22258		
R^2 / R^2 adjusted	0.010 / 0.001		0.832 / 0.830			0.000 / -0.000		

Panel C

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	0.00	0.00 – 0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	0.00	-0.00 – 0.00	0.603
Market capitalisation	-0.00	-0.02 – 0.02	0.939	-0.05	-0.10 – 0.01	0.097	-0.00	-0.12 – 0.12	0.987
(Intercept)				0.04	0.03 – 0.04	**<0.001**	-0.00	-0.00 – 0.00	0.630
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.002 / -0.012			0.091 / 0.086			0.000 / -0.000		

Table 31: Equation: quoted bid-ask spread ~ odd-lot volume ratio + market capitalisation.

Panel A

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*

Odd-lot volume ratio	0.00	-0.00 – 0.01	0.500	-0.01	-0.04 – 0.02	0.427	0.00	-0.00 – 0.01	0.431
Market capitalisation	-0.00	-0.02 – 0.02	0.906	-0.01	-0.03 – 0.01	0.359	-0.01	-0.12 – 0.11	0.919
(Intercept)				0.01	0.00 – 0.01	**0.001**	0.00	-0.00 – 0.00	0.948
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.009			0.013 / -0.002			0.000 / -0.000		

Panel B

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.01	0.00 – 0.01	**0.017**	-0.01	-0.04 – 0.01	0.262	0.00	0.00 – 0.01	**0.029**
Market capitalisation	-0.00	-0.01 – 0.01	0.823	-0.01	-0.02 – 0.01	0.267	-0.00	-0.07 – 0.07	0.950
(Intercept)				0.01	0.01 – 0.01	**<0.001**	-0.00	-0.00 – 0.00	0.945
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.000 / -0.008			0.021 / 0.010			0.000 / 0.000		

Panel C

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.01	0.00 – 0.01	**0.005**	-0.06	-0.13 – -0.00	**0.042**	0.00	-0.00 – 0.01	0.089
Market capitalisation	-0.00	-0.02 – 0.02	0.901	-0.04	-0.10 – 0.01	0.134	-0.00	-0.12 – 0.12	0.987

(Intercept)				0.05	0.04 – 0.05	**<0.001**	-0.00	-0.00 – 0.00	0.631
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.020 / 0.015			0.000 / 0.000		

Table 32: Equation: quoted bid-ask spread ~ trade-to-order vol. ratio + market capitalisation.

Panel A

		Fixed model			**Between model**			**First differences model**	
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.01	-0.02 – -0.00	**0.005**	-0.08	-0.19 – 0.03	0.138	-0.01	-0.01 – -0.00	**0.044**
Market capitalisation	-0.00	-0.02 – 0.02	0.917	-0.01	-0.04 – 0.01	0.165	-0.01	-0.12 – 0.10	0.916
(Intercept)				0.01	0.00 – 0.02	**0.005**	0.00	-0.00 – 0.00	0.947
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.008			0.025 / 0.010			0.000 / 0.000		

Panel B

		Fixed model			**Between model**			**First differences model**	
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.01	-0.02 – -0.01	**<0.001**	0.03	-0.03 – 0.09	0.364	-0.01	-0.02 – -0.01	**<0.001**

Market capitalisation	-0.00	-0.01 – 0.01	0.846	-0.01	-0.02 – 0.00	0.154	-0.00	-0.07 – 0.07	0.945
(Intercept)				0.01	0.00 – 0.01	**0.010**	-0.00	-0.00 – 0.00	0.950
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.002 / -0.007			0.018 / 0.008			0.001 / 0.001		

Panel C

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.02	-0.03 – -0.02	**<0.001**	0.22	0.09 – 0.36	**0.001**	-0.02	-0.02 – -0.01	**<0.001**
Market capitalisation	-0.00	-0.02 – 0.02	0.910	-0.04	-0.09 – 0.02	0.167	-0.00	-0.12 – 0.12	0.982
(Intercept)				0.02	0.01 – 0.04	**0.003**	-0.00	-0.00 – 0.00	0.636
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.003 / -0.011			0.037 / 0.032			0.002 / 0.002		

Table 33: Equation: effective bid-ask spread ~ average trade size + market capitalisation.

