UNIVERSITY OF TRENTO

DOCTORAL THESIS

Tick size regulation and the liquidity of UK venues: Three market microstructure essays

Author:
Maria Francesca NUZZO

Supervisor:
prof. Flavio BAZZANA
Co-supervisor:
prof. Andrew LEPONE

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in cotutelle with

Macquarie University of Sydney
Department of
Economics and Management

Declaration of Authorship

I, Maria Francesca NUZZO, declare that this thesis titled, "Tick size regulation and the liquidity of UK venues: Three market microstructure essays" and the work presented in it are my own. I confirm that:

- This work was done wholly while in a cotutelle program between University of Trento and Macquarie University of Sydney.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 23/10/2020

"The alchemists in their search for gold discovered many other things of greater value."

Arthur Schopenhauer

UNIVERSITY OF TRENTO

Abstract

Finance
Economics and Management

Doctor of Philosophy

Tick size regulation and the liquidity of UK venues: Three market microstructure essays

by Maria Francesca NUZZO

This dissertation contributes to the research in the applied market micro-structure field, aiming to investigate the impact of a specific article of the MiFID II enforced on the 3rd of January 2018: the so-called tick size regime. It is constituted by three papers that see in the market regulators and policy-makers their optimal target.
The first paper evaluates the consequences of the new regulation on UK minor venues in terms of liquidity and price discovery and highlights minor unintended consequences in the implementation of the new grid. The second paper builds on these conclusions and promotes an alternative to ESMA grid, a recalibration of the tick size that might lead to a greater orderliness of UK order books.
The third paper endogenously investigates the behaviour of the market participants in the time frame around the MiFID II enforcement, simulating liquidity breakdowns thus providing the regulators with new simple metrics to detect and monitor abnormal market participants interactions.

Acknowledgements

I would like to thank Toby Wallis and the Market Analysis and Risk Team for all the assistance and interest shown in my research.
My acknowledgements go to Prof. Flavio Bazzana and Prof. Andrew Lepone for all the guidance and support in the production of this work. I ought to extend my gratitude to the Rozetta Institute (Sydney), which sponsored my three-year partnership with the Financial Conduct Authority in London.
I am also grateful to Fabio Braga for the extensive work conducted together on transparency requirements and to Alessandro Chiozzi, Teresa Bono and Adam Gilles for all the food (not only) for thoughts and their lovely invaluable curiosity and geekiness.

Contents

List of Figures

List of Tables

List of Abbreviations

ADNT	**A**verage **D**aily **N**umber (of) **T**ransactions
AMF	**A**utorité (des) **M**archés **F**inanciers
ANN	**A**rtificial **N**eural **N**etwork
AS	**A**bsolute **S**pread
AQX	**AQ**uis E**X**change
BATS	**B**etter **A**lternative **T**rading **S**ystem
CCP	**C**learing **C**ounter **P**party
CHX	**CH**i **X** Exchange
FCA	**F**inancial **C**onduct **A**uthority
HFT	**H**igh **F**requency **T**raders
LEI	**L**egal **E**ntity **I**dentifier
LSE	**L**ondon **S**tock **E**xchange
MiFID	**M**arkets (in) **F**incancial **I**nstruments **D**irective
MLP	**M**ulti **L**ayer **P**receptron
MRM	**M**ost **R**elevant **M**arket
OB	**O**rder **B**ook
NCN	**N**on **C**learing **M**embers
QRS	**Q**uoted **R**elative **S**pread
SEC	**S**ecurity (and) Exchange Commission
SI	**S**ystematic Internalizers
SIX	**S**wiss EXchange
SNA	**S**ocial **N**etwork **A**nalysis
SVM	**S**upport Vector Machine
TRQ	**T**u**RQ**uoise Exchange

To Anna Giulia

Introduction

Improve market quality: A public policy objective

Enhance market quality is undoubtedly valuable for all the listed firms and should also be the stated goal of public policy-oriented to shape the market design. High market quality is mirrored by a fluid and fast price discovery process and by low trading costs for the investors. This dissertation focuses on three fundamental and interconnected attributes of market quality: transparency, liquidity, and stability.

A transparent market allows market participants to achieve the right level of information about the current trading conditions. It is possible to differentiate between *pre-trade* transparency and *post-trade* transparency. The former pertains to quotes and quote sizes, whilst the latter applies to transaction prices and trade sizes.

It is not a coincidence that transparency is the cornerstone of MiFID II (2014/65/EU) that, compared to MiFID I, extends transparency requirements to all financial instruments other than shares only, i.e., equity instruments and non-equity instruments.

Transparency impacts investors' tactical decisions, such as their orders timing, order sizing, and optimal bid and ask prices. It also affects market makers ability to monitor the quality of executions they receive, and it is generally thought of as ultimately mirroring in greater liquidity, stability and fairness. Against this background, we try to empirically assess the impact of the new regime on the liquidity and price formation of UK most fragmented minor venues.

In doing so, we test one of the most challenging aspects of the new directive, namely the setting of thresholds for assessing the liquidity of equity and equity-like financial instruments.

Thanks to a unique MiFID II transaction report database provided us by the Financial Conduct Authority; we are able to observe not only the overall impact on the liquidity of different venues but also to gauge changes in the behaviour of market participants.

The research context: Regulatory background

Tick sizes represent the minimum increment at which an instrument is quoted. Take a StockABC currently bought at 20 GBP. If we fix its tick size at 0.002, it means we can place a buy order at 20.002, but we can never trade this stock at 20.001. Tick size is thus the minimum meaningful cost that an investor must pay to see their order executed ahead of the best offer presented in the order book.

The need of a mandatory tick size steamed by a perceived risk of a race to the bottom of prices in the two years before the implementation of the MiFID II directive. Trading venues were constantly decreasing their prices in order to offer tighter prices and win market share.

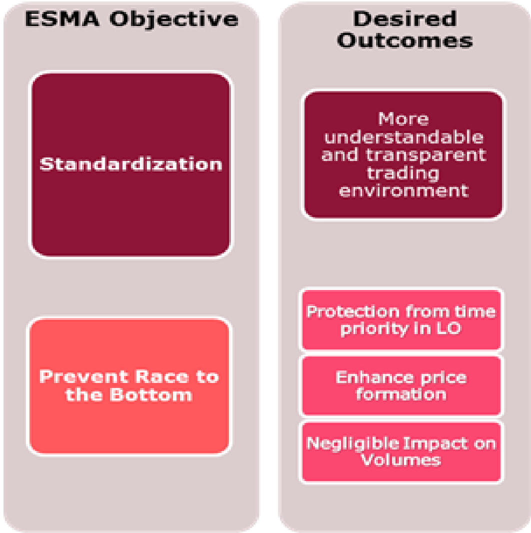


FIGURE 1: ESMA Stated Goals

This attitude produced negative effects on the overall market quality, as measured by higher noise in the order book. A too small tick size in fact, entails a too high number of prices in the order book and consequently, impairs the price formation. In particular High Frequency Traders (HFTs) can step in front in the price priority queue by offering frequent and negligible price improvements.

Although the main reason for a tick size regime is the perceived risk of a too low tick-size, MiFID II did not prescribe a larger tick-size for all the instruments.

Instead, the new regulation presents a new standardised grid to compute tick sizes updating for each instrument according to two parameters: the price of the instrument and its liquidity band. The liquidity band is given by the average daily number of transactions in the most liquid market, where the equity instrument is traded and is updated on yearly basis. The higher the price, the higher the tick size while the more liquid an instrument is, the lower its tick size.

The new grid produced different effects on the tick size of equity instruments: for some instruments tick sizes increased, while for others there was no change or tick sizes were narrowed further. Indeed, this is the case for the majority of the security traded in the FTSE100.

In our first paper [1], we investigated which were the consequences of these non-uniform changes in tick sizes on the overall market quality. The purpose is to assess the impact on market quality, by accounting for several of its dimensions (volumes, transaction costs and price informativeness), on the new tick size regime. We quantify to what extent ESMA effectively satisfied its stated objective. The purpose of the ESMA tick size regime, in fact, is to ensure the orderliness of the order books. Its scope is of promoting the effective formation of prices and sustaining deep order books whilst allowing spreads to fluctuate.

Although standardization is required to create a more understandable and transparent trading environment, we can see it as a secondary and auxiliary objective: a more standardised grid is necessary to grant the same tick-size across venues and prevent the perceived risk of a race to the bottom. Therefore, the first research focuses on the main ESMA objective and tries to assess if and to what extent the new tick size contributed to enhancing price formation, without adversely impacting the volumes and reducing HFTs incentives to trade aggressively, meanwhile granting a liquid market to the least liquid equity instruments.

The stated goals of the Art. 49 were essentially two (see Figure 1)

(a) the standardization of the tick size grid across venues and

(a) prevent the perceived risk of a race to the bottom of prices.

We identified gaps arising both from the (partial) implementation of the regime and from the proper implementation on UK fragmented minor venues. The reasons to investigate this specific aspect of the new directive were both practical and theoretical.

First, we believe the research question is relevant to evaluate the policy. Exchanges have been able to set their own (narrower) tick sizes for a long time in the past and a harmonization was necessary to limit the risks of a race to the bottom where a narrower tick size would have led to a higher share of the market. However, the new tick size grid was not a simple standardisation contrary to ESMA stated goals. It had consequences on market quality of UK trading venues, affecting spreads and depths, and could open up questions on the calibration of the transparency calculations. Second, the bulk of past research on the impact of changes in tick size on market quality is not conclusive and left us with inconsistent findings.

From transparency to liquidity: Evaluation of the tick size regime

The new tick size regime introduced in Europe gave us the opportunity to examine tick size changes in both directions at the same time, thus naturally controlling for overall market conditions. Tick size represents the minimum implicit cost of transaction (spread) and can be seen as the compensation for market makers for holding an asymmetric information risk of trading with an informed trader. First, we consider this intuitive, mechanical effect of tick size changes and we estimate the impact of tick sizes on market markets' behaviour. Too narrow tick sizes can discourage market makers from supporting liquidity, especially for smaller capitalization stocks.

A second way in which different tick sizes can affect the behaviour of market participants is that, according to theoretical models, institutions trading large blocks and small retailers have different preferences about optimal tick sizes with the former favouring larger tick sizes compared to the latter. Too small a transaction tick size can be detrimental to the ability of order-books to enhance price discovery and reduce the level of noise (we call this *order book viscosity*). We have investigated the impact of tick sizes on viscosity proxied by transactions mid-size and quote duration (i.e. the time a quote remains alive in the order book).

High Frequency traders (HFTs) could have modified their behaviour according to different tick sizes. We therefore try to answer this third research question investigating HFTs' market shares expressed in terms of turnover of instruments.

The last spotlight of Chapter 1 regards the price discovery process. We built a metric robust to different levels of noises, namely the Information Leadership Share (ILS). A market is typically considered to dominate the price formation process if available prices reflects the new information around the fundamental value (see Putniņš 2013). Then we compared the mean value of this metric for each group of equal, widened and narrowed tick size.

We assessed the impact of the tick-sizes changes on several dimensions of market quality: the activity of market makers and HFTs, the order books stability and the price discovery.

Do wide tick sizes constitute an incentive for market makers to provide liquidity?

With our first research question we aim at verifying whether tick sizes constitute an incentive for market makers to provide liquidity on the market, especially for the most illiquid instruments.

We investigate the incentive for market makers to provide liquidity using two proxies: realised spreads and effective spreads. All else being equal, wider spreads lead to greater profits for market makers. We used effective spreads to measure implicit trading cost, whilst realised spreads are used to measure the difference between the transaction price and the quote midpoint at some time after the trade. We computed them with 30 seconds and 30 minutes delayed mid-quote, in order to give to the market a shorter and a longer time to incorporate the information contained in the buy or in the sell and try to gauge the activity of HFTs.

We have found that the main lit market and the minor UK MTFs differ in the structure of their transaction costs. In particular, it seems that profits for of proprietary traders, high frequency traders (HFTs) are on average higher on the smaller venues compared to the main lit exchange. In relation to the tick size, we have found that realised spreads of Market Makers and Proprietary traders, high frequency traders are higher for securities with a narrower tick size. These results indicate a first unintended consequence of the ESMA regime reflected in a decline in cost of transactions in favour of HFTs and a minor incentive in the provision of liquidity on the main lit market.

Can tick sizes be detrimental to order book viscosity?

Secondly, we tried to gauge if the order book viscosity is deteriorated by too narrow tick sizes, that can thereby expose the order book to disordered price formation processes.

Viscosity is a desired characteristic of order books, because it represents their ability to enhance price discovery and reduce the level of noise. We found, in line with what theory prescribed, that on minor UK venues the life of orders (i.e. how long the best bid(ask) remains in the order book before being executed or substituted by a better quote) associated with a trade from Market Makers and HFTs is higher when the instrument is allowed to be traded at a lower tick size. Indeed, a too small tick size in fact, entails a too high number of prices in the order book and consequently impair the price formation. In particular High Frequency Traders (HFTs) can step in front in the price priority queue by offering frequent and negligible price improvements. This result suggests that ESMA tick size applied on the London Stock Exchange was not optimal for order book stability.

Is high frequency traders (HFTs) activity intensified in relation to narrower tick sizes?

Third, we are interested in spreading light onto the role of high frequency traders (HFTs), especially, we would like to observe if their activity is intensified with respect to narrower tick sizes, reflecting a greater exposure to the undercutting risk.

The market shares of market makers were higher for instruments with narrower tick sizes across all venues. However, looking at the diff-in-diff, HFTs market share is 2.78% larger for instruments with a narrower tick size compared to the instruments in the control on CBOE, whilst their market share for the treatment group is lower on Aquis and Turquoise (respectively -1.81% and -0.49%). Thus, the impact of the HFTs activity is not univocal and it differs across venues, namely is intensified only on the larger ones.

Does the change in tick sizes undermine the leadership of UK lit venues in the price discovery role?

Finally, we considered the extraterritoriality effect on a specific sub-sample of Swiss cross-listed securities whose most liquid market is outside the EU, but whose Average Daily Number of Transaction (ADNT) is determined in the most relevant market (MRM) within the Union. We evaluate the price formation process of these cross-traded instruments on UK venues and on the Swiss exchange finding that the price for cross listed securities updated first on the Swiss Exchange followed by movements on UK venues.

Our results show that Information Leadership Share (ILS) for the instruments that have narrowed their tick sizes has improved with respect to the SIX Exchange in the period from pre-MiFID II to post-MiFID II implementation. As far as narrower tick sizes are concerned, all the UK MTFs contribute less than before to price information in the presence of different levels of noise produced by the new discrete grid of prices under Art. 49 and their widened tick sizes. We confirm our expectations on the correlation between higher spreads and slower and noisier price series.

This constitutes further evidence of an unwanted impact of the new ESMA regime on market efficiency. In terms of price discovery, UK venues resulted less efficient and timely in the incorporation of information about the fundamental value into the price.

Overall conclusions

We believe that the new tick size regulation entered into force with MiFIDII cannot be considered as a simple standardization, it had rather had a significant impact on UK venues' market quality.

In the first paper, we investigate the impact of Art. 49 on market participants' behaviour, especially, on their incentive to act as liquidity providers. The purpose of capital markets is to allocate risk/capital effectively and this is done through the mechanism of liquidity. We show that ESMA tick sizes provide only a small incentive to the provision of liquidity on the main lit market and in a decline in cost of transactions in favour of HFTs.

We have found that for some instruments, the tick size is too narrow to ensure the order books stability and to protect investors from undercutting risks by high frequency traders. We have investigated UK minor venues, because we believe that the impact of a tick-size change is particularly sharp for fragmented markets, which are already onerously inclined to display higher trading costs and a minor depth compared to centralised exchanges on which the market's ability to match buyers and sellers increases. They also tend to be less price informative because of their increasing searching costs. On the other hand, the emblematic advantage of a fragmented structure is the increased competition between trading venues, which may result in lower trading costs. Thus, if this competitive advantage is harmed by a change in the tick-size regime, their market quality can be impaired.

All in all, our finding suggests that minor unintended consequences of the new regime impacted the market quality of UK venues and could open-up questions on the calibration of the transparency calculations.

Recalibration of the tick size regime

The recalibration is therefore the cornerstone of our second work. We aimed to propose a better methodology to improve the effectiveness of the regime. We adopt a supervised machine learning approach to propose a better calibrated alternative to the ESMA grid. Our approach is based on: (i) market capitalization; and (ii) quoted spread. We show how our calibration for the regime would achieve optimal tick sizes for equities 3 times more frequently than the current ESMA regime. This allows us to outline an idealized grid for determining an equity's minimum tick size for this proposed regime. This paper is especially relevant for UK policy makers in the context of the UK leaving the EU and suggests the ESMA grid can be abandoned. At the best of our knowledge, it is also the first time a supervised machine learning model is adopted to evaluate policy implications of a financial regulation.

From liquidity to stability: A market participants' network

In our last paper, we simulate liquidity breakdowns on UK equity market adopting a direct network approach. Compared to previous studies in the literature we focused on the link between the structure of the network and the role of dealer's inventories. In our simulations, the risk of liquidity breakdown stemmed directly from the dealer's inventory optimization problem. The ability of a system to deal with the shock changes with different financial agents and different liquid/illiquid instruments as well as with different levels of competition among dealers and fragmentation among UK venues.

Along these lines, the dynamic of the contagion in our approach is subject to the dynamic of dealer's inventories. As far as the relationship between the topology of the network and the risk of contagion is regarded, we show that agency brokers present the highest centrality score in the network. This means that participants within this firm type tend to lead more price information than others as well as to spread faster the risk of liquidity disruption. Our findings also suggest that central nodes are linked with other stable nodes whose inventory levels are hard to get altered.

We have found a positive correlation between the degree of centrality of a node and the speed of contagion. All in all, our study contributes to the existing empirical market micro-structure literature adopting a novel, distinctive propagation algorithm based on market participant inventories. The direct-network approached based on dealer's inventory can provide regulators an extra tool to monitor the risk of liquidity breakdowns in the equity market, identifying who are the participants who can spread the risk faster and wider.

Chapter 1

Tick size regime impact on UK equity market participants

Co-authors: This research article was co–authored with my supervisors, Prof. Flavio Bazzana and Prof. Andrew Lepone, who contributed in designing the study and in interpreting the results.

Abstract

This paper evaluates the impact in the immediate aftermath of the MiFID II tick size regime implementation (Art 49) as observed on four of the UK trading venues that have implemented it in January 2018: the London Stock Exchange, Aquis, Turquoise and CBO Europe. The purpose of this work is to assess the impact on market quality, determined across several of its dimensions (volumes, cost of transactions and price informativeness), of the new tick size regime which entered into force in January 2018 as part of the Markets in Financial Instruments directive (MiFID II). We have found that the main lit market and the minor UK MTFs differ in the structure of their transaction costs. In particular, realised spreads for HFTs are higher for securities with a narrower tick size. These results indicate a first unintended consequence of the ESMA regime reflected in a decline in transaction costs, which favours HFTs and a minor incentive in the liquidity provision on the main lit market. Moreover, the market shares of market makers were higher for instruments with narrower tick sizes across all venues. Our results show that Information Leadership Share (ILS) for the instruments that have narrowed their tick sizes has improved with respect to the SIX Exchange in the period from pre-MiFID II to post-MiFID II implementation. As far as narrower tick sizes are concerned, all the UK MTFs contribute less than before to price information in the presence of different levels of noise produced by the new discrete grid of prices under Art. 49 and their widened tick sizes. We confirm our expectations on the correlation between higher spreads and slower and noisier price series. This constitutes further evidence of an unwanted impact of the new ESMA regime on market efficiency. In terms of price discovery, we find that UK venues are less efficient in timely incorporating information about the fundamental value into the price. All in all, we believe that the new tick size regulation entered into force with MiFID II cannot be considered only as a simple standardization procedure, since it had rather had a significant impact on UK venues' market quality.

Keywords MiFID II – Tick Size Regime – Market Quality – Liquidity Provision – Market Makers profits - Cost Benefits Analysis – Policy Evaluation – Information Leadership Share (ILS) – Price Discovery

1.1 Introduction

Tick sizes represent the minimum increment at which an instrument is quoted. Take a Stock ABC currently bought at 20 GBP. If we fix its tick size at 0.002, it means we can place a buy order at 20.002, but we can never trade this stock at 20.001. This 0.002 is the minimum meaningful cost that an investor must pay to see their order executed ahead of the best offer presented in the order book.

The need of a mandatory tick size steamed by a perceived risk of a *race to the bottom* of prices in the two years before the implementation of the MiFID II directive [1]. Trading venues were steadily decreasing their prices to make them tighter and win market share.

This attitude produced adverse effects on the overall market quality, as measured by higher noise in the order book. A too-small tick size, in fact, entails an excessively high number of prices in the order book and consequently impairs the price formation. In particular, High-Frequency Traders (HFTs) can step in front of the price priority queue by offering frequent and negligible price improvements.

Although the main reason for having a tick size regime is the perceived risk of a too low tick-size, MiFID II did not prescribe a larger tick-size for all the instruments. Instead, the new regulation presents a new standardised grid to calculate and update tick sizes for each instrument according to two parameters: the price of the instrument and its liquidity band. The liquidity band is given by the average daily number of transactions in the most liquid market where the equity instrument is traded and is updated on yearly basis. The higher the price the higher the tick size while the more liquid an instrument is the lower its tick size.

The new grid produced different effects on the tick size of equity instruments: for some instruments tick sizes increased, while for others there was no change or tick sizes were narrowed further. Indeed, this is the case for the majority of the security traded in the FTSE100. Therefore, we investigated which were the consequences of these non-uniform changes in tick sizes on the overall market quality.

The purpose is to assess the impact on market quality, determined across several of its dimensions (volumes, cost of transactions and price informativeness), of the new tick size regime. We quantify to what extent ESMA effectively satisfied its stated objective. The purpose of the ESMA tick size regime, in fact, is to ensure the orderliness of the order books. Its scope is promoting the effective formation of prices and sustaining deep order books whilst allowing spreads to fluctuate.

Although standardization is required to create a more understandable and transparent trading environment,it can be considered as a secondary and auxiliary objective: a more standardized grid is necessary to grant the same tick-size across venues and prevent the perceived risk of a race to the bottom. Therefore, this research focuses on the main ESMA objective and tries to assess if and to what extent the new tick size contributed to enhancing price formation, without adversely impacting the volumes and reducing HFTs incentives to trade aggressively, meanwhile granting a liquid market to the least liquid equity instruments.

The stated goals of the Art. 49 were essentially two: a) the standardization of the tick size grid across venues and b) prevent the perceived risk of a race to the bottom of prices. We identified gaps arising both from the (partial) implementation of the regime and other arising from the proper implementation on UK fragmented minor venues.

[1] Risk and Trends 2018

The reasons to investigate this specific aspect of the new Directive were both practical and theoretical. First, the research question is relevant to evaluate the policy. Exchanges have been able to set their own (narrower) tick sizes for a long time in the past and a harmonization was necessary to limit the risks of a race to the bottom where a narrower tick size would have led to a higher share of the market. However, the new tick size grid was not only a standardisation procedure, in which it had consequences on the market quality of UK trading venues and it affected spreads and depths. Indeed, these unintended consequences could open up questions on the calibration of the transparency calculations as proposed by ESMA.

Second, the bulk of past research on the impact of tick size changes on market quality is not conclusive and left us with inconsistent findings. The new tick size regime introduced in Europe gave us the opportunity to examine tick size changes in both directions at the same time, thus naturally controlling for overall market conditions.

Tick size represents the minimum implicit cost of transaction (spread) and can be seen as compensation for market makers for holding an asymmetric information risk of trading with an informed trader. Firstly, we consider this intuitive, mechanical effect of tick size changes and we estimate the impact of tick sizes on market makers' behaviour. Too narrow tick sizes can therefore discourage market makers from supporting liquidity, especially for smaller capitalization stocks. Second, different tick sizes can affect the behaviour of different market participants. According to theoretical model, in fact, institutions trading large blocks and small retailers have different preferences about optimal tick sizes, namely the former favouring larger tick sizes compared to the latter. A Too small transaction tick size can be detrimental for the ability of order-books to enhance price discovery and reduce the level of noise (we call this *order book viscosity*).

We have investigated the impact of tick sizes on order books viscosity proxied by transactions mid-size and quote duration (i.e. the time a quote remains alive in the order book). High Frequency traders (HFTs) are another firm type of market participants that could have modified their behaviour according to different tick sizes. We therefore try to answer this third research question investigating HFTs' market shares expressed in terms of turnover of instruments from the treatment and control group. The last spotlight of our analysis regards the price discovery process understood in its "efficient and timely incorporation of new information into fundamental values" (Lehmann 2002) interpretation. However, since the first updates can also be the noisiest, we need not only to evaluate who moves first but we also need a metric robust to different levels of noises. We computed the Information Leadership Share (ILS) for each UK market included in our analysis and for the Swiss Exchange for each 75 securities in our sample. A market is typically considered to dominate or lead price discovery if it is the first to reflect new information about the fundamental value. Then we compared the mean value of this metric for each group of equal, widened and narrowed tick size.

To evaluate the tick size impact on market quality we have computed a series of proxies for market liquidity and price information based on two different datasets. First, we looked into Refinitv order book data for our cross-temporal analysis on price information. Second, thanks to FCA proprietary data, we could investigate the different impact of a tick size change on venues distinguishable by a different level of activity by market makers, High Frequency Traders (HFTs) and regulated liquidity providers[2]. Third, to provide a comprehensive picture of the overall impact on market quality, we summarise previous known findings around the tick size impact on order book depth and spreads.

The implementation of the new regulation gave us the opportunity to examine tick size changes in both directions at the same time (widened and narrowed tick sizes), thus naturally controlling for overall market conditions. We designed difference-in-differences analysis on a cross-sectional dataset including the 4 UK venues. Our control group is represented by the set of shares that have not changed their tick size. In addition, we introduced a control market (the Swiss exchange) that did not experience any regulatory change. We try to assess the overall impact of the new regulation focusing on the following research questions:

a *Do wide tick sizes constitute an incentive for market makers to provide liquidity on the market, especially for the most illiquid instruments?*

b *Can narrower tick sizes be detrimental for order book ability to enhance price discovery and reduce the level of noise through a significant quote duration and transaction size therefore exposing the order book to chaotic price formation processes (viscosity)?*

c *Is high frequency traders (HFTs) activity intensified in relation to narrower tick sizes reflecting a greater exposure to the undercutting risk?*

d *Does the change in the tick sizes affect the ability of UK lit venues to predominantly anticipate the long-term price changes (price information leadership)?*

The unique opportunity offered by the new tick-size regime compared to other natural experiment (e.g. US Tick-Size Pilot in 2016 that led to an increase in the tick size from one to five cent for smaller capitalization stock[3])is to investigate the impact of tick-size changes in both directions: equity instruments started to be traded with wider or narrower tick sizes according to their liquidity band computed by ESMA and according to their prices. It's worthy to notice here that ESMA did not prescribe *per se* a wider/narrower tick for instruments in different liquidity bands. This meaning that we can find instrument that narrowed/widened the tick within the same liquidity band [4].

We run two different analysis. To assess the implications of the new tick sizes on market participants' behaviour, we perform a cross-venue analysis, considering as treatment group the instruments traded with a different tick size on different venues[5]. Second, for price discovery investigation, we perform a cross-temporal analysis (before-after) for the securities with wider tick sizes after the implementation of the new regime (Up-Tick), securities with narrower tick size (Down-Tick) and securities that did not change their tick size (Equal)[6]. For this latter purpose, our dataset consists of Refinitiv Times and Sales [7] order books for cross-listed equities

[2]These are trading participants registered in the Equities Liquidity Provider Program (LPP) designed to ensure continuous bid and ask prices and deep liquidity to a variety of securities among which there are also instruments from FTSE 100 and FTSE250.

[3]It was approved by the SEC on May 6, 2015 and it begun on the 3rd of October 2016.

[4]For robustness check, to show the true exogeneity of the effect, we have analysed instruments within the same liquidity band.

[5]This choice was led by data availability, a cross-temporal analysis pre-post MiFID II on market participants behaviour would have also required a pre-MiFID II transaction reports dataset not available to the writer.

[6]Appendix A presents the complete list of instruments used in or analysis.

[7]Formerly Thomson Reuters Thick History

on four UK venues and one Swiss venue that has not implemented the MiFID II new calculations regime (Swiss Exchange).

This design allow to disentangle the effect of the tick size and answer to the research questions. This paper is organised as follows: Section 2 presents the literature review on tick sizes, their impact on market quality and their relationship with market structure. Section 3 draws hypotheses from this literature, highlighting the elements of novelty of our research design. It also describes the dataset and broaches the adopted statistical models. Results of our analysis can be found in Section 4, including robustness checks to verify our hypotheses while we control for firm-size and exchange-rates.

1.2 Research Context

Tick sizes represent the minimum increment at which an instrument is quoted but they can be also thought of as the implicit cost of transactions and structural constituents of quoted bid-ask spreads, which cannot be smaller than the minimum, prescribed tick size. Consequently, if an instrument is traded at a narrower (binding) tick size, should be cheaper for the investor to buy or sell it. On the contrary, we can expect that a wider tick size will mechanically lead to higher bid-ask spread.

Before the implementation of MiFID II, each trading venue could decide a tick size applicable on its platform. However, this led to a *race to the bottom* where a narrower tick size would have led to a higher share of the market. Regulators did intervene by imposing a minimum tick size to protect the investors who submit limit orders. Binding the spread to a minimum tick size harmonized across all venues they also increase the cost of undercutting. This means they aim to discourage High Frequency Traders (HFTs) to step in front in the price priority queue by offering frequent and negligible price improvements.

Minimum tick sizes have been criticized by some authors (see Ricker 1996 or Peake 1993) because they can increase market-maker's pay-offs as well as investor's trading-costs. Binding tick sizes can also, as shown in Budish, Cramton, and Shim 2015, result in an increased competition on speed that ultimately eliminates undercutting effects.

The impact of an increased tick size has been observed for the first time in the US tick size pilot [8] in 2001, when SEC adopted a decimal pricing and increased tick sizes for small caps.

The overall effect of this policy was an improvement in the liquidity of these instruments. This happened because, in addition to the mechanical effect on spreads, the wider tick sizes increased also the incentive for market makers to support small caps and to provide liquidity.

Several studies have documented a correlation between a reduction of tick sizes and a decrease in quoted spreads (Bessembinder 2000; Goldstein and A. Kavajecz 2000; Jones and Lipson 2001). Yet, the overall impact of a change in tick sizes is not conclusive. A trade-off between depth and spreads has been extensively illustrated. Whilst the spreads decrease following a tick size reduction, at the same time also the depth suffers of a contraction. This is because market orders become cheaper and more appealing to investors rather than limit ones. This contraction for larger trades

[8] In technical details: 1 control group and 3 tests (each 400 w ISINs) TG1: quote at 0.05$, trade at current increments TG2 quote and trade at 0.05$ + exceptions: executions at midpoint, retail investor orders, with price improvement of at least 0.005$, negotiated trades TG3 as TG2 + trade at prohibition (apply exception from Rule 611) Control: quote and trade at 0.01$.Source: https://www.sec.gov/ticksizepilot

may even imply an increase in transaction costs (see Hsieh, Chuang, and Lin 2008 for an example in order-driven emerging market or Bollen and Whaley 1998 for the analysis of a quote-driven developed exchange).

