Republic of South Africa
Computer Science & Applied Mathematics
Department

University of the Witwatersrand
Johannesburg

Bank Seta & NRF

NRF UGN-111755
Bank Seta ETQA/SDLN°: CHE/L94022731

APPLICATIONS OF MACHINE LEARNING TO ESTIMATING THE MULTIVARIATE ADAPTIVE PRICE IMPACT FUNCTIONS OF THE JOHANNESBURG STOCK EXCHANGE TRADED FINANCIAL INSTRUMENTS

presented at

Wits University
(Johannesburg)

in order to obtain the

Degree of Master of Science

by

Witness Maake

Dedication

My grandfathers Piet Patholo and Frans Kududu ,
my grandmothers Casa Boledi and Mmetsi Twaalf,
my father Dimo, my mother Hunadi, Gods and God.

To all of you,
I dedicate this work.

Witness Maake

Thanks

The research in case would not have been possible without the invaluable contributions of the following participants: Professor Terence Van Zyl, Wycliffe Ouma, Warren Carlson and Marike Kluyts. Thanks to my supervisor Prof. Terence Van Zyl for making the project a possibility. Moreover for the conceptual and programming inputs that have been applied to drive the project towards success. Thanks to Wycliffe Ouma for providing access to the data portal. Thanks to Warren Carlson for suggestions on the technical approach. The author appreciates the work done by Marike Kluyts on the incredible editorial and review of the research. Thanks to Dr. Lumkile Mondi, Bank Seta and the National Research Foundation for providing financial aid.

Declaration

The work in the dissertation was based on the research conducted by the Big Data Science Team, in the School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg. No part of the dissertation has been submitted elsewhere for any other degree or qualification. It is my unaided work, with the exception of the input of my supervisor Prof. Terence Van Zyl, unless referenced to the contrary in the text.

Signature : W.R Maake

Date: October 24, 2020

Publication

Part of the work included herein has been published. The details of the publication are shared subsequently.

Article : Applications of Machine Learning to Estimating the Sizes and Market Impact of Hidden Orders in the BRICS Financial Markets.

Publisher : Journal of Advanced Studies in Finance.

Issue : Vol 11 No 1 (2020): JASF Volume XI Issue 1(21) Summer 2020 .

DOI : https://doi.org/10.14505//jasf.v11.1(21).04 .

Authors : Witness Maake and Terence Van Zyl .

Abstract

The research aims to investigate the accuracy of the functional form of market impact as disclosed in the literature. The concept of market impact is necessary for achieving cost-effective portfolio acquisitions or liquidations. The research replicates the literature. The research extracted numerous market features from the trading data with the purpose of establishing areas of exploitation to which machine learning techniques could be applied. The research implemented machine learning to uncover hidden orders. The investigation concludes with a replication of hidden orders inclusive of research. The research discovers that the market impact function in literature is conditionally true. Generalized Linear Models and Support Vector Machines produced low $\widehat{\text{MSE}}$ and high R^2 each. The $\widehat{\text{MSE}}$ and R^2 fits of the literature model have improved subsequent to the inclusion of hidden orders. The market impact of both visible and invisible trades closely resembles the model of the literature. Machine learning techniques, in the context of the research, imply that hidden orders execute within similar market conditions.

Contents

List of Figures

List of Tables

Glossary

Fee restructure	Adjustments that pertain to fixed costs of trading that are market maker decided.
Insider trader	A market participant that actively trades a security by leveraging on the confidential information of the security.
Market maker	An institution that facilitates the buying and selling of securities.
Noise trader	A zero intelligent market participant that trades a security without information.
Normalized volume	Volume traded scaled by average volume traded.
Power-law distribution	A structure that resembles a power function.
Price adjustment rule	The process by which security prices adjust to new information announcement.
Price manipulation	A way of lobbying financial securities prices with the sole intention of making profit.
Quasi-arbitrage	Potential positive return realizations on investment that exceed the amount of risk required to match such returns.

Chapter

1 Introduction

Financial markets are exchange platforms that are intended to facilitate the exchange of financial securities at minimal transaction costs. Evidence, quoted within writings such as Brooks et al.(1968), suggests that financial markets were first discovered in the early 12^{th} century, and were formalized as institutions since the early 16^{th} century. The evolution of financial markets since then has seen the emergence of other cost structures that are solely attributable to excess demand. One such cost structure is commonly referred to as "market impact".

The research articulates to the unfamiliar researcher as to how market impact would ideally emerge. A buy initiating agent that wishes to execute a large order may not be able to execute the order on one occasion, because there may not be a counterpart to the transaction of that size. As such the buy initiating agent ideally subdivides the single large order execution into multiple child order executions. Each receding child order is likely to be completed at an excessive price relative to the completed preceding child order as a consequence of the impact on the price that the preceding child order had. To complete the execution of the large order, the buy initiating agent would have to make provisions to account for the price slippage throughout time. When such provisions are integrated they are referred to as the "implementation shortfall" as per Bertsimas et al.(1998). It is the average price slippage following the execution of an order of a specified size that is referred to as the "market impact" of the order.

Numerous publications, including the Obizhaeva et al.(2005), have suggested optimal trading strategies to decide on the sizes of child orders that minimize price slippage. It was Lillo et al.(2003) that discovered an approach to reveal the average price shift associated with specified normalized order size.

Meanwhile, technological change that has revolutionized the operations of financial markets produced publications that seek to resolve practical problems. These include cost-effective trading algorithms, risk measures, derivatives pricing and many others. In addition to the existing publications, the current research shows that the research of Lillo et al.(2003) could be reproduced via machine learning.

The implication of the results of this research would be that analysts would not need to revisit the methodology of Lillo et al.(2003), whenever they wish to conduct studies on market impact. Consequently, machine learning could be employed to inform analysts with up to date empirical information regarding the average market impact of orders. In addition, this research shows that even special types of orders, namely: hidden orders, could have their characteristics, such as their sizes, uncovered via machine learning. Consequently, machine learning can approximate the average market impact of a complete historical trading session.

The research extracted market features from the order book. The features are used to train machine learning models to estimate the sizes of hidden orders. Machine learning in the context of the study is supervised learning in which both inputs and targets are known.

The research aims to answer the research questions:

- What role do hidden orders play on the average market impact of orders?,
- What role does market capitalization of firms play on the average market impact?

- Can machine learning appropriately reveal hidden orders and the average market impact thereof?

The research formulates a set of objectives to answer the research questions, which are:

- The research groups the firms into quantiles to investigate the precise role of market capitalization.
- The research fits the market impact function of Lillo et al.(2003) on the data of quantile averaged price changes and averaged normalized volumes before and after the inclusion of hidden orders.
- The research trains machine learning techniques namely : Generalized Linear Models of Seber et al.(1997), Artificial Neural Networks of Nguyen et al.(1990), Support Vector Machines of Vapnik et al.(1995), and Random Forests of Breiman et al.(2001), to uncover hidden orders and the average market impact functions.

A hidden order is defined as an order that is not visible to the market. The order has quantity that is greater than or equal to minimum reserve size, has a visible size equal to zero in the trading data and must have a minimum execution size. Readers are referred to the Johannesburg Stock Exchange trading manual for further specifications regarding the use of terms which may not be clear in the context of the definition of a hidden order. Other readings which readers may revisit include Moro et al.(2009)

The dissertation is structured in the following format: Chapter 2 provides a thorough review on the work of the relevant scholars on the research topic, Chapter 3 presents data sourcing and pre-processing, Chapter 4 outlines a logical approach through which the investigation was conducted, Chapter 5 produces the results of the investigation, and rigorous discussions of the results. The conclusion in Chapter 6 states the implication of the results. The Appendix discloses the supporting results.