Panel A

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*

Average trade size	0.00	0.00 – 0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**0.005**
Market capitalisation	0.00	-0.01 – 0.01	0.955	0.00	-0.00 – 0.00	0.262	0.00	-0.06 – 0.06	0.933
(Intercept)				-0.00	-0.00 – -0.00	**<0.001**	0.00	-0.00 – 0.00	0.965
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.001 / -0.008			0.992 / 0.992			0.000 / 0.000		

Panel B

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	0.00	0.00 – 0.00	**0.003**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**0.038**
Market capitalisation	0.00	-0.01 – 0.01	0.978	0.00	-0.00 – 0.00	0.369	0.00	-0.05 – 0.05	0.979
(Intercept)				-0.00	-0.00 – -0.00	**<0.001**	-0.00	-0.00 – 0.00	0.991
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.000 / -0.008			0.914 / 0.913			0.000 / 0.000		

Panel C

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	0.00	-0.00 – 0.00	0.225	0.00	0.00 – 0.00	**0.004**	-0.00	-0.00 – 0.00	0.148

Market capitalisation	0.00	-0.01 – 0.01	0.933	-0.00	-0.02 – 0.01	0.480	0.01	-0.07 – 0.08	0.879
(Intercept)				0.00	0.00 – 0.01	**0.001**	-0.00	-0.00 – 0.00	0.772
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.023 / 0.018			0.000 / 0.000		

Table 34: Equation: effective bid-ask spread ~ odd-lot volume ratio + market capitalisation.

Panel A

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.00	-0.00 – 0.00	0.850	-0.00	-0.02 – 0.01	0.585	0.00	-0.00 – 0.00	0.844
Market capitalisation	0.00	-0.01 – 0.01	0.965	-0.00	-0.01 – 0.01	0.709	0.00	-0.06 – 0.06	0.934
(Intercept)				0.00	-0.00 – 0.00	0.182	0.00	-0.00 – 0.00	0.965
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.009			0.004 / -0.011			0.000 / -0.000		

Panel B

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*

Odd-lot volume ratio	0.00	-0.00 – 0.00	0.115	-0.00	-0.01 – 0.01	0.580	0.00	-0.00 – 0.01	0.076
Market capitalisation	-0.00	-0.01 – 0.01	0.991	-0.00	-0.01 – 0.00	0.895	0.00	-0.05 – 0.05	0.979
(Intercept)				0.00	-0.00 – 0.00	0.172	-0.00	-0.00 – 0.00	0.994
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.000 / -0.009			0.002 / -0.008			0.000 / 0.000		

Panel C

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	-0.00	-0.00 – 0.00	0.491	-0.00	-0.02 – 0.01	0.597	0.00	-0.00 – 0.00	0.706
Market capitalisation	0.00	-0.01 – 0.01	0.924	-0.01	-0.02 – 0.01	0.461	0.01	-0.07 – 0.08	0.880
(Intercept)				0.00	0.00 – 0.01	**<0.001**	-0.00	-0.00 – 0.00	0.773
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.003 / -0.003			0.000 / -0.000		

Table 35: Equation: effective bid-ask spread ~ trade-to-order vol. ratio + market capitalisation.

Panel A

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.00	-0.00 – 0.00	0.823	-0.06	-0.11 – -0.01	**0.024**	-0.00	-0.00 – 0.00	0.432
Market capitalisation	0.00	-0.01 – 0.01	0.958	-0.00	-0.01 – 0.00	0.316	0.00	-0.06 – 0.06	0.936
(Intercept)				0.00	0.00 – 0.01	**0.011**	0.00	-0.00 – 0.00	0.965
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.009			0.039 / 0.025			0.000 / -0.000		

Panel B

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.00	-0.01 – 0.00	0.054	-0.03	-0.05 – 0.00	0.058	-0.00	-0.01 – -0.00	**0.022**
Market capitalisation	0.00	-0.01 – 0.01	0.986	-0.00	-0.01 – 0.00	0.456	0.00	-0.05 – 0.05	0.981
(Intercept)				0.00	0.00 – 0.00	**0.023**	-0.00	-0.00 – 0.00	0.995
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.000 / -0.009			0.020 / 0.009			0.000 / 0.000		

Panel C

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*

Trade-to-order volume ratio	-0.00	-0.01 – 0.00	0.092	0.03	-0.00 – 0.06	0.083	-0.00	-0.01 – -0.00	**0.032**
Market capitalisation	0.00	-0.01 – 0.01	0.939	-0.00	-0.02 – 0.01	0.564	0.01	-0.07 – 0.08	0.881
(Intercept)				0.00	-0.00 – 0.01	0.267	-0.00	-0.00 – 0.00	0.774
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.010 / 0.004			0.000 / 0.000		

Table 36: Equation: std. dev. (arithmetic) ~ average trade size + market capitalisation.