Our study adds fresh insights to the previous literature: the new tick size regime introduced in Europe with Art.49 of MiFID II gave us the opportunity to examine tick size changes in both direction at the same time(i.e. some securities narrowed their tick size, whilst others widened it), thus naturally controlling for overall market conditions. The impact of a new tick size regime has to be especially assessed taking into account the unique fragmented nature of UK-venues. A fragmented market is already inclined to display higher trading costs and a reduced depth compared to a centralised exchange on which the market's ability to match buyers and sellers increases. It also tends to be less price informative because of its increasing searching costs (see Hendershott and Jones 2005).

On the other hand, the advantage of a fragmented market structure is the increased competition between trading venues, which may result in lower trading costs. Thus, if this competitive advantage is harmed by a change in the tick size regime, cash outflows can occur and liquidity can decrease as a consequence. Our study shows how a change in tick sizes has different impact on market quality depending on the market characteristics and that it should not be considered as a simple standardisation as in the stated scope of ESMA regulators. In its potential implications, our study is of value to policy makers and can be relevant beyond academic audiences.

Since 2007, Europe's regulation of financial markets is founded on the MiFID I [9] aimed at enhancing European Exchanges competitiveness through the creation of a unified market for financial services. It was designed to deploy regulatory reporting to avoid market abuse and improve trade transparency for shares. Its scope, however, has been notably enlarged by MiFID II that achieves a greater market transparency introducing pre-trade and post-trade requirements for non-equity instruments and strengthens the existing regime for equity and equity-like instruments.

Article 49 of MiFID II [10] required trading venues to adopt minimum tick sizes in relation to equity and equity-like instruments [11]. The minimum tick size to be applied depends on the Equity liquidity band and price level. ESMA liquidity band are computed based on the average daily number of transactions (ADNT) on the most relevant market (MRM) in the EU. The ADNT is automatically calculated and published by the Financial Instruments Transparency System (FITRS), a database operated by ESMA, based on quantitative information received from EU trading venues and national competent authorities (NCAs)[12].

The legislative aim was to "ensure the orderly functioning of the market" [13], but concerns arose from some venues about unintended consequences of this standardization:

a Concerns around the contraction of liquidity for cross-listed securities as a result of a transfer of trading on non-EU venues. According to Reuters, in fact, the beginning of 2018 has been characterised by a 75% jump (Month over

[9] Markets in Financial Instruments Directive (2004/39/EC)

[10] "MiFID II/MiFIR is a new legislative framework that will strengthen investor protection and improve the functioning of financial markets making them more efficient, resilient and transparent." *Source: ESMA website*

[11] Specified by the Regulatory Technical Standard (RTS) 11

[12] See Appendix B of Article 49 `https://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160714-rts-11-annex_en.pdf`

[13] ESMA Q & A paragraph 4) Tick Size Regime – Last Update (18/12/2017)

Month) in cross-traded Swiss shares on the Swiss Bourse "- up 10 percentage points from a year previously"(Year over Year)[14] . However, there is no evidence that the inflows on Swiss Exchange were associated to correspondent outflows from UK venues.

b Concerns about the determination of instruments' liquidity by ESMA, that calculates the average daily number of transactions (ADNT) on a yearly basis. Aquis CEO Alasdair Haynes sent a notice to the exchange members in which he claimed to follow SIX ADNT for Swiss instruments to price them more competitively than other European venues starting from the 12th of January 2018. An analogous measure has been followed by CBOEurope and Turquoise Exchange that effectively aligned tick size tables to the Swiss requirements, starting respectively by the 12th and by the 15th of January 2018.

Because of these concerns, the first months after the new regime enforcement were characterised by inconsistencies in the application of the tick size scheme among different venues. This allowed us to create a treatment group made by instruments traded with a different tick size on different venues. To address this issue, a second mandatory harmonized regime was enforced on the 1st of April 2018, followed by a proposal for amendments at the end of 2018[15].

This second implementation and the adjustments required by European multilateral trading facilities, are signals that Art. 49 was not a simple "standardisation" as was intended by ESMA regulators.

Before the implementation of MiFID II, tick sizes were computed referring to FESE tables of prices of April 2011, harmonized across all MTFs according to MiFID I regulation[16]. For Swiss venues, some indications were provided by the SIX Exchange website[17].

The most relevant difference with the new tick size regime introduced with MiFID II is the dependency on the Equity liquidity band as well as price levels used before. ESMA liquidity bands are computed based on the average daily number of transactions (ADNT) on the most relevant market (MRM) in the EU. The concept of most relevant market opened up to some concerns regarding the so-called *extraterritoriality*. In its Q & A, ESMA clarified that the most relevant market to refer to in the computation of the liquidity band had to be one within the Union. Therefore, ADNT for cross listed stocks was determined based on liquidity band observed in the Union only, even if their main primary liquidity was outside the Union.

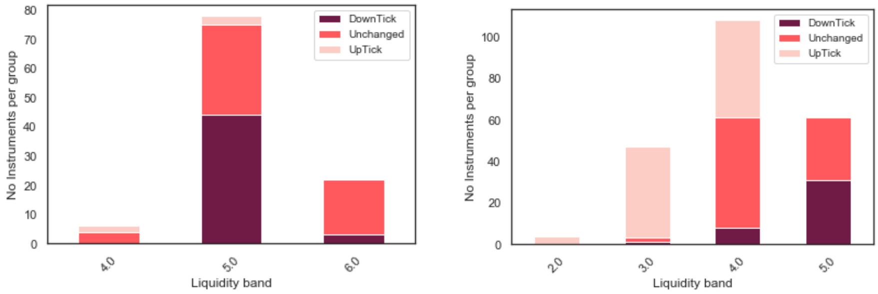
Figure A.1 shows which equity instruments have been affected by the new tick size changes, and in which direction (i.e., they widened, narrowed or maintained the same tick size as before). Instruments in FTSE100 and the most liquid instruments in FTSE250 ($2000 < \text{ADNT} \leq 9000$) saw their tick sizes decrease or remain unchanged. On the contrary, less liquid instruments ($10 < \text{ADNT} \leq 2000$) saw their tick sizes increase. We have analysed the impact of the tick size regime immediately after the implementation of the regime with the purpose of identifying cases of tick size mismatch across venues and to evaluate the impact of a different tick size isolated from

[14] https://www.reuters.com/article/eu-exchanges-align-with-zurich-after-mif-idCNL8N1P74BD

[15] Most respondents reiterated the need to apply the tick size regime in a consistent and harmonised way throughout the EU and across all possible execution venues – including Systematic Internalisers (SIs)". Source: https://www.esma.europa.eu/sites/default/files/library/esma70-156-834_final_report_on_the_proposed_amendments_to_rts_11.pdf

[16] http://www.fese.eu/images/documents/UPDATED_FESE_TICK_SIZE_TABLES_AS_OF_OCT_2012.pdf

[17] https://www.six-exchange-regulation.com/en/site/regulatory-changes/tick-sizes.html



(A) FTSE100 (B) FTSE250

FIGURE 1.1: Direction of change per liquidity band

any other conditions. We use order book data from Refinitiv (formerly Thomson Reuters Tick History v2).

In this paper, we refer to the columns of the ESMA grid reported in Annex II in RTS 11 as to the liquidity bands of the equity instrument. They go from Liquidity 1 which includes the less liquid instruments with an ADNT ≤ 10 and liquidity band 6 which includes instruments with an ADNT ≥ 9000.

Our analysis performed on UK venues exhibits very different result from those observed on CAC and SMI securities in France. The new regime led 74% of French blue chips and mid-cap companies to increase their tick size, with no change for the remaining 26% companies [18]. At the same time, 21% of French smallest caps showed an increased tick size, 64% of them displayed no change, and 15% reduced their ticks. This stark difference was simply due to the ADNT of instruments composing FTSE100 and FTSE250, compared to the ADNT of instruments composing CAC40 (majority with $2000 < \text{ADNT} \leq 9000$).

1.3 Research design

1.3.1 Dataset samples and securities selection

We run two main different analysis. First, for an overview on market quality metrics and for price discovery investigation, we perform a cross-temporal analysis (before-after the implementation of MiFID II on the 3rd of January 2018) for the securities with wider tick sizes after the implementation of the new regime (Up-Tick), securities with narrower tick size (Down-Tick) and securities that did not change their tick size (Equal). For this purpose, our dataset consists of Refinitiv Times and Sales order books. All the spreads, depth and order imbalances were computed for 309 instruments of the FTSE100 and FTSE250 excluding instruments not traded in every business day and for which ESMA didn't publish an ADNT. As far as price discovery is concerned, our dataset from Refinitv consists of order books for 20 cross-listed Equities on four venues that have implemented the new regime (LSE, AQX, TRQ, BTE) and one Swiss venue that has not implemented the MiFID II new calculations regime (SIX).

[18] It can seem counter-intuitive that the majority of the CAC40 securities increased their tick whilst the majority of FTSE100 securities reduced them since w can reasonably expect that they are both traded very frequently. In absolute terms, we have 13 CAC40 securities that increased their tick (versus 4 securities in the FTSE100), 23 securities over 37 securities included in the AMF study. At the same time 24 securities did not changed their tick in the CAC40.

We have considered only the trading days in a sample-period from the 1st of November 2017 to the 28th of February 2018. Our market metrics were based on millisecond time-stamped quotes grouped in 1-second interval time. Our study includes only quotes and trades data from each venue in the continuous opening section i.e.: from 8:00 to 16:30 for Aquis Exchange, LSE, BTE and Turquoise and 9:00–17:30 for SIX Exchange. We drop the first hour from 8:00 to 9:00 and the last one from 16.30 to 17:30 to obtain the same number of hours in the Swiss market and on UK venues. The whole compositions of the two groups is reported in Appendix A.

Second, to assess the implications of the new tick sizes on market participants' behaviour, we used MiFID II transactions reports and perform a cross-venue analysis, considering, as treatment group, the instruments traded with a different tick size on different venues. For this analysis, we have computed a series of proxies for market liquidity and price information based on Refinitiv order book data. Also, thanks to FCA proprietary data (MDP) we could observe all transactions with millisecond timestamps from official market makers, Systematic Internalisers (SI), and liquidity takers (buy-side asset managers, fund managers). We choose the shares that had not changed their tick sizes across venues as our control group. We matched our proprietary dataset with Refinitiv (formerly Thomson Reuters v2) order book data. We focus on BATS, CHIX, AQX, TRQ and LSE. Unfortunately, we could not observe the counter-party associated to each quote, nevertheless we can look at the correlation between the order flow and trades being executed by a certain counter-party.

We use 35 equity instruments included in our cleaned dataset, because they exhibited consistent mismatch of the tick across different venues. The majority of these ISINs were classified as moderately liquid according to ESMA (15 instruments from liquidity band 5 (42%) and 13 instruments from liquidity band 4 (37%)). The control group includes the same number of instruments per liquidity band as the treatment group. This helps to mitigate the impact of volatility in our analysis. It is incorrect to assume an equal impact of exogenous shocks (other than the new tick size regime) on different groups (down-Tick, up-Tick and control): the impact of volatility on less liquid instruments (i.e. with a lower ADNT) is larger than the impact on more liquid ones. Within the same liquidity band, we reasonably expect the volatility to be equivalent for all the analysed instruments. We thus, create the two groups (control and treatment) including the same number of ISINs per each liquidity band. Where possible we tried to match firms from the same business sector or with a similar market cap.

1.3.2 Experiment design

A cross-venue natural experiment for market quality proxies

The tick size regime applied differently across different trading platforms. This means we have the chance to investigate the impact of different ticks in a cross-venue natural experiment, where the control group consists of instrument that presents the same tick sizes (TS) on the LSE and other MTFs and the treatment group that includes instruments with different TS on different venues. In this way, we have the chance to isolate venue-specific effects and disentangle how different tick sizes affect the behaviour of market participants.

We matched our proprietary dataset with Refinitiv (formerly Thomson Reuters v2) order book data. We focus on BATS, CHIX, AQX, TRQ and LSE. Our objective here is to verify if the expected mean change in market quality metrics (depth,

TABLE 1.1: Number of observations included in the analysed MiFID II transaction reports dataset, control and treatment group. Source: Author's calculations based on FCA proprietary dataset and Refinitv data

	All Obs.		Control		Treatment	
	Abs.	%	Abs.	%	Abs.	%
Agency Broker	720,618	*2.52*	50,971	*1.55*	81,801	*2.19*
Asset Manager	348,055	*1.22*	40,641	*1.24*	28,247	*0.76*
Central bank	12	*0*		*0*		*0*
Charity Fund	1	*0*		*0*		*0*
Clearer, Custodian or Fund Manager	11,648	*0.04*	6,877	*0.21*	119	*0*
Commercial Bank	214,011	*0.75*	17,418	*0.53*	33,402	*0.89*
Corporate	39,562	*0.14*	989	*0.03*	2,488	*0.07*
Exchange Operator	59,500	*0.21*	4,234	*0.13*	3,496	*0.09*
Financial Intermediary	70,506	*0.25*	15,244	*0.46*	8,795	*0.24*
Fund Manager	737	*0*				
Fund of Funds	3	*0*				
Hedge Fund	379,941	*1.33*	52,514	*1.6*	44,319	*1.19*
IB - Large Dealer	13,447,763	*47.07*	1,588,712	*48.39*	1,739,781	*46.52*
IB - Small or Medium	1,351,785	*4.73*	121,405	*3.7*	165,197	*4.42*
Interdealer Broker	5,835	*0.02*		*0*		*0*
Market Maker	329,172	*1.15*	13,681	*0.42*	12,155	*0.33*
Other Funds	717,485	*2.51*	66,740	*2.03*	83,007	*2.22*
PLC	17	*0*		*0*		*0*
Pension Funds	4,645	*0.02*	90	*0*	81	*0*
Private Bank	26,440	*0.09*	1,618	*0.05*	2,198	*0.06*
Prop Trader - HFT	10,714,624	*37.51*	1,296,396	*39.49*	1,528,292	*40.87*
Unclassified	125,707	*0.44*	5,713	*0.17*	6,400	*0.17*
Total	28,568,067	*100*	3,283,243	*100*	3,739,778	*100*

spread, realised spread) from the "Pre-MiFID II" period to the "Post-MiFID II" one was different in the clusters of instruments that have widened or narrowed their tick-sizes and the cluster of instruments that have maintained them unchanged. In other words, we are disentangling the sole Art. 49 impact from the overall effect of MiFID II. As in Chakravarty, Panchapagesan, and Wood 2005 we performed our analysis for the cluster of shares that have maintained the same tick-size pre-post MiFID II and the clusters of shares that have either narrowed or widened them. We test this hypothesis of a significant difference in the three groups performing a cross-section analysis per each market affected by the MiFID II and we have run a Fixed Effect Least-Squares Dummy Variable (LSDV) regression as in 1.1

$$
\begin{gathered}
Y_{it} = \delta_1 + \delta_2 * dummy_{widened} + \delta_3 * dummmy_{narrowed} + \delta_4 * dummy_{time} + \\
+ \delta_5 * dummy_{widened} * dummy_{time} + \delta_6 * dummy_{narrowed} * dummy_{time} + \\
+ \delta_{v1} * dummy_{venue} + \delta_{v2} * dummy_{increased} * dummy_{venue} + \\
+ \delta_{v3} * dummy_{narrowed} * dummy_{venue} + \delta_{v4} * dummy_{increased} * dummy_{venue} + \\
+ \delta_{v5} * dummy_{time} * dummy_{increased} * dummy_{venue} + \\
+ \delta_{v6} * dummy_{time} * dummy_{narrowed} * dummy_{venue} + \delta_n * ControlVariables_{n,i,t} + \epsilon_{it}
\end{gathered} \tag{1.1}
$$

Where the dependent variable, Y_{it} is the series of realised spreads (effective spreads, order imbalances or depths) 1-second means by instrument on the UK-MTF, $dummy_{widened}$ is a dummy variable that takes value of 1 for the group of instruments that have increased their tick-sizes and 0 otherwise, $dummy_{narrowed}$ is a dummy variable that takes value of 1 for the group of instruments that have reduced their tick-sizes and 0 otherwise, $dummy_{time}$ is a dummy variable that takes value of 1 in the period "post-MiFID II" and 0 otherwise. For the purpose of our analysis we will focus in particular on the behaviour of the coefficient δ_4 of interaction term $dummy_{widened} * dummy_{time}$.

Coefficients δ_v account for the venue specific effects in order to understand if minor venues were impacted with a different extent compared to the main lit market. There are several reasons for venues to differ among each other. For instance, they can be adversely affected by fragmentation, or they are characterised by the presence of more aggressive traders. Note that we introduce only 3 dummy venues (one for Turquoise, one for Cbo Europe, one for Aquis) to avoid multicollinearity. Therefore δ_{v1} is the differential intercept for each venue.

With $\delta_n\ ControlVariables_{n,i,t}$ we are controlling for Brexit announcements and ECB, FED and BOE monetary and inflation announcements. ϵ_{it} is the error term. In Appendix B we have also investigated our liquidity metrics through a time series analysis to determine the presence of a structural break in our time series at point in time fixed at the 21st of December 2017. The advantage of this approach is that we can test for multiple hypothesis at the same time. In Appendix B we compare this Fixed Effect model for our proxies computed at a different time horizon.

Metrics for implicit costs of transactions and viscosity

Transaction costs are an important component of the investor return and are characterized by an internal (spreads) and an external dimension (commissions/fees). We don't know the complete breakdown of commissions fees per exchange. As a result, our proxy for trading costs includes only implicit costs[19] and ignores dimensions of explicit, and opportunity costs. We measure them computing effective spreads and realised spreads.

We use effective spreads to measure implicit trading cost. They are defined as the absolute difference between the transaction price and the midpoint of the contemporaneous best bid offer (BBO). We calculate them using all valid quotes available in Refinitiv order books.

We identify if the trade was initiated by the buyer or seller using the Lee and Ready algorithm [20] in order to obtain a more accurate measure of the effective spread. As in Hendershott and Jones 2005, if the trade is initiated by the seller, the effective spread is the prevailing quote midpoint minus the execution price. For this reason, it can potentially be negative.

Realised spreads measure the difference between the transaction price and the quote midpoint at some time after the trade. We computed them with a 30 seconds and 30 minutes delayed mid-quote, in order to give to the market a shorter and a longer time to incorporate the information contained in the buy or in the sell. For liquidity takers, the realised spreads can be seen as an ex post measure of trading costs, whilst, for liquidity providers they are an ex post measure of gross profitability. The positive or negative sign is given again by the direction of the trade. Realised spreads are often used as a measure of a market's competitiveness.

[19] Implicit costs arise directly from trading because a large buy or sell order can temporarily push the price up or down.

[20] The algorithm classifies a trade as buyer initiated (+1) when the price is higher than the immediately previous midpoint, or if it is equal to midpoint (-1 if, otherwise is seller initiated). Price has to be higher than last price. Lee and Ready (1991)

When the tick size is too narrow, we can expect an increase in the passive execution latency. This is because investors can be discouraged from placing orders in the book. We attempt to measure the extent of this impact on the willingness of different market participants to provide liquidity. To do so, we estimate: profits, or costs, for passive and aggressive participants; High frequency traders [21] (HFTs) aggressiveness; and viscosity (lifetime of quotes, transaction sizes). The description of these metrics is reported in the appendix A.

Panel OLS regression for market participants

In the second part of our analysis we are interested in separating the impact of different tick size on the behaviour of market participants (MPs). Does a narrower tick-size decrease the probability of execution of a limit order and consequently reduce the incentive for a MPs to provide liquidity? Or do MPs submit more market orders when a narrower tick-size mechanically reduce their costs (decreased BBO)? We evaluate the net effect of a different tick size on a minor venue and on London Stock Exchange by performing a difference in differences (diff-in-diff). To do so we run a Panel OLS regression of the following form:

$$\begin{aligned} Y_{it} = \beta_0 &+ \delta_1 * dummy_{narrowed} + \delta_2 * dummy_{widened} + \delta_3 * dummy_{participant} \\ &+ \delta_4 * dummy_{narrowed} * dummy_{venue} + \delta_4 dummy_{widened} \times \ dummy_{venue} \\ &+ \delta_n \ ControlVariables_{n,i,t} + \varepsilon_{it} \end{aligned} \tag{1.2}$$

where our dependent variable represents each time a different proxy of market quality (realised spreads, the proportion of aggressive trades, mid-size transactions and quote duration) and where i denotes the stock, t denotes time, and n denotes the number of control variables. On this regard, it is worthy to remind that the trading volume of a cross-listed stock is proportionally higher on the exchange in which the cross-listed asset returns have greater correlation with returns of other assets traded on that market Baruch, Karolyi, and Lemmon 2007. We used few Control Variables to identify a specific day in which news could have influenced the market. Such variables include: the dummy for Brexit announcements, the dummy variable for ECB, FED and BOE monetary and inflation announcements. Appendix A presents the RHS variables description, sources and the computation methodology.

The coefficient δ_3 captures the interaction of venue specific effect and divergences in the tick size applied to the instrument and thus estimates any incremental effects of the new minimum tick size regime. Hence, δ_3 reflects the change in profit/costs for liquidity provision in instrument traded at a smaller tick size compared to profit/costs for liquidity provision in instrument traded at the prescribed ESMA tick size.

Measuring Price Discovery

We also investigated if UK venues contribute in a different proportion to price discovery following the implementation of MiFID II in general and the tick size regime (Art.49) in particular. We refer to price discovery in terms of inter-market price adjustments among the UK venues and the Swiss Exchange. This analysis exploits an important property of time series: the so called co-integration. Two-time series are

[21] "Professional market participants trading in a proprietary capacity. They are a sub category of algorithmic traders characterised by high-speed in generating, routing, and executing orders and usually end the trading day in as close to a flat position as possible" (SEC, 2010)

TABLE 1.2: Market share by firm type across venue and control and treatment group.
Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

	LSE		CBOE		AQX		TRQ	
Market share (%)	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**	**Treatment**	**Control**
Agency Broker	1.29	0.59	1.88	0.72	1.63	1.74	0.91	0.56
Asset Manager	0.34	0.69	0.28	0.77	0.4	0.26	0.32	0.48
Fund Manager	2.36	2.04	1.66	2.39	1.25	1.41	1.76	1.43
IB - Large Dealer	17.99	20.19	10.28	12.66	11.26	12.29	18.3	19.57
IB - Small or Medium	0.22	0.34	0.49	0.37	0.39	0.56	0.27	0.29
INTC	0.38	0.52	0.7	0.62	0.83	1.14	0.25	0.4
Interdealer Broker	1.82	2.05	0.79	1.45	0.85	0.87	0.75	1.12
Market Maker	12.66	11.31	11.67	11.07	9.01	5.17	7.06	7.52
Private / Commercial Bank	0.7	0.66	0.42	0.28	0.4	0.26	0.32	0.3
Prop Trader - HFT	19.27	16.82	27.67	22.45	17.54	16.9	23.82	21.86
RSP	0.27	0.17	0.29	0.32	0.11	0.21	0.22	0.15
Retail	0.01	0.01	0	0	0	0	0	0
SI	42.68	44.6	43.86	46.91	56.34	59.19	46.03	46.3

TABLE 1.3: Panel A: Summary Statistics. Trading characteristics. Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

		XLON	CBOE	TRQ	AQXE
Price	No. Obs.	3,935,842	646,815	1,214,484	198,499
	Avg.	15.7	17.05	16.16	17.78
	St. D.	15.82	17.09	16.35	16.56
	25thQ	4	3.97	3.94	4.87
	75thQ	21.82	25.64	22.66	25.68
Volume	Avg.	3,463	2,359	1,409	1,017
	St. D.	38,961	7,377	8,245	2,256
	25thQ	357	267	154	118
	75thQ	2,913	2,028	1,181	1,007
Aggressive Rate	Avg.	0.29	0.33	0.37	0.38
	St. D.	0.22	0.21	0.24	0.18
	25thQ	0.12	0.17	0.17	0.23
	75thQ	0.4	0.43	0.51	0.51
Transaction Midsize	Avg.	885	583	570	480
	St. D.	1,260	816	862	554
	25thQ	242	145	154	130
	75thQ	957	700	621	685

said to be co-integrated if a long-run relationship exists between them. Since we are considering the same equity instrument traded on multiple venues, we can expect its price on one venue to share the same stochastic trend with its price on any other venue due to the low of one price.

We built a price discovery metric called Information Leadership Share (ILS) as described in Putniņš 2013. It builds up from the common view of price discovery as the *"who moves first"* interpretation and it combines Hasbrouck Information Share (IS) Hasbrouck 1995 and Harris-McInish-Wood Component Share (CS) (Deb. Harris, Mcinish, and Wood 2002). The latter metric embodies the contribution of each exchange in the update of price given a weighted factor contribution to innovations underlying the common trend. ILS improves on the aforementioned measurements for it has the ability to correctly attribute contributions to price discovery in the presence of different levels of noise. Noise can arise for different reasons. We are particularly interested in detecting the noise produced by the new discrete grid of prices under Art. 49. We expect exchanges with higher spreads to be also slower and noisier, and to follow innovations rather than contribute on new information. Two preliminary steps are required by the ILS: first we needed to compute the Information Share that focuses on the variance of innovation to the common factor. Second, we followed Gonzalo-Granger PT analysis based on the error correction process. The measurements for each pair of markets analysed are obtained, we have followed Yan and Zivot (Yan and Zivot 2010), using IL to measure which price leads the price innovation adjustment in the fundamental value:

$$IL_i = \left| \frac{IS_i}{IS_j} \frac{CS_j}{CS_i} \right| \tag{1.3}$$

$$IL_j = \left| \frac{IS_j}{IS_i} \frac{CS_i}{CS_j} \right| \tag{1.4}$$

Where i and j are pairs of the analysed market in the list [AQX, TRQ, LSE, CBOE, SIX]. We have calculated the information leadership share of each pair and we have compared it in the sample pre- and post-MiFID II. Technical derivations of the VECM model are derived in Appendix A.

TABLE 1.4: Panel B: Summary Statistics. Order book characteristics. Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

		XLON	CBOE	TRQ	AQXE
Bid Price	**Avg.**	15.7	17.05	16.15	17.78
	St. D.	15.82	17.09	16.35	16.55
	25thQ	4	3.96	3.94	4.86
	75thQ	21.82	25.64	22.65	25.71
Bid Size	**Avg.**	18,170	7,601	3,431	3,985
	St. D.	77,285	22,350	10,919	9,200
	25thQ	1,386	653	325	434
	75thQ	13,722	5,929	2,678	3,796
Ask Price	**Avg.**	15.71	17.06	16.16	17.79
	St. D.	15.83	17.1	16.36	16.56
	25thQ	4	3.97	3.94	4.87
	75thQ	21.83	25.64	22.67	25.72
Ask Size	**Avg.**	18,541	7,659	3,544	4,104
	St. D.	74,995	23,485	12,589	9,746
	25thQ	1,389	644	325	445
	75thQ	13,992	5,885	2,696	3,892
Mid-quote	**Avg.**	15.7	17.06	16.16	17.78
	St. D.	15.82	17.09	16.35	16.56
	25thQ	4	3.97	3.94	4.87
	75thQ	21.83	25.64	22.66	25.72
Quoted Spread	**Avg.**	5.51	6.74	8.01	5.79
	St. D.	5.97	22.26	15.81	7.88
	25thQ	2.45	2.7	2.98	2.9
	75thQ	6.51	7.28	8.61	7.08
Spread	**Avg.**	0.01	0.01	0.01	0.01
	St. D.	0.01	0.02	0.03	0.03
	25thQ	0	0	0	0
	75thQ	0.01	0.01	0.01	0.01
First Level Depth	**Avg.**	6,097	3,441	2,262	3,749
	St. D.	11,914	5,933	4,220	5,548
	25thQ	1,201	633	427	672
	75thQ	6,383	3,471	2,249	4,409
Effective Spread	**Avg.**	5.38	4.29	5.22	194.57
	St. D.	6.12	16.61	12.31	419.33
	25thQ	2.27	1.4	1.47	3.33
	75thQ	6.41	4.63	5.77	264.44

TABLE 1.5: Panel B: Summary Statistics. Order book characteristics cont. Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

		XLON	CBOE	TRQ	AQXE
Simple Price Impact	**Avg.**	5.16	4.72	5.29	192.16
	St. D.	17.55	23.47	26.41	427.1
	25thQ	-	-0.37	-0.52	2.05
	75thQ	9.48	8.1	9.22	264.77
Price Impact (5min)	**Avg.**	5	5	5	192
	St. D.	31	34	36	427
	25thQ	-6	-3.98	-4.85	2.53
	75thQ	15	12	14	266
Order Imbalance	**Avg.**	-111	3	-26	-45
	St. D.	7688	3365	2597	2191
	25thQ	-828	-397	-273	-400
	75thQ	765	402	264	348
Quote Duration	**Avg.**	0	0	1	0
	St. D.	9	9	22	18
	25thQ	0	0	-	0
	75thQ	0	0	0	0
Realised Spread (30sec)	**Avg.**	0.22	-0.48	-0.1	2.3
	St. D.	17.06	19.42	24.67	84.71
	25thQ	-4.69	-4.35	-4.64	-3.71
	75thQ	5.56	3.79	4.8	5.96
Realised Spread (30min)	**Avg.**	0.36	-0.41	-0.08	2.05
	St. D.	31.12	30.64	34.03	78.98
	25thQ	-9.69	-7.97	-8.85	-8.49
	75thQ	10.68	7.47	8.95	10.88

1.4 Results

Narrower Tick Sizes are correlated with higher incentives for Market Makers to provide liquidity. On the other side, narrower Tick Sizes are detrimental to order book stability. We found that narrow tick sizes are associated with shorter orders' lifetime and a lower transaction mid-size. We have also found that all the UK MTFs contribute less than before to price information in the presence of different levels of noise produced by the new discrete grid of prices under Art. 49 and their widened tick sizes.

1.4.1 Overall impact on costs and liquidity

As we explained in the literature review, one first effect we expect to find is a trade-off between depth at touch and quoted spreads. Whilst the spreads decrease following a tick size reduction, at the same time also the depth suffers of a contraction. This is because market orders become cheaper and more appealing to investors rather than limit ones. This contraction for larger trades may even imply an increase in transaction costs. Table 5 reports the results for the Fixed Effect Least-Squares Dummy Variable Model (LSVD) described in equation 1.1. The advantage of this approach is that we can test for multiple hypothesis at the same time. There are 67094 included observations for four venues in 80 trading days for 318 equity instruments. The overall R-squared ranges between 0.2 and 0.7.

The headings of the columns represent the dependent variable we are considering. The constant (δ_1 in 1.1) is the intercept value of instruments that had not change their tick sizes on the London Stock Exchange. Dummy widened (δ_2) and dummy narrowed (δ_3) tells us how much the intercept value of the instrument with a widened (narrowed) tick size differed from the securities that maintained it equals already before the change. The results are significant with a p-value < 0.001 and tell us that instruments with a widened tick were already around 2.7 ticks larger than instruments that had not changed the tick and that instruments with a narrowed tick were already around -0.7 ticks.

The depth at touch is narrower for instruments with a wider tick size and wider for instruments with a narrower one. The intercept of the overall change before the pre-post MiFID II is expressed by the differential intercept dummy δ_4. It tells us that both the effective spreads and the spreads expressed in ticks reduced overall respectively of around -3.4 bps and -2 ticks. The change in the depth at touch is not statistically or economically significant.