Chapter 2

Literature review

This chapter provides a rigorous; however, brief discussion of the contribution of familiar authors on the subject matter. Readers are advised to refer to the glossary for clarity on concepts used herein.

Kyle et al.(1985) investigated the concept of market impact in the context of its structural form and characteristics. They established that the price adjustment rule that is implemented by the market maker, subsequent to the clearing of orders from an insider trader and a noise trader at a specified time window, is linear in the aggregate signed order size. The implication is that the price adjustment between the inception and maturity of complete order execution is permanent, assuming that the market impact persists at least for the whole duration of incremental order execution. Consequently, Kyle et al.(1985) postulated that the structure of market impact is linear in the traded volume and permanent in time.

O' Hara et al.(1995) introduced a theoretical framework on micro-structure concepts including the notion of market impact. The theory aided fellow researchers such as Li et al.(2016) to investigate the impact of transaction costs on hedging portfolios. Jones et al.(1996) adopted a different approach to studying the concept of market impact, by studying the relationship between volatility and the order size. They discovered that order size significantly influences market impact; however, when the volume is applied as a proxy to the rate of daily information arrivals, then it is the number of executed trades that indicate a strong influence on the market impact rather than the sizes of the orders.

An independent investigation of Barra et al.(1997) postulated that the structure of market impact follows a square root law, in which the average price shift is proportional to the square root of the average normalized transaction sizes. Gabaix et al.(2006) also independently supported the findings of the Barra et al.(1997) model subsequent to the investigation of price increments that are effected by trades that belong to large institutions in illiquid markets. Gabaix et al.(2006) concluded that price changes are power-law distributed.

Plerou et al.(2002) proposed the use of two demand metrics to study the notion of market impact. The research discovered that market impact has a structure of power $\frac{1}{3}$ for 5 minutes aggregated normalized transaction sizes. Meanwhile, for at least 15 minutes of aggregated transactions, the market impact has in between $[\frac{3}{2}, 1]$ power structure. Separate research conducted by Hasbrouck et al.(2002) discovered a sub-linear power-law function between the market impact and volume traded.

Huberman et al.(2004) studied the concept of quasi-arbitrage and the role played by market impact in the context of price manipulation. Huberman et al.(2004) discovered that quasi-arbitrage can only be diminished provided permanent market impact is linear in time.

Almgren et al.(2000) argued that the structure of market impact has two components namely: permanent market impact and temporary market impact. Evans et al.(2002) discovered that the market impact of aggregated transactions is an increasing power function of increasing time window with a power of 0.5 for daily logged returns of individual stocks, futures, and currencies.

Bouchaud et al.(2009) argued that the Kyle et al.(1985) model would imply that an insider trader is able to submit uncorrelated orders such that prices are perfectly random; however, real market data reveals correlations of orders to an extent. Hence the impact cannot be permanent and linear at all times. Bouchaud et al.(2017) defined market impact as the correlation between an incoming order to buy or sell, and a subsequent price change.

Lillo et al.(2003) revealed the volume dependence of the market impact of single trades, and reported a power-law dependence. The research indicated that financial assets that differ with class, and possibly many other factors, are governed by similar statistical laws. Harvey et al.(2016) adopted the methodology of Lillo et al.(2003) with significant modifications of parameter choices. The research proved that the average market impact subsequent to fee restructure could be derived directly from the average market impact prior to fee restructure, taking into account the effect of fee restructure.

Lillo et al.(2003) produces answers to the research questions and objectives. Harvey et al.(2016) is being used as a benchmark to judge some of the results of the research. Consequently the research adopts the methodology of Lillo et al.(2003) and Harvey et al.(2016) to calculate the market impact of financial trades. The author does not assume a particular stance with regard to the various reports which were previously presented. Furthermore, information available to the author at the time of research may not be sufficient to critique the various structures which were reported in the literature. The adoption of the Harvey et al.(2016) is informed by the accessibility to its pseudo codes, and also to strategically avoid the literature arguments that could prohibit the research from maturing into the machine learning aspect.

Chapter 3

Data processing

When liquidity providers and takers seek to exchange shares, by auctioning the shares in search of quotations for buying and selling prices per share, a series of records which are collectively called ”order book” are realized. The research requires passive information of the order book, which we refer to as ”tick history”. Tick history has information regarding trades and quotes of a particular stock. In addition, tick history has the stock Reuters Instrument Code, dates and times of trades and quotes, bid prices, bid sizes, ask prices, ask sizes, trade prices, and trade sizes as shown in Figure 2 and Figure 3.

The research sourced the tick history from Thomson Reuters Eikon. We articulate to the unfamiliar researcher as to how the tick history was acquired and processed. Through a personal account[1], we received access to the Thomson Reuters Eikon. The sourcing of data was achieved in 7 steps which are outlined in Figures 1 - 3. The preliminary step is to log into the Thomson Reuters Eikon portal. The subsequent step is to enter the Reuters Instrument Code (RIC), of a specified firm e.g JSEJ.J, in the search engine marked as position number 1, followed by ”Enter”. The next step is to click on ”Price & Charts” marked as position number 2, to find ”Time and Quotes”. The subsequent step is to specify the time period from the start to the end at position numbers that

[1], We thank Wycliffe Ouma for allowing us access to the Thomson Reuters Eikon portal via his personal credentials. Wycliffe is a Ph.D. candidate at Wits Business School, 2^{nd} St Davids PI, & St Andrew Rd, Park-town, Johannesburg, 2193

are marked as 4 and 5. The final step is to click at position number 7 to download the quotes data.

Similarly, the trade data for the same period could be accessed by changing "Time and Quotes" to "Time and Sales" at position number 3. Followed by a click on "Update View", and a click at position number 7 to download.

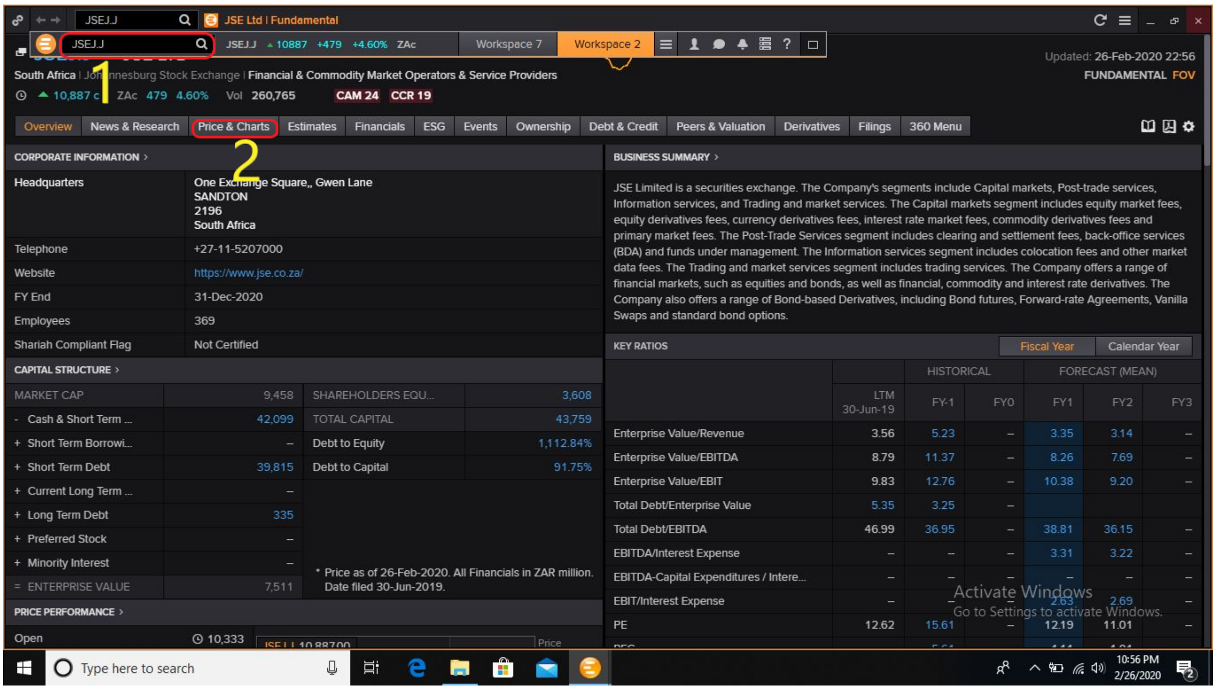


Figure 1. The website page and the search engine of the Thomson Reuters Eikon portal.