Panel A

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	-0.00	-0.00 – -0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**<0.001**
Market capitalisation	0.00	-0.01 – 0.01	0.447	-0.00	-0.00 – 0.00	0.229	0.02	-0.03 – 0.06	0.446
(Intercept)				0.00	0.00 – 0.00	**<0.001**	0.00	-0.00 – 0.00	0.988
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.006 / -0.002			0.988 / 0.988			0.012 / 0.012		

Panel B

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	-0.00	-0.00 – -0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**<0.001**
Market capitalisation	0.00	-0.00 – 0.01	0.283	-0.00	-0.00 – 0.00	0.120	0.00	-0.03 – 0.03	0.969
(Intercept)				0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – 0.00	0.838
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.005 / -0.004			0.954 / 0.953			0.009 / 0.009		

Panel C

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	-0.00	-0.00 – -0.00	**0.036**	0.00	0.00 – 0.00	**0.003**	-0.00	-0.00 – -0.00	**0.004**
Market capitalisation	0.00	-0.01 – 0.01	0.575	-0.02	-0.09 – 0.04	0.441	0.00	-0.07 – 0.07	0.995
(Intercept)				0.02	0.01 – 0.03	**<0.001**	-0.00	-0.00 – 0.00	0.220
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.025 / 0.020			0.000 / 0.000		

Table 37: Equation: std. dev. (arithmetic) ~ odd-lot volume ratio + market capitalisation.

Panel A

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	-0.00	-0.00 – 0.00	0.740	-0.01	-0.04 – 0.02	0.496	0.00	-0.00 – 0.00	0.829
Market capitalisation	0.00	-0.01 – 0.02	0.436	-0.01	-0.03 – 0.01	0.561	0.02	-0.03 – 0.06	0.451
(Intercept)				0.00	0.00 – 0.01	**0.019**	0.00	-0.00 – 0.00	0.989
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.008			0.007 / -0.008			0.000 / -0.000		

Panel B

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.00	-0.00 – 0.00	0.751	-0.01	-0.03 – 0.01	0.389	0.00	-0.00 – 0.00	0.467
Market capitalisation	0.00	-0.00 – 0.01	0.287	-0.00	-0.01 – 0.01	0.557	0.00	-0.03 – 0.03	0.970
(Intercept)				0.00	0.00 – 0.01	**<0.001**	-0.00	-0.00 – 0.00	0.839
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.000 / -0.009			0.009 / -0.002			0.000 / -0.000		

Panel C

	Fixed model	Between model	First differences model

Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	-0.00	-0.00 – 0.00	0.929	-0.04	-0.11 – 0.03	0.320	-0.00	-0.00 – 0.00	0.367
Market capitalisation	0.00	-0.01 – 0.01	0.574	-0.02	-0.09 – 0.04	0.471	0.00	-0.07 – 0.07	0.995
(Intercept)				0.02	0.01 – 0.04	**<0.001**	-0.00	-0.00 – 0.00	0.221
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.005 / -0.000			0.000 / -0.000		

Table 38: Equation: std. dev. (arithmetic) ~ trade-to-order vol. ratio + market capitalisation.

Panel A

	Fixed model			**Between model**			**First differences model**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.00	-0.01 – -0.00	**0.012**	-0.09	-0.20 – 0.01	0.070	-0.00	-0.01 – -0.00	**0.045**
Market capitalisation	0.00	-0.01 – 0.01	0.463	-0.01	-0.03 – 0.01	0.263	0.02	-0.03 – 0.06	0.455
(Intercept)				0.01	0.00 – 0.02	**0.009**	0.00	-0.00 – 0.00	0.987
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.008			0.028 / 0.013			0.000 / 0.000		

Panel B

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.01	-0.01 – -0.00	**<0.001**	-0.01	-0.07 – 0.04	0.645	-0.01	-0.01 – -0.00	**<0.001**
Market capitalisation	0.00	-0.00 – 0.01	0.297	-0.01	-0.02 – 0.00	0.293	0.00	-0.03 – 0.03	0.976
(Intercept)				0.01	0.00 – 0.01	**0.014**	-0.00	-0.00 – 0.00	0.848
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.002 / -0.007			0.006 / -0.004			0.002 / 0.002		

Panel C

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.01	-0.01 – -0.01	**<0.001**	0.01	-0.14 – 0.16	0.919	-0.01	-0.01 – -0.01	**<0.001**
Market capitalisation	0.00	-0.01 – 0.01	0.595	-0.03	-0.09 – 0.03	0.376	0.00	-0.07 – 0.07	1.000
(Intercept)				0.02	0.01 – 0.04	**0.008**	-0.00	-0.00 – 0.00	0.224
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.002 / -0.012			0.002 / -0.003			0.002 / 0.002		

Table 39: Equation: std. dev. (logarithm) ~ average trade size + market capitalisation.