We are particularly interested in the interaction terms δ_5 and δ_6 to disentangle the effect of the tick over time. We notice how the orderliness requested by the new grid is obtained through the reduction of the spreads expressed in tick for instruments that have widened their tick size (-2.6 ticks) and, vice versa, the increase for the instruments that have narrowed the tick (+1 tick). This orderliness is not reflected in the effective spreads change that results in a +2.14 bps after the MiFID II for the group of widened instruments and -2.38 bps for the group of narrowed securities. As we were expecting the depth at touch change accordingly: it improves after the 3^{rd} of January for instruments that have widened their tick and narrowed for the instruments that have narrowed their tick sizes.

With the same LSDV fixed-effect model we investigated the venue specific effect of the change, to test the hypothesis that minor venues are more adversely impaired by the change in tick. We found that these venues have increased their effective spreads in both groups of instruments.

1.4.2 Market makers' incentives

Hypothesis 1.1 *Tick sizes constitutes an incentive for market makers to provide liquidity on the market, especially for the most illiquid instruments.*

We investigate the impact on realised spreads and effective spreads. All else being equal, wider spreads lead to greater profits for market makers. We used effective spreads to measure implicit trading cost, whilst realised spreads are used to measure the difference between the transaction price and the quote midpoint at some time after the trade.

Table 1.7 presents differences in mean of realised spreads at 30 minutes and 30 seconds for Market Makers, Fund Managers and HFTs for securities that traded with the same (control) and different (treatment) tick size across venues. In this way, we aim to disentangle the *real* information that may be contained in trade execution, whilst incorporating the predominance of the HFTs in the market. To read this table, we need to remember that the treatment groups include instruments whose tick size on minor MTFs presented consistent mismatch (narrower) compared to the tick size on London Stock-Exchange.

We have found that realised spreads after the change are on average lower on the London Stock Exchange than on Turquoise, Aquis and CBOE. In relation to the tick size we have found that 30 seconds realised spreads of Proprietary traders, high frequency traders are higher for securities in the treatment group with a narrower tick size and lower for the same group of securities for Market Makers and Buy Side participants. The difference in differences reported in Table 1.7 shows consistently

TABLE 1.6: Parameter Estimation of LSVD model.Each column refers to a different regression based on the liquidity metric considered. Robust standard errors are in italic font. Superscripts ***, ** and * denote significance at 1%, 5% and 10%, respectively.

Fixed Effect LSVD Parameters Estimation	**Spread in tick**		**1L depth**		**Effective spread**		**5 min price impact**	
constant	3.0979***		6603.7***		7.5559***		0.42***	
	0.4985		*111.5*		*0.46*		*0.0299*	
dummy widened	2.7427	***	-2252.7	***	9.8831	***	0.1356	***
	0.142		*214.62*		*0.3189*		*0.0274*	
dummy narrowed	-0.7135	***	4319.3	***	0.6776	***	0.055	
	0.116		*221.35*		*0.1328*		*0.0338*	
dummy Post MiFID II	*-2.2672*	***	*-18.845*		*-3.4043*	***	*-0.1532*	**
	1.3008		*78.287*		*1.2839*		*0.0514*	
interaction widened post	-2.6835	***	2015.6	***	2.1399	***	0.0466	*
	0.2404		*262.21*		*0.4565*		*0.0283*	
interaction narrowed post	1.0049	***	-5002.9	***	-2.3754	***	-0.0703	***
	0.2264		*224.6*		*0.2227*		*0.0334*	
Aquis	0.5403	***	-3775	***	20.431	***	-0.1657	***
	0.1242		*147.51*		*2.3632*		*0.0561*	
Turquoise	2.8222	***	-4510.6	***	8.8753	***	0.0003	
	0.2093		*100.84*		*1.8725*		*0.0768*	
CBOE	3.5379	**	-3741.4	***	0.956	***	-0.1311	***
	2.0862		*104.8*		*0.1986*		*0.014*	
interaction Aquis widened	-0.1037		1494.4	***	-21.781	***	-0.3729	***
	0.4394		*335.52*		*3.3731*		*0.0765*	
interaction Turquoise widened	5.2681	***	1371.4	***	5.2079	**	1.1408	***
	0.7745		*217.56*		*2.422*		*0.2639*	
interaction CBOE widened	27.782	**	1495.2	***	26.07	***	0.6955	***
	14.394		*224.54*		*2.2207*		*0.2131*	
interaction Aquis narrowed	-0.8757	***	-3009.1	***	-21.042	***	-0.2921	***
	0.1245		*286.8*		*2.4139*		*0.0644*	
interaction Turquoise narrowed	-2.2557	***	-3553.2	***	-10.416	***	-0.2418	***
	0.2141		*227.85*		*1.8756*		*0.0831*	
interaction CBOE narrowed	-1.879		-3123.2	***	-2.8668	***	-0.0955	***
	2.5375		*235.07*		*0.23*		*0.0344*	
interaction Aquis widened post	2.9264	***	-351.17		163.18	***	2.8807	***
	0.9875		*1114.3*		*23.448*		*0.6488*	
interaction Turquoise widened post	-4.9686	***	-1622.6	***	-6.4099	***	-1.1176	***
	0.7497		*263.29*		*1.7129*		*0.2581*	
interaction CBOE widened post	-27.838	**	-1578.3	***	-9.3535	***	-0.0304	
	14.24		*274.65*		*2.4646*		*0.2249*	
interaction Aquis narrowed post	0.969	***	4224.7	***	239.53	***	2.4216	***
	0.2168		*414.65*		*17.414*		*0.212*	
interaction Turquoise narrowed post	0.8761	***	4026.9	***	1.8624	***	0.0219	
	0.0776		*229.7*		*0.1724*		*0.0328*	
interaction CBOE narrowed post	-1.2168		3497.7	***	1.1207	***	0.0239	
	1.4453		*236.86*		*0.1589*		*0.0325*	
VIX	-0.0501		-13.015	***	0.0033		0.0044	***
	0.037		*3.9372*		*0.0339*		*0.002*	
F-statistic robust	2390		1456.8		3300.2		28620	
R-overall	0.5		0.2		0.6		0.5	
No of Obs	460899		460899		460899		460899	

an increase in 30 seconds realised spread for HFTs on minor venues compared to HFTs on the main lit market when they trade securities in the treatment group.

These results indicate a first possible unintended consequence of the prescribed ESMA tick size implemented by the London Stock Exchange. They result in a decline in cost of transactions in favour of HFTs and a minor incentive in the provision of liquidity on the main lit market.

Table 1.7 shows that:

a Realised spreads after the change are on average lower on the London Stock Exchange than on Turquoise, Aquis and CBOE.

b Profits for of proprietary traders, high frequency traders (HFTs) are on average higher on the smaller venues compared to the main lit exchange for securities in the treatment group.

c 30 minutes realised spreads of Market Maker and Buy Side Participants, are higher for securities with a narrower tick size on minor venues compared to the London Stock Exchange.

Hypothesis 1.2 (H1.2) *Tick sizes are detrimental to order book viscosity, thereby exposing the order book to chaotic price formation processes.*

Viscosity is a desired characteristic of order books, because it represents their ability to enhance price discovery and reduce the level of noise. Table 1.8 shows the mean differences in quote duration and mid-size transaction for Market Makers, Fund Managers and HFTs in respect of securities that traded with the same (control) and different (treatment) tick size across venues.

We found that on Aquis, CBOE and Turquoise the life of orders (i.e. how long the best bid(ask) remains in the order book before being executed or substituted by a better quote) associated with a trade from Market Makers is higher when the instrument is allowed to trade at a lower tick size. On Aquis, CBOE the life duration of quotes associated to proprietary traders – HFTs is however shorter.

This result suggests that ESMA tick size applied on the London Stock Exchange was not optimal for order book stability.

Hypothesis 1.3 (H1.3) *High frequency traders (HFTs) activity is intensified in relation to narrower tick sizes, reflecting a greater exposure to the undercutting risk.*

Table 1.9 displays the percentage of overall turnover by firm type across venues. The market shares of market makers were higher for instruments with narrower tick sizes across all venues. However, looking at the diff-in-diffs, HFTs market share is 2.78% larger for instruments in the treatment group compared to the instruments in the control on CBOE, whilst their market share for the treatment group is lower on Aquis and Turquoise (respectively -1.81% and -0.49%). Thus, the impact of the HFTs activity is not univocal.

See also parameter estimation for Panel OLS model in Appendix B.

1.4.3 Price discovery

Hypothesis 1.4 (H1.4) *UK MTFs contribute less than before to price information in the presence of different levels of noise produced by the new discrete grid of prices under Art. 49 and their widened tick sizes.*

TABLE 1.7: Mean differences in realised spreads (at 30 sec and 30 min)
Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

	Buy Side	Market Maker	Prop Trader - HFT	Buy Side	Market Maker	Prop Trader - HFT	Buy Side	Market Maker	Prop Trader - HFT
	Control			**Treatment**			**Diff-in-diffs**		
	30min Realised Spread (bps)								
XLON	8	13	21	5	3	12			
AQXE	15	33	48	14	26	34	2	3	-5
CBOE	1	19	37	9	13	22	11	4	-6
TRQ	13	24	50	9	20	38	-1	6	-3
	30sec Realised Spread (bps)								
XLON	7	13	36	6	12	30			
AQXE	22	34	47	13	26	55	-8	-7	14
CBOE	12	24	25	7	17	26	-4	-6	7
TRQ	14	29	37	9	22	41	-4	-6	10

TABLE 1.8: Mean differences in quote duration and mid-size transaction.
Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

	Fund Manager	**Market Maker**	**Prop Trader - HFT**	**Fund Manager**	**Market Maker**	**Prop Trader - HFT**	**Fund Manager**	**Market Maker**	**Prop Trader - HFT**
	Control			**Treatment**			**Diff-in-diff**		
	Quote duration								
XLON	0.07	0.27	0.14	0.06	0.18	0.15			
Aquis	0.04	0.06	0.07	0.05	0.09	0.05	0.02	0.11	-0.03
CBOE	0.09	0.13	0.08	0.32	0.33	0.08	0.24	0.29	-0.01
Turquoise	0.59	1.32	0.35	0.8	1.34	0.43	0.21	0.11	0.06
	Midsize transaction								
XLON	763.73	829.09	864.57	1653.38	1578.33	1343.73			
Aquis	246.05	428.5	511.37	684.12	752.95	523.01	-451.58	-424.8	-467.52
CBOE	509.19	744.46	551.26	1164.33	1805.35	1053.63	-234.51	311.65	23.21
Turquoise	519.75	580.6	567.72	1054.18	1292.77	1089.54	-355.22	-37.08	42.66

TABLE 1.9: Market Share (%) by firm type across venues.
Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

	Fund Manager	**Market Maker**	**Prop Trader - HFT**	**Fund Manager**	**Market Maker**	**Prop Trader - HFT**	**Fund Manager**	**Market Maker**	**Prop Trader - HFT**
	Control			**Treatment**			**Diff-in-diff**		
	Market Share (% Turnover)								
XLON	2.04	11.31	16.82	2.36	12.66	19.27			
Aquis	1.41	5.17	16.9	1.25	9.01	17.54	-0.48	2.5	-1.81
CBOE	2.39	11.07	22.45	1.66	11.67	27.67	-1.05	-0.75	2.78
Turquoise	1.43	7.52	21.86	1.76	7.06	23.82	0.01	-1.82	-0.49

Table 1.10 presents the contribution to price creation across all markets considered in pairs. ILS for the instruments that have narrowed their tick sizes has improved with respect to the SIX Exchange in the period from pre-MiFID II to post-MiFID II implementation (compare panel C and Panel F). The results for the control group (compare Panel A and Panel D) are ambiguous. Information leadership share of Aquis Exchange and BATS-CBOE seem to have improved over that of SIX. Also, LSE has maintained its predominance, while on Turquoise, the ILS is reduced. The most interesting comparison for the purpose of our research question regards the comparison between Panel B and panel E. All the UK MTFs contribute less than before to price information in the presence of different levels of noise produced by the new discrete grid of prices under Art. 49 and their widened tick sizes. We confirm our expectations on the correlation between higher spreads and slower and noisier price series.

1.5 Conclusions

In this study, we have evaluated the impact of the MiFID II tick size regime (Art 49) on four of the UK trading venues that have implemented it in January 2018. Using FCA data, we have investigated the different impact of a tick size change on venues distinguishable by a different level of activity of market makers, HFTs and liquidity providers.

The implementation of the new regulation gave us the opportunity to examine tick size changes in both directions at the same time, thus naturally controlling for overall market conditions. We found that narrower tick sizes, all other things equal, coincide with higher incentives for market makers to provide liquidity, as well as with a deterioration of a desired order viscosity. They were in fact associated with shorter orders' lifetime (i.e. how long a quote remains in the order book before the order is fulfilled) and a lower transaction mid-size.

As far as the activity of market makers and HFTs is regarded we found higher shares of turnovers associated with narrower tick sizes. Finally, we considered the price formation process of cross-traded instruments on UK venues and on Swiss exchange finding that the price for cross listed securities updated first on the Swiss Exchange followed by movements on UK venues.

Our finding suggests that minor unintended consequences of the new regime impacted the market quality of UK venues and could open-up questions on the calibration of the transparency calculations. For some instruments, the tick size is too narrow to ensure the order books stability and to protect investors from undercutting risks by high frequency traders.

TABLE 1.10: Information Leadership Share (ILS) across venues pre and post MiFID II.
Source: Author's calculations based on FCA proprietary dataset and Refinitiv data.

Information Leadership Share (ILS)											
Pre-MiFID II						**Post-MiFID II**					
	AQX	**TRQ**	**LSE**	**CHI-X**	**SIX**		**AQX**	**TRQ**	**LSE**	**CHI-X**	**SIX**
Control											
AQX		96.3	0.09	99.84	19.14	**AQX**		96.41	62.27	12.72	40.87
TRQ	3.7		59.95	99.5	37.63	**TRQ**	3.59		64.51	99.97	32.16
LSE	99.91	40.05		99.95	100	**LSE**	37.73	35.49		97.57	100
CHI-X	0.16	0.5	0.05		92.45	**CHI-X**	87.28	0.03	2.43		99.98
SIX	80.86	62.37	0	7.55		**SIX**	59.13	67.84	0	0.02	
Widened											
AQX		97.07	75.18	77.2	8.48	**AQX**		4.27	98.68	56.67	1.26
TRQ	2.93		24.82	82.62	74.04	**TRQ**	95.73		95.5	96.91	0
LSE	99.8	0.17		79.8	96.55	**LSE**	1.32	4.5		0.3	81.41
CHI-X	22.8	17.38	20.2		100	**CHI-X**	43.33	3.09	99.7		16.01
SIX	100	25.96	3.45	0		**SIX**	98.74	100	18.59	83.99	
Narrowed											
AQX		22.33	51.01	42.18	12.28	**AQX**		83	0.17	98.43	91.7
TRQ	77.67		46.37	78.97	2.01	**TRQ**	17		0.02	100	11.16
LSE	48.99	53.63		39.65	69.19	**LSE**	99.83	99.98		99.67	100
CHI-X	57.82	21.03	60.35		83.68	**CHI-X**	1.57	0	0.33		80.82
SIX	87.72	97.99	30.81	16.32		**SIX**	8.3	88.84	0	19.18	

Chapter 2

Is a better tick-size grid possible for UK venues? A recalibration based on a supervised learning approach

Co-authors: This research article was co–authored with my supervisors, Prof. Flavio Bazzana and Prof. Andrew Lepone, who contributed in designing the research hypotheses of this study and in interpreting the results.

Abstract

In the last two years, market regulators have been trying to quantify the MiFID II impact on market quality. One particularly controversial aspect of this regulation has been the adoption of a new tick size regime (Art. 49). The Financial Conduct Authority (FCA) in the UK, as well as the Autorité des Marchés Financiers (AMF) in France, have conducted assessments on the tick size regime, which entered into force in January 2018 as part of MiFID II, concluding the regime had an opposite impact on the overall market quality of UK and EU venues. Unintended consequences on UK venues have been measured, and the regime resulted not perfectly calibrated for UK equities. This consideration open-up to whether a better methodology can be constructed to improve the effectiveness of the regime. In this paper, we adopt a supervised machine learning approach to propose a better-calibrated alternative to ESMA grid. Our approach is based on: (i) market capitalization; and (ii) quoted spread. Having defined an optimal tick sizes in terms of market orderliness, we show how our proposed calibration for the regime would achieve an optimal tick size for equities three times more frequently than the current ESMA regime. This allows us to outline an idealized grid for determining an equity's minimum tick size for this proposed regime. This paper is especially relevant for UK policy makers in the context of the UK leaving the EU and suggests the ESMA grid can be abandoned. At the best of authors' knowledge, it is also the first time a supervised machine learning model is adopted to evaluate policy implications of financial regulation on secondary markets.

Keywords
MiFID II – Tick Size regime – Multi-classification problem – Parameter Tuning

2.1 Introduction

Electronic market platform fixes a grid on which traders can place their prices. The grid step represents the minimum increment equity or equity liked instruments are quoted and is called the tick value (measured in the currency of the asset). To intuitively understand its mechanism, imagine that a StockABC is currently bought at 20 GBP. If we fix its tick size at 0.002, it means we can place a buy order at 20.002, but we can never trade this stock at 20.001. This 0.002 is the minimum meaningful cost that an investor must pay to see their order executed ahead of the best offer presented in the order book.

Before MiFID II coming into force on the 3rd of January 2018, for equity instruments on UK venues, as in many other markets, the spacing of the grid was dependent on the instrument's price: the higher the price, the wider the tick value. Prior to MiFID II, trading venues had discretion over the sizes of the ticks. In practice, they may have been incentivised to reduce their tick sizes to attract HFTs and increase their market share. However, the price is not the only important variable to consider when we are interested in determining an instrument tick size. There are more empirical and qualitative aspects traders take into account when they estimate if a tick size is too narrow or too wide: such as the tick value, the price, the daily volume and value traded in the asset and their own trading strategy. Tick sizes influence both the liquidity and the price formation process, therefore, in order to provide a more representative grid, ESMA proposed a system based not only on prices but also on the average daily number of transactions (ADNT).

The overall objective of the tick size regime under MiFID II is to maintain the orderliness of equity secondary trading. This is done by setting the tick sizes on equities. Orderliness is maintained by ensuring that orders do not become undercut by only very fine amounts, for example by a high frequency trader (HFT). If allowed to take place, traders can continue to undercut each other within the order book, at minimal cost to themselves.

Article 49 of MiFID II [1] requires trading venues to adopt minimum tick sizes in relation to equity and equity-like instruments [2]. The minimum tick size to be applied depends on the Equity liquidity band and price level. ESMA liquidity band are computed based on the average daily number of transactions (ADNT) on the most relevant market (MRM) in the EU. The ADNT is automatically calculated and published by the Financial Instruments Transparency System (FITRS), a database operated by ESMA, based on quantitative information received from EU trading venues and national competent authorities (NCAs) [3].

Tick-size grid by ESMA was designed following two rules: on the one hand, maintaining the ADNT unchanged, the tick-size is expected to be directly correlated with the price, on the other hand, when the price is kept constant, tick-sizes should show an inverse correlation with the ADNT.

In this work, we argue that neither the ADNT nor the price, are the best factors to determine the tick sizes. First, the ADNT is computed on yearly base, consequently impairing its representativeness of the daily liquid status of an equity instrument. Second, the relationship among tick sizes and these factors individually taken is non

[1]MiFID II/MiFIR is a new legislative framework that will strengthen investor protection and improve the functioning of financial markets, making them more efficient, resilient and transparent. Source: ESMA website

[2]Specified by the Regulatory Technical Standard (RTS11)

[3]See Annex 2 of Article 49 https://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160714-rts-11-annex_en.pdf

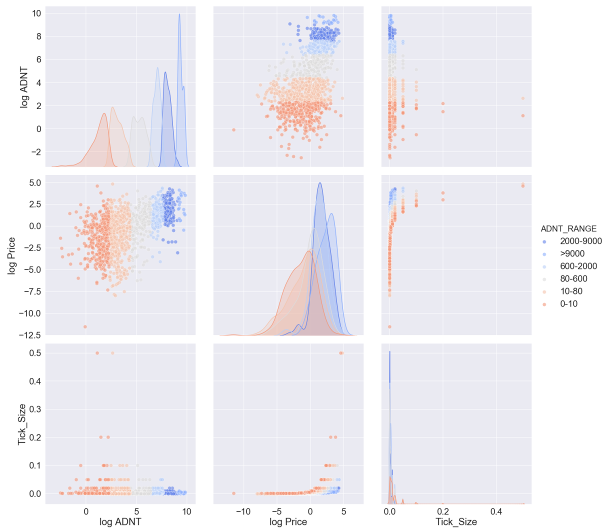


FIGURE 2.1: Non Linearities between tick sizes, ADNT and prices. Source: Author's calculations based on ESMA and Bloomberg data.

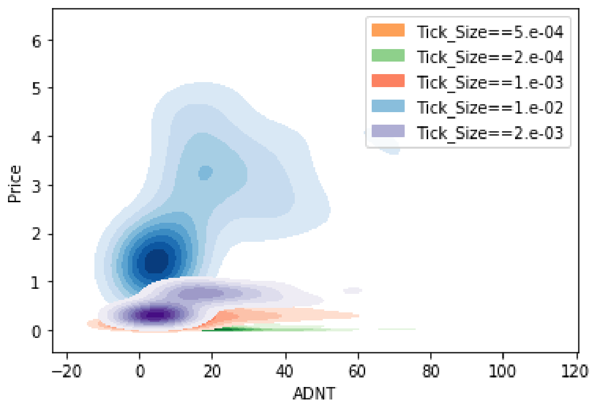


FIGURE 2.2: Bivariate kernel density estimate of tick sizes shows an overlapping in the region of the most illiquid instruments. Source: Author's calculations based on ESMA and Bloomberg data.

linear(see Figure 2.1) and can not, therefore, be described by linear models. There is a wide dispersion in the ADNT distribution that shows an average of approximatively 776 daily transactions and a more than twice as large standard deviation of 1968. The price displays a similar pattern with values ranging from a minimum of £ 0.15 to a maximum of roughly £ 121.

Third, the kernel density estimation in Figure 2.2 shows how the probability distribution of the most illiquid instrument against the ADNT and the price overplots and makes effectively impossible to determine the distribution from which the tick for these instruments are extracted.

Finally, the legislative aim was to "ensure the orderly functioning of the market" [4]. Nevertheless, concerns arose from some venues about unintended consequences of this standardization.

a Concerns around the contraction of liquidity for cross-listed securities as a result of a transfer of trading on non-EU venues.

[4]ESMA Q&A paragraph 4) Tick Size Regime – Last Update (18/12/2017)

b Concerns about the determination of instruments' liquidity by ESMA, that calculates the average daily number of transactions (ADNT) on yearly base.

Because of these concerns, the first months after the new regime enforcement were characterized by inconsistencies in the application of the tick size scheme among different venues. To address this issue, a second mandatory harmonized regime was enforced on the 1st of April 2018, followed by a proposal for amendments at the end of 2018 . This second implementation and the adjustments required by European multilateral trading facilities, are signals that Art. 49 was not a simple "standardization" as was intended by ESMA regulators.

In this work we treat the tick size calibration as a multi-class classification problem. We train five different supervised learning models(K-Nearest Neighbours, Support Vector Machines, Artificial Neural Network, Random Forest and Gradient Boosted Decision Trees) to select the most relevant features that can be used in order to determine for each instrument, a tick size that maximizes the market orderliness. We are consequently able to redesign a tick-size grid that can be used on UK venues as an alternative to the one proposed by ESMA.

The reminder of this paper is organized as follow: Section 2 will review three streams of literature we looked at to design our study: paragraph 2.1. will focused on the tick size literature, with a focus on optimum tick size in paragraph 2.2. Section 2.3. will discuss machine learning techniques applied to the resolution of classification problems in the field of policy evaluation. Section 3 will describe data and methodology applied to test the two hypotheses and will examine the base model and the alternatives. Findings will be finally presented along with preprocessing and cross-validation steps undertaken in Section 4. Section 5 sums up the conclusions and quantifying the costs and benefits of the new proposed tick-size grid.

2.2 Literature review

In this section, we depict the technical and regulatory background. First, we briefly summarise, the empirical literature findings on the impact of a change in tick sizes. These are case studies, primarily focusing on the US pilot in 2001. Second, we present the theoretical microstructure model for optimum tick-size calibration. Third, we mention the most recent machine learning applications to policy evaluation.

2.2.1 Tick size literature

Minimum tick sizes have been criticized by some authors (see Ricker 1996 or Peake 1993) because they can increase market-maker's pay-offs as well as investor's trading-costs. Binding tick sizes can also, as shown in Budish, Cramton, and Shim 2015, result in increased competition on the speed that ultimately eliminates undercutting effects. The bulk literature explores the impact of the change in the tick-sizes. The impact of an increased tick size was observed for the first time in the US tick size pilot in 2001, when SEC adopted a decimal pricing and increased tick sizes for small caps (Rindi and Werner 2019). The overall effect of this policy was an improvement in the liquidity of these instruments. This happened because, in addition to the mechanical effect on spreads, the wider tick sizes also increased the incentive for market makers to support small caps and to provide liquidity (Werner et al. 2019).

Several studies have documented a correlation between a reduction of tick sizes and a decrease in quoted spreads (Bessembinder 2000, Goldstein and A. Kavajecz 2000, Jones and Lipson 2001, O., Schossmann, and Veverka 2004). Nevertheless, the overall impact of a change in tick sizes is not conclusive. A trade-off between depth and spreads has been extensively illustrated. Whilst the spreads decrease following a tick size reduction, at the same time also the depth suffers from a contraction. This is because market orders become cheaper and more appealing to investors rather than limit ones. This contraction for larger trades may even imply an increase in transaction costs (see Hsieh, Chuang, and Lin 2008 for an example in an order-driven emerging market or Bollen and Whaley 1998 for the analysis of a quote-driven developed exchange).

The impact of a new tick size regime has to be especially assessed considering to what extent a market is fragmented and consequently inclined to display higher trading costs and reduced depth. Hendershott and Jones 2005 have also highlighted how fragmented exchanges tend to be less price informative because of their increasing searching costs.

In a previous study, we have shown how a change in tick sizes has a different impact on market quality beyond the well-known trade-off between depth at touch and quoted spreads (Nuzzo 2020). That study was preparatory to this research, in which we show how the impact of a change in tick sizes depends on the market intrinsic characteristics, and it should not be considered as a simple standardization as in the stated scope of ESMA regulators.

2.2.2 Optimum tick size

MiFID II aimed to harmonize tick values across exchanges and asset classes in the EEA and this harmonization required standardisation of tick sizes around an optimum tick size level. Optimum tick size is a tick size that is not too wide nor too narrow (Harris 1999). When the tick value is too small, one tick is insignificant and market participants are incentivised to frequently change marginally the price of their limit orders to gain price priority, which can be very discouraging for market makers and can ultimately lead to more unstable order books and disperses liquidity through too many price points, delaying liquidity provision. On the other hand, the issue with a wider tick size is that it makes crossing the spread too expensive and, consequently, it leads to longer and slower queues. This queue effect is discouraging for investors who can repeatedly get beaten to the top by HFTs with a faster connectivity. According to Eisler, Bouchaud, and Kockelkoren 2012 a tick size can be considered large when the bid-ask spread for the stock is almost always equal to one tick.

The tick value on the exchange is correlated with many properties of the assets. Some quantities such the volatility or the daily traded volume are macroscopic and should remain constant after the change in the tick value, other fundamentals variables, instead, change along with the tick value as it is the case of the daily number of trades. The dimension of the implicit spread, is of great importance for large tick assets. Has been shown that for these securities the effective spread is almost always equal to one tick (Dayri and Rosenbaum 2012). According to the authors the optimum tick size can be derived from the intrinsic market micro-structure properties of the asset. They moved from two seminal models that link the spreads to the volatility, namely Madhavan, Richardson, and Roomans 1997 and Wyart et al. 2007 and derive their econometric relationship between the tick size and the implicit spread. They consider a tick value to be optimal if (i) the average cost of a market and a limit order are equals and are close to zero, (ii) the quoted spread is stable and close to one tick. This situation in fact can be desirable both for liquidity takers and liquidity providers.

In order to design the optimum tick size, the policy makers or platform designers have (i) to clarify which are the desired effects of a tick size changes (ii) decide how to achieve the desired tick sizes. Another reason why it can be hard to determine an optimum tick is that for each type of market participants there can be a different concept of what a good tick value is. Tick values can hardly be determined in advance just according to theoretical models, more often they are determined by trials and errors and their success or failure can only be assessed by an ex post evaluation of their impacts. In testing our second hypothesis we will recalibrate the ESMA grid under the light of a previous empirical study of the impact of the new tick size regime on UK venues (Nuzzo 2020). The models we build and compare in our analysis are statistical models, in which they are designed with the intent to reproduce the stylized facts observed on the UK venues and to be useful for practitioners and regulators.

2.2.3 Machine Learning and Policy Evaluation

Whilst machine learning has gained a foothold in economic and financial academic literature in the last ten years, its applications regards mainly the prediction domain. However, also in the domain of policy evaluation efforts have been made to build supervised learning methods to estimate casual parameters (Varian 2014). These methods accomplished the goal of soften misspecification issues and improve the model selection transparency.

The application of ML methods to policy evaluation has been discussed by Athey and Imbens who highlighted how the usage of observational data to draw inference about the casual effect of a policy is at the same time necessary and more challenging compared to a pure randomized experiment, due to unobserved "confounders" that lead to spurious correlations (Athey and Imbens 2016).

One application of ML to policy evaluation to an empirical, economical problem can be found in Kreif and Diazordaz 2019.

Supervised learning models, are among the most popular machine learning techniques and are used both for classification and regression problems. The aim of supervised learning classification is to create models able to learn from a given and known set of labels how to classify new instances with new and unseen labels. The simplest way to formalize this problem it is in its binary classification application.

2.2.4 Model comparison

This study contributes also to the comparative literature, comparing the performance and the consistency of our findings using five different models. We refer to the work of Patel et al. 2015, where four machine learning algorithms are compared to evaluate their precision in predicting the stock price indices CNX Nifty and S&P Bombay Stock Exchange (BSE). The authors used accuracy and f-measures to evaluate the performance of their proposed models. As in their approach, we will first identify the best combination of parameters of our proposed models and, secondly, we will compare the prediction performance of these models at their best parameter combination. We will report the limits of these metrics in the Methodology Section, where we will describe the adopted evaluation metrics.

We will also consider how much accuracy changes as a function of these key parameters and we will present the sensitivity analysis in the final paragraph of Section 4. In Huang and Tsai 2009, the prediction accuracy of a different classification algorithm for Taiwan stock prediction has been computed. This paper exploits, different feature selection methods such as Wrapper, χ^2- statistic, information gains, CFS that we will also discuss in the methodology.

Another useful piece of work we will refer to, is the three ensemble methods benchmarked against four single classifiers in an article by Ballings et al. 2015. In particular, we follow their approach in model evaluation, involving twofold cross-validation techniques and AUC curves to gauge the models' performance. We will use the area under the receiver operating characteristic (ROC) curve to assess the discriminatory power of our classifier. The ROC is obtained by plotting Sensitivity versus Specificity for various cut-off values (Rodriguez and Rodriguez 2004). We will discuss it in details in the methodology section.

Table 2.1 systematize the different machine learning algorithms compared in the reviewed literature.