Firms that belong to the same markets were grouped into identical folders. Collectively, the various folders were merged into one folder to form a database. To create a flat file, of a stock, the data sets of quotes and trades of the stock were imported into MatLab. A short coded program was applied to concatenate the two data sets. Events were rearranged according to their historical time of occurrence inline with the data structures that are disclosed in the literature such as Nonyane et al.(2018). The flat files[2] of firms per market were used in the implementation of the method of section 4.3.

[2]The database and flat files creation approach was recommended by Professor Timothy Gebbie of the University of Cape Town, Rondebosch, Cape Town, 7700.

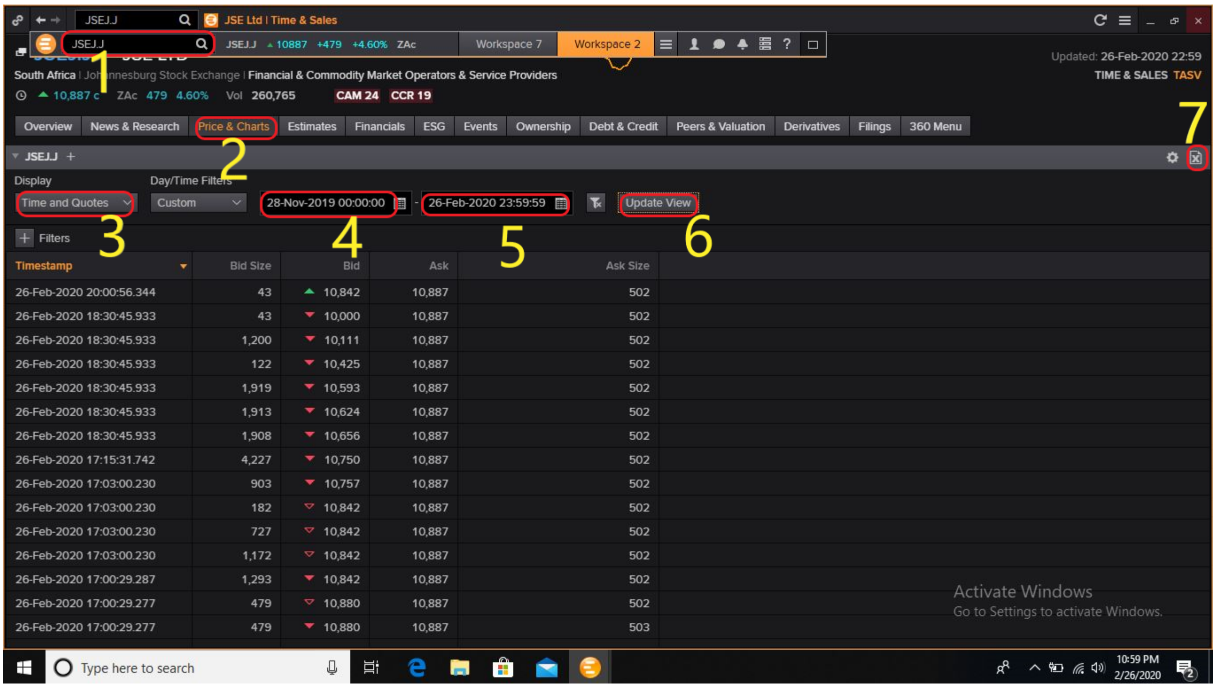


Figure 2. The overview of the data of historic quotes in Thomson Reuters Eikon portal.

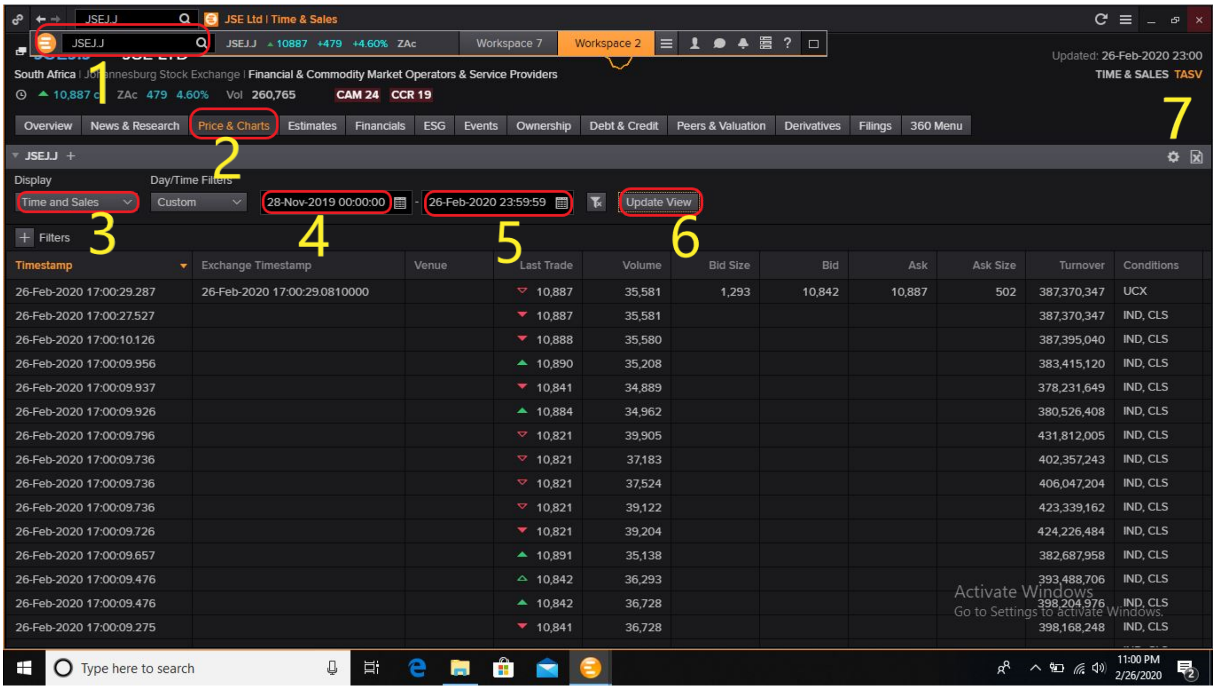


Figure 3. The overview of the data of historic trades in Thomson Reuters Eikon portal.

Chapter 4

Methodology

This chapter discloses the data, research methods and models that were implemented. The chapter also provides to the reader the source of all the information which is disclosed herein. The research studied two types of markets namely: (1) Emerging markets, and (2) Developed markets. Frontier markets have been omitted due to insufficient data. The research discloses the specific markets that were studied in section 4.1 and section 4.2. The markets of section 4.1 and section 4.2 were chosen based on the criteria that: (1) the number of missing observations is minimal, and/ or (2) there exist related literature that has performed similar analysis, that allows factual comparisons of current research results.