Panel A

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	-0.00	-0.00 – -0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**<0.001**
Market capitalisation	-0.00	-0.01 – 0.01	0.928	-0.00	-0.00 – 0.00	0.067	0.00	-0.03 – 0.04	0.966
(Intercept)				0.00	0.00 – 0.00	**<0.001**	0.00	-0.00 – 0.00	0.984
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.011 / 0.003			0.989 / 0.989			0.023 / 0.023		

Panel B

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	-0.00	-0.00 – -0.00	**<0.001**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**<0.001**
Market capitalisation	0.00	-0.00 – 0.00	0.950	-0.00	-0.00 – 0.00	0.058	-0.00	-0.02 – 0.02	0.840
(Intercept)				0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – 0.00	0.795
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.008 / -0.000			0.953 / 0.953			0.015 / 0.015		

Panel C

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Average trade size	-0.00	-0.00 – -0.00	**0.002**	0.00	0.00 – 0.00	**<0.001**	-0.00	-0.00 – -0.00	**<0.001**
Market capitalisation	0.00	-0.01 – 0.01	0.964	-0.02	-0.06 – 0.02	0.285	-0.00	-0.06 – 0.05	0.912
(Intercept)				0.02	0.01 – 0.02	**<0.001**	-0.00	-0.00 – 0.00	0.270
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.013			0.061 / 0.056			0.001 / 0.000		

Table 40: Equation: std. dev. (logarithm) ~ odd-lot volume ratio + market capitalisation.

Panel A

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.00	-0.00 – 0.00	0.209	-0.01	-0.04 – 0.02	0.473	0.00	0.00 – 0.00	**0.032**
Market capitalisation	-0.00	-0.01 – 0.01	0.867	-0.01	-0.02 – 0.01	0.525	0.00	-0.03 – 0.04	0.977
(Intercept)				0.00	0.00 – 0.01	**0.017**	0.00	-0.00 – 0.00	0.979
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.000 / -0.008			0.008 / -0.007			0.000 / 0.000		

Panel B

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.00	-0.00 – 0.00	0.122	-0.01	-0.03 – 0.01	0.361	0.00	0.00 – 0.00	**0.039**
Market capitalisation	0.00	-0.00 – 0.00	0.980	-0.00	-0.01 – 0.01	0.522	-0.00	-0.03 – 0.02	0.840
(Intercept)				0.00	0.00 – 0.01	**<0.001**	-0.00	-0.00 – 0.00	0.799
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.000 / -0.009			0.010 / -0.000			0.000 / 0.000		

Panel C

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Odd-lot volume ratio	0.00	-0.00 – 0.00	0.630	-0.03	-0.07 – 0.01	0.180	-0.00	-0.00 – 0.00	0.405
Market capitalisation	0.00	-0.01 – 0.01	0.971	-0.02	-0.06 – 0.02	0.307	-0.00	-0.06 – 0.05	0.911
(Intercept)				0.02	0.01 – 0.03	**<0.001**	-0.00	-0.00 – 0.00	0.271
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.000 / -0.014			0.009 / 0.004			0.000 / -0.000		

Table 41: Equation: std. dev. (logarithm) ~ trade-to-order vol. ratio + market capitalisation.

Panel A

		Fixed model			**Between model**			**First differences model**	
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.00	-0.01 – -0.00	**<0.001**	-0.09	-0.19 – 0.01	0.070	-0.00	-0.00 – -0.00	**0.002**
Market capitalisation	-0.00	-0.01 – 0.01	0.900	-0.01	-0.03 – 0.01	0.239	0.00	-0.03 – 0.04	0.978
(Intercept)				0.01	0.00 – 0.02	**0.009**	0.00	-0.00 – 0.00	0.981
Observations	16184			136			16048		
R^2 / R^2 adjusted	0.001 / -0.008			0.029 / 0.014			0.001 / 0.000		

Panel B

		Fixed model			**Between model**			**First differences model**	
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.01	-0.01 – -0.01	**<0.001**	-0.01	-0.07 – 0.04	0.670	-0.01	-0.01 – -0.00	**<0.001**
Market capitalisation	0.00	-0.00 – 0.00	0.981	-0.01	-0.01 – 0.00	0.263	-0.00	-0.03 – 0.02	0.833
(Intercept)				0.01	0.00 – 0.01	**0.014**	-0.00	-0.00 – 0.00	0.808
Observations	22450			192			22258		
R^2 / R^2 adjusted	0.003 / -0.006			0.007 / -0.004			0.003 / 0.003		

Panel C

	Fixed model			Between model			First differences model		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Trade-to-order volume ratio	-0.01	-0.01 – -0.01	**<0.001**	0.04	-0.05 – 0.13	0.411	-0.01	-0.01 – -0.01	**<0.001**
Market capitalisation	0.00	-0.01 – 0.01	0.992	-0.02	-0.06 – 0.02	0.255	-0.00	-0.06 – 0.05	0.906
(Intercept)				0.02	0.01 – 0.03	**0.002**	-0.00	-0.00 – 0.00	0.275
Observations	28654			389			28265		
R^2 / R^2 adjusted	0.003 / -0.011			0.006 / 0.001			0.003 / 0.003		