TABLE 2.1: Machine learning algorithm for applied finance compared in literature

	Machine learning methods compared:				
	LR	**SVM**	**NN**	**K-NN**	***ensembles***
Kara, Acar Boyacioglu, and Baykan 2011		x	x		
Leung, Daouk, and Chen 2000	x		x		
Patel et al. 2015		x	x		x
Ballings et al. 2015	x	x	x	x	x
Rodriguez and Rodriguez 2004	x		x		x
Huang and Tsai 2009	x	x	x	x	
Kumar, Meghwani, and Thakur 2016	x	x	x		x
This study		x	x	x	x

The studies reviewed in this section can be grouped in two main categories: regression (non linear) models to forecast the stock prices and classification methods to forecast stock prices direction. For a comprehensive review of the machine learning techniques applied to financial markets see (Henrique, Sobreiro, and Kimura 2019). Our study shared with this stream of literature the model evaluation techniques and data analytic tools exploited but diverges from these reviewed works in applying machine learning technique for calibration purposes rather than forecasting reasons.

2.3 Data and methodology

2.3.1 Dataset

Our dataset includes 1232 unique equity instruments whose main relevant market (MRM) is the London Stock Exchange. For these instruments we have collected qualitative or categorical variables such as the index the equity belongs to, the sector, industry, and quantitative variables such as market variables, firm level variables, closing prices as published by the exchange (updated daily) and the daily average number of transactions as published by ESMA (once per ISIN), their market cap, P/E, ROE, Number of Shares Outstanding, daily implied volatility, number of outstanding shares, quoted spreads, spreads expressed in tick sizes and tick sizes for the month of January 2018, February 2018. All the variables have been aggregated on a daily basis to build a 1232 x 44 panel for a total of 54,208 observations per feature.

Figure 2.3 reports the Tick size table as published in ESMA regulatory technical standard (RTS11). The heatmap highlights the number of ISINs in our dataset corresponding to a specific liquidity band, price and consequently tick sizes. Remarkably, this data frame is wider than others used to evaluate the impact of the tick size

	0≤ADNT<10		10≤ADNT<80		80≤ADNT<600		600≤ADNT<2000		2000≤ADNT<9000		9000≤ADNT	
Price Range	No ISIN	Tick Size	No ISIN	Tick Size	No ISIN	Tick Size	No ISIN	Tick Size	No ISIN	Tick Size	No ISIN	Tick Size
0-0.1	109	0.0005	106	0.0002	27	0.0001			1	0.0001		
0.1-0.2	37	0.001	37	0.0005	6	0.0002	3	0.0001	3	0.0001		
0.2-0.5	53	0.002	46	0.001	16	0.0005	5	0.0002	1	0.0001		
0.5-1	51	0.005	63	0.002	24	0.001	8	0.0005	1	0.0002	1	0.0001
1.0-2.0	47	0.01	58	0.005	28	0.002	24	0.001	19	0.0005	2	0.0002
2.0-5.0	35	0.02	66	0.01	36	0.005	38	0.002	41	0.001	3	0.0005
5.0-10.0	11	0.05	22	0.02	15	0.01	30	0.005	30	0.002	3	0.001
10.0-20.0	6	0.1	14	0.05	11	0.02	16	0.01	20	0.005	2	0.002
20-50	2	0.2	4	0.1	4	0.05	11	0.02	19	0.01	8	0.005
50-100	1	0.5			1	0.1	3	0.05	2	0.02	1	0.01
100-200			1	0.5								
Grand Total	**352**		**417**		**168**		**138**		**137**		**20**	

FIGURE 2.3: Distribution of the UK analysed equity instruments in the ESMA tick size grid. Source: Author's calculations based on ESMA and Bloomberg data.

regime in previous studies, in which it does not only includes the most liquid instruments in the FTSE 100 and FTSE 250 but also instruments in the most illiquid bands. This allow us to investigate how the regime is calibrated in all 6 liquidity bands [5] with ADNT ranging from 0 to an ADNT larger than 9000.

In the following two sections we will state the research hypothesis. The last section is instead dedicated to presenting the robustness check implemented to validate our findings.

2.3.2 Stating the research hypotheses

In this study we treat the tick sizes calibration as a multi-class classification problem. We train five different supervised learning models(K-Nearest Neighbours, Support Vector Machines, Artificial Neural Network, Random Forest and Gradient Boosted Decision Trees) to select the most relevant features that can be used in order to determine for each instrument, both the tick size as it has been prescribed by ESMA, and the tick size that maximizes the market orderliness for each one of the UK instruments included in our analysis. The reason to compare different models, is to verify the robustness of our findings. If two features perform consistently better across all the considered models, these are the features we want to take into account in the construction of the recalibrated grid.

The two hypotheses under exam differ for the target they are trying to classify. In Hypothesis 2.1 this is represented by the tick size as prescribed by the ESMA grid, whilst in Hypothesis 2.2 this is given by the tick size that maximize the orderliness of UK order-books. Therefore, to test both hypotheses, we need to follow the same stages. First, we need to optimize the five models and set their parameters accurately. In a second stage we need to perform a feature selection and PCA. Finally, we can evaluate the performance of the fine-tuned models trained on the most relevant selected variables.

Hypothesis 2.1 *There are other variables more relevant in determining the tick size as it is assigned by ESMA*

[5] ESMA liquidity bands are computed based on the average daily number of transactions (ADNT) on the most relevant market (MRM) in the EU. The ADNT is automatically calculated and published by the Financial Instruments Transparency System (FITRS), a database operated by ESMA, based on quantitative information received from EU trading venues and national competent authorities (NCAs).

We treat the problem as a multi-class classification problem. We have 12 tick size classes that we want to predict (0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0. 2, 0.5). Our goal at this stage of the study is to build the best classifier able to predict the tick size as assigned in the ESMA grid. Using ADNT and prices to train the classifier, we expect the machine learning algorithm to predict the new instances with an accuracy close to 100% if these two variables are representative of the assigned tick-size. We compare the performance of a K-NN, ANN, GBDT, RF and SVM models in correctly classifying the tick-size as in the ESMA RTS-11 grid.

Hypothesis 2.2 *There are other variables more relevant in determining a tick size that also grants a better market orderliness*

Quoting RTS11, a tick size regime or minimum tick size has to be set out in respect of equity and equity-linked instruments to ensure the orderly functioning of the markets. In particular, the risk of an ever-decreasing tick size for shares and its impact on orderliness of the market had to be controlled according to ESMA regulators by means of a tick size regime. First, we defined the orderliness of the market using as proxy the spread expressed in tick sizes (S_{ts}):

$$S_{ts} = \frac{ask - bid}{tick\ size} \tag{2.1}$$

If two stocks are traded with a certain price and are roughly equally liquid, the difference between their spreads must be close to zero. Consequently, what it matters here is not only granting the same tick size within the same liquidity band and price range. Instead, the tick size (TS) should be such that the spreads expressed in tick are the same:

$$S_{ts}(a) = S_{ts}(b) \tag{2.2}$$

We believe the ESMA grid failed to reach a full harmonization in the level of spreads and, more in general we don't think the ADNT is the best parameter to include in the tick-size selection. First, following our cost-analysis' findings, we perform a simulation to force all the considered instruments to reach the same level of spread-in-tick, determining the size of the tick itself backward. This simulation included also the 36 instruments of FTSE250 in the Down-Tick group. On the other hand, the analysis of the distribution of FTSE350 instruments across different bands of liquidity and prices shows that more than a half of the area was not populated by any instrument in our sample. On the other hand, we were aware that the instruments outside the index were also the less liquid: in fact, the 44% of instruments with XLON as the most relevant market, laid in liquidity band 1 and 30% were in liquidity band 2. This means that proposing a more economic grid based on our calibration of spread in tick can present a weak external validity. Instead, we focused on general characteristics of instruments and the relationship between ADNT-AVT-Price and Market-Cap. in order to modulate the mismatch at an ADNT level.

The adjustments proposed following these two different approaches will be discussed in the Findings Section. Following the first approach we forced all instruments in the same price range and the same liquidity band to reach the same spread in terms of ticks. Once we got the benchmark per each group, we compared it with the spread of each instrument, and we obtained the new tick-size and consequently a new forecast for the ADNT/liquidity band. We found a 1-band mismatch for 485 over 1045 considered instruments (46%) plus a further 2-band mismatch for the 21% of the instruments.

Please note how, for 187 ISINs, the ADNT is not an injective function of the tick-size and they find themselves in mixed/wider liquidity bands: 148 ISINs in liquidity bands 3 to 6, further 19 in range 5-6 and 20 in range 4-6. This is the case of the instruments with a very low price (between 0 and 0.5) and high liquidity with a 0.0001 tick size in the top right corner of ESMA grid. Following the second approach we investigated the relationship between ADNT and prices and Market-Cap. and prices, finding a non-linear relationship, suggesting that the ADNT proposed by ESMA increases non-linearly with the Market Cap up to a certain value and decreased afterwards. We tried to predict a tidy derivation of tick size based on different parameters rather than ADNT and Price as in ESMA model.

2.3.3 Hyper-parameters Settings

The ultimate goal of our classifier is to minimize a certain loss over the i.i.d. variables from a distribution (X_{train}), mapping this set to a function f through the optimization of a training criterion with respect to a set of parameters. Beyond these parameters the algorithm typically presents a set of hyper-parameters. These are characteristics of the model that are external and are estimated before the learning phase. The learning algorithm itself is obtained once they are selected. For example, in the support vector classifiers the hyper-parameter C is the hyper-parameter controlling the margin (regularization penalty). Selected hyper-parameters for each model are summarized in Table 2.2 and Table 2.3 respectively for Hypothesis 2.1 and 2.2 and discussed in the following paragraphs in this section.

TABLE 2.2: Hyperparameters tuning for each classifier in Hypothesis 2.1. Source: Author's calculations based on ESMA and Bloomberg data.

Parameter	Value			Optimization Method
				KNN
k	3			min MAE & Grid Search algorithm
distance metric	Manhattan distance			
weighting function				
				ANN
Number of hidden layers	1			min MSE & Grid Search algorithm
alpha	0.001			
weight optimization	stochastic gradient-based optimizer			
epochs	1000			
activation function	Softmax			
Loss	log-loss			
				SVM
	Sigmoid	***Radial Basis***	***Linear***	***Sigmoid*** ***Radial Basis*** ***Linear***
Gamma (Γ)	1	0.1	ignored	min MSE & Exhaustive Grid Search
Regularization parameter (c)	10	10	10	
				RF
Number of Trees	100			Accuracy & Exhaustive Grid Search
minimum samples at split	3			
Depth	8			
				GBDT
Learning Rate	0.2			Exhaustive Grid Search
No of Trees	10			
minimum sample at split	7			
Loss	deviance			

K-nearest neighbour

K-nearest neighbour (k-NN) classifier, is a non-parametric algorithm that finds a group of k points in the training set that are the closest to the test target without using any assumptions on the data distribution. It evaluates the predominance of a

TABLE 2.3: Hyperparameters tuning for each classifier in Hypothesis 2.2. Source: Author's calculations based on ESMA and Bloomberg data.

Parameter		Value		Optimization Method
				KNN
k		11		min MAE & Grid Search algorithm
distance metric		Minkowski Euclidean distance		
weighting function				
				ANN
Number of hidden layers		1		min MSE & Grid Search algorithm
alpha		0.0001		
weight optimization		quasi-Newton		
epochs	1000			
activation function		f(x) = max(0,x)		
Loss		log-loss		
				SVM
	Sigmoid	***Radial Basis***	***Linear***	***Sigmoid*** ***Radial Basis*** ***Linear***
Gamma (Γ)	1	1	ignored	min MSE & Exhaustive Grid Search
Regularization parameter (c)	10	10	10	
				RF
Number of Trees		100		Accuracy & Exhaustive Grid Search
minimum samples at split		3		
Depth		8		
				GBDT
Learning Rate				Accuracy & Exhaustive Grid Search
No of Trees				
minimum sample at split				
Loss				

particular class in a defined neighbourhood and it assigns labels on the basis of this predominance. The performance of this classifier is affected by four key decisions:

(i) the number of nearest neighbours (K)

(ii) a distance or similarity metric to compute distance between objects.

(iii) a weighting function on the neighbour points

(iv) an aggregating method for the classes of the neighbour point

We select 3 nearest neighbours [6] that minimize the error rate (i) and the Manhattan distance (ii) for the case of the two-feature model.

We begin creating a 3-NN classifier as baseline model. We cross-validated this parameter by trying all the values of K = 1, . . . ,40. As we can see from Figure 1. This is the number that seems to minimize the error rate. We have 12 tick size classes that we want to predict (0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0. 2, 0.5).

We used the Manhattan distance that is based on the absolute value distance as opposed to squared error as the Euclidean distance. This should return more robust results since it is not influenced by unusual values. Also, the neighbour points are weighted by the inverse of their distance, with the closer neighbours of the query point being more influential than further away neighbours.

K-nearest neighbours are also known for their fast and economical training phase, in which they approximate their learning functions locally, and all computation is deferred until the classification. K-NN classifiers do not use the training data points to do any generalization. This means the training phase is minimal and consequently

[6] It is common to choose an odd number of neighbours for binary classification problems in order to avoid ties. This is not the case for multi-class problems but it's however a good starting point for a base model.

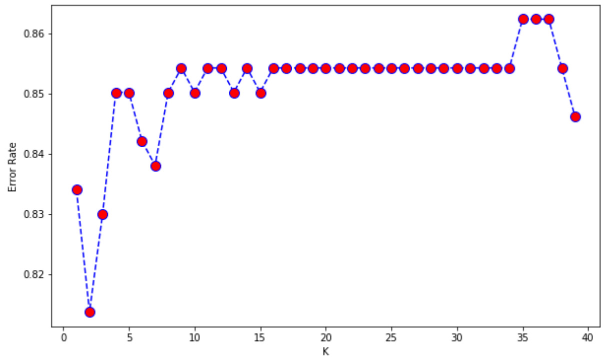


FIGURE 2.4: K-selection. The error rate is minimized for K =3. Source: Author's calculations based on ESMA and Bloomberg data.

fast. Lack of generalization means that all the training data is needed during the testing phase.

K-NN performs better with a lower number of features than a large number of features and consequently reduce the opportunity of over fit the data as in high dimensional circumstances. To avoid over-fitting, the needed data will have to grow exponentially as the number of exogenous variables increases. To tackle the so-called curse of dimensionality, principal component analysis (PCA) or feature selection approach is usually performed before applying any machine learning algorithm. This will be also our approach in applying these models when we test our second hypothesis. At this stage, instead, we deal with ADNT and prices only, the two features defined by the RTS11 grid.

Random Forest and Gradient Boosted Decision Trees (GBDT)

Random Forests, also known as *ensembles* are excellent classifiers that perform very well on a variety of problems. They were first developed in 1995 by Tin Kam Ho using the random subspace method and exploring the idea of random selection of features (Ho 1995). The main objective of a random forest is to build up on decision trees fast algorithm and create a collection of trees to control for the variance. Compared to other methods, random forests do not require an extensive tuning or features normalization and therefore we select them as baseline models for our study.

However, the number of trees and the depth of the forest affect profoundly the accuracy on the test set. Generalization is achieved by the tree inferring a branch split based on the features values observed in the training set. The split at each node is based on the feature that gives the maximum information gain, i.e. the change in information entropy from a prior state that takes information (I) as given:

$$IG_{X,A}\left(X,a\right) = D_{kl}(P_X(x|a)||P_X(x|I) \tag{2.3}$$

$$IG\left(X_{train},\ a\right) = H\left(X_T\right) - H(T|a) \tag{2.4}$$

A new example is classified by following a path from the root node to a leaf node, where at each node a test is performed on some feature of that example. The leaf node reached is considered the class label for that example. The leaf nodes can refer to each of the K classes concerned. Using a grid search method, we tuned the parameters of the random forest. To reduce the over-fitting, risk we tuned the maximum number of features to be included at each split (30% of total). We also cross-validated the minimum size of the sample at each split and the maximum depth.

Another family of ensemble models is represented by gradient boosted decision trees (GBDT). Compared to random forests, GBDT models build a series of trees in which each tree is trained so that to attempt to correct the mistakes of the previous tree in the series. The performance of this algorithm is affected by the following three factors: a. by tree-specific parameters, as the discussed random forest parameters, b. by boosting parameters such as the learning rate and the fraction of observations to include in each tree in order to control the variance and c. by the loss function that we want to minimize at each split. As far as the learning rate is regarded, a relatively low value of 0.05 has been chosen given that a large number of trees 80 is involved in the training phase.

Artificial Neural Network

We have implemented a simple multi-layer perceptron (MLP) supervised algorithm that is trained using gradient descent back-propagation on a training set to learn the function $f(\bullet) : \mathbb{R}^m \rightarrow \mathbb{R}^o$ where m is the number of input dimensions and o is the number of output dimensions (**Kingma2017**). Given the set of features $X = x_1,\ x_2, \ldots, x_m$, MPL can learn a non-linear function approximator for the classification problem. The set of features is known as the input layer, whilst the classes of possible tick sizes are called output layer. A non-linear hidden layer connects the input and the output layers. Each node (neuron) in the hidden layer transforms the input values assigning a weight w to each feature (i.e. $w_1x_1 + w_2x_2 + \ldots + w_mx_m$). The back-propagation in the training phase, minimize a series of cross-entropy loss functions as in (1):

$$-\Sigma_{k=1}^{m} y_{o,k} \log p_{o,k} \tag{2.5}$$

returning a vector of probability estimates $P(y|x) \forall x$. In order to avoid overfitting, we have tuned the L2 regularization parameter (α) so that large magnitude weights are penalised (see Figure 2.5). Because we are dealing with a multi-classification problem, the f(x) function could not be a logistic function. Instead, f(x) itself is a vector of size 12 that passes the softmax function:

$$P(y = j|x) = \frac{\exp x^{\tau} w_j}{\Sigma_{l=1}^{k} \exp x^{\tau} w_k} \tag{2.6}$$

Where k is the number of classes and j represents the j^{th} element of the input to the softmax function. The high dimensional probability space is transformed in a vector containing the probabilities that x is part of each given class and the output is the class with the highest probability.

We allow for a maximum of 1000 iterations (epochs). After testing different learning rates we decide too keep the learning function constant.

Support Vector Machines

Support Vector Machines (SVMs) aim to identify the maximum margin hyper plane in a high dimensional feature space. The hyper plane assigns data points to the two disjointed half spaces and therefore classify them maximizing the separation between negative and positive examples **Xu2009**. The value added of this method lays onto the so-called *kernel trick*. The algorithm does not have to internally perform the transformation to the new high-dimensional feature space. Instead, it can compute the decision boundaries in terms of *similarity* between pairs of points in the high dimensional space. In our analysis we use the three most common kernels:

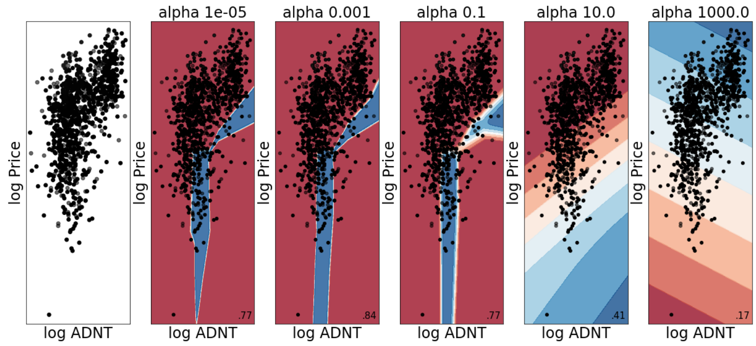


FIGURE 2.5: Regularization of alpha parameter for multi-class MLP classifier. Five different decision functions are reported in the figure with the correspondent alpha. Source: Author's calculations based on ESMA and Bloomberg data.

sigmoid, linear and radial basis function kernel (RBF). The similarity in the RBF is a decaying function of the distance between vectors and the original input space.

The performance of this classifier is particularly influenced by two parameters: the regularization parameter (c) and the kernel parameter (γ). The kernel parameter controls for the influence of a single training example, which, in turns, affects how highly the decision boundaries ends up surrounding points in the input space. A smaller value of gamma allows for large similarity radius, whilst with larger values, the kernel value decay more quickly and the example points (i.e. the observations collected for each instruments with their relative features, prices, ADNT and so on...) have to be very close to be considered similar. The regularization parameter controls the trade-off between satisfying the maximum margin criterion to find the simple decision boundary and avoiding misclassification on the training set.

We tuned these two parameters jointly using a grid search cross validation criterion. The reason being, that with a large gamma, changes in c have little or no effect (validation curves are reported in Appendix A).

For the linear Support Vector Classifier (SVC), in order to optimally adjust c to account for different sizes of the training data, we minimize the risk for equation:

$$Z = C\Sigma_{i=1}^{n}\mathcal{L}\left(f\left(X_i\right), y_i\right) + \Omega(w) \tag{2.7}$$

where C is the amount of regularization, $\mathcal{L}$ is a loss function and Ω is the penalty function (i.e. the individual error per sample)

2.3.4 Performance evaluation

Accuracy is probably the most adopted metric for the evaluation of a classifier performance. Although its simplicity, this metric has several drawbacks. In particular, it performs very bad with imbalanced classes scenarios as it is the case of our dataset. Therefore, we generalized the binary performance f-metrics and Matthew's Correlation Coefficients (MCC) as follow:

$$Precision_{micro} = \frac{tp}{tp + fp} \tag{2.8}$$

$$Recall_{micro} = \frac{tp}{fn + tp} \tag{2.9}$$

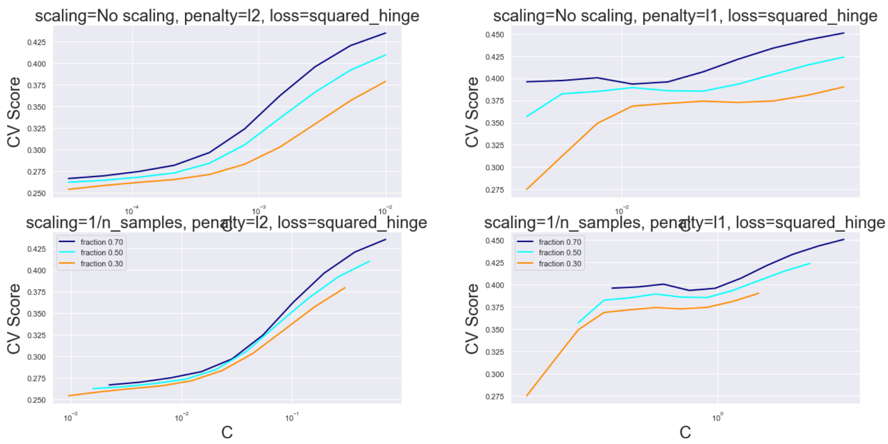


FIGURE 2.6: Optimization of C parameter per amount of training sample. Source: Author's calculations based on ESMA and Bloomberg data.

$$F_1 micro = \frac{2\ Precision_{micro}\ Recall_{micro}}{Precision_{micro} + Recall_{micro}} \tag{2.10}$$

$$MCC_{micro} = \frac{tp\ tn\ - fp\ fn}{\sqrt{(tp+fp)\ (tp+fn)\ (tn+fp)\ (tn+fn)}} \tag{2.11}$$

$$FPR_{micro} = \frac{fp}{fp+tn} \tag{2.12}$$

Where $tp = (TP_1 + \ldots + TP_k)$, $tn = (TN_1 + \ldots + TN_k)$, $fp = (FP_1 + \ldots + FP_k)$ and $fn = (FN_1 + \ldots + FN_k)$. Second, we will look at AUC metrics and the ROC area under the curve. AUC is defined as:

$$AUC = \int_0^1 \frac{tp}{tp+fn} d\frac{fp}{fp+tn} = \int_0^1 \frac{tp}{p} d\frac{fp}{n} \tag{2.13}$$

Where *fp*, *fn*, *tp*, *tn* are defined as above and n is the number of the non-event per class. AUC ranges between 0.5 and 1, with a value of 0.5 meaning a random guess and 1 a perfect prediction. ROC or Receiver Operating Characteristic curves are usually used as visualization tool for binary classifier performance illustration. ROC shows on x-axis the false positive rate (FPR) in the range [0-1] and the TPR on the y-axis also going from 0 to 1. The optimal point in the ROC space is one where the classifier scores a TPR of 1 and a FPR of 0. To apply ROC on our multi-class classification problem and get a visualization, we convert our problem into a One Versus Rest problem.

2.3.5 Feature selection

Multicollinearity, inclusion of irrelevant and redundant variables in a model, has always been worrying for econometricians. In machine learning and data mining, to handle this risk of redundancy, features selection techniques have been developed. Feature selection algorithms enhance interpretability of the model, simplifying it (i.e. reducing its computational complexity) and increase model generalization, reducing overfitting. We use a range of techniques to identify the most relevant variables that can be used in order to determine for each instrument, a tick size that maximizes the market orderliness. These are: correlation criteria, PCA, RF features importance.

Correlation Criteria are the simplest techniques for feature selections. They are usually based on Pearson's correlation coefficients (PCC) between the dependent and independent variable.

$$\rho_{xy} = \frac{cov(xy)}{\sigma(x)\sigma(y)}^2 \tag{2.14}$$

However, since the PCC assumes normality distributed variables, we have adopted Spearman's rank correlation (SRC), defined as the Pearson Correlation coefficient between the rank variables.

$$r_s = (cov(rg_x, rg_y))/(\sigma_(rg_x)\sigma_(rg_y)) \tag{2.15}$$

As we have seen in the literature review, Random Forests (RFs) are very popular classification algorithms. We have also explored the core idea behind the RFs and we saw how, to be really effective as generalization tools, they have to be diverse. This diversity is based on the construction of many unpruned trees with each tree using bootstrap training data. We now highlight how RFs rather than determining the best split among all features, only use a sub-sample of the available independent variables. Denoting the number of training units by m and the features by n, m examples are selected with replacement for each k decision tree. During the process of decision tree construction, the best split is decided among the m samples in the k^{th} tree. The final score is determined aggregating the results from all k decision trees. In the training phase only 2/3 of randomly split bootstrapped data is used. The remaining part is known as Out-of-a-Bag (OOB) sample is used to test the accuracy of the built estimator and the Mean Squared Errors (MSEs) are computed. Perturbated OOB are produced for each tree and MSE's are computed for each tree (see Equation 2.6). Feature Importance is also computed as "the decrease in label homogeneity at the node weighted by the probability to reach that node. The latter is computed by the number of samples that reach that node divided by m. The higher the value the more important the feature" (Hastie et al. 2017)

$$I_x^{RF} = \frac{1}{K}\Sigma_{\tau=1}^{k}((MSE_\tau) - MSE'_\tau) \tag{2.16}$$

In k-NN and SVM models, different weights for different features are assigned with respect to classification importance, then the information gain is computed to get the importance of each features and obtain the weights as in Chen and Hao 2017.

2.3.6 Scaling, train-validation-test split and cross-validation

One element of pre-processing we were particularly sensitive was scaling. Our prices can range from 0 to 200 whilst the ADNT can by higher than 9000, spreads are in basis points and market cap can be expressed in millions or thousands. We use a Min Max Scaler because we want to preserve the shape of the original distribution of the data. This scaler ranges from zero to one and it does not reduce the importance of outliers too. Therefore, being the least disruptive to the information in the original data it is adequate for our base model. In formula we have:

$$m = \frac{x - x_{min}}{x_{MAX} - x_{min}} \tag{2.17}$$

That is, we subtract the minimum value in the column and then divide by the differences between the original maximum and the original minimum. Doing this, both the prices and the ADNTs range between 0 and 1 with no alteration in their density functions.

As far as the train-test split is regarded, we are aware that doing only cross-validation on a test set to do model selection may lead to a subtle over-fitting and a more optimistic generalization estimate. This is because the more observations we see about the data frame as part of repeated cross-validation passes in choosing the model, the more influence any potential held up test data has played into selecting the final model not merely evaluating it. For this reason, we split our data in a train-validation-test set, and we use train test in building the models, the validation set for model selection and parameter settings and the test set for the model evaluation. To account for the imbalanced nature of our classes, we have adopted a stratified sampling technique, i.e. we have divided the examples in our dataset into homogeneous subgroups before sampling in order to achieve collectively exhausted strata. We also set the random state to zero in order to be able to replicate our analysis. We use a 60-20-20 stratified split and the composition of the three set is described in Table and we can see how all the tick sizes are represented in the same proportion in each set.

TABLE 2.4: Stratified train-validation-test split.
Source: Author's calculations based on ESMA and Bloomberg data.

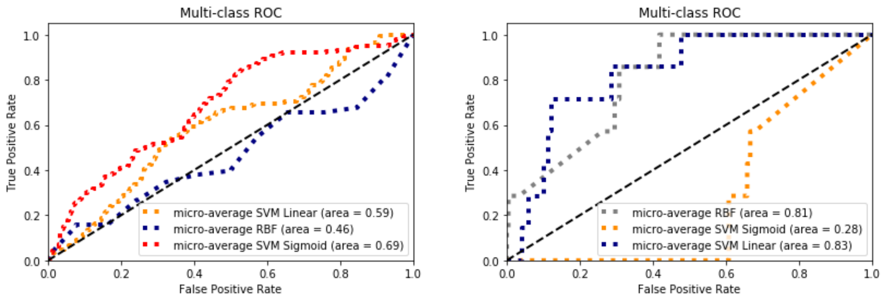
	0.005	0.02	0.001	0.0005	0.05	0.0002	0.002	0.01	0.0001	Total
Train Set	121	49	105	115	20	72	128	97	22	729
	17%	*7%*	*14%*	*16%*	*3%*	*10%*	*18%*	*13%*	*3%*	*60%*
Validation Set	41	16	35	39	6	24	43	33	7	244
	17%	*7%*	*14%*	*16%*	*2%*	*10%*	*18%*	*14%*	*3%*	*20%*
Test Set	41	16	35	39	6	24	43	33	7	244
	17%	*7%*	*14%*	*16%*	*2%*	*10%*	*18%*	*14%*	*3%*	*20%*
Total										1217

Another aspect we have carefully considered, regarding the detection of imbalanced classes is the adoption of micro and macro averages in multi-class evaluation. Micro averages assign equal weight to each instance. In this way, largest classes are the most influencing. On the other hand, with macro averages, each class is given an equal weight. In our dataset, some tick size classes are much larger than others, therefore, to weight our metric toward the largest class we use the micro average. Finally, we computed the macro average for sensitivity checks.

As far as high dimensionality is regarded, we have seen how this issue does not affect the test of the Hypothesis 2.1 and we tackled it in the Hypothesis 2.2 involving the series of techniques described in Section 2.3.5. Here, we highlight how we took this issue in consideration in setting classifiers' parameters. In K-NN models, research has shown that, in large dimensions, Euclidean distance is not useful any more. Therefore, we preferred other measures such as cosine similarity, which get decidedly less affected by high dimension.

2.4 Implementation of classification models

The empirical computations where performed on the authors' local machine having 2.11 GHz processor with 16.0 GB RAM and algorithms were implemented using Python 3.7.6 Scikit-learn 0.22.2 package Pedregosa et al. 2011



(A) SVM (class 0.001)

(B) SVM (class 0.01)

FIGURE 2.7: ROC for models based on ADNT and prices. Source: Author's calculations based on ESMA and Bloomberg data.

2.5 Findings

2.5.1 Testing hypothesis 1

We have evaluated the ability of the ESMA grid based on ADNT and prices to assign a calibrated tick size to 1232 instruments traded on the UK venues. The simple assumption behind our approach is that if two instruments are traded with a similar frequency at close prices the grid should assign them the same tick size. This would be the case if the ADNT computed once per year by ESMA was a good proxy for the liquidity of the equity instrument. We treat the problem as a multi-class classification problem. We have 12 tick size classes that we want to predict (0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0. 2, 0.5). Using the parameters discussed in the Methodology Section we have trained the best classifier able to predict the tick size as assigned in the ESMA grid. Using ADNT and prices to train the classifier, we expect the machine learning algorithm to predict the new instances with an accuracy close to 100% if these two variables are representative of the assigned tick-size. Table 5 compares the performance of a K-NN, ANN, GBDT, RF and SVM models in correctly classifying the tick-size as in the ESMA RTS-11 grid.