The author would like to acknowledge a series of developments which have lead to the creation of the research. The notion of the research project was first introduced in a lecture session in 2014 by Professor Diane Wilcox, Dr Timothy Gebbie and Mr Dieter Hendricks. The base code was developed by (Mutswari, 2014). The functionality of the program code was upgraded using functions which were provided by (Harvey, 2014). Such work has been documented in (Maake, 2014). (Maake, 2014) replicated the algorithms of (Mutswari, 2014) and (Harvey, 2014), and consequently identified shortcomings of code such as algorithms which were supposed to step in real time, which were not in the revised base program code. The works of (Du Preez, 2016) and (Nonyane,

2018) have also been recognised and therefore played critical role in the development of the research.

4.1 Emerging markets

South Africa (The Johannesburg Stock Exchange, JSE), Brazil (The Bolsa de Valores de Sao Paulo, BOVESPA), Russia (The Moscow Exchange, MOEX), India (The National Stock Exchange of India, NSE), and China (The Shanghai Stock Exchange, SSE).

4.2 Developed markets

United States of America (The New York Stock Exchange, NYSE), United Kingdom of Britain (The London Stock Exchange, LSE), United Arab Emirates (The Dubai Financial Markets, DFM), France (The Paris Stock Exchange, Euronext Paris), Singapore (The Singapore Exchange, SGX), Spain (The Bolsa de Madrid, BME), Australia (The Australian Securities Exchange, ASX), Austria (The Vienna Stock Exchange, WBAG), Portugal (The Lisbon Stock Exchange, Euronext Lisbon), Belgium (The Brussels Stock Exchange, Euronext Brussels), Norway (The Oslo Stock Exchange, OSE), Finland (The Helsinki Stock Exchange, OMXH), and Germany (The Frankfurt Stock Exchange, FWB).

4.3 Literature methodology

The research presents the methodology of Lillo et al.(2003) to the unfamiliar researcher in order to calculate market impact curve. Each row is regarded as a data point regardless of the columns, because of the time-series nature of the data as depicted in Figure 3. Data points with errors such as zero prices, zero volumes, missing values and not a number figures were cleared from the data sets. The first 10 minutes and the last 10 minutes of data were removed to reduce the effect of noise as per Harvey et al.(2016). Trades with the same date and same time were compacted using the volume-weighted average price algorithm of Berkowitz et al.(1988). Volume traded per trade was normalized by dividing by the mean volume traded.

Trades were classified as buyer initiated or seller initiated using the method of Lee et al.(1991). The data of buyer initiated trades was separated from the data of seller initiated trades; however, the same quotes were kept in both data sets. The research assumed the mid quote price of each quote to be the midpoint between the ask price and the bid price of the quote. The research considered either the data set with seller initiated trades or buyer initiated trades one at a time. For each trade, the price shift was calculated as: $\Delta P_{i+1} = P_{i+1} - P_i$, if $P_i = \ln m_i$ where $m_i = \frac{a_i + b_i}{2}$, a_i and b_i being the time i ask price and bid price respectively. For consecutive trades, the price change was set to zero.

The research arbitrarily constructed twenty bins in a logarithmic space where the left edge and the right edge were negative 3.2 and positive 1.0 respectively. Trades were added to bins based on the trade's normalized volume. The price shifts were binned according to the volume bin classification in a separate, but corresponding twenty bins. The averages of the volume bins and the price shift bins were calculated. A logarithmic plot of the averages of the volume bins, on the x-axis, and averages of the price shifts, on the y-axis was produced. The resulting curve is the average market impact curve of an ideal firm in consideration. The example of a logarithmic plot is shown in chapter 5, Figure 5.

An increased number of firms in a specified market made it impossible to plot clear graphical results of all firms on the same set of axes. The research grouped firms into quantiles using the average daily value traded (ADVT) of Harvey et al.(2016) as a proxy to the firms market capitalizations. Four quantiles were used. Following the approach of Harvey et al.(2016) the axes were re-scaled in the manner: $x \longrightarrow x/ADVT^{\delta}$ and $y \longrightarrow y \times ADVT^{\gamma}$, to collapse the individual average market impact curves onto a single scale. Following the approach of Lillo et al.(2003) the author investigated the values of δ and γ that would bring the aggregated curves close to one another without contrary directions of curves evolution. The parameters δ and γ are determined by minimizing the function:

$$\epsilon = \left(\frac{\delta_x}{\mu_x}\right)^2 + \left(\frac{\delta_y}{\mu_y}\right)^2 \tag{4.1}$$

where δ_x and δ_y are the standard deviations of the re-normalized volumes and price changes respectively. The μ_x and μ_y are the means of the normalized volumes and price changes respectively. Readers are referred to [33, 46, 49] for further explanation of the 2 dimensions variance minimization problem that is described in equation 4.1.

Our definition of hidden orders is in accordance with Moro et al.(2009). Visible orders have empty strings; meanwhile, hidden orders have none empty strings under the column "Conditions" as shown in Figure 3. Hence, via a string detection technique, we are able to separate hidden orders from visible orders if needs be.

4.4 Research methodology

Following the methodology of section 4.3, the researchers analyzed the data of visible orders only. In a separate investigation, the researchers analyzed the data of visible and hidden orders. The results and discussion of the analysis are provided in chapter 5. The research fitted the linearized market impact model of Lillo et al.(2003) over the data of aggregated quantile average price changes and aggregated quantile average normalized transactions sizes. Scientific measures such as the R^2 and the $\hat{\text{MSE}}$ were applied to evaluate the performance of the average market impact model of Lillo et al.(2003). The research discuss the relevance of the Lillo et al.(2003) model on a specified market, and whether or not the inclusion of hidden orders implied the improvement of the relevance of the Lillo et al.(2003) model.

The core of the investigation is the use of machine learning to retrieve the sizes of orders and the implied aggregated averaged market impact. We apply appropriate methods from the literature to extract features from the order book data. The features discussed in subsection 4.4.1 include normalized transactions sizes, price, bid price, bid size, ask price, ask size, turnover ratio, price change, spread prior to trade, spread subsequent to trade, mid quote price prior to trade, mid quote price subsequent to trade, moving average value traded, volatility, momentum, order sign, signed volume, order imbalance, liquidity, execution time, probability of informed trading, Kullback - Leibler divergence counts and Kullback - Leibler divergence volume.

4.4.1 Features

- Price, P_i. The matching of a limit order and a market order of opposite signs as a consequence of the interplay between the order book, and order flow. Potters et al. (2003).
- Bid price, b_i. The price in cents at which the market maker buys a specified- security at a specified time instant. Potters et al. (2003)
- Bid size, v_i^b. The number of shares in lots that are quoted at the bid price. Potters et al. (2003)
- Ask price, a_i. The price in cents at which the market maker sells a specified security at a specified time instant. Potters et al. (2003)
- Ask size, v_i^a. The number of shares in lots that are quoted at the asking price. Potters et al. (2003)
- Turnover ratio, TR_i. Measures a firm's trading frequency. Amihud et al. (1986)
- Price change, $\Delta P_{i+1} = \ln\left(\frac{m_{i+1}}{m_i}\right)$. The natural logarithm of the ratio of the mid quote price subsequent to the trade, to the mid quoted price before the trade. Lillo et al. (2003)
- Spread, $s_i = a_i - b_i$. The difference between the ask and the bid. Glosten et al. (1988)

- Mid quote price, $m_i = \frac{a_i+b_i}{2}$. The mid quote price is the midpoint of the ask and the bid. Potters et al. (2003)