The evaluation performance, in line with our expectations, returns an accuracy ranging between a minimum of 0.83 for support vector machines to 0.98 of Random forest Classifier. However, the accuracy in not the most reliable metrics to take into account in our multi-class problem where imbalanced classes can lead to a situation where the same classifier performs very well in a certain class, and at the same time very poorly in another one. This is well displayed by the areas under the ROCs of the evaluated models (Figure 2.7 and 2.8). Ensables, k-NN and MLP models are better in classifying larger tick-sizes rather than narrower. On the other side the SVM with sigmoid functions returns more precise classifications on narrower tick. A summary of the ROC AUC score and the other evaluation metrics is reported in Table 2.5.

To test our first hypothesis, we allowed these classifiers to be trained on a training set with multiple features (see first paragraph of Section 2.3. Table 2.6 compares the performance of the same models trained to predict the same ESMA tick-sizes using Market Cap and Quoted Spread.

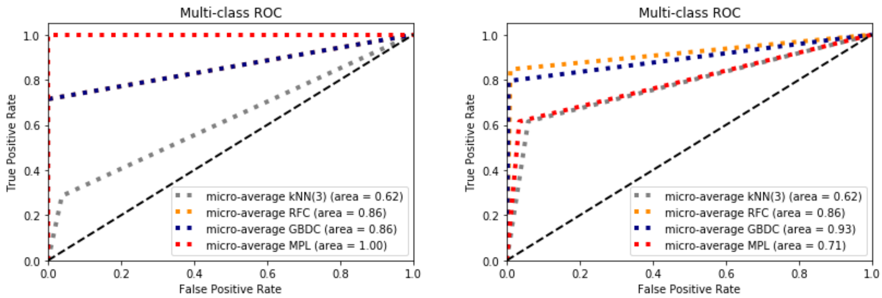
We looked at the correlation among our features and the outcomes, performing a principal component analysis (PCA). If two features are highly correlated it means they embody the same information , therefore since we want to deliver a parsimonious model with no more than two features, we ideally want to maximize the

TABLE 2.5: Evaluation of Classifiers based on ADNT and prices performance. Source: Author's calculations based on ESMA and Bloomberg data.

Model	Accuracy	ROC AUC Score	Micro Precision	Micro Recall	F1 Micro	Micro MCC	FPR micro
K-NN	0.95	0.7	0.79	0.79	0.79	3.70E-03	0.02
ANN	0.93	0.87	0.68	0.68	0.68	2.00E-04	0.04
Linear SVM	0.83	0.69	0.233	0.233	0.233	1.86E-05	0.09
Sigmoid SVM	0.83	0.64	0.237	0.237	0.237	1.91E-05	0.09
RBF SVM	0.83	0.84	0.27	0.27	0.27	2.35E-05	0.09
RF	0.98	0.87	0.91	0.91	0.91	2.00E-03	0.01
GBDT	0.97	0.77	0.87	0.87	0.87	1.00E-03	0.015

TABLE 2.6: Evaluation of Classifiers based on Market Cap. and Quoted Spread. Source: Author's calculations based on ESMA and Bloomberg data.

Model	**Accuracy**	**ROC AUC Score**	**Micro Precision**	**Micro Recall**	**F1 Micro**	**Micro MCC**	**FPR micro**
K-NN	0.95	0.82	0.83	0.83	0.83	3.70E-03	0.02
ANN	0.97	0.89	0.68	0.68	0.68	2.00E-03	0.01
Linear SVM	0.86	0.77	0.47	0.47	0.47	2.00E-04	0.1
Sigmoid SVM	0.86	0.75	0.47	0.47	0.47	4.00E-04	0.1
RBF SVM	0.86	0.85	0.47	0.47	0.47	4.00E-04	0.1
RF	0.98	0.9	0.91	0.91	0.91	4.60E-03	0.01
GBDT	0.97	0.86	0.87	0.87	0.87	4.50E-03	0.015



(A) MLP, ensambles and k-NN (class 0.001)

(B) MLP, ensambles and k-NN (class 0.01)

FIGURE 2.8: ROC for models based on ADNT and prices. Source: Author's calculations based on ESMA and Bloomberg data.

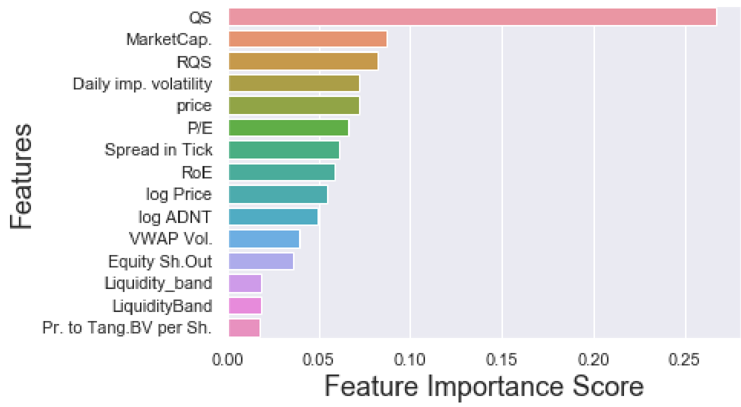


FIGURE 2.9: Gini Impurity reduction due to each relevant feature. Source: Author's calculations based on ESMA and Bloomberg data.

correlation between our features and the label whilst we minimize the correlation among our features.

These results were in line with our findings from the most relevant feature approach. In Figure 2.9, we plotted the feature importance of each variable included in our classification study. The importance of each feature is the reduction in Gini Impurity produced by each feature in the random forest fitted model. As we can see the Quoted Spread (QS) is the most relevant measures, followed by the relative quoted spread and the spread in tick and the current Market-Cap.

2.5.2 Testing hypothesis 2

The focus of Hypothesis 2.2 is the market orderliness. With Hypothesis 2.1 we gauged the ability of ESMA liquidity bands and price ranges of being descriptive of the tick sizes of equity instruments traded on UK venues as they were assigned by ESMA and as they were enforced on UK venues. With Hypothesis 2.2, instead, we have a theoretic optimum tick-size as the target to reach market orderliness. The orderliness of the market is defined in our approach using the spread expressed in tick sizes.

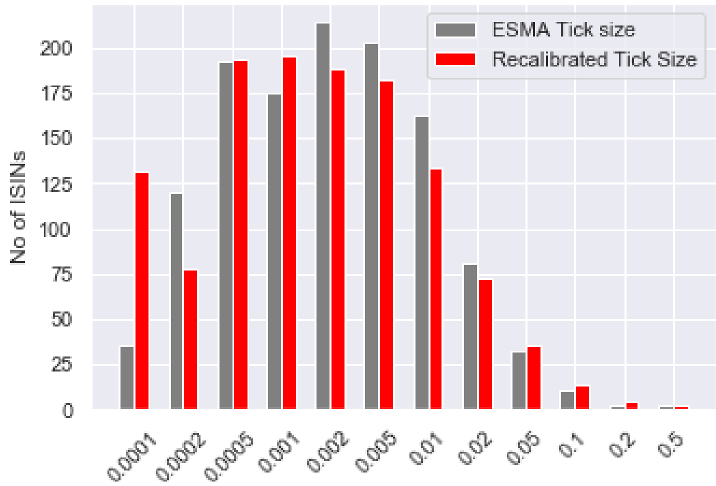


FIGURE 2.10: Recalibration of Tick Sizes to obtain the harmonization of spreads in tick.Source: Author's calculations based on ESMA and Bloomberg data.

Table 2.7 reports the ability of classifiers based on ADNT and prices of labelling these recalibrated tick sizes.

We believe the ESMA grid failed to reach a full harmonization in the level of spreads and, more in general we don't think the ADNT is the best parameter to include in the tick-size selection. First, following our cost-analysis' findings, we perform a simulation to force all the considered instruments to reach the same level of spread-in-tick, determining the size of the tick itself backward. This simulation included also the 36 instruments of FTSE250 in the Down-Tick group. On the other hand, the analysis of the distribution of FTSE350 instruments across different bands of liquidity and prices shows that more than a half of the area was not populated by any instrument in our sample.

Table 2.8 reports the same performance evaluation for models based on Market Cap. and Quoted Spread. As we can see the precision of the prediction improves notably.

2.6 Conclusion

In the last two years market regulators have been trying to quantify the MiFID II impact on market quality. One particularly controversial aspect of this regulation has been the adoption of a new tick size regime (Art. 49). The Financial Conduct Authority (FCA) in UK as well as the Autorité des Marchés Financiers (AMF) in France have conducted assessments on the tick size regime, which entered into force in January 2018 as part of MiFID II, concluding the regime had an opposite impact on the overall market quality of UK and EU venues. Unintended consequences on UK venues have been measured and the regime resulted not perfectly calibrated for UK equities. This consideration open-up to whether a better methodology can be constructed to improve the effectiveness of the regime. In this paper, we adopt a supervised machine learning approach to propose a better calibrated alternative to ESMA grid. Our approach is based on: (i) market capitalization; and (ii) quoted spread.

TABLE 2.7: Evaluation of Classifiers based on ADNT and prices performance on recalibrated tick sizes. Source: Author's calculations based on ESMA and Bloomberg data.

Model	Accuracy	ROC AUC Score	Micro Precision	Micro Recall	F1 Micro	Micro MCC	FPR micro
K-NN	0.87	0.52	0.37	0.37	0.37	3.60E-03	0.07
ANN	0.88	0.5	0.39	0.39	0.39	3.93E-03	0.068
Linear SVM	0.84	0.59	0.19	0.19	0.19	1.49E-03	0.087
Sigmoid SVM	0.84	0.49	0.19	0.19	0.19	1.54E-03	0.09
RBF SVM	0.84	0.46	0.21	0.21	0.21	1.73E-03	0.087
RF	0.88	0.56	0.41	0.41	0.41	4.34E-03	0.065
GBDT	0.88	0.58	0.39	0.39	0.39	3.99E-03	0.06

TABLE 2.8: Evaluation of Classifiers based on Market Cap. and quoted Spread on recalibrated tick sizes.
Source: Author's calculations based on ESMA and Bloomberg data.

Model	Accuracy	ROC AUC Score	Micro Precision	Micro Recall	F1 Micro	Micro MCC	FPR micro
K-NN	0.98	0.94	0.89	0.9	0.79	3.70E-05	0.1
ANN	0.99	0.98	0.94	0.79	0.68	2.00E-05	0.18
Linear SVM	9.98	0.96	0.63	0.78	0.233	1.00E-05	0.1
Sigmoid SVM	0.99	0.96	0.63	0.78	0.237	1.00E-05	0.09
RBF SVM	0.97	0.96	0.78	0.78	0.27	1.35E-05	0.09
RF	0.98	0.97	0.98	0.98	0.91	1.20E-05	0.16
GBDT	0.99	0.97	0.99	0.99	0.87	1.00E-05	0.16

Our proposed calibration for the regime would achieve optimal tick sizes for equities 3 times more frequently than the current ESMA regime. This allows us to outline an idealized grid for determining an equity's minimum tick size for this proposed regime. This paper is especially relevant for UK policy makers in the context of the UK leaving the EU and suggests the ESMA grid can be abandoned. At the best of authors' knowledge, it is also the first time a supervised machine learning model is adopted to evaluate policy implications of a financial regulation.

Chapter 3

Simulating the risk of liquidity breakdown in the UK equity market using a network approach

Abstract

In this paper, we simulate liquidity breakdowns on the UK equity market, adopting a direct network approach. Compared to previous studies in the literature, we focused on the link between the structure of the network and the role of dealer's inventories. In our simulations, the risk of liquidity breakdown stemmed directly from the the dealer's inventory optimisation problem. The ability of a system to adapt to the shock changes with different financial agents and different liquid/illiquid instruments as well as with different levels of competition among dealers and fragmentation among UK venues. Along these lines, the spread of the contagion in our approach is subject to the dynamic of the dealer's inventories. As far as the relationship between the topology of the network and the risk of contagion is accounted for, we show that agency brokers present both the highest in and out degree of centrality in the network. This means that participants within this firm type tend to lead more price information than others as well as to spread faster the risk of liquidity disruption. Our findings also suggest that central nodes are linked with other nodes whose inventory levels are hard to get altered. We have found a positive correlation between the degree of centrality of a node and the speed of contagion. All in all, our study contributes to the existing empirical market microstructure literature adopting a novel, distinctive propagation algorithm based on market participant inventories. The direct-network approached based on dealer's inventory can provide regulators an extra tool to monitor the risk of liquidity breakdowns in the equity market, identifying the participants who can spread the risk faster and broader.

Keywords

UK equity market – Liquidity breakdown – Inventory models – Contagion models

Introduction

In this paper, we simulate liquidity breakdowns on the UK equity market, adopting a direct network approach. Compared to previous studies in the market microstructure literature, we focused on the link between the structure of the network and the role of dealer's inventories. In our simulations, the risk of liquidity breakdown stemmed directly from the dealer's inventory optimisation problem. Inventory models address the dealer's problem of maintaining inventories on both sides of the market. Since order flows are not synchronized, dealers face the possibility of running out of cash (bankruptcy) or out of inventory (failure). Bid/ask spreads in this context are seen as compensations. We fixed a cut-off under which is too costly for agents to execute transactions (namely, when their inventory levels reach zeros on one side of the order book) and we infected their nodes. The untraded turnover is distributed proportionally among other participants and their inventory positions are modified accordingly. The ability of a system to deal with the shock changes with different classes of financial agents and different liquid/illiquid instruments as well as with different levels of competition among dealers and fragmentation among UK venues. Along these lines, the dynamic of the contagion in our approach is subject to the dynamic of dealer's inventories.

The aim of this approach is to explain the market efficiency and quality of stock exchanges, in the most structural aspect of their functioning: the interaction among their participants. The way participants interact on the exchange, and the way they compete, ultimately affects the price of the traded instruments, the liquidity offered on the exchange and the *fair treatment* of their clients. We studied the interconnectedness of financial agents meaning the number and magnitude of trading links between them. Interconnectedness is relevant for policy-makers because it is a vehicle for contagion. We simulate a series of liquidity breakdowns on the basis of real-world MiFID II transaction report dataset to detect which are the largest interconnected entities who can spread the contagion widely and quickly and under which circumstances the shock can even cause the liquidity breakdown of the whole financial network. Notably, we try to answer four fundamental research questions.

i How can we define the risk of contagion and measure its intensity?

ii Who are the market participants bearing the highest risk of contagion for the network? Do they cover a specific position in the financial network? (i.e. are they more central or more peripheral nodes in the network?)

iii Does fragmentation among venues exacerbate or mitigate the risk of contagion?

iv Does the network analysis raise concerns around concentration of trading and a subsequent impairment of best execution?

We believe these research questions are highly policy relevant. Assuring that markets are fair and efficient is a core responsibility for the FCA and investigating how intermediaries interact in a market is fundamental to calibrate risks and design policies: a system where only few big companies react to shocks needs a different treatment compared to a fragmented environment made up by small different companies Allen and Gale 2000. A system can either react to a shock and absorb it or collapse and spread it. All the network-related studies following the financial crisis in 2008 (Iori et al. 2005, Battiston et al. 2012, Bardoscia et al. 2017) agreed that financial networks are able to absorb shocks up to a certain threshold, but they tend to spread risk, rather than contain it, once this cut-off is crossed (Haldane and May 2011). Fixing the threshold is therefore a crucial decision in the simulation of contagion.

Firstly, we define the contagion as the dynamic mechanism across which, altering the inventories of market participants, their incentives to buy(sell) an equity instruments result also altered. When these incentives are altered to the point that the inventories on one side of the order book are zero (cut-off point), market participants stop trading. This decision not to trade affects other counter-parties' inventory levels and their future transactions.

We have approached the network stability problem setting up a flow model to monitor the maximum level of liquidity in the network of market participants. We have simulated a series of shocks in the network. In each simulation only one market maker/dealer/high frequency trader was originally removed from the network to control if and how the shock percolate in the system (()Radicchi2016). We built four metrics to proxy for the stability of the market.

Secondly, we have investigated the position of each market participant in the network and the risks associated. We have ranked each market participant importance in the network exploiting several measures of centrality. Centrality, in graph theory, identifies the most important, influential nodes in the network. The meaning of importance changes according to the decision to measure centrality in terms of flow in the network or as the role of the node in the cohesiveness of the network.

Thirdly, we have analysed the structure of the UK equity market network and we have addressed separately the role of competition among venues with its potential fragmentation issue.

Fourthly, we looked at the role of competition among dealers highlighting potential side-effects deriving from the risk of concentration. The reminder of this paper is organised as follow: Section 2 (research context) provide a description of the interconnectedness of broker-dealers on UK equity markets and present the literature related to broker-dealers' inventory models. Section 3 (research design) describe the dataset and the metrics used to proxy for broker-dealer's competition, concentration and network stability. Lastly, Section 4 (results) provide our findings around competition among venues and dealers, centrality of different participants and our simulation of liquidity breakdowns.

3.1 Literature Review

3.1.1 Financial network stability

Interconnectedness: a key feature of financial markets

The classic view of a financial market is a place where buyers and sellers come together to facilitate exchanges among parties and to exchange goods at a price determined by the level of their aggregate supply and demand. To ensure that trades are fair and efficient, the investigation of the interactions among these intermediaries on the secondary market is fundamental to calibrate and monitor the risk of liquidity breakdown. A system where only a few large companies react to shocks needs different treatment compared to a fragmented environment made up by different small companies. We use the word interconnectedness to refer to the interaction among trading firms.

Interconnectedness is characterised by several dimensions. It can be studied as exposure to common assets, haircuts, shadow banking or information spillovers. The literature on interconnectedness con be broadly divided into two macro-categories (for a complete taxonomy of studies on interconnectedness see Kara and Tian 2015): network approaches (Direct and Indirect) and not-network approaches (Principal Component Analysis (PCA), Regressions, Default model). We propose a direct network approach. This literature mainly focuses on systemic risk (Allen and Babus 2008 Iori and Porter 2016 Benoit et al. 2017). We analyse an aspect intimately linked with systemic risk, liquidity risk. It arose from the underlying financial network. To the best of our knowledge, financial agents' interconnectedness has already been investigated in relation to liquidity flow (Cifuentes et al. 2005, Tasca and Battiston

2016, Caccioli et al. 2014).Levent Uslu and Evren 2018 correlate the liquidity flow based on the network topology of 20 equity instruments on the Borsa Istanbul and draw-downs between peaks and trough in a time series of daily prices. In Chang and Zhang 2016 the heterogeneity in the volatility of their liquidity needs makes the most volatile banks willing to trade with the more stable banks, creating a multipartite network with the most stable banks in the core.

The dynamic of contagion

Interconnectedness can be a conduit of contagion. The first occurrence of the word contagion in the dictionary of academic financial language dates back to 1997 after the Asian Financial Crisis. However, already ten years earlier, the market crash of October 1987 triggered the first empirical study on a financial contagion and an attempt to model it as the transmission of volatility resulting from rational agents' mistakes to infer information (King and Wadhwani 1990).

According to the European Central Bank (ECB), financial contagion can emerge from physical exposure and asymmetric information (Bank 2018). The contagion stemming from a physical exposure, spreads from one market to another in the form of declining prices, declining liquidity, increased volatility and increased correlation associated with financial intermediaries (Kyle and Xiong 2001).

Seth and Panda 2018 provide an extensive literature review of financial contagion. Yet, all studies that followed the financial crisis of September 2008 agreed that financial markets are able to absorb shocks up to a critical point, whilst they tend to spread risk rather than contain it once this threshold is crossed (Gray and Leibrock 2015).

Although the importance of the network structure in propagating a shock is recognised, however, the effect of connectivity on impairing or enhancing the system stability is not straightforward.

In the literature, the problem of network resiliency has been usually studied employing simulations. The first step in simulations requires the generation of an appropriate shock able to describe the risk we want to assess. Once the structure of the network is designed, the shock propagates according to certain assumptions, either iteratively (Caccioli et al. 2015; Nier et al. 2007) or simultaneously, affecting many nodes at the same time (Battiston et al. 2016, Serri, Caldarelli, and Cimini 2016, Tressel 2010).

Moreover, the shock can be endogenous or exogenous. Figure 3.1 shows the positioning of our research in the reviewed literature on financial interconnectedness. Our study contributes to the existing empirical market micro-structure literature adopting a novel, distinctive propagation algorithm based on market participant inventories.

First, we define the contagion as the dynamic mechanism through which, altering the inventories of market participants, their incentives to buy(sell) an equity instruments result also altered, and they can be altered to the point that the market participant doesn't trade any more, affecting other counter-parties' inventory levels and their future transactions. We have approached the network stability problem setting up a flow model to monitor the maximum level of liquidity in the network of market participants, detecting the total capacity, expressed in terms of inventories, of the minimal subset of participants that would impair the feasible market functioning. We have simulated a series of shocks, removing one by one all the nodes representing official market makers from the network to control if and how the shock *percolate* in the system (Radicchi and Castellano 2016).

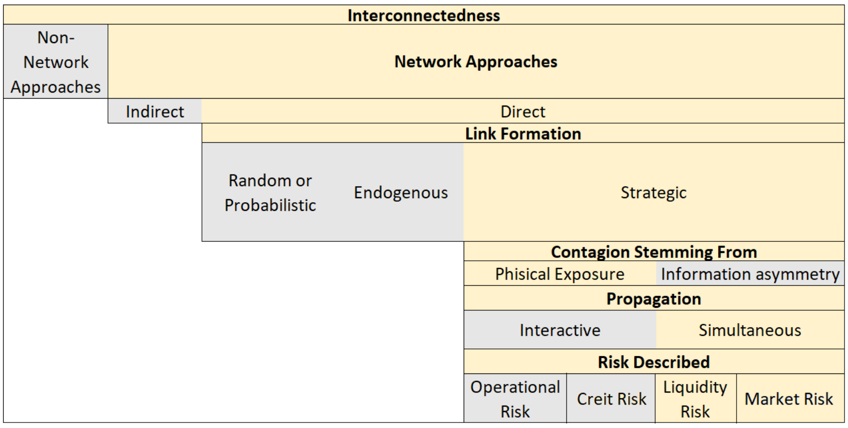


FIGURE 3.1: The figure shows the positioning (in yellow) of our research in the reviewed literature on financial interconnectedness. Our study contributes to the existing empirical market micro-structure literature simulating a shock in a direct network of market participant based on their inventories.

3.1.2 Dogs, wolves and Baltos: the role of broker-dealers

The nodes of our network represent market participants. Two very important subgroups of these are the dealers' nodes and the brokers' nodes. *Dealers* can be firms or people acting on financial markets as principals in trading for their own account, as opposed to broker who acts as agents who executes orders on behalf of their clients. A dealer is a market maker in the security who seeks to profit from the bid-ask spread while providing liquidity for the instrument. They don't trade on behalf of a client as brokers and they don't facilitate transactions as facilitators.

The dealer's role is critical for the financial market functioning. They become crucial especially in relation to less liquid instruments. Participants trading illiquid securities have more difficulties in getting their orders matched and they require an agent to avoid markets dysfunction and breakdowns. In other words, they are important liquidity suppliers and for this activity they are remunerated with a a fee on either bid and ask quotes, exactly as market makers.

As far as it regards *brokers*, also called *agency-brokers*, they differ from dealers in which they own the responsibility to find the best execution for their clients. Even if markets have become less physical/immaterial and dis-intermediation is a key aspect of many recent business model, we are far from cutting the middleman out. The relative importance of brokers depends on their role as facilitators. Agency-brokers are totally committed to grant the best execution to their clients (i.e. fill a client order at the lowest price available and in the fastest possible way).

To this main classification we must add another category of market participant: the broker-dealers. The role of *broker-dealers* is twofold: as brokers they buy and sell securities on behalf of their clients. As dealers, they act on their own account, selling their own products (warehouse brokers) or products from outside sources (independent brokers). This complicate the view of the stock exchange as a place where each participant plays fixed, determined roles (buy-side and sell-side agents).

Different participants come together on the market acting each time under different risk profiles, based on their exposures, profit opportunity and incentive to liquidity provision. It is critical to keep in mind that each intermediary can act on both deal and agency capacity, as well as undertake their main business activity. For this latter purpose, we distinguish among firms with different levels of risk exposure on the base of their core activities: are they custodians or are they putting their own or their client's capital at risk?

3.1.3 A dealer's inventory problem

Dealers incur in different types of costs: cost of inventory, cost of execution and information asymmetry to trade with informed traders. In this paper, we focus on the cost of inventory in the determination of the dealer's behaviour. The dealer's inventory problem is an optimization problem. Inventory models address the dealer's problem of maintaining inventory on both sides of the market. Since order flows are not synchronized, dealers face the possibility of running out of cash (bankruptcy) or out of inventory (failure). Bid/ask spreads in this context are seen as compensations. The most popular inventory models in the literature are essentially:

a risk-neutral models: based on the Walrasian framework according to which lower (higher) price drives (depresses) demand. They demonstrate that rational dealers attempting to maximize their profits must establish certain bid/ask spread and manipulate its size for maintaining preferred inventory. (Garman 1976, Amihud and Mendelson 1980)

b models with risk aversion that yields the bid/ask spread that depends linearly on the dealer's risk aversion and the asset volatility (Stoll 1978).

3.1.4 Competition among venues

The universe of financial markets after MiFID I [1] is dotted with an increasing number of trading places. Abolishing the concentration rule, MiFID I allowed trading on alternative trading venues not only with some positive effects for global and local liquidity (Gresse 2015), but also with side effects for some equity instruments who suffered from the new fragmented environment.

When several traders use a specific market, the market's ability to match buyers and sellers increases with the resulting reduction in trading costs, which attracts more traders. Another reason why more liquidity, ultimately, reduces specialists' spreads is the increased efficiency apparently explained by inter-dealer arbitrage: i.e. spreads depend on volume in the entire market not only on dealers' own volumes (Hamilton 1979). It can thus increase the informativeness of prices.

Referring to competition among venues, we believe that there is a trade-off between competition and consequent possible issues around fragmentation.

On the one hand, increased competition between trading venues, may result in lower trading costs (O'Hara and Ye 2011) and can allow for increasingly tailored trading platforms able to suit the needs of different clients (Hendershott and Mendelson 2000). When a security trades in two markets with a similar structure and type of investors, orders concentrates in the market where a larger number of traders is expected (Pagano 1989). In the absence of price priority, consolidated depth can be larger in competing markets as proven theoretically and empirically by Foucault and Menkveld 2008.

On the other hand, fragmentation may impinge on liquidity when the security is cross-listed and different exchanges compete for the same order flow or when a portion of the order flow is internalised. When a stock is traded on different markets

[1] Markets in Financial Instruments Directive entered into force on the 1st of November 2007 providing an harmonised regulatory framework for 30 countries in the European Economic Area (EEA).

adverse selection costs may increase (Chowdhry and Nanda 1991). Also, fragmentation may harm price discovery (Hendershott and Jones 2005) by increasing search costs and thus decreasing competition between liquidity providers.

3.1.5 Competition among dealers

Fourthly, we looked at the role of competition among dealers highlighting potential side-effects deriving from the risk of concentration. Assuming perfect competition among risk-neutral liquidity providers, Back and Baruch 2007 shows that each equilibrium on the floor exchange involves at least a partial pooling.

In order to maximize social welfare within the exchange the regulators' interests is to promote a proper level of competition among dealers. One expected benefit of competition is, in fact, a lowering of transaction fees, which is desirable for investors.

However, it has been noted how the nature of competition can result in a different impact on spreads. Experimental financial studies show that bid–ask spreads are wider and price discovery is lower (slower prices' responses to order flow) when multiple dealers compete on the same asset (direct competition) compared to where different assets compete with a monopolistic dealer in each one (indirect competition) (Lamoureux and Schnitzlein 2004).

The role of competition in determining the bid-ask spreads is also supported by empirical studies. Huang & Masulis (1999) show that the level of competition is time-varying, highly predictable, and displays a strong seasonal component that in part is induced by geographic concentration of business activity over the 24-hour trading day. It actually follows the typical spreads U-shape curve (Huang and Masulis 1999). The author estimates that the addition of one more competing dealer would lower the average quoted spread of 1.7%.

How the interaction among asymmetrically-informed agents affects assets prices has also been extensively studied in market micro-structure literature(Ho and Stoll 1981, Garman 1976, Akerlof 1970). This asymmetry is reflected in the existence of a positive bid-ask spread, even when the dealer is risk neutral and the expected profit is null (Glosten and Milgrom 1985).

The spreads' U-shape intra-day pattern has been linked to the degree of competition in market making and the extent of informed trading (Chan, Chung, and Johnson 1995). At the same time, the dealers' probability of entering and exiting the market, therefore competing on a certain venue, is affected by volatility and spreads that affects the level of their returns (Ho and Macris 1984). For earlier studies, we can point to Cohen-Cole, Kirilenko, and Patacchini 2014.

Network analysis can be useful to highlight how difficult is for certain categories of client to access the market via multiple market makers, whilst it is much more frequent that they execute all their trades with the same provider. This might call a best execution rationale in question. Therefore, in the last part of our analysis we concentrate on the risk of concentration due to a lack of competition among dealers.

3.2 Research design

3.2.1 Datasets and firm classification

We selected transaction reports timestamped at the microsecond for 317 equity instruments in FTSE350 that traded 21 trading days in January 2018. We exploit the legal entity identifiers (LEIs) codes to determine the unique buyers and sellers presented in our dataset and link final buyers to final sellers. We could classify 11 different firm types and 54 trading venues including lit, dark and Systematic Internalisers.

We classify market participant according to different layers of classifications. To identify market makers, we used the official list reported by the London Stock Exchange. Second, we mapped the Systematic Internalisers (SI) using an FCA official list and then we moved to the Buy-Side using Orbis data-set, containing financial information on companies, scrapping the text and obtaining a wide classification of Asset Managers, Funds Managers, Investment Banks (Large or small-medium) and Private/Commercial Banks. Proprietary-Trading Firms and High Frequency Traders (HFTs) have been, for a long-time, object of research for FCA Economic Department (ED) and we decided to exploit their classification to identifies these subjects.

Although some categories are broadly defined (e.g. under the label *Asset Manager* there are different buy-side individuals or under the label *Fund Manager* there are both active and passive fund managers, ETFs as well as mutual funds), we believe that we are not losing any meaningful insight for our economic analysis of the market structure. Appendix A provides further technical details on the dataset.

3.2.2 Design the market participants network

We model the intraday liquidity demand on the sell side (buy side) and we design the simple graph of market participant according to Social Network Analysis (SNA) tools. The set of all the transactions on a given day in a given venue is given by the network:

$$G_{it} = (V_{it}, E_{iit}) \tag{3.1}$$

where each node (vertex) v represent a market-participant buying or selling a given stock in day t. Each node is also provided with a given (attribute) weight w_v representing the end(start) of day inventories.