- Average value traded, $AVT_i = (V_i - V_{i-1}) \times \left(\frac{2}{n+1}\right) + V_{i-1}$, where V_i is time i values traded and n is the total number of values traded at time i. The moving mean of the individual share values exchanged over a specified duration. Harvey et al. (2016)

- Volatility, $\sigma_i = std\,(r_i)$, where $r_i = \ln(P_i) - \ln(P_{i-1})$. The standard deviation of natural logged returns. Micciche et al. (2002)

- Momentum, $M_i = \frac{P_i}{P_{i-1}}$. The ratio of the prevailing price to the lag one prevailing price. Jegadeesh et al. (1993)

- KL divergence, kl_i. The degree of change in the empirical distribution of counts or sizes of orders, which are ordered according to price. Kullback et al. (1951)

- Order sign, $sign(\omega_i)$. Classification of order as either buyer initiated or seller initiated. Lee et al. (1991)

- Signed volume, $sign(\omega_i) \times \omega_i$. The signed order is the product of the order size and the order sign. Bouchaud et al. (2017)

- Order imbalance, δv_i. Order imbalance is the sum of signed orders over a specified duration. Lillo et al. (2003)

- Liquidity, λ_i. The slope of the linearized version of the price impact model. Lillo et al. (2003)

- Execution time, $\delta t_i = t_i - t_{i-1}$. Execution time is the time latency between order match instant and the appearance of the subsequent event. Moro et al. (2009)

- PIN, pin_i. The probability of informed trading. Easely et al. (1996)

- Normalized volume, ω_i. The normalized volume is measured as the size of each order scaled by the mean volume traded. Lillo et al. (2003)

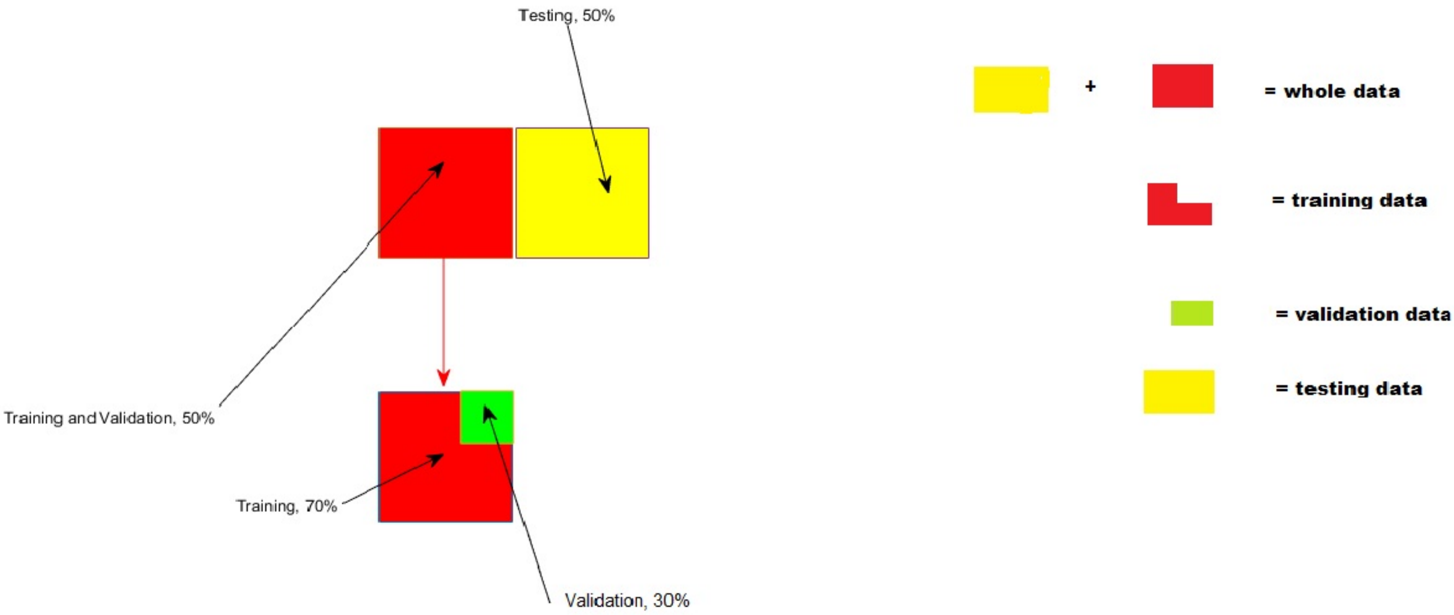


Figure 4. The diagram illustrates the utilization of data among training, validation and testing.

Many of the features which are described in section 4.4.1 were obtained directly from the data. These features include: traded prices, bid prices, bid sizes, ask prices, ask sizes and turnover ratio. The other miscellaneous features such: price change, spread, mid quote price, average value traded, volatility, momentum, order sign, signed volume, order imbalance, and execution time were calculated using the equations that are provided in section 4.4.1. Advanced features such as the probability of informed trading and Kullback - Leibler divergence require detailed explanations. The research refers readers to the literature of Easely et al. (1996) for calculating the probability of informed trading. The research shares with readers on how Kullback - Leibler divergence was applied to create a feature. An empirical distribution was designed in such a way that the traded price range of securities were aligned with the bin range of the distribution. For every order that arrives, it belongs to a specific bin. The number of volumes in a particular bin divided by the total number of volumes as at a particular time instant yields the probability outcome of such volume. Alternatively the research could use arbitrary bins instead of price bins. Using the method of Kullback et al. (1951) the research determined the number that represents the divergence of the probability distribution subsequent to the arrival of an order from the probability distribution prior to the arrival of the order. The time series of such numbers create a new feature which surprisingly seem to be of importance for determining the hidden order sizes in some markets.

We trained machine learning techniques to uncover the sizes of hidden orders. In the context of the current investigation, the research tests the reliability of machine learning techniques on the estimation of sizes of orders by hiding many more orders including visible orders. The research provides details to the unfamiliar researcher on the implementation of machine learning techniques that were used in subsection 4.4.3 to 4.4.6.

4.4.2 Data Splitting

The data is separated into in-sample data (50%) which is used for training and validation, and out-of-sample data (50%) which is used for testing as indicated in Figure 4. To allow for comparison, the research used only 70% of the data for training. Of the 30% data remaining, the research deployed only 15% for validation. The validation could have been performed with the 30% data; however, the same results could be achieved with about 15% of the data. The research selected 15% of data to improve time efficiency of the investigation. Normalized volume is retrieved as the response variable and the other market features serve as input variables. The results of the performance of the models are presented in Table 2 and Tables 27+ - 43+.

The research is not concerned with finding the best tools for the prediction problem, rather the research is concerned with discovering as to whether or not machine learning techniques which are applied herein could be of importance to tackling problems of this nature. Thus the use of the techniques herein was based purely on abstract selection. Hence, the results that are presented herein constitute the trial version of this type of architecture. Furthermore, the author did not program the machine learning techniques used herein. The reasons being that given the sizes of data which amounted to millions of transactions per market, it is very likely that unjustified loss of data and/ or features during the training of the algorithms would occur. To the knowledge of the author, there has not been a relevant literature that discusses the use of the architecture specifically for this particular problem. Therefore, the authors have resorted to using built in programs, which the authors assumed already have good defaults.

The research is not concerned with showcasing in-depth machine learning techniques, rather the research is concerned with the use of such techniques. Readers who may wish to review the in-depth of the techniques that are used herein are advised to consult the disclosed references. Given that the initial attempt is to uncover the applicability of the machine learning techniques, the authors used a validation set to manually set hyper-parameters. Defaults where sufficient for most algorithms. For specific explanation of these parameters and what they are, readers are referred to the disclosed references.