For our purpose, market transactions are visualized through the edges that link different nodes and are weighted by their turnovers (size of the arch). Thus, the underlying network of liquidity is created and updated at the day start and end of day, and is represented formally by the squared (vxv) adjacency matrix G_t:

$$[{}^{v}A(G)]_{ij} = \begin{bmatrix} 0 & a_{ij}w_iw_jk_{ij} \\ a_{ji}w_jw_ik_{ji} & 0 \end{bmatrix} \tag{3.2}$$

Where i and j are the buyers and sellers trading on each venue in each day. The element a_{ij} is 1 when there is a transaction between i and j, (i.e. $i,j \in$ E) or is 0 otherwise (i.e. $i,j \notin$ E). The diagonal of this matrix contains zeros since there are not transactions between a participant and itself. w_i and w_j represent the weights of the nodes i and j, that is the end (start) of day inventory level of the market participant (node). k_{ij} represents the weight of the edge between i and j. Economically, this is interpreted as the total daily turnover traded between i and j.

Once the network is designed, we show how the contagion spread across edges and infects nodes.

3.2.3 Risk and intensity of contagion

In this section we aim to answer to our first research question: *How can we define the risk of contagion and measure its intensity?*

As in Cohen-Cole, Kirilenko, and Patacchini 2014 we focus on the link between the network structure and liquidity risk, defining the network as the given number of transactions executed among traders in one day time. Compared to them we had not looked at the profitability of traders, instead we exploited an old determinant of broker-dealers model, focusing on a risk shared by all the market participants: inventory risk. Inventory risk can be simply explained as the risk for buy order flows and sell order flows of being unbalanced (see Section 3.1.3).

Most studies of direct financial networks focused only on the reliability of financial institutions because they tend to be the most interconnected nodes. Levent Uslu and Evren 2018 for instance mapped traders into informational categories, suggesting that financial institutions are more likely to act as informed traders compared to individuals. A possible drawback of these definitions is that in these cases the analysis is strictly linked to the classification used to label all different firms.

We propose a direct network approach in which we include the whole market participant's cohort.

We assume direct competition among dealers, thinking of them as they were competing each time on the same ISIN. We think the liquidity as a flow in the network. It is determined in each moment, endogenously by:

a the Demand (D) of the flow: this parameter indicates how much flow (in our case turnover) a node wants to send (negative demand) or receive (positive demand).

b the Capacity (C) of the edge: each edge of the graph (Broker-Dealer) can support a determined amount of flow, this amount is expressed by the attribute capacity. In our case this parameter is set to the value of dealer's Inventories.

If the market is sufficiently interconnected, buy-side participants will always be able to find their demand satisfied by market makers and liquidity providers. Once designed, we can simulate shocks in this network model: What happen when we remove one of the liquidity provider? Is the network still able to satisfy the liquidity demand? At which point the market can incur in a liquidity breakdown?

To answer these questions, we simulate a series of shocks, removing one by one all the nodes representing official market makers from the network. We simulate liquidity breakdowns in our network and studied the contagion produced endogenously by the exposure of one market participant (node) trading with (linked to) an infected node.

To understand the dynamic of the contagion we present an example.

TABLE 3.1: Illustrative example of the contagion dynamic

Buyer	Turnover	Total sellers demand	Starting Inventories	Total Turnover w/o node	%per node	Surplus Turnover per node	EoD. Inventories
node removed	300		2000				
other node1	150	1000	700		21%	64	636
other node2	210		800		30%	90	710
other node3	340		100	700	49%	146	**-46**

Assume a buyer A is trading with a set of sellers who demand £1000mln in total. At the beginning of the day (T0) this demand is satisfied by the buyer A (for £300mln) buyer 1 (for £150), buyer 2 (£210ml) and buyer 3 (£340).

In T0 we inject a liquidity shock in node A, that is, we remove it from the network. Consequently, the aggregated demand can now be satisfied only by buyers 1,2 and 3. The way we decide to redistribute the turnover among them is proportional to the share of demand they are already satisfied. In other words, the £300mln not traded any more by buyer A are redistributed 21% (£64mln) to node 1, 30% (£90mln) to node 2 and 49% (£146mln) to node 3. Once the simulation ends, we repeat it removing the node of another market maker/HFTs/dealer.

The extra demand they are required to satisfy, affects the level of their end of day inventories. If the EoD inventories is still positive they are able to buy (satisfy more demand). Otherwise, as in the case of node 3, they would be able to buy only borrowing money and therefore they would have less incentive to buy and would be cheaper for them to sell a quantity of the accumulated stock.

This mean the node 3 in T_1 results infected and the contagion start spreading in the network. If the total demand finds a channel in the network to be satisfied the network is said to be stable. If nodes are infected to a level the demand cannot be satisfied any more the network is experimented a liquidity breakdown. The same reasoning follows for the market participant who is trading as a seller. A node is said to be infected ($V_{infected}$)when its inventories in T_1 result altered due to the removal of another shocked node. Formally the node is said to be shocked (in breakdown) when:

$$V_{Shocked} = \begin{cases} w_{v,t_0} + w_{v,t_1} \leq 0 \; if w_{v,t_0} > 0 \\ w_{v,t_0} + w_{v,t_1} \geq 0 \; if w_{v,t_0} < 0 \end{cases} \tag{3.3}$$

We built four metrics to proxy for the stability of the market:

- Rate of contagion: Measures the number of nodes that have to change their levels of inventories to cope with the new demand of the market.

$$RI = \frac{V_{infected}}{|V|} \tag{3.4}$$

- Speed of contagion: Measures the speed at which more nodes (greater portion of the market participants) have been infected (altered their turnovers) after the shock. It tells us if the contagion is spreading among market participants at a fast or at a slow pace.

$$\Delta vc = \frac{\partial s}{\partial t} = \frac{RC}{T} \tag{3.5}$$

- Rate of Nodes in Breakdown (RS): Measures the number of nodes that have cross the threshold in their level of inventories and for who is too costly/costlier to trade on this side of the book.

$$RS = \frac{V_{shocked}}{|V|} \tag{3.6}$$

- Alacrity: Measures the dynamic increment in the rate of contagion from day to day, that is when market participant changes their inventory levels after the shock. Is the exchange able to satisfy the total demand for the instrument after the shock?

$$a(t) = s''(t) = \frac{\Delta vc}{\partial t} \tag{3.7}$$

3.2.4 Infected and influencers: contagion and centrality

Our second research question concerns the relationship among the intensity of contagion and the specific topology of the network as defined by data. We aim to discover who are the market participants bearing the higher risk of contagion. Do they cover a specific position in the financial network? (i.e. are they more central or more peripheral nodes in the network?)

We investigate the relationship among our measures of contagion intensity and several measures of centrality of the originally shocked market participant node from which the contagion start being propagated.

A node is said to be central in a network if it is highly ranked, highly important. This importance is determined according a value-function on the vertices of the network. The way we determine this function, of course, also determines our definition of importance. Four key dimensions have been identified in the classification of centrality metrics: a. type of nodal involvement., b. type of walk considered, c. property of walk assessed and d. choice of summary metrics (Borgatti and Everett 2006). We have exploited different real-valued functions:

- Average Degree of Centrality of Infected and not Infected Region.

  The degree of centrality is the most immediate proxy for the risk of a node to be hit by a shock, an information or, generally, whatever is flowing through the network. It is determined by the number of edges each node (v) has. Following **wasserman1994** notation we write:

  $$CD(v) = nk_i = \Sigma_{j=1}^{n} a_{ij} \tag{3.8}$$

  where a_{ij} are elements of the adjacency matrix of the network as we saw in Section 3.2.2. In a certain sense the degree is also a measure of the popularity of vertex.

  We compute the average degree of centrality of the nodes included in the infected region and the average degree of centrality of nodes not affected by the contagion. It quantifies the proximity among market participants in the infected and non infected region. Dealers with higher centrality scores trade with more counter-parties than those with lower centrality scores. This measure can also be interpreted as an index of the capability of a dealer to lead information to other agents in the network.

- Current-flow betweenness centrality of infected and not infected region.

  The class of metrics under the label *betweenness centrality* describes the degree of participation of edges or nodes in communication between different region of the network. Compared to the shortest-path betweenness centrality (that counts shortest paths through a node or an edge) current-flow betweenness centrality is the amount of current flowing through the node, averaged over all the source-target pairs. It is also known as random-walk betweenness centrality. It accounts also for the quite longer paths which can however have edges that are important for communication processes in the network (Newman 2005).

  $$b_i = \frac{\Sigma_{s<t} currentflow_i^{st}}{\frac{1}{2}n(n-1)} \tag{3.9}$$

Betweenness centrality can be interpreted as a measure of the extent to which a vertex has control over information flowing between others. We compute the average edge current flow betweenness centrality of the edges included in the infected region and the average edge current flow betweenness centrality of the edges not affected by the contagion.

- Global Clustering Coefficient of infected and not infected region.

  The local clustering coefficient is the number of triangles in which vertex v_i participates normalised by the maximum possible number of these triangles (Kaiser 2008). In easy words, we can think to the sentence friends of friends are often friends, formally we have:

$$GCC(G) = \frac{2Tr(v)}{CD(v)(CD(v)-1)} \Sigma_{hj} \widehat{w_{ij}} \widehat{w_{ih}} \widehat{w_{hj}} \tag{3.10}$$

  Where $Tr(v) = \frac{1}{6} \Sigma_{h,i,j} a_{ij} a_{ih}, a_{jh}$ is the number of triangles through node v and $CD(v)$, as in Equation 3.8, is the degree of centrality of the considered vertex. The edge weights w_{ij} are normalised by the maximum weight in the network (Saramaki et al. 2006). We compute the average global clustering coefficient for the infected region and the average global clustering coefficient for the region not affected by the contagion.

- Transitivity of infected and not infected region.

  In mathematics, transitivity is a relation R over a set X of elements a,b,c. If R relates a to b and b to c then R also relates a to c. We can interpret network transitivity in the same way. Considering the simple network G with only 3 nodes h, i, j, if there is an edge between node i and node h, and an edge between node h and node j, and also j and i are linked then the network is transitive. We say there is a closed triangle. In larger networks, transitivity is the fraction of all possible triangles in the network i.e. the number of possible triads. Triads are two edges with a shared vertex, we can also call them *open triangles* (Wasserman Faust, 1994)

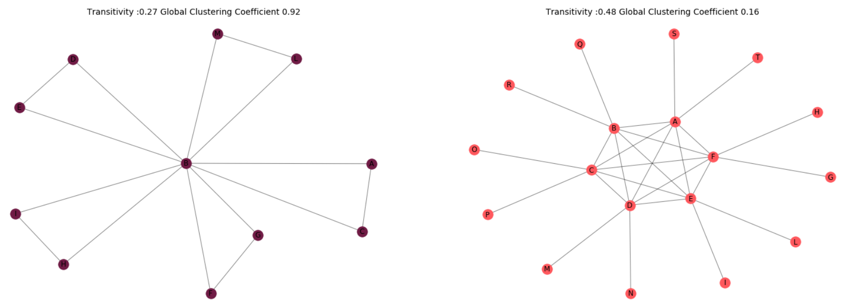
$$T(G) = \frac{3\Sigma Tr(v)}{\Sigma Triads} \tag{3.11}$$

  Transitivity and global clustering coefficient measures both the number of triangles in the network, but they differ in the way they sample random two-stars and for this reason they can produce different results Rohe 2015 and we want to include both in our analysis. Figure 2 presents a visual discrepancy between Transitivity and global clustering coefficient (GCC). The network in 3.2 (A) where only one node closes all the possible triangles, has a low transitivity (0.27) and a high GCC (0.92). On the contrary, network in 3.2 (B), where all the closed triangles are central and the open ones are peripheral, presents a transitivity higher then GCC.

Once centrality metrics are computed, we run a Panel OLS regression with fixed effects to find any relation with the metrics of contagion obtained as in Section 3.2.3.

$$y_{i,t} = \alpha_i + \beta_1 X_{1,i,t} + \beta_2 X_{2,i,t} + ... + \beta_k X_{k,i,t} + \varepsilon_{it} \tag{3.12}$$

Where we assume the error term $\varepsilon_{it} = e_{it} + \eta_i$ to account for firm specific effect. Each centrality metrics averaged in the region of infected and non-infected nodes



(A) GCC higher than Transitivity

(B) Transitivity higher than GCC

FIGURE 3.2: Transitivity and GCC

is embodied in the dependent variable $y_{i,t}$ and computed for each instrument i, in each time stamp t (daily aggregated in the results) for $t = 1, \ldots, T$ and $i = 1, \ldots, N$. The coefficient α_i represent the unobserved time-invariant individual effect. β is the $(k\ x\ 1)$ vector of unknown coefficients that we want to estimate.

3.3 Fragmentation or competition among venues?

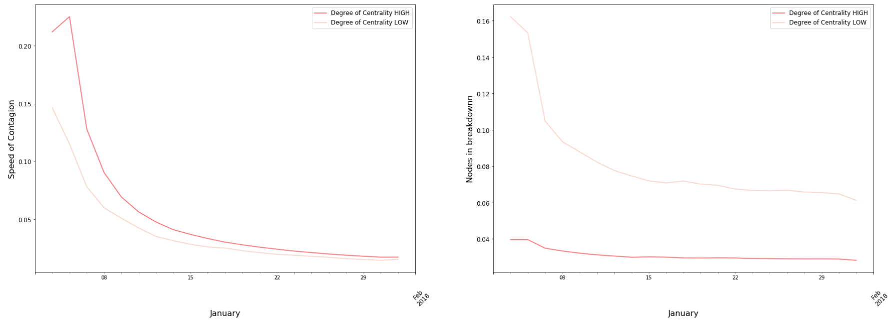
With our third research question, we investigated the impact of competition among venues on market quality. We run five distinct Panel OLS regressions to evaluate the impact of the current level of fragmentation on market quality proxies such as log returns, trading costs (effective and quoted spreads), price impact, first level depth (See A for a full description of different metrics on the left-hand-side of the panel regression). The model has the following form:

$$y_{i,p,t} = \alpha + \beta_1 NoVenues_{i,p,t} + \beta_2 NoCounterparties_{i,p,t} + \beta_3 var\left(P_{it}\right) + \beta_4 VIX + \\ + \beta_5 Turnover_{i,p,t} + \beta_6 Turnover_{it}\ +\ \varepsilon_{it} \quad (3.13)$$

Where we assume the error term $\varepsilon_{it} = e_{it} + \eta_i$ to account for firm specific effect. $y_{i,p,t}$ is the dependent variable computed for each instrument i, in each time stamp t (daily aggregated in the results), for each market participant p. $NoVenues_{i,p,t}$ is the number of unique different venues the participant is actively trading on. $NoCounterparties_{i,p,t}$ represents the number of unique sellers the considered market participant is trading with. We added some controls: $var\left(P_{it}\right)$ is the variance of the price of instrument p in time t. VIX is the fear index closing price, to proxy for the volatility, $Turnover_{i,p,t}$ is the total turnover per participant, per instrument per day and $Turnover_{it}$ is the total turnover of the instrument per day.

3.4 It's a wild, sell-side world: competition among dealers

Fourthly, we looked at the role of competition among dealers and others sell-side participants highlighting potential side-effects deriving from the risk of concentration.



(A) Speed of contagion (%)

(B) Nodes in breakdown (%)

FIGURE 3.3: Contagion intensity metrics

Some niches of buy-side participants, in fact, execute all their trades with only one sell-side participant. Network analysis highlights how difficult is for them to access the market via multiple market makers, whilst it is much more frequent that they execute all their trades with the same one.

A very useful tool provided by the application of graph theory to our dataset is the degree distribution analysis. Our network, as many real-world ones, is characterised by a low degree in the large majority of its nodes and a high degree in few ones known as hubs.

In Annex 2 (privileged distribution) we apply this analysis to the specific case of Retail Service Providers (RSP) showing how this might call a best execution rationale in question.

3.5 Results

In Section 3.2.2 we defined the risk of contagion and derived four metrics to measure its intensity. The cornerstone of our model is the inventory update mechanism of each dealer/sell-side participant/ market maker. Here we answer the three remaining research questions.

3.5.1 Who might impair liquidity in the network?

We investigate the relationship among the intensity of contagion and measures of centrality of the originally shocked market participant described in Section 3.2.3 and 3.2.4 from which the contagion start being propagated.

We fund a positive correlation between the degree of centrality of a node and the speed of contagion: the most peripheral nodes are also slower in spreading the contagion across the network, whilst the most central ones infect most of their neighbours in the immediate aftermath. These nodes represent the market participants bearing the highest risk of contagion for the network as proxied by both the contagion alacrity and the number of nodes in breakdown.

Figure 3.3 presents the average daily speed of contagion and percentage of nodes in breakdown for the group of market participant with a high degree of centrality (>=10) and the participants with a low one. Interestingly, there is a direct relationship between the speed of contagion and the degree of centrality.

Parameter estimations of the regression are reported in 3.2.

TABLE 3.2: Infected and Influencers: Parameters estimation of PanelOLS model

	Alacrity	% of Nodes in Breakdown	Rate of Contagion	Speed of Contagion
Avg. b infected	-0.0004	-0.041	-0.37 *	0.06
	0.0002	*0.021*	*0.1*	*0.05*
Avg. b not infected	-0.0005	-0.04	0.19	-0.09
	0.0004	*0.06*	*0.14*	*0.12*
Avg. CD infected	0.0006	0.05 *	-0.3	-0.17
	0.0003	*0.043*	*0.152*	*0.1*
Avg. CD not infected	0.001	0.135	1.8 *	0.26
	0.0004	*0.05*	*0.44*	*0.13*
GCC infected	0.002 **	0.027 *	0.012	0.3
	0.0005	*0.014*	*0.03*	*0.24*
GCC not infected	0.0015 ***	-0.24 *	-0.92	-0.05
	0.00019	*0.09*	*0.494*	*0.02*
Transitivity not infected	-0.001	-0.02	0.093	-0.08
	0.0004	*0.04*	*0.11*	*0.3*
Transitivity infected				-0.03
				0.09
Covariance estimator				Clustered
R-overall	95%	84%	86%	48%
No obs.	460899	460899	460899	460899

Finally, we performed a 2-sample-Kolmogorw-Smironv statistic on the average transitivity, global cluster coefficients, current-flow betweenness centrality and degree of centrality for the region of infected and non-infected nodes for all the 313 equity instruments included in our analysis. We could reject the null-hypothesis that the two sample are drawn from the same distribution since the pp-values was always below 1%.

3.5.2 Competition among venues

Fragmentation of market venues is significantly correlated with higher implicit costs of transactions. In contrast, the ability of a participant to interact with multiple different counter-parties leads to a reduction in quoted spreads. we found that when a buyer trades on 1 more trading venue there is an increase of 0.102 bps in quoted spread, whilst when a buyer can match one more seller we see a reduction of -0.594, confirming findings as in Huang and Masulis 1999. P-values are significant even including the VIX control. The contribution of traded turnover, is significant but extremely small.

Our findings indicate that costs of transactions are correlated with the number of venues on which our traders are active. Considering a financial firm active on multiple venues, that trades multiple instruments in different days, we found that an increase of one unit in the number of venues the trader is active on, is positively and significantly correlated with an increase in the cost of transactions. On the contrary, the transaction with 1 more counter-party, reduces the quoted spread by -0.594 bps.

Figure 3.4 shows the positive slope relation between average spreads by buyer per instrument and the number of venues the buyer is active on. Our results seem to confirm Hendershott and Jones 2005 rather than O'Hara and Ye 2011.

Annex B reports the parameter estimations from the base line Pooled-OLS regression and a Fixed Entity Effect model with clustered covariance for several liquidity proxies. Our findings show higher implicit costs of transactions associated with a more fragmented trading environment (more venues). Contextually, for dealers competing on multiple venues for the same ISINs we have found a reduced first level depth of the order books. We had not find significant impacts on returns and price impacts. Profits for liquidity provision (proxied by realised spreads) are instead slightly higher when a dealer is trading on more venues.

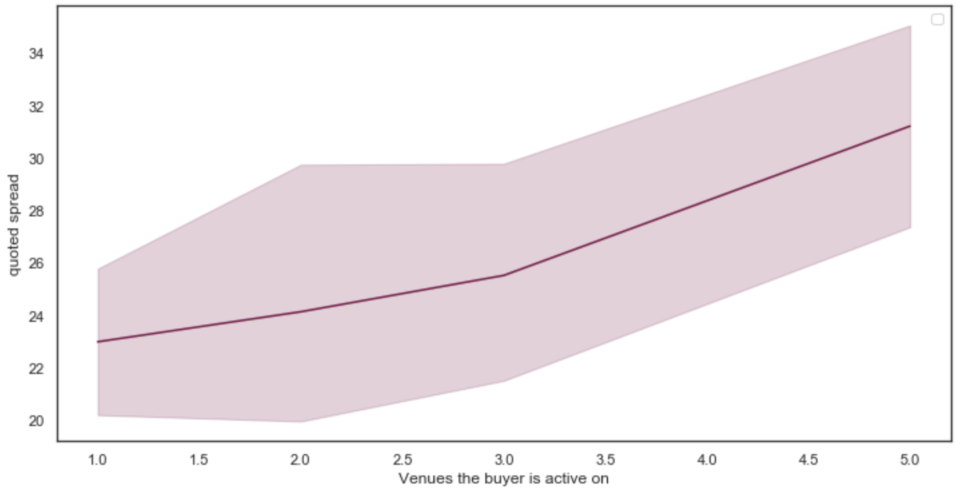


FIGURE 3.4: Positive correlation between spreads and number of venues the trader is active on

3.5.3 Centrality scores by firm type

Does the network analysis raise concerns around concentration of trading and a subsequent impairment of best execution? Agency-brokers tend to lead more information than other firm type as we can see from the centrality scores reported in Table 3.3.

TABLE 3.3: Centrality Scores

Buyer Firm	Degree Centrality	In Degree Centrality	Out Degree Centrality
Agency Broker	0.22	0.09	0.13
Interdealer Broker	0.2	0.08	0.11
SI	0.14	0.07	0.07
Prop Trader - HFT	0.11	0.06	0.05
RSP	0.08	0.05	0.03
IB - Large Dealer	0.07	0.04	0.03
market maker	0.07	0.04	0.03
Fund manager	0.09	0.05	0.03
IB - Small or Medium	0.05	0.03	0.02
Private/Commercial Bank	0.02	0.01	0.01
Asset Manager	0.02	0.01	0.01

3.6 Conclusion

In this paper, we simulate liquidity breakdowns on UK equity market adopting a direct network approach. Compared to previous studies in the literature we focused on the link between the structure of the network and the role of dealer's inventories.

In our simulations, the risk of liquidity breakdown stemmed directly from the dealer's inventory optimisation problem. We fixed a cut-off under which is too costly for the agent to execute the transaction and we *infected* its node. The untraded turnover is distributed proportionally among other participants and their inventory positions are modified accordingly. The ability of a system to deal with the shock changes with different financial agents and different liquid/illiquid instruments as well as with different levels of competition among dealers and fragmentation among UK venues. Along these lines, the dynamic of the contagion in our approach is subject to the dynamic of dealer's inventories.

The aim of this approach is to explain the efficiency and quality of stock exchanges, in the most structural aspect of their functioning: the interaction among their participants. The way participants interact on the exchange, and the way they compete, ultimately affects the price of the traded instruments, the liquidity offered on the exchange and the *"fair treatment"* of their clients.

First, we model the intra-day liquidity demand on the sell side (buy side) as a simple direct graph of market participant according to Social Network Analysis (SNA) tools. The set of all the transactions on a given day in a given venue is given by the network where each node represents a market-participant buying or selling a given stock in a given day.

Each node is also provided with a given (attribute) weight w_v representing the end(start) of day inventories. For our purpose, market transactions are visualized through the edges that link different nodes and are weighted by their turnovers (size of the arch). Thus, the underlying network of liquidity is created and updated at the day start and end of day.

Once the network is designed, we define the contagion as the alteration in a dealer's inventory level due to a liquidity breakdown produced endogenously in the market by its exposure with an infected node. An infected node is a sell-side participant for who is too expensive to keep its position (buy/sell) and is therefore removed (shocked) from the network.

The analysis displays that agency brokers present the highest centrality score in the network. This means that participants within this firm type tend to lead more price information than others as well as to spread faster the risk of liquidity disruption.

We have fund a positive correlation between the degree of centrality of a node and the speed of contagion: the most peripheral nodes are also slower in spreading the contagion across the network, whilst the most central ones infect most of their neighbours in the immediate aftermath. These nodes represent the market participants bearing the highest risk of contagion for the network as proxied by both the speed of contagion and its alacrity.

On the other side, this study reveals that nodes with a higher degree of centrality presents lower rates of contagion compared to more peripheral nodes. This finding suggest that more central nodes are linked with other stable nodes whose inventory levels are hard to get altered, but when the contagion start it spread faster than when the shock starts in the periphery of the network.

Thirdly, we looked at the competition among venues and the competition among dealers. We show that fragmentation of market venues is significantly correlated with higher costs of transactions. In contrast, the ability of an agent to interact with multiple different counter-parties leads to a reduction in quoted spreads.

Fourthly, the network analysis highlights how difficult it is for some niches of buy-side participants to access the market via multiple market makers, whilst it is much more frequent that they execute all their trades with the same one. This might call a best execution rationale in question.

All in all, our study contributes to the existing empirical market micro-structure literature adopting a novel, distinctive propagation algorithm based on market participant inventories. The direct-network approached based on dealer's inventory can provide regulators an extra tool to monitor the risk of liquidity break-downs in the equity market, identifying who are the participants who can spread the risk faster and wider.

Conclusions

Market quality can be declined in a plurality of way. In this dissertation we focused on three dimensions: liquidity, transparency, and stability. The way market participants access to the market, the way available prices are aggregated on the exchange, and the way they reflect the available information around the stocks fair value, all of these directly affect liquidity and credit costs. Regulation shape the quality of the market. In the last two years market regulators have been trying to quantify the MiFID II impact on market quality. One particularly controversial aspect of this regulation has been the adoption of a new tick size regime (Art. 49). The Financial Conduct Authority (FCA) in UK as well as the Autorité des Marchés Financiers (AMF) in France have conducted assessments on the tick size regime, which entered into force in January 2018 as part of MiFID II, concluding the regime had an opposite impact on the overall market quality of UK and EU venues.

In the first paper, we have evaluated the impact of the MiFID II tick size regime (Art 49) on four of the UK trading venues that have implemented it in January 2018.Although the new regulation was carachterised by the goal of improving transparency, the ultimate quality of the UK venues did not improved. In a transparent market, public participants can easily obtain good information about current market condition. Knowing the quotes and trade sizes and prices, better enables public buyers and sellers to monitor and assess the quality of the execution they have received. Greater transparency should mirror better market quality. Nevertheless, we believe transparency is not a value per se. Too much transparency can discourage the provision of dealers' capital and in so doing cause a market to be less liquid. We found that narrower tick sizes, all other things equal, coincide with higher incentives for market makers to provide liquidity, as well as with a deterioration of a desired order viscosity. They were in fact associated with shorter orders' lifetime (i.e. how long a quote remains in the order book before the order is fulfilled) and a lower transaction mid-size. The first paper was preparatory for the second. Unintended consequences on UK venues have been measured and the regime resulted not perfectly calibrated for UK equities. This consideration open-up to whether a better methodology can be constructed to improve the effectiveness of the regime.

In the second paper, we adopt a supervised machine learning approach to propose a better calibrated alternative to ESMA grid. Our approach is based on: (i) market capitalization; and (ii) quoted spread. Our proposed calibration for the regime would achieve optimal tick sizes for equities 3 times more frequently than the current ESMA regime. This allows us to outline an idealized grid for determining an equity's minimum tick size for this proposed regime. This paper is especially relevant for UK policy makers in the context of the UK leaving the EU and suggests the ESMA grid can be abandoned. At the best of authors' knowledge, it is also the first time a supervised machine learning model is adopted to evaluate policy implications of a financial regulation.

At the centre of our investigations in all the three essays, is the activity of market makers and HFTs. In relation to the tick size change we found higher shares of turnovers associated with narrower tick sizes. In our last paper we saw all markets can be seen as network participants who come together to trade. The larger is the network the greater is the value it can offer to the participants. This is because there are more opportunity to match orders and the large is the number of traders using the same venue the more order converge in the marketplace the greater is the liquidity. “No man is an island” quoting John Donne and this is particularly true in the network of financial markets. Order flows attracts order flows. Market makers trades from their own inventory as principals. When public investors want to buy, market makers firms sell from their own portfolio either using long positioning or going short. When public investors want to sell market makers firms buy for their own portfolio reducing a short position or going long. This reflects the broad market desire to buy in to sell shares. Liquidity provision is a service that is commonly attributed to market makers but they cannot be the ultimate source of liquidity: after buying shares from a public seller the market maker hopes to sell the shares to a public buyer. If the prices are set properly and if the public buy and sell orders are reasonably balanced the market makers inventories will stay reasonably flat or close to zero. But if the public buy and sell orders do not off-set each other sufficiently, an inventory imbalance will develop. When it does the market makers is forced to re liquefy by adjusting the quotes or by the inter-dealer trading. The market maker is not a typical investor, in which they don’t buy or sell for investment purposes and supplies shares to others to absorb them from others and in so doing the market maker calmly acquires a poorly diversified portfolio and accepts risk that could have been diversified away. The reward for bearing this risk is the bid ask spread. A difficult inventory control automatically translates into more costly market making. In the last paper we simulate liquidity breakdowns on UK equity market adopting a direct network approach. Compared to previous studies in the literature we focused on the link between the structure of the network and the role of dealer’s inventories.

In our simulations, the risk of liquidity breakdown stemmed directly from the dealer’s inventory optimisation problem. We fixed a cut-off under which is too costly for the agent to execute the transaction and we *infected* its node. The untraded turnover is distributed proportionally among other participants and their inventory positions are modified accordingly. The ability of a system to deal with the shock changes with different financial agents and different liquid/illiquid instruments as well as with different levels of competition among dealers and fragmentation among UK venues. Along these lines, the dynamic of the contagion in our approach is subject to the dynamic of dealer’s inventories.

The aim of this approach is to explain the efficiency and quality of stock exchanges, in the most structural aspect of their functioning: the interaction among their participants. The way participants interact on the exchange, and the way they compete, ultimately affects the price of the traded instruments, the liquidity offered on the exchange and the *"fair treatment"* of their clients.

First, we model the intra-day liquidity demand on the sell side (buy side) as a simple direct graph of market participant according to Social Network Analysis (SNA) tools. The set of all the transactions on a given day in a given venue is given by the network where each node represents a market-participant buying or selling a given stock in a given day.

Each node is also provided with a given (attribute) weight w_v representing the end(start) of day inventories. For our purpose, market transactions are visualized through the edges that link different nodes and are weighted by their turnovers (size of the arch). Thus, the underlying network of liquidity is created and updated at the day start and end of day.

Once the network is designed, we define the contagion as the alteration in a dealer's inventory level due to a liquidity breakdown produced endogenously in the market by its exposure with an infected node. An infected node is a sell-side participant for who is too expensive to keep its position (buy/sell) and is therefore removed (shocked) from the network.

The analysis displays that agency brokers present the highest centrality score in the network. This means that participants within this firm type tend to lead more price information than others as well as to spread faster the risk of liquidity disruption.

We have fund a positive correlation between the degree of centrality of a node and the speed of contagion: the most peripheral nodes are also slower in spreading the contagion across the network, whilst the most central ones infect most of their neighbours in the immediate aftermath. These nodes represent the market participants bearing the highest risk of contagion for the network as proxied by both the speed of contagion and its alacrity.

On the other side, this study reveals that nodes with a higher degree of centrality presents lower rates of contagion compared to more peripheral nodes. This finding suggest that more central nodes are linked with other stable nodes whose inventory levels are hard to get altered, but when the contagion start it spread faster than when the shock starts in the periphery of the network.