4.4.3 Generalized Linear Models (GLM).

The basic assumption regarding the application of linear regression models[3] was that each observed input market feature contributes linearly to the output response variable. Thus the response variable was assumed to be a linear function of the weights of the respective input market features; however, not necessarily a linear function of the market features. Mathematically a generalized linear regression model is a model of the form:

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_p x_{ip} + \epsilon_i, i = 1, \cdots, n, \tag{4.2}$$

$$y_i = \alpha_0 + \sum_{m=1}^{M} \alpha_m f_m(x_{i1}, x_{i2}, \cdots, x_{ip}) + \epsilon_i, i = 1, \cdots, n, \tag{4.3}$$

where

- y_i is the i^{th} response
- α_m is the m^{th} coefficient, and α_0 is the bias term of the model

[3]Copyright 2013-2014 The MathWorks, Inc.

• $f(\cdot)$ is a scalar valued function which resembles a variety of forms such as nonlinear functions and polynomials

• x_{ij} is the j^{th} explanatory variable realization, $j = 1, \cdots, p$. Collectively the x_{ij}s form a design matrix when concatenated horizontally

• ϵ_i is the noise factor.

Additionally assumptions were that each data point was independent of its fellow data points, all data points were members of the same family and maximizing the likelihood yielded best results for the optimization routine. The assumptions of linear regressions models in a mathematical construct are:

- The noise factors, ϵ_i's, are uncorrelated
- The noise factors, ϵ_i's, are independent and identically normally distributed with zero mean and constant variance, σ^2, such that

$$\mathbb{E}\left(y_i\right) = \mathbb{E}\left(\alpha_0 + \sum_{m=1}^{M} \alpha_m f_m(x_{i1}, x_{i2}, \cdots, x_{ip}) + \epsilon_i\right) \tag{4.4}$$

$$\mathbb{E}\left(y_i\right) = \alpha_0 + \sum_{m=1}^{M} \alpha_m f_m(x_{i1}, x_{i2}, \cdots, x_{ip}) + \mathbb{E}\left(\epsilon_i\right) \tag{4.5}$$

$$\mathbb{E}\left(y_i\right) = \alpha_0 + \sum_{m=1}^{M} \alpha_m f_m(x_{i1}, x_{i2}, \cdots, x_{ip}) \tag{4.6}$$

$$\mathbb{V}\left(y_i\right) = \mathbb{V}\left(\alpha_0 + \sum_{m=1}^{M} \alpha_m f_m(x_{i1}, x_{i2}, \cdots, x_{ip}) + \epsilon_i\right) \tag{4.7}$$

$$\mathbb{V}\left(y_i\right) = \mathbb{V}\left(\epsilon_i\right) = \sigma^2. \tag{4.8}$$

where the symbols $\mathbb{E}$ and $\mathbb{V}$ represent the expectation and variance operators respectively.

If we substitute the features of subsection 4.4.1 into Equation 4.2, then the resultant research model is :

$$\begin{gathered}\hat{\omega_{i-1}} = \alpha_0 + \alpha_1 P_i + \alpha_2 b_i + \alpha_3 v_i^b + \alpha_4 a_i + \alpha_5 v_i^a + \alpha_6 TR_i + \alpha_7 \Delta P_i + \\ \alpha_8 s_i^{before} + \alpha_9 s_i^{after} + \alpha_{10} m_i^{before} + \alpha_{11} m_i^{after} + \alpha_{12} AVT_i + \\ \alpha_{13}\sigma_i + \alpha_{14} M_i + \alpha_{15} kl_i^{counts} + \alpha_{16} kl_i^{volume} + \alpha_{17} sign(\omega_i) + \\ \alpha_{18} sign(\omega_i) \times \omega_i + \alpha_{19}\delta v_i + \alpha_{20}\lambda_i + \alpha_{21}\delta t_i + \alpha_{22} pin_i.\end{gathered}$$

Tables 3 - 4 present the results of the training of the generalized linear model on the Johannesburg Stock Exchange Market. Figure 13 provides the results of the performance of the generalized linear model.

4.4.4 Artificial Neural Networks (ANN).

Following Nguyen et al.(1990), the research fitted a two-layer feed-forward network[4] with a sigmoid transfer function in the hidden layer, and a linear transfer function in the output layer. The validation set was deployed to impose the early stopping of the training when generalization stopped improving. The testing set provided an independent measure of network performance. The study set the initial number of neurons in the hidden layer to 10, which could be increased or decreased depending on the network performance per stock. The study trained the network via the Levenberg - Marquardt back-propagation algorithm which is partly introduced in Marquardt et al.(1963).

For the neural network the authors supposed that the hidden layers, optimization algorithm and the number of neurons were unknown. The authors have resorted to the finding of literature which suggested a 2 layers, 10 neurons and a first degree linear basis function in the event a researcher may not be certain regarding the nature of the problem at hand (Nguyen, 1990). The author verified that using 2 layers, 10 neurons and a first degree linear transfer function yielded acceptable results. The author verified out of curiosity that increasing or decreasing neurons in the hidden layer lead to good performance on the training and poor performance on the testing. Changing the linear basis function to a sigmoid transfer function or increasing the layers of the network has yielded the out of expectation observations. For instance there were gains on the exploitation of the nonlinear relationship among order sizes; however, the gains created disproportionate losses of the linear relationship among order sizes. This was trained on training and validated on validation set.

The neural network package allows for the monitoring of the progress of the training by returning the following variables: Epoch, Time, Performance, Gradient, Mu, and Validation Checks. Table 5 presents the results of the training of the neural network on the Johannesburg Stock Exchange Market. Figure 13 provides the results of the performance of the neural network.

4.4.5 Support Vector Machines (SVM).

Conceptually, the investigation applied the idea of Vapnik et al.(1995). Of the j market features namely : trade price, bid price, bid size, ask price, ask size, turnover ratio, price change, spread, mid quote price, average value traded, volatility, liquidity, order sign, signed order, momentum, order imbalance, divergence, execution time, and pin, which are denoted by , x_{ij}, that are quantified at time i, the research defined the response feature normalized volume, which we denoted by, ω_i. The investigation determined the function $f_i(x_{i1}, x_{i2}, ..., x_{ip})$ such that the distance between ω_i and the function f_i is less

[4] Copyright 1992-2015 The MathWorks, Inc.

than some well defined tolerance measure ϵ_i, and at the same time $f_i(x_{i1}, x_{i2}, ..., x_{ip})$ is as flat as possible. In closed form the research minimized:

$$\mathbf{J}(\alpha) = \frac{1}{2}\alpha'\alpha \tag{4.9}$$

subject to:

$$|\omega_i - f_i(x_{i1}, x_{i2}, ..., x_{ip})| \leq \epsilon_i \tag{4.10}$$

where

$$f_i(x_{i1}, x_{i2}, ..., x_{ip}) = \alpha_0 + \sum_{m=1}^{M} \alpha_m f_m(x_{i1}, x_{i2}, \ldots, x_{ip}) = \alpha\mathbf{x} + \alpha_0. \tag{4.11}$$