Thirdly, we looked at the competition among venues and the competition among dealers. We show that fragmentation of market venues is significantly correlated with higher costs of transactions. In contrast, the ability of an agent to interact with multiple different counter-parties leads to a reduction in quoted spreads.

Fourthly, the network analysis highlights how difficult it is for some niches of buy-side participants to access the market via multiple market makers, whilst it is much more frequent that they execute all their trades with the same one. This might call a best execution rationale in question.

All in all, our study contributes to the existing empirical market micro-structure literature adopting a novel, distinctive propagation algorithm based on market participant inventories. The direct-network approached based on dealer's inventory can provide regulators an extra tool to monitor the risk of liquidity break-downs in the equity market, identifying who are the participants who can spread the risk faster and wider.

Appendix A

Chapter 1: Data and methodology

A.1 Chapter 1: Dataset Description

Our analysis exploits two datasets. A Refinitiv dataset (formerly Thomson Reuters Tick History) including quotes and trades for different venues timestamped at the millisecond and MiFID II Transaction Reports an FCA proprietary dataset including transactions with information related to reporting firms and their counterparty.

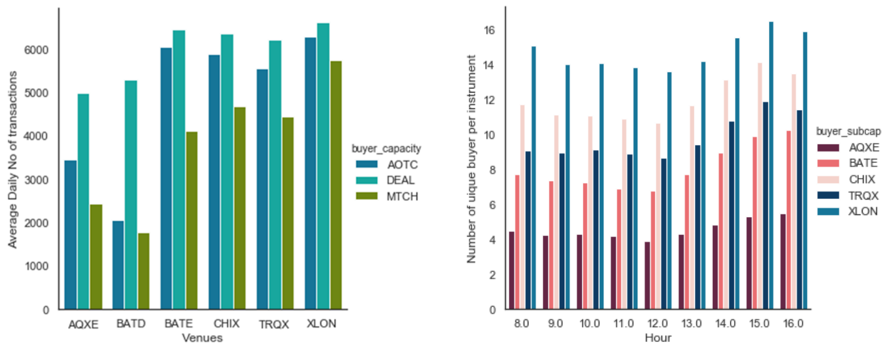
TABLE A.1: Total Turnovers by firm type across tick size group

	All Obs		**Control**		**Treatment**	
	Abs £m	**%**	**Abs £m**	**%**	**Abs £m**	**%**
Agency Broker	6347.38	*3.97*	243.25	*1.66*	327.11	*2.12*
Asset Manager	3628.11	*2.27*	231.04	*1.58*	140.25	*0.91*
Central bank	5.96	*0*				
Charity Fund	0.87	*0*				
Clearer, Custodian or Fund Manager	727.17	*0.45*	54.09	*0.37*	1.21	*0.01*
Commercial Bank	2030.6	*1.27*	103.85	*0.71*	146.37	*0.95*
Corporate	331.71	*0.21*	5.23	*0.04*	10.36	*0.07*
Exchange Operator	1109.44	*0.69*	27.58	*0.19*	13.02	*0.08*
Financial Intermediary	1144.16	*0.71*	174.73	*1.19*	36.02	*0.23*
Fund Manager	706.85	*0.44*				
Fund of Funds						
Hedge Fund	1688.69	*1.06*	255.69	*1.75*	186.05	*1.21*
IB - Large Dealer	65176.55	*40.72*	6939	*47.39*	6830.62	*44.26*
IB - Small or Medium	9260.76	*5.79*	585.51	*4*	722.43	*4.68*
Interdealer Broker	230.4	*0.14*				
Market Maker	10340.57	*6.46*	89.07	*0.61*	66.59	*0.43*
Other Funds	8432.88	*5.27*	350.43	*2.39*	390.09	*2.53*
PLC	11.5	*0.01*				
Pension Funds	95.63	*0.06*	0.32	*0*	0.29	*0*
Private Bank	164.11	*0.1*	11.93	*0.08*	12.29	*0.08*
Prop Trader - HFT	45244.36	*28.27*	5546.81	*37.88*	6518.72	*42.24*
Unclassified	3378.11	*2.11*	25.13	*0.17*	29.77	*0.19*
Total	160055.8	*100*	14643.67	*100*	15431.21	*100*

Our RHS and LHS metrics are computing using quotes from Refinitiv. A cleaning phase was necessary in order to remove outliers and misreported prices. We

TABLE A.2: Total turnovers by firm type across venues

	AQX		BATE		CHIX		TRQ		XLON	
	Abs £m	**%**	**Abs £m**	**%**	**Abs £m**	**%**	**Abs £m**	**%**	**Abs £m**	**%**
Agency Broker	59.16	*2.41*	128.72	*1.78*	412.88	*1.96*	185.83	*1.7*	910.89	*1.52*
Asset Manager	22.62	*0.92*	87.36	*1.21*	251.56	*1.19*	98.99	*0.9*	696.85	*1.17*
Clearer, Custodian or Fund Manager	0.44	*0.02*	7.68	*0.11*	9.06	*0.04*	7.11	*0.06*	32.46	*0.05*
Commercial Bank	22.92	*0.93*	45.08	*0.62*	134.78	*0.64*	61.81	*0.56*	540.83	*0.9*
Corporate	1.29	*0.05*	27.33	*0.38*	54.4	*0.26*	38.91	*0.36*	171.82	*0.29*
Exchange Operator	3.01	*0.12*	3.12	*0.04*	17.43	*0.08*	8.23	*0.08*	135.68	*0.23*
Financial Intermediary	9.84	*0.4*	10.56	*0.15*	48.51	*0.23*	15.95	*0.15*	104.85	*0.18*
Fund of Funds	0	*0*		*0*	0	*0*		*0*		*0*
Hedge Fund	0.03	*0*	62.23	*0.86*	181.51	*0.86*	186.03	*1.7*	1162.23	*1.94*
IB - Large Dealer	824.58	*33.53*	3022.65	*41.83*	9674.97	*45.94*	5187.55	*47.39*	29388.6	*49.16*
IB - Small or Medium	211.29	*8.59*	238.48	*3.3*	726.24	*3.45*	492.17	*4.5*	2470.65	*4.13*
Market Maker	23.62	*0.96*	13.91	*0.19*	96.41	*0.46*	24.9	*0.23*	205.12	*0.34*
Other Funds	71.15	*2.89*	152.15	*2.11*	568.51	*2.7*	188.69	*1.72*	1142.38	*1.91*
Pension Funds	0.2	*0.01*	0.53	*0.01*	5.2	*0.02*	1.19	*0.01*	13.02	*0.02*
Private Bank	5.43	*0.22*	5.16	*0.07*	22.16	*0.11*	7.32	*0.07*	50.92	*0.09*
Prop Trader - HFT	1194.01	*48.56*	3407.68	*47.16*	8812.21	*41.84*	4429.88	*40.47*	22662.02	*37.91*
Unclassified	9.3	*0.38*	13.3	*0.18*	44.53	*0.21*	12.15	*0.11*	97.25	*0.16*
Total	2458.9	*100*	7225.94	*100*	21060.35	*100*	10946.7	*100*	59785.59	*100*



(A) Avg. no of transactions by venue and buyer capacity

(B) Intraday participation per venue

FIGURE A.1: Summary statistics

used end of prices as benchmark and we removed all of the observations in non-continuous market hours. We also removed all cross spread quotes and invalid quantity and volumes.

We use 35 instruments among 109 venues and 314 equity instruments included in our cleaned dataset, because they exhibited consistent mismatch of the tick across different venues. The majority of these ISINs were classified as moderately liquid according to ESMA (15 instruments from liquidity band 5 (42%) and 13 instruments from liquidity band 4 (37%)). The control group includes the same number of instruments per liquidity band as the treatment group.

A.2 Chapter 3: Dataset Description

For our network analysis, our dataset contains 21 trading days, 314 ISINs from FTSE350 index, 54 venues, 1280 unique buyers and 1302 unique sellers that we have classified in 11 firm type. Figure 6 is a directed graph displaying the average daily traffic on one instrument for one month. The size of the nodes represents the firm type's inventories, whilst the width of the edges represents the average exchanged turnover. Fund managers handle the highest inventories levels, whilst the largest turnovers are exchanged between Investment banks (large dealers), official Market makers, Systematic Internalisers and high frequency traders.

Figure A.1 shows that the great part of transactions are executed by dealers on their own account (deal capacity).

Table **??** and Table **??** show how even if the buy-side of the market is much more diversified (1) (greater number of asset managers and fund managers) of course it is the sell-side, SI, proprietary traders and inter-dealer brokers that executes the greater number of transactions (2).

TABLE A.3: List of instruments in the treatment and control group

Treatment	Market capitalisation (in £m)	Sector	Liquidity band
GB0008706128	41,286	Banks	6
GB00B03MLX29	96,503	Oil, Gas and Coal	6
GB00B1XZS820	27,792	Industrial Metals and Mining	6
GB0000811801	6,541	Household Goods and Home Construction	5
GB0005603997	16,201	Life Insurance	5
GB0007197378			5
GB0008782301	5,556	Household Goods and Home Construction	5
GB0031698896	1,510	Travel and Leisure	5
GB0032089863	8,641	Retailers	5
GB0033195214	4,366	Home Improvement Retailers	5
GB00B01C3S32			5
GB00B0HZP136			5
GB00B18V8630	3,866	Gas, Water and Multi-utilities	5
GB00B1KJJ408	5,594	Travel and Leisure	5
GB00BK1PTB77	2,034	Electrical Components	5
GB00BMJ6DW54	9,914	Media Agencies	5
GB00BYZWX769			5
GB00BZ4BQC70	6,133	Chemicals	5
GB0007668071	2,149	Banks	4
GB0009292243	2,030	Chemicals	4
GB0009887422	1,279	Chemicals	4
GB00B63QSB39	2,082	Food Retailers and Wholesalers	4
GB00BGLP8L22	2,997	Electronic and Electrical Equipment	4
GB00BKRC5K31	946	Construction and Materials	4
GB00BLT1Y088	590	Property and Casualty Insurance	4
GB00BVC3CB83	1,765	Household Goods and Home Construction	4
GB00BYM8GJ06	1,363	Software and Computer Services	4
GB00BYRJH519	1,159	Property and Casualty Insurance	4
GB00BYXJC278	1,037	Construction and Materials	4
GI000A0F6407	588	Travel and Leisure	4
IE00B1RR8406	6,470	General Industrials	4
GB00B128J450	615	Asset Managers and Custodians	3
GG00B4ZPCJ00	935	Real Estate Holding and Development	3
GG00BBHX2H91	2,055	Closed End Investments	3
JE00BVRZ8S85	796	Administration, reporting and fiduciary service	3

TABLE A.4: List of instruments in the treatment and control group Cont.

Control	Market capitalisation (in £m)	Sector	Liquidity band
GB0031348658	29,379	Banks	6
GB00B03MM408	181,825	Integrated Oil and Gas	6
GB00B24CGK77	41,125	Nondurable Household Products	6
GB0006825383	7,754	Household Goods and Home Construction	5
GB0002162385	17,183	Life Insurance	5
GB0001411924			5
GB00B02L3W35	5,762	Household Goods and Home Construction	5
GB00B7KR2P84	5,142	Travel and Leisure	5
GB0001367019	5,132	Diversified REITs	5
GB0005576813	348,892	Home Improvement Retailers	5
GB00B83VD954			5
GB00BD8QVH41			5
GB00B033F229	4,293	Gas, Water and Multi-utilities	5
GB0031215220	5,869	Travel and Leisure	5
GB0003308607	3,029	Electrical Components	5
GB0006776081	5,370	Media Agencies	5
GB00BD8YWM01			5
GB00B1WY2338	6,468	General Industrials	5
GB00BD6GN030	2,031	Banks	4
GB0009633180	2,858	Pharmaceutical	4
GB00B012BV22	1,453	Building Materials: Other	4
GB00B2PDGW16	2,646	Speciality Retailers	4
GB0006027295	787	Electronic and Electrical Equipment	4
GB00B3Y2J508	754	Construction and Materials	4
BMG4593F1389	3,698	Property and Casualty Insurance	4
GB00BYPHNG03	1,626	Household Goods and Home Construction	4
GB00BYZFZ918	2,783	Software and Computer Services	4
BMG5361W1047	1,399	Property and Casualty Insurance	4
GB0008025412	657	Construction and Materials	4
GB00B6YTLS95	732	Travel and Leisure	4
GB00BWFGQN14	6,151	Machinery: Industrial	4
GB0008829292	1,903	Closed End Investments	3
GB00BD7XPJ64	576	Real Estate Investment Trusts	
GB0030517261	2,000	Closed End Investments	3
GB00BYWWHR75	819	Industrial Support Service	3

TABLE A.5: Dataset for network flow analysis

No obs.	ISIN	No days	No Venues	No Buyers	No Sellers	Firm type
308492	315	21	54	1280	1302	11

A.3 Chapter 3: Firm type Classification

Whilst in the US there are more than 3,700 firms acting as broker-dealers[1], FCA, in line with ESMA provides a list of authorised market makers and primary dealers[2]. For the purpose of this analysis, we try to identify an agency broker, according to the definition, as a broker that acts as a middleman on the stock exchange, and places orders on behalf of clients. We matched end buyers and sellers across venues. This is not trivial because each market participant reports to FCA only its *leg* of a transaction and there are always Clearing Houses (CCPs) or brokers between two end users that must be removed without losing the information they convey about their clients. Moreover, bilateral agreements off-book can be executed on exchange [3]. The cleaning and de-duplication procedure allows us to unpack the transactions executed on so-called "MATCH" capacity [4] and access to information related to Non-Clearing Members (NCNs [5]).

For the latter purpose, the role of the LEI, Legal Entity Identifier, is fundamental. It allows a further distinction among firms with divers levels of risk exposure and among core activities of a firm. Another piece of preliminary research regards the classification of counter-parties. In 2 months of analysed trades, we deal with more than 30,000 different LEIs (Legal Entity Identifiers).

However, 419 of them accounts for 95% of exchanged turnovers, thus for the sake of simplicity we decided to focus only on these participants, and labels all the remaining according their reported capacity as *Other Dealers* or *Other Agents*. We tried to be more granular and go further than a simple buy-side vs sell-side classification.

First, we started with the classification of LEI for RSP and market makers using the official list reported by LSE[6]. Second, we mapped the Systematic Internalisers (SI) using an FCA official list and then we moved to the Buy-Side using Orbis

[1]Source: https://www.finra.org/about/firms-we-regulate

[2]Source: https://www.esma.europa.eu/sites/default/files/library/list_of_market_makers_and_primary_dealers.pdf

[3]Rules 3000.1 and 3000.2. Source: https://www.londonstockexchange.com/traders-and-brokers/rules-regulations/rules-lse.pdf

[4]Article 4(38) of Directive 2014/65/EU: "*matched principal trading* means a transaction where the facilitator interposes itself between the buyer and the seller to the transaction in such a way that it is never exposed to market risk throughout the execution of the transaction, with both sides executed simultaneously, and where the transaction is concluded at a price where the facilitator makes no profit or loss, other than a previously disclosed commission, fee or charge for the transaction."

[5]This is an entity that has or has not direct market access, but rather wishes to use another financial intermediary (Clearing Member, CM) to execute orders. This is not a transmission of an order, as the NCM is actively trading on venue. Source: Reporting Guidelines.

[6]There are more than 20 registered market makers providing continuous pools of liquidity on LSE LOBs. Source: https://www.lseg.com/markets-products-and-services/our-markets/london-stock-exchange/exchange-traded-funds/marketmakers

TABLE A.6: Number of buyer and number of transactions per firm type

	AQXE	BATD	BATE	CHIX	TRQX	XLON	AQXE	BATD	BATE	CHIX	TRQX	XLON
						Buy Side						
Asset Manager	3%	3%	2%	2%	2%	2%	0%	1%	1%	1%	0%	1%
Fund Manager	11%	0%	9%	8%	9%	8%	2%	6%	4%	4%	5%	5%
Other buy side agents	67%	59%	71%	73%	70%	72%	4%	3%	2%	2%	2%	2%
Other buy side dealers	1%	1%	1%	2%	1%	2%	0%	0%	0%	0%	0%	0%
Private/Commercial bank	3%	2%	2%	2%	2%	2%	1%	0%	0%	0%	0%	1%
Custodian/Fund M.	0%	14%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Total Buy Side	**85%**	**79%**	**85%**	**87%**	**84%**	**86%**	**7%**	**10%**	**7%**	**7%**	**7%**	**9%**
						Sell Side						
Agency Broker	2%	3%	2%	2%	2%	2%	1%	1%	1%	1%	1%	1%
IB - Large Dealer	3%	3%	0%	0%	1%	1%	14%	16%	14%	14%	18%	16%
IB - Small or Medium	1%	1%	3%	3%	1%	3%	1%	1%	1%	1%	0%	1%
Interdealer Broker	5%	6%	4%	4%	4%	4%	2%	3%	1%	1%	1%	2%
Market Maker	2%	3%	2%	1%	2%	1%	1%	14%	6%	8%	5%	10%
Prop Trader - HFT	1%	1%	1%	1%	1%	1%	13%	2%	18%	18%	16%	16%
RSP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
SI	2%	3%	2%	2%	2%	2%	62%	52%	51%	49%	49%	44%
Total Sell Side	**16%**	**20%**	**14%**	**13%**	**13%**	**14%**	**94%**	**89%**	**92%**	**92%**	**90%**	**90%**

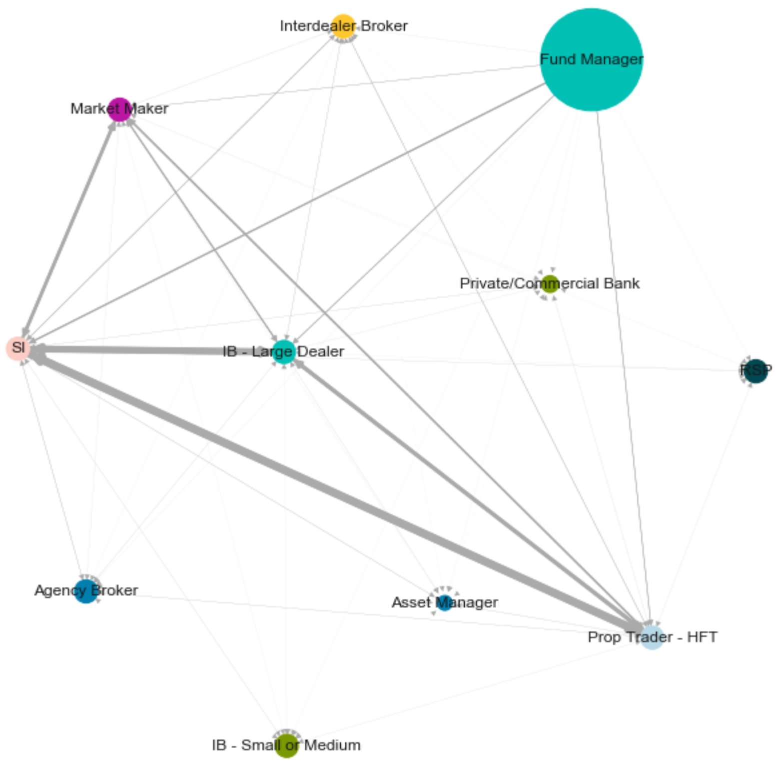


FIGURE A.2: Direct competition among market participant on one equity instrument

data-set, containing financial information on companies[7], scrapping the text and obtaining a wide classification of Asset Managers, Funds Managers, Investment Banks (Large or small-medium) and Private/Commercial Banks. Proprietary-Trading Firms and High Frequency Traders (HFTs) have been, for a long-time, object of research for FCA Economic Department (ED) and we decided to exploit their classification to identifies these subjects [8]. Although some categories are broadly defined (e.g. under the label "Asset Manager" there are different buy-side individuals or under the label "Fund Manager" there are both active and passive fund managers, ETFs as well as mutual funds), we believe that we are not losing any meaningful insight for our economic analysis of the market structure.

As far as the time horizon, we selected 21 trading days in January and February 2018. In Appendix 2 we focused on a particular subset of market participants isolating all the transactions executed against Retail Service Providers (RSPs, also referred to as market makers. We also selected only trades executed OTC, ignoring hedging strategies on lit because these trades would never be executed against retailers, instead they involve large dealer investment banks. We aggregated the total turnover executed by each counterpart, and we weighted each node (size of the node) by its number of counterparts (degree of the node).

[7]Orbis is a subscription-based service containing data on more than 310 million companies around the world. Source: https://www.bvdinfo.com/en-gb/our-products/data/international/orbis

[8]Acknowledgements to Peter O'Neil, PhD and Technical Specialist in FCA ED.

We removed all the buy-sell pairs with a turnover of less the 1 million £ in order to exclude one-time/small-size traders from the analysis. Finally, we highlighted in red nodes(retailers) that execute with only one RSP.

Figure A.2 is a directed graph displaying the average daily traffic on one instrument for one month. The size of the nodes represents the firm type's inventories, whilst the width of the edges represents the average exchanged turnover.

A.4 Chapter 1: Market quality metrics

We investigated if UK venues contribute in a different proportion to price discovery following the implementation of MiFID II in general and the tick size regime (Art.49) in particular.

To do so, we built a price discovery metric called Information Leadership Share (ILS) as described in Putniņš Putniņš 2013. It moves away from the common view of price discovery as the "who moves first" interpretation and it combines Hasbrouck Information Share (IS) (**Hasbrouk1995**) and Harris-McInish-Wood Component Share (CS) (**Harris2002**). ILS improves on the just mentioned measurements in as much as it has the ability to correctly attribute contributions to price discovery in the presence of different levels of noise. Noise can arise for different reasons. We are particularly interested in detecting the noise produced by the new discrete grid of prices under Art. 49. We expect exchanges with higher spreads to be also slower and noisier, and to follow innovations rather than contribute on new information. Two preliminary steps are required by the ILS: first we needed to compute the Information Share that focuses on the variance of innovation to the common factor. Second, we followed Gonzalo-Granger PT analysis based on the error correction process. The measurements for each pair of markets analysed are obtained, we have followed Yan and Zivot (Yan and Zivot 2010), using IL to measure which price leads the price innovation adjustment in the fundamental value:

$$IL_i = \left| \frac{IS_i}{IS_j} \frac{CS_j}{CS_i} \right| \tag{A.1}$$

$$IL_j = \left| \frac{IS_j}{IS_i} \frac{CS_i}{CS_j} \right| \tag{A.2}$$

Where i and j are pairs of the analysed market in the list [AQX, TRQ, LSE, BTE, SIX]. We have calculated the information leadership share of each pair and we have compared it in the sample pre- and post-MiFID II. Both PT and CS are based on a VECM model. For each security quoted on the five markets under analysis (103 instruments), we have computed and matched the one-second-time-stamped log-return on the five different exchanges. The assumption of a co-integration of order I is consistent for time series referring to the same instrument on five different exchange. Nevertheless, we checked our economic intuition through an eigenvalue test (Johansen, 1991). From the VAR(p) with non-stationary unit root in A.3, we have calculated the error correction terms A.4 and error correction vectors.

$$x_t = \varphi + \varphi_1 x_{t-1} + \ldots + \varphi_p x_{t-p} + \varepsilon_t \tag{A.3}$$

Note that Δx_t is I (0) and $\Pi = -\Phi(1)$ is singular.

$$\Delta x_t = \phi + \Pi x_{t-1} + \Sigma_{i=1}^{p-1} = \Phi_i^* \Delta x_{i-t} + \epsilon_t \tag{A.4}$$

We have run a seemingly unrelated regression (SUR) of VECM at 200lags and saved the residuals in order to derive α and the elements m_{11}, m_{12},m_{22} of Cholesky factorised matrix. Thus, we derived

$$CS_i = \frac{\alpha_i}{\alpha_i + \alpha_j} \tag{A.5}$$

and

$$S_i = \frac{\left(CS_i * m_{11} + CS_j * m_{21}\right)^2}{\left(CS_i * m_{11} + CS_j * m_{21}\right)^2 + \left(CS_j * m_{22}\right)^2}. \tag{A.6}$$

A.5 Chapter 3: Market quality metrics

Where: p is the price, a is the ask price, b is the bid price in the order book, D_{iy} is the number of days for which data are available.

TABLE A.7: Variables description

Variable	Units	Source	Definition	Note
Midquote	GBP	Refinitiv	$\frac{a+b}{/}2$	Computed at quote update, timestamped at millisecond per instrument per venue
Price	GBP	Refinitiv	$\Sigma_{t=1}^{\tau} p$	where T is the trade
Quoted Spread	bps	Refinitiv	$10000\frac{(a-b)}{m}$	
First Level Depth	GBP	Refinitiv	$Q_b id + Q_a sk$	
Effective Spread	2bps	Refinitiv	$20000d\frac{p-m}{m}$	where d is the buy indicator as described by Lee and Ready (1991). It takes value +1 and -1 if the trade is buy initiated or seller initiated
Realised Spread (1min)	2bps	Refinitiv	$\Sigma_{t=1}^{\tau} 20000d\frac{p-m_{1min}}{m}$	where T is the trade and the metric is weighted per second and matched with MiFID II transaction reports
Price Impact (5min)	2bps	Refinitiv	$\Sigma_{t=1}^{\tau} 20000d\frac{m_{5min}-m_0}{m_0}$	where T is the trade and the metric is weighted per second and matched with MiFID II transaction reports
Order Imbalance	GBP	Refinitiv	$Q_b id - Q_a sk$	
Transaction Midsize	GBP	Refinitiv	$\frac{1}{T}\Sigma_{t=1}\tau Q_t$	where T is trade
Quote Duration	fraction of second	Refinitiv	$quote_{t+1} - quote_0$	is the time the same ask and bid remains in the order book before an ask or bid update

TABLE A.8: Liquidity metrics

Liquidity proxy	**Formula**
Amihud ratio	$\frac{1}{D_{iy}} \sum_{i=1}^{D_{iy}} \frac{\lvert R\rvert_{iyd}}{\text{Vol}_{iyd}}$
Amivest measure	$\frac{\Sigma_{i=1}^{D_{iy}} \text{VOLD}_{iyd}}{\Sigma_{i=1}^{D_{iy}} \lvert R_{iyd}\rvert}$
Effective spread	$2b(p-m)$
Price impact (bps)	$E[(p_{5min}-p_0)\epsilon_0]-E[p_{5min}-p_0]E[\epsilon_0]$
Absolute quoted spread	$a-b$

Appendix B

Results

B.1 Further insights from chapter 1

To ensure our diff-in diff analysis remain robust we included in our regression of a control dummy for BoE announcements and a continuous variable for Brexit announcements. Also, we correct for heteroskedasticity and we applied a clustered covariance estimator.

B.2 Further insights from chapter 3

TABLE B.1: Fixed Effect LSVD Parameters Estimation. Each column refers to a different regression based on the choice of the market quality metric used as dependent variable. Robust standard errors are in italic font. Superscripts ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	OI 1 min	**OI 15 min**	**Spread % Price**	**Realised Spread 1 min**
constant	4014.8 ***	57210 ***	0.089 ***	2.4318 ***
	1545	*23000*	*0.0045*	*0.33*
dummy widened	1118.1	8204.9	0.1076 ***	6.7902 ***
	2196	*29390*	*0.0035*	*0.2677*
dummy narrowed	-8772 ***	-128900 ***	0.0088 ***	0.4955 ***
	2069.9	*30850*	*0.0013*	*0.1015*
Dwidned Dpost	-3515.2 *	-36440	0.0196 ***	1.0759 ***
	1885.4	*23520*	*0.0048*	*0.383*
Dnarrowed Dpost	8856.9 ***	130700 ***	-0.0264 ***	-1.36 ***
	1764.2	*26290*	*0.0022*	*0.1603*
Daqx	6913 *	105600 **	-0.0013	1.3158 ***
	3946.7	*59230*	*0.0086*	*0.4969*
Dtrq	-1341.6	-18890	0.143 ***	5.0522 ***
	1431.7	*21330*	*0.0166*	*1.2966*
Dcboe	-1640.1	-23350	0.0576 ***	0.366 ***
	1496.3	*22290*	*0.0021*	*0.1335*
Daqx Dwidened	-9627.7 **	-137100 ***	0.0121	-3.9858
	4314.8	*63050*	*0.0267*	*2.4406*
Dtrq Dwidened	-1529.4	-14570	0.1514 ***	-2.0009
	2210.7	*29630*	*0.0274*	*1.8649*
Dcboe Dwidened	-1199.6	-8036.9	0.4197 ***	4.879 ***
	2260.5	*30430*	*0.0322*	*1.4459*
Daqx Dnarrowed	-109.41	-4648.2	-0.0013	-0.0859
	4237.6	*63530*	*0.0087*	*0.5258*
Dtrq Dnarrowed	7975.7 ***	117700 ***	-0.1046 ***	-6.1759 ***
	2103.5	*31340*	*0.0166*	*1.2989*
Dcboe Dnarrowed	9685.7 ***	142500 ***	-0.0305 ***	-1.6484 ***
	2177.6	*32440*	*0.0026*	*0.1504*
Daqx Dwidened Dpost	15980	222900	0.1501 ***	6.3391
	9986.9	*148600*	*0.0579*	*3.9877*
Dtrq Dwidened Dpost	4369.4 ***	49510 ***	-0.0224	-2.7281 **
	1861.1	*23100*	*0.0245*	*1.5187*
Dcboe Dwidened Dpost	4661.8 ***	47960 ***	-0.1207 ***	1.9111
	1890.1	*23450*	*0.0356*	*1.7032*
Daqx Dnarrowed Dpost	-2959.1	-42870	0.0254 ***	1.4636 ***
	6295.1	*94310*	*0.0066*	*0.2946*
Dtrq Dnarrowed Dpost	-7786.6 ***	-115700 ***	0.0115 ***	1.9567 ***
	1766.1	*26300*	*0.0025*	*0.0986*
Dcboe Dnarrowed Dpost	-8926.3 ***	-132400 ***	0.0013	1.8591 ***
	1820	*27100*	*0.0021*	*0.0858*
VIX	-214.45 ***	-3079.9 ***	-0.0006 **	-0.094 ***
	51.01	*759.55*	*0.0003*	*0.0243*
F-statistic robust	3390	1456.8	3300.2	28620
R-overall	0.6	0.2	0.6	0.5
No of Obs	460899	460899	460899	460899