We introduced the slack variable η_i and η_i^* for each point to allow regression error to bypass the ϵ_i threshold into $\epsilon_i + \eta_i$ threshold soft margin. With the additional conditions, the new objective is to minimize:

$$\mathbf{J}(\alpha) = \frac{1}{2}\alpha'\alpha + \mathbf{A}\sum_{i=1}^{N}(\eta_i + \eta_i^*) \tag{4.12}$$

subject to

$$\forall i : \omega_i - f_i(x_{i1}, x_{i2}, ..., x_{ip}) \leq \epsilon_i + \eta_i \tag{4.13}$$

$$\forall i : f_i(x_{i1}, x_{i2}, ..., x_{ip}) - \omega_i \leq \epsilon_i + \eta_i^* \tag{4.14}$$

$$\forall i : \eta_i^* \geq 0 \tag{4.15}$$

$$\forall i : \eta_i \geq 0 \tag{4.16}$$

where $\mathbf{A} \in \mathbb{R}^+$ is the regularization parameter which determines the degree of trade-off between the amount by which deviations over the tolerance ϵ_i are accepted and the flatness of the function $f_i(x_{i1}, x_{i2}, ..., x_{ip})$. Considering the fact that the function $f_i(x_{i1}, x_{i2}, ..., x_{ip})$ is non linear due to its adaption to the non linear series of normalized volume, the investigation drawn on the idea of Vapnik et al.(1995) to introduce Lagrange dual formula. Meanwhile, the study replaced the linear dot product $x'_{i1}x_{i2}$ with the kernel function $\mathbf{G}(x_{i1}, x_{i2}) = \langle\psi(x_{i1}), \psi(x_{i2})\rangle$ where ψ is a mapping that transforms to high dimensional space. The research determined the Gram matrix $\mathbf{G}$ directly and the optimal solution in the predictor space. The idea which is supported by Huang et al.(2005). In closed form the study minimized:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*)\mathbf{G}(x_i, x_j) + \epsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{N}\omega_i(\alpha_i - \alpha_i^*)$$

subject to

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \tag{4.17}$$

$$\forall i : 0 \leq \alpha_i < \mathbf{A} \tag{4.18}$$

$$\forall i : 0 \leq \alpha_i^* < \mathbf{A} \tag{4.19}$$

with the subsequent Karush Kuhn Tucker conditions:

$$\forall i : 0 \leq \alpha_i(\epsilon + \eta_i - \omega_i + f_i(x_{i1}, x_{i2}, ..., x_{ip})) = 0 \tag{4.20}$$

$$\forall i : 0 \leq \alpha_i^*(\epsilon + \eta_i^* + \omega_i - f_i(x_{i1}, x_{i2}, ..., x_{ip})) = 0 \tag{4.21}$$

$$\forall i : \eta_i(\mathbf{A} - \alpha_{\mathbf{i}}) = \mathbf{0}$$

$$\forall i : \eta_i^*(\mathbf{A} - \alpha_{\mathbf{i}}^*) = \mathbf{0}.$$

All observations inside the epsilon radius posses the multipliers of zero, and the observations outside posses non-zero multipliers. They are regarded as support vectors[5] which are solely the dependent variables of the function which is applied to predict new values which are provided by :

$$f_i(x_{i1}, x_{i2}, ..., x_{ip}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{G}(x_i x_{ip}). \tag{4.22}$$

Of the twenty-two features that the research applied as inputs, the study refined the training approach following Fan et al.(2005) to determine the minimal number of features that would obtain optimal training outcomes. Table 6 presents the results of the training of the support vector machines on the Johannesburg Stock Exchange Market. Figure 14 provides the results of the performance of the support vector machines.

4.4.6 Random Forests (RF).

In line with the literature of Breiman et al.(2001) the research grew a regression bagged ensemble of 200 trees. The study sampled all variables at each node. To counteract the loss of accuracy due to missing information the research specified surrogate splits. The split predictors were determined via the interaction test. The R^2 of the model was determined via out-of-bag predictions. The research determined the significance of each market feature by permuting out-of-bag observations; otherwise, the study could have determined the significance of each market feature by summing the gains in the mean squared error due to splits on each market feature.

Bagging, which stands for "bootstrap aggregation," is a type of ensemble learning. To bag a weak learner such as a decision tree on a data set, generate many bootstrap replicas of this data set and grow decision trees on these replicas. Obtain each bootstrap replica by randomly selecting N observations out of N with replacement, where N is the data set size. To find the predicted response of a trained ensemble, take an average over predictions from individual trees.

Bagging works by training learners on re-sampled versions of the data. This re-sampling is usually done by bootstrapping observations, that is, selecting N out of N observations with replacement for every new learner. In addition, every tree in the ensemble can randomly select predictors for decision splits—a technique known to improve the accuracy of bagged trees.

By default, the minimal leaf sizes for bagged trees are set to 5 for regression. Trees grown with the default leaf size are usually very deep. These settings are close to optimal for the predictive power of an ensemble. Often you can grow trees with larger leaves without losing predictive power. Doing so reduces training and prediction time, as well as memory usage for the trained ensemble.

The research selected bagged regression ensembles as a consequence of the availability of attractive features such as the proximity matrix. Every time two observations land on the same leaf of a tree, their proximity increases by 1. For normalization, sum these proximities over all trees in the ensemble and divide by the number of trees. The resulting matrix is symmetric with diagonal elements equal to 1 and off-diagonal elements ranging from 0 to 1. The matrix could be used for finding outlier observations and discovering clusters in the data through multidimensional scaling. Table 7 presents the results of the training of the random forests on the Johannesburg Stock Exchange Market. Figure 14 provides the results of the performance of the random forests.

Chapter 5

Results and Discussion

5.1 Emerging Markets

Emerging markets are the economies that are invested in productive economic operations which improve society. The results of individual market impact curves of the emerging markets are provided in Figures 5, 16, 17, 18 & 19. The results suggest that market impact is a power law function of normalized transaction sizes in some markets. Meanwhile, in other markets market impact is a convex function of normalized transaction sizes.

The research analyzed emerging markets to investigate answers to the research questions. The results of the analysis of the emerging markets are provided in subsection 5.1.1 to subsection 5.1.5.

5.1.1 South Africa (Johannesburg Stock Exchange, JSE)

The JSE is an order-driven free market which operates on the central order book trading system with opening, intra-day and closing auctions. The study performed an investigation of the research questions on the JSE. The results of the investigations are provided in subsubsection 5.1.1.1 to subsubsection 5.1.1.2.

5.1.1.1 Literature Results (Johannesburg Stock Exchange, JSE)

Figure 6 provides the results of the analysis of seller initiated and buyer initiated aggregated average market impact curves of the Johannesburg Stock Exchange, both before and after the inclusion of hidden orders. The research discovers that large-capitalization stocks tend to have small price changes; meanwhile, small-capitalization firms have large price shifts. Consider Figure 5, and observe that small-capitalization firms tend to cluster at the top; meanwhile, large capitalization firms tend to cluster at the bottom. Figure 6 shows that market capitalization determines the vertical locations of the aggregated average market impact curves. Similar observations are reported in the Lillo et al.(2003).