TABLE B.2: Fixed Effect LSVD Parameters Estimation cont. Each column refers to a different regression based on the choice of the market quality metric used as dependent variable. Robust standard errors are in italic font. Superscripts ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Spread in tick	**simple price impact**	**1L depth**	**effective spread**	**5 min price impact**
constant	3.0979 ***	0.4389 ***	6603.7 ***	7.5559 ***	0.42 ***
	0.4985	*0.0232*	*111.5*	*0.46*	*0.0299*
dummy widened	2.7427 ***	0.0911 ***	-2252.7 ***	9.8831 ***	0.1356 ***
	0.142	*0.0191*	*214.62*	*0.3189*	*0.0274*
dummy narrowed	-0.7135 ***	0.0263	4319.3 ***	0.6776 ***	0.055
	0.116	*0.0191*	*221.35*	*0.1328*	*0.0338*
Dwidned Dpost	-2.6835 ***	0.0571 **	2015.6 ***	2.1399 ***	0.0466 *
	0.2404	*0.0208*	*262.21*	*0.4565*	*0.0283*
Dnarrowed Dpost	1.0049 ***	-0.0302	-5002.9 ***	-2.3754 ***	-0.0703 ***
	0.2264	*0.0201*	*224.6*	*0.2227*	*0.0334*
Daqx	0.5403 ***	-0.1748 ***	-3775 ***	20.431 ***	-0.1657 ***
	0.1242	*0.0554*	*147.51*	*2.3632*	*0.0561*
Dtrq	2.8222 ***	0.0111	-4510.6 ***	8.8753 ***	0.0003
	0.2093	*0.0633*	*100.84*	*1.8725*	*0.0768*
Dcboe	3.5379 **	-0.1457 ***	-3741.4 ***	0.956 ***	-0.1311 ***
	2.0862	*0.0095*	*104.8*	*0.1986*	*0.014*
Daqx Dwidened	-0.1037	-0.3418 ***	1494.4 ***	-21.781 ***	-0.3729 ***
	0.4394	*0.0666*	*335.52*	*3.3731*	*0.0765*
Dtrq Dwidened	5.2681 ***	0.9706 ***	1371.4 ***	5.2079 **	1.1408 ***
	0.7745	*0.2425*	*217.56*	*2.422*	*0.2639*
Dcboe Dwidened	27.782 **	0.6673 ***	1495.2 ***	26.07 ***	0.6955 ***
	14.394	*0.1959*	*224.54*	*2.2207*	*0.2131*
Daqx Dnarrowed	-0.8757 ***	-0.2642 ***	-3009.1 ***	-21.042 ***	-0.2921 ***
	0.1245	*0.0583*	*286.8*	*2.4139*	*0.0644*
Dtrq Dnarrowed	-2.2557 ***	-0.2353 ***	-3553.2 ***	-10.416 ***	-0.2418 ***
	0.2141	*0.0657*	*227.85*	*1.8756*	*0.0831*
Dcboe Dnarrowed	-1.879	-0.0596 ***	-3123.2 ***	-2.8668 ***	-0.0955 ***
	2.5375	*0.0206*	*235.07*	*0.23*	*0.0344*
Daqx Dwidened Dpost	2.9264 ***	2.9002 ***	-351.17	163.18 ***	2.8807 ***
	0.9875	*0.6542*	*1114.3*	*23.448*	*0.6488*
Dtrq Dwidened Dpost	-4.9686 ***	-0.9866 ***	-1622.6 ***	-6.4099 ***	-1.1176 ***
	0.7497	*0.2374*	*263.29*	*1.7129*	*0.2581*
Dcboe Dwidened Dpost	-27.838 **	-0.0932	-1578.3 ***	-9.3535 ***	-0.0304
	14.24	*0.2043*	*274.65*	*2.4646*	*0.2249*
Daqx Dnarrowed Dpost	0.969 ***	2.4038 ***	4224.7 ***	239.53 ***	2.4216 ***
	0.2168	*0.2125*	*414.65*	*17.414*	*0.212*
Dtrq Dnarrowed Dpost	0.8761 ***	-0.0116	4026.9 ***	1.8624 ***	0.0219
	0.0776	*0.0191*	*229.7*	*0.1724*	*0.0328*
Dcboe Dnarrowed Dpost	-1.2168	-0.0105	3497.7 ***	1.1207 ***	0.0239
	1.4453	*0.0196*	*236.86*	*0.1589*	*0.0325*
VIX	-0.0501	0.0031 **	-13.015 ***	0.0033	0.0044 ***
	0.037	*0.0016*	*3.9372*	*0.0339*	*0.002*
F-statistic robust	1275.4	5673	67321	28620	67453
R-overall	0.2	0.6	0.7	0.5	0.7
No of Obs	460899	460899	460899	460899	460899

TABLE B.3: Parameter Estimation of Pooled OLS for London Stock Exchange. Each column refers to a different regression based on the choice of the market quality metric used as dependent variable. Robust standard errors are in italic font. Superscripts ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Return (10-6)	Simple price impact	5 min price impact	First Level Depth	Effective Spread	Spread	Amihud (10-6)	Quoted Spread	Realised Spread (1 min)	Realised Spread (5 min)
Constant	0.207 ***	1.110 ***	1.173	7778.4 ***	12.18	0.015 ***	5.727 ***	16.534 ***	0.376 ***	0.311 ***
	0.005	0.009	0.01	102.96	0.083	0	0.233	0.119	0.008	0.009
No venues buyer	0.001	0.011 ***	0.011	-190.76 ***	-0.037	-0.000 ***	0.03329 *	-0.0188 *	-0.010 ***	-0.010 ***
	0	0.001	0.001	8.697	0.007	0	0.02	0.01	0.001	0.001
No Sellers per buyer	0.001 *	-0.005 ***	-0.005	40.393 ***	-0.051	-0.000 ***	-0.036 ***	-0.0771 ***	-0.003 ***	-0.002 ***
	0	0	0	2.492	0.002	0	0.006	0.003	0	0
Var(p)	-0.017 ***	0.059 ***	0.077	-3401.2 ***	1.047	0.007 ***	0.28	1.3744 ***	0.063 ***	0.044 ***
	0.006	0.012	0.013	137.38	0.111	0	0.311	0.159	0.01	0.011
VIX	-0.0245 ***	0.036 ***	0.031	-72.96 ***	0.239	0.0002 ***	-0.1648 ***	0.2967 ***	0.002 ***	0.008 ***
	0	0.001	0.001	9.032	0.007	0	0.02	0.011	0.001	0.001
Turnover per buyer	0.000 ***	-0.000 ***	0	-0.0034 ***	0	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***
	0	0	0	0	0	0	0	0	0	0
Total turnover ISIN	-0.000 ***	0.000 ***	0	0.000 ***	0	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***
	0	0	0	0	0	0	0	0	0	0
F-statistic robust	500.62	2390	21050	1456.8	2980	3300.2	500.62	28620	5120.2	3852.3
R-overall	0.01	0.2	0.2	0.02	0.3	0.04	0.006	0.3	0.06	0.05
No of Obs										460899
Cov Estimator										Unadjusted

TABLE B.4: Base Line : Pooled OLS for LSE cont. Each column refers to a different regression based on the choice of the market quality metric used as dependent variable. Robust standard errors are in italic font. Superscripts ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Return (10-6)	**Simple price impact**	**5 min price impact**	**First Level Depth**	**Effective Spread**
Constant	0.207 ***	1.110 ***	1.173	7778.4 ***	12.18
	0.005	*0.009*	*0.01*	*102.96*	*0.083*
No venues buyer	0.001	0.011 ***	0.011	-190.76 ***	-0.037
	0	*0.001*	*0.001*	*8.697*	*0.007*
No Sellers per buyer	0.001 *	-0.005 ***	-0.005	40.393 ***	-0.051
	0	*0*	*0*	*2.492*	*0.002*
Var(p)	-0.017 ***	*0.059 ****	0.077	-3401.2 ***	1.047
	0.006	*0.012*	*0.013*	*137.38*	*0.111*
VIX	-0.0245 ***	*0.036 ****	0.031	-72.96 ***	0.239
	0	*0.001*	*0.001*	*9.032*	*0.007*
Turn. per buyer	0.000 ***	*-0.000 ****	0	-0.0034 ***	0
	0	*0*	*0*	*0*	*0*
Tot. turn. ISIN	-0.000 ***	*0.000 ****	0	0.000 ***	0
	0	*0*	*0*	*0*	*0*
Cov Estimator					Unadjusted
R-overall	0.01	0.2	0.2	0.02	0.3
No of Obs	460899	460899	460899	460899	460899
F-statistic robust	500.62	2390	21050	1456.8	2980

TABLE B.5: Base Line : Pooled OLS for LSE cont. Each column refers to a different regression based on the choice of the market quality metric used as dependent variable. Robust standard errors are in italic font. Superscripts ***, ** and * denote significance at 1%, 5% and 10%, respectively.

	Spread	**Amihud (10-6)**	**Quoted Spread**	**Realised Spread (1 min)**	**Realised Spread (5 min)**
Constant	0.015 ***	5.727 ***	16.534 ***	0.376 ***	0.311 ***
	0	*0.233*	*0.119*	*0.008*	*0.009*
Venues per buyer	-0.000 ***	0.03329 *	-0.0188 *	-0.010 ***	-0.010 ***
	0	*0.02*	*0.01*	*0.001*	*0.001*
Sellers per buyer	-0.000 ***	-0.036 ***	-0.0771 ***	-0.003 ***	-0.002 ***
	0	*0.006*	*0.003*	*0*	*0*
Var(p)	0.007 ***	0.28	1.3744 ***	0.063 ***	0.044 ***
	0	*0.311*	*0.159*	*0.01*	*0.011*
VIX	0.0002 ***	-0.1648 ***	0.2967 ***	0.002 ***	0.008 ***
	0	*0.02*	*0.011*	*0.001*	*0.001*
Turn. buyer	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***
	0	*0*	*0*	*0*	*0*
Tot. Turn. ISIN	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***
	0	*0*	*0*	*0*	*0*
Cov Estimator					Unadjusted
R-overall	0.04	0.006	0.3	0.06	0.05
No of Obs	460899	460869	460899	460899	460899
F-statistic robust	3300.2	500.62	28620	5120.2	3852.3

Appendix C

An application of the network approach to detect potential threat to best execution

Broker retail market in the UK is grounded on the concept of Retail Service Providers. These are platforms that offer online access to equity instruments, but also intermediaries, including financial advisers who use the platform to access retail investment providers on behalf of their clients. We estimated that a turnover of £2.2bn for the month of January 2018 was linked to this retail market over a total £190bn turnover (1.2%) for all the transactions we classified as Off-Book/OTC [1] .

Why do we care? In a previous market study FCA "found that investment platforms who provide stockbroking services to retail investors could do more to ensure consistent compliance with their best execution obligations" [2]. The present study offers some insights into the effectiveness of the execution arrangements.

C.1 How do Retail Service Providers work?

According to an FCA market study published in July 2018, around 95% of retail orders are executed through a network called the Retail Service Provider (RSP) system in the UK. In the same study, FCA found that over 90% of listed security transactions are carried out via RSPs rather than through stock exchange order books. Different platforms are connected into the RSPs system. A client willing to place an order, typically goes to one of these platforms connected with the RSPs system and its request for quote (RFQs) is sent to market makers in the network who accepts or rejects the request. All market makers' quotes are aggregated and displayed on the platform where the consumer is connected. He has usually between 10 and 30 seconds to accept a quote and finalize the order [3].

Most of retailers execute all their trades with only one RSP. Network analysis highlights how difficult is for retail clients to access the market via multiple market makers, whilst it is much more frequent that they execute all their trades with the same Retail Service Provider(RSP). This might call a best execution rationale in question. The network is bipartite, which means that it can be divided into two disjoint sets of nodes, in our case RSPs and Retailers, such that every edge connects a vertex

[1] Basically, OTC and Off Book trades are classified based on the field of Waivers and the deduplication process (i.e. one counterpart reports on XLON whilst the other XOFF and there is no CCP between them.)

[2] Source: https://www.fca.org.uk/publication/market-studies/ms17-1-2.pdf

[3] Investment Platform Market Study. Interim Report. Market Study, MS17/1.2. July 2018. Paragraph 6.61. Source: https://www.fca.org.uk/publication/market-studies/ms17-1-2.pdf

in the set of RSPs to one in the set of Retailers. We can easily notice how the retailers can be classified into two categories: retailers who trade with multiple RSPs and traders with only one RSPs.

Analysing all the transactions in the sample, we found 1,044 out of 2,886 retailers traded with more than one RSPs while 1842 traded with only one (64%). When we focused on the largest trades per turnover (> GBP 1mn) we still found 16 out of 45 retailers (35%) were buying from only one RSP .

All buyers in Table C.1, trade with one or more (up to 5) RSPs available on the market, still their shares of executed turnover are quite concentrated ranging from 38.89% for xxx with yyy, up to 100% of zzz with www and xzy with hgh. The smaller are the turnovers per counterpart (left side in C.1) the larger the number of counterparts who trades only with one retailer.

The observed over-reliance of clients on one RSP makes it unluckily that best execution is achieved. Using transactions from MDP we tried to ascertain if RSPs were consistently obtaining the best possible results for their clients, as required by COBS 11.2A (Best Execution – MiFID provisions)[4]. We highlighted all the cases where, at the same time on the same instrument, different RSPs were trading with their clients at different prices. It serves as a proxy for best execution, measuring the rate at which each platform could offer the best price. Based on the assumption that the lowest price is the best price, Table 8 shows that RSP are very far from offering the best possible price at any time.

C.2 Best Execution Monitoring?

Orders are not always executed at best price. The over-reliance by retail client on 1 RSP makes it unlikely that best execution is achieved consistently. The role of best execution (measured as the percentage of trades executed at the lowest price) is below 32%.

C.3 Further investigations and future research

To make our results robust, it would be ideal to verify the order book (SMARTS) and best execution should be monitored at a tight price variation tolerance (currently 1 second) and a large sample of trades must be scrutinized. Clearly the way we classify and aggregate our data affects the final findings in which, using a different classification of firm type, each group can include a different number of firms.

[4]Source: https://www.handbook.fca.org.uk/handbook/COBS/11/2A.html

TABLE C.1: Concentration of Sellers' turnover (top and bottom 10)

Buyer	**Seller (RSP)**	**Turnover %**	**Turnover (mln GBP)**	**Buyer**	**Seller (RSP)**	**Turnover %**	**Turnover (mln GBP)**
Retail1	RSP1	61.28%	10,09	Retail11	RSP1	54.55%	1.09
Retail2	RSP1	85.21%	8,41	Retail12	RSP3	67.12%	1.08
Retail3	RSP1	71.31%	7,44	Retail13	RSP4	100.00%	1.07
Retail4	RSP2	69.27%	6,66	Retail14	RSP2	87.65%	1.02
Retail5	RSP1	38.89%	5,91	Retail15	RSP4	100.00%	1
Retail6	RSP3	100.00%	5,03	Retail16	RSP1	59.81%	1
Retail7	RSP1	72.14%	4,40	RSP3	RSP3	85.37%	0.48
Retail8	RSP2	100.00%	4,34	WSL	RSP1	100.00%	0.12
Retail9	RSP1	45.27%	3,58	RSP4	RSP4	100.00%	70944.88
Retail10	RSP1	43.98%	3,48	RSP2	RSP2	100.00%	60316.99

TABLE C.2: Rates of transactions at best price available

RSPs sell to retailers	Times at best price available %	RSPs buy from retailers	Times at best price available %
RSP1	52.79%	**RSP1**	53.85%
RSP3	48.57%	**RSP3**	100%
RSP4	82.60%	**RSP4**	47.42%
RSP2	69.86%	**RSP2**	62.50%
RSP5	31.81%	**RSLP5**	76.19%

Bibliography

Akerlof, George (1970). "The Market for "Lemons": Qualitative Uncertainty and Market Mechanisms". eng. In: *The Quarterly Journal of Economics* 84, p. 488. ISSN: 0033-5533. URL: http://search.proquest.com/docview/1296899575/.

Garman, M.B (1976). "Market Microstructure". eng. In: *Journal of Financial Economics* 5.5(3), pp. 277–308.

Stoll, Hans R. (1978). "THE PRICING OF SECURITY DEALER SERVICES: AN EMPIRICAL STUDY OF NASDAQ STOCKS". In: *Journal of Finance* 33.4, pp. 1153–1172. ISSN: 0022-1082.

Hamilton, J.L. (1979). "Market place fragmentation, competition, and the efficiency of the stock exchange". In: *Journal of Finance* 34, pp. 171–187. ISSN: 0022-1082.

Amihud, Yakov and Haim Mendelson (1980). "Dealership market: Market-making with inventory". eng. In: *Journal of Financial Economics* 8.1, pp. 31–53. ISSN: 0304-405X.

Ho, Thomas and Hans R Stoll (1981). "Optimal dealer pricing under transactions and return uncertainty". eng. In: *Journal of Financial Economics* 9.1, pp. 47–73. ISSN: 0304-405X.

Ho, T.S. and R.G. Macris (1984). "Dealer Bid-Ask Quotes and Transaction Prices: An Empirical Study of Some Amex Options". In: *Journal of Finance* 39, pp. 23–45. ISSN: 0022-1082.

Glosten, Lawrence R and Paul R Milgrom (1985). "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders". eng. In: *Journal of Financial Economics* 14.1, pp. 71–100. ISSN: 0304-405X.

Pagano, Marco (1989). "Trading Volume and Asset Liquidity". eng. In: *The Quarterly Journal of Economics* 104.2, pp. 255–274. ISSN: 00335533.

King, Mervyn and Sushil Wadhwani (1990). "Transmission of Volatility between Stock Markets". eng. In: *The Review of Financial Studies* 3.1, p. 5. ISSN: 0893-9454. URL: http://search.proquest.com/docview/1292181407/.

Chowdhry, Bhagwan and Vikram Nanda (1991). "Multimarket Trading and Market Liquidity". eng. In: *The Review of Financial Studies* 4.3, p. 483. ISSN: 0893-9454. URL: http://search.proquest.com/docview/1292181824/.

Peake, Junius W. (1993). "Brother can you spare a dime? Let's decimalize the U.S. equity markets!" (Global Equity Markets: Technological, Competitive and Regulatory Challenges).

Chan, Kalok, Y. Peter Chung, and Herb Johnson (1995). "The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options". In: *Journal of Financial and Quantitative Analysis* 30.3, pp. 329–346. ISSN: 0022-1090.

Hasbrouck, Joel (1995). "One Security, Many Markets: Determining the Contributions to Price Discovery". In: *Journal of Finance* 50.4, pp. 1175–1199. ISSN: 0022-1082.

Ho, T.K. (1995). "Random Decision Forests." In: (3rd International Conference on Document Analysis and Recognition), pp. 278–282.

Ricker, J. (1996). "Decimal Quotes: the price is right!" (1730 Filbert Street No 105). Working Paper.

Madhavan, Ananth, Matthew Richardson, and Mark Roomans (1997). "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks". In: *The Review of Financial Studies* 10.4, pp. 1035–1064. ISSN: 0893-9454.

Bollen, Nicolas and Robert Whaley (1998). "Are "teenies" better?" eng. In: *Journal of Portfolio Management* 25.1, pp. 10–24. ISSN: 00954918. URL: http://search.proquest.com/docview/195582816/.

Harris, Lawrence E. (1999). "Trading in Pennies?" In: *U.S. Equity market in transition*.

Huang, Roger D. and Ronald W. Masulis (1999). "FX Spreads and Dealer Competition Across the 24-Hour Trading Day". In: *The Review of Financial Studies* 12.1, pp. 61–93. ISSN: 0893-9454.

Allen, F. and D. Gale (2000). "Financial contagion." In: *Journal of Political Economy, 108(1)*, pp. 1–33.

Bessembinder, Hendrik (2000). "Tick Size, Spreads, and Liquidity: An Analysis of Nasdaq Securities Trading near Ten Dollars". eng. In: *Journal of Financial Intermediation* 9.3, pp. 213–239. ISSN: 1042-9573.

Goldstein, Michael A and Kenneth A. Kavajecz (2000). "Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE". eng. In: *Journal of Financial Economics* 56.1, pp. 125–149. ISSN: 0304-405X.

Hendershott, Terrence and Haim Mendelson (2000). "Crossing Networks and Dealer Markets: Competition and Performance". In: *Journal of Finance* 55.5, pp. 2071–2115. ISSN: 0022-1082.

Leung, Mark T, Hazem Daouk, and An-Sing Chen (2000). "Forecasting stock indices: a comparison of classification and level estimation models". eng. In: *International Journal of Forecasting* 16.2, pp. 173–190. ISSN: 0169-2070.

Jones, Charles M and Marc L Lipson (2001). "Sixteenths: direct evidence on institutional execution costs". eng. In: *Journal of Financial Economics* 59.2, pp. 253–278. ISSN: 0304-405X.

Kyle, Albert S. and Wei Xiong (2001). "Contagion as a Wealth Effect". In: *Journal of Finance* 56.4, pp. 1401–1440. ISSN: 0022-1082.

Deb. Harris, Frederick H, Thomas H Mcinish, and Robert A Wood (2002). "Security price adjustment across exchanges: an investigation of common factor components for Dow stocks". eng. In: *Journal of Financial Markets* 5.3, pp. 277–308. ISSN: 1386-4181.

Lehmann, Bruce N (2002). "Some desiderata for the measurement of price discovery across markets". eng. In: *Journal of Financial Markets* 5.3, pp. 259–276. ISSN: 1386-4181.

Lamoureux, Christopher G and Charles R Schnitzlein (2004). "Microstructure with multiple assets: an experimental investigation into direct and indirect dealer competition". eng. In: *Journal of Financial Markets* 7.2, pp. 117–143. ISSN: 1386-4181.

O., Loistl., B. Schossmann, and A. Veverka (2004). "Tick size and spreads: The case of Nasdaq's decimalization". In: *European Journal of Operational Research* 155 (2), pp. 317–334.

Rodriguez, Pedro N. and Arnulfo Rodriguez (2004). "Predicting Stock Market Indices Movements". In: *Computational Finance and its Applications*. Ed. by Marco Constantino and Carlos Brebia. Wessex Institute of Technology.

Chakravarty, Sugato, Venkatesh Panchapagesan, and Robert A Wood (2005). "Did decimalization hurt institutional investors?" eng. In: *Journal of Financial Markets* 8.4, pp. 400–420. ISSN: 1386-4181.

Cifuentes, R et al. (2005). "Liquidity risk and contagion". In: *Journal of the European Economic Association* 3(2-3), pp. 556–566.

Hendershott, Terrence and Charles M Jones (2005). "Trade-through prohibitions and market quality". eng. In: *Journal of Financial Markets* 8.1, pp. 1–23. ISSN: 1386-4181.

Iori, Giulia et al. (Aug. 2005). "A Network Analysis of the Italian Overnight Money Market". In: *Journal of Economic Dynamics and Control* 32, pp. 259–278. DOI: 10.1016/j.jedc.2007.01.032.

Newman, M.E. J (2005). "A measure of betweenness centrality based on random walks". eng. In: *Social Networks* 27.1, pp. 39–54. ISSN: 0378-8733.

Borgatti, Stephen P and Martin G Everett (2006). "A Graph-theoretic perspective on centrality". eng. In: *Social Networks* 28.4, pp. 466–484. ISSN: 0378-8733.

Saramaki, J et al. (2006). "Generalizations of the clustering coefficient to weighted complex networks". eng. In: *arXiv.org* 75.2 Pt 2. ISSN: 15393755. URL: http://search.proquest.com/docview/2085396651/.

Back, K. and S. Baruch (2007). "Working Orders in Limit-Order Markets and Floor Exchanges". In: *Journal of Finance* 62, pp. 1589–1621. ISSN: 0033-5533.

Baruch, S., A. Karolyi, and M. Lemmon (2007). "Multi-Market Trading and Liquidity: Theory and Evidence". In: *Journal of Finance* 62, pp. 2169–2200.

Nier, Erlend et al. (2007). "Network models and financial stability". eng. In: *Journal of Economic Dynamics and Control* 31.6, pp. 2033–2060. ISSN: 0165-1889.

Wyart, Matthieu et al. (2007). "Relation between Bid-Ask Spread, Impact and Volatility in Double Auction Markets". eng. In: *arXiv.org*. ISSN: 2331-8422. URL: http://search.proquest.com/docview/2090731025/.

Allen, F. and A. Babus (2008). "Networks in Finance". In: Working Paper No. 08-07. URL: https://ssrn.com/abstract=1094883.

Foucault, Thierry and Albert J. Menkveld (2008). "Competition for Order Flow and Smart Order Routing Systems". In: *Journal of Finance* 63.1, pp. 119–158. ISSN: 0022-1082.

Hsieh, Tzung-Yuan, Shaung-Shii Chuang, and Ching-Chung Lin (2008). "Impact of Tick-Size Reduction on the Market Liquidity — Evidence from the Emerging Order-Driven Market". eng. In: *Review of Pacific Basin Financial Markets and Policies* 11.4, pp. 591–616. ISSN: 0219-0915.

Kaiser, Marcus (2008). "Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks". eng. In: *New Journal of Physics* 10.8, p. 083042. ISSN: 1367-2630.

Huang, Cheng-Lung and Cheng-Yi Tsai (2009). "A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting". eng. In: *Expert Systems With Applications* 36.2, pp. 1529–1539. ISSN: 0957-4174.

Tressel, Thierry (2010). "Financial Contagion Through Bank Deleveraging: Stylized Facts and Simulations Applied to the Financial Crisis". eng. In: *IMF Working Papers* 10.236, p. 1. ISSN: 1018-5941.

Yan, Bingcheng and Eric Zivot (2010). "A structural analysis of price discovery measures". eng. In: *Journal of Financial Markets* 13.1, pp. 1–19. ISSN: 1386-4181.

Haldane, A. and R. May (2011). "Systemic risk in banking ecosystems". In: *Nature 469(7330)*.

Kara, Yakup, Melek Acar Boyacioglu, and Ömer Kaan Baykan (2011). "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange". eng. In: *Expert Systems With Applications* 38.5, pp. 5311–5319. ISSN: 0957-4174.

O'Hara, Maureen and Mao Ye (2011). "Is market fragmentation harming market quality?" eng. In: *Journal of Financial Economics* 100.3, pp. 459–474. ISSN: 0304-405X.

Pedregosa, Fabian et al. (2011). “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12.85, pp. 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

Battiston, Stefano et al. (2012). “Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk”. eng. In: 36.8, pp. 1121–1141. ISSN: 0165-1889.

Dayri, Khalil and Mathieu Rosenbaum (2012). “Large tick assets: implicit spread and optimal tick size”. In:

Eisler, Zoltán, Jean-Philippe Bouchaud, and Julien Kockelkoren (2012). “The price impact of order book events: market orders, limit orders and cancellations”. eng. In: *Quantitative Finance* 12.9, pp. 1395–1419. ISSN: 1469-7688. URL: http://www.tandfonline.com/doi/abs/10.1080/14697688.2010.528444.

Putniņš, Tālis J (2013). “What do price discovery metrics really measure?” eng. In: *Journal of Empirical Finance* 23.C, pp. 68–83. ISSN: 0927-5398.

Caccioli, M. et al. (2014). “Stability analysis of financial contagion due to overlapping portfolios”. In: *Journal of Banking and Finance* 46, 233–245.

Cohen-Cole, Ethan, Andrei Kirilenko, and Eleonora Patacchini (2014). “Trading networks and liquidity provision”. eng. In: *Journal of Financial Economics* 113.2, pp. 235–251. ISSN: 0304-405X.

Varian, Hal R. (2014). “Big Data: New Tricks for Econometrics †”. eng. In: *Journal of Economic Perspectives* 28.2, pp. 3–28. ISSN: 0895-3309.

Ballings, Michel et al. (2015). “Evaluating multiple classifiers for stock price direction prediction”. eng. In: *Expert Systems With Applications* 42.20, pp. 7046–7056. ISSN: 0957-4174.

Budish, Eric, Peter Cramton, and John Shim (2015). “The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response”. eng. In: *The Quarterly Journal of Economics* 130.4, p. 1547. ISSN: 00335533. URL: http://search.proquest.com/docview/1729392433/.

Caccioli, Fabio et al. (2015). “Overlapping portfolios, contagion, and financial stability”. eng. In: *Journal of Economic Dynamics and Control* 51, p. 50. ISSN: 0165-1889.

Gray, Andrew and Michael Leibrock (2015). “Understanding interconnectedness risks. To build A more resilient financial system”. DTCC White Paper.

Gresse, C. (2015). “Effects of competition between multiple trading platforms on market liquidity. Evidence from the MiFID experience”. Working Paper.

Kara, G and M Tian (2015). *Taxonomy of Studies on Interconnectedness*. Tech. rep. FED.

Patel, Jigar et al. (2015). “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques”. eng. In: *Expert Systems With Applications* 42.1, pp. 259–268. ISSN: 0957-4174.

Rohe, K. (2015). *The difference between the transitivity ratio and the clustering coefficient*. Tech. rep. University of Wisconsin-Madison. URL: http://pages.stat.wisc.edu/.

Athey, Susan and Guido Imbens (2016). “The State of Applied Econometrics - Causality and Policy Evaluation”. In:

Battiston, Stefano et al. (2016). “Leveraging the network: a stress-test framework based on DebtRank”. eng. In:

Chang, C. and S. Zhang (2016). “Endogenous Market Making and Network Formation”. In: *Systemic Risk Centre Discussion Paper* 50. URL: https://ssrn.com/abstract=2600242.

Iori, G. and J. Porter (2016). “Agent based modelling for financial markets”. In: *The Oxford Handbook on Computational Economics and Finance*. Ed. by S.-H. Chen and M. eds. Kaboudan. Oxford University Press.

Kumar, Deepak, Suraj S Meghwani, and Manoj Thakur (2016). "Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets". eng. In: *Journal of Computational Science* 17, pp. 1–13. ISSN: 1877-7503.

Radicchi, Filippo and Claudio Castellano (May 2016). "Leveraging percolation theory to single out influential spreaders in networks". In: *Physical Review E* 93. DOI: 10.1103/PhysRevE.93.062314.

Serri, Matteo, Guido Caldarelli, and Giulio Cimini (2016). "How the interbank market becomes systemically dangerous: an agent-based network model of financial distress propagation". eng. In: *arXiv.org* 3.1. ISSN: 20557795. URL: http://search.proquest.com/docview/2075092856/.

Tasca, P. and S. Battiston (2016). "Market procyclicality and systemic risk". In: *Quantitative Finance* 16(8), 1219–1235.

Bardoscia, Marco et al. (2017). "Pathways towards instability in financial networks". In: *Nature Commnications (8)14416*.

Benoit, S. et al. (2017). "Where the risks lie: a survey on systemic risk". In: *Review of Finance* 21(1), 109–152.

Chen, Yingjun and Yongtao Hao (2017). "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction". eng. In: *Expert Systems With Applications* 80, pp. 340–355. ISSN: 0957-4174.

Hastie, T. et al. (2017). "The Element of Statistical Learning". In: Springer. Chap. 15. Random Forest, p. 593.

Bank, European Central (2018). Tech. rep. ECB.

Levent Uslu, Çağrı and Burak Evren (2018). "Liquidity flows, drawdowns and trading networks in order driven markets: An application to Borsa Istanbul". eng. In: *Borsa Istanbul Review* 18.3, pp. 176–190. ISSN: 2214-8450. URL: https://doaj.org/article/d26cf40f4a5c4d35b0a7a534b95dd7cb.

Risk and Trends (2018). *MiFID II: Impact of the new tick size regime*. Tech. rep. Authorithé des marchés financiers.

Seth, Neha and Laxmidhar Panda (2018). "Financial contagion: review of empirical literature". eng. In: *Qualitative Research in Financial Markets* 10.1, pp. 15–70. ISSN: 1755-4179.

Henrique, Bruno Miranda, Vinicius Amorim Sobreiro, and Herbert Kimura (2019). "Literature review: Machine learning techniques applied to financial market prediction". eng. In: *Expert Systems With Applications* 124, pp. 226–251. ISSN: 0957-4174.

Kreif, Noemi and Karla Diazordaz (2019). "Machine learning in policy evaluation: new tools for causal inference". eng. In: *arXiv.org*. ISSN: 2331-8422. URL: http://search.proquest.com/docview/2187712146/.

Rindi, B. and I.M. Werner (2019). "US Tick Size Pilot, Third Draft". Working Paper.

Werner, I.M. et al. (2019). "Tick Size, Trading Strategies and Market Quality". Working Paper.

Nuzzo, M. Francesca (2020). "Tick size regime impact on UK equity market participants." (FCA Occasional Paper).