Figure 7 provides the re-scaled aggregated average market impact curves. The axes have been re-scaled by the aggregated average daily value traded to the exponent 0.1 on the vertical axes, and exponent -0.1 on the horizontal axes. In Figure 7 the research collapsed,

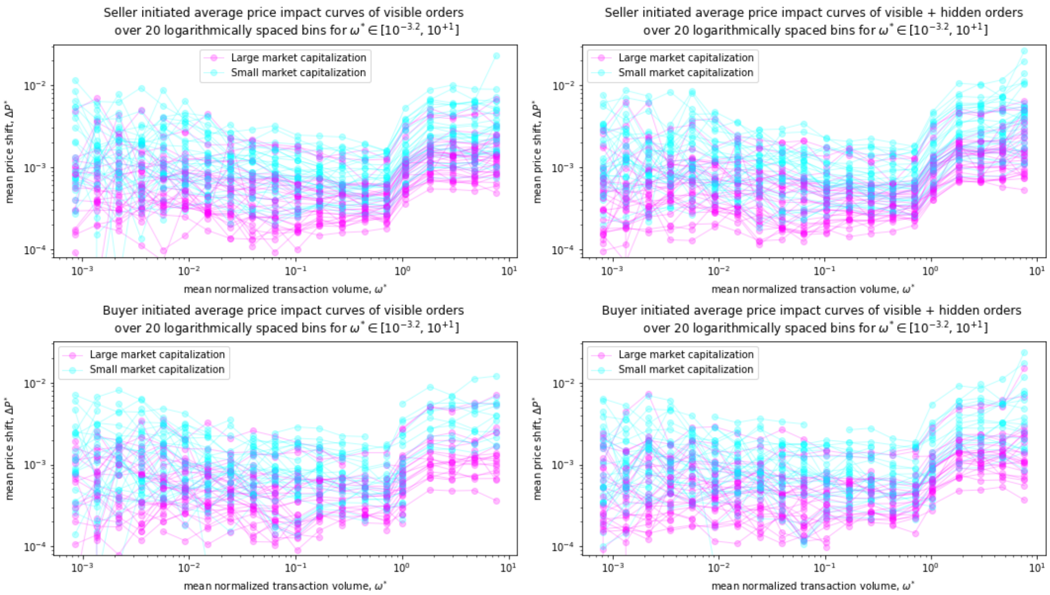


Figure 5. JSE. Price impact curves of single trades of 63 relatively liquid firms. The curves are computed using the analytical literature approach. Magenta curves represent firms with the average daily value traded larger than the median average daily value traded. Cyan curves represent firms with the average daily value traded smaller than the median average daily value traded.The analysis period is from 28-09-2018 to 28-06-2019.

the aggregated averaged market impact curves onto a single scale. The approach assists in bringing the individual curves closer to one another such that the general shape of the aggregated average market impact curves can be deduced.

Table 1 provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes, before and after the inclusion of hidden orders.

In Table 1 the R^2 of visible seller initiated orders of 0.5811, which is greater than 50%, indicates that there is a sufficient degree of variation of the data of re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes, which is explained by the Lillo et al.(2003) average market impact function. The $\hat{\text{MSE}}$ of visible seller initiated orders is low at 8×10^{-04}, which indicates that the distances between the individual nodes of the data of re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes are located close enough to the line that describes the Lillo et al.(2003) average market impact function. Hence, the Lillo et al.(2003) average market impact model accurately represents the observed nodes.

As expected, the R^2 of the fit of the Lillo et al.(2003) model over visible seller initiated orders after the inclusions of hidden orders has improved by 37.21% from 0.5811 to 0.7973. The $\hat{\text{MSE}}$ of the fit of the Lillo et al.(2003) model over seller initiated orders has improved by 99.91% from 8×10^{-04} to 7×10^{-07} after the inclusion of hidden orders. The alpha

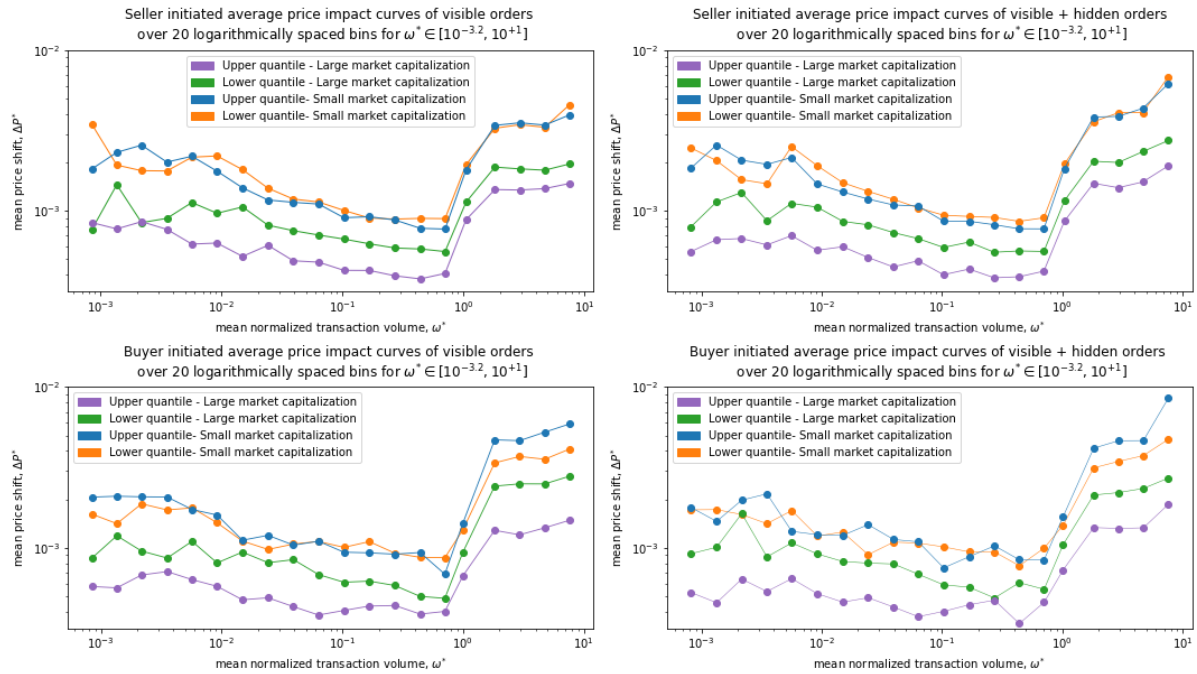


Figure 6. JSE. Aggregated price impact curves of single trades of 64 liquid firms. The data composition consists of a total of 10, 977,280 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

term of seller initiated orders has improved by 65.67% from 8×10^{-04} over visible orders to 0.0013 when hidden orders are included. Harvey et al.(2016) reported similar numerical values.

Similarly, the R^2 of the fit of the Lillo et al.(2003) model over buyer initiated orders after the inclusions of hidden orders has improved by 11.55% from 0.7361 of visible buyer initiated orders to 0.8211 of visible and hidden buyer initiated orders. The $\hat{\text{MSE}}$ of the fit of the Lillo et al.(2003) model over buyer initiated orders has improved by 46.85% from 8×10^{-07}, before the inclusion of hidden orders, to 6×10^{-07}, after the inclusion of hidden orders. The alpha term of buyer initiated orders has improved by 18.18% from 0.0011, before the inclusion of hidden orders, to 0.0013, after the inclusion of hidden orders. Other notable observations are the high tStat ranging in between $[5.0, 9.7]$, low SE ranging in between $[1 \times 10^{-04}, 2 \times 10^{-04}]$, and low pValue ranging in between $[4 \times 10^{-08}, 1 \times 10^{-04}]$. The residuals, "SE", are small. The tStats of the intercept terms, "Int.", and the gradient term, "$\ln(\omega)$", are at least greater than 5.0. The pValues of the intercept terms, "Int.", and the gradient term, "$\ln(\omega)$", are at most less than 0.05. The three statistics agree with one another. The implication is that the liquidity and the sizes of the normalized transactions are important for understanding the average market impact of the Johannesburg Stock Exchange in terms of the Lillo et al.(2003) average market impact model. The additional parameters including: standard error (SE), tStatistics (tStat) and pValue are there to provide readers with some additional information of the significance of particular terms of the fit of the Lillo et al.(2003) model. Furthermore, the statements that the alpha terms have improved indicate much closer to power law behavior of the average market impact curves than before. Same arguments apply to all similar tables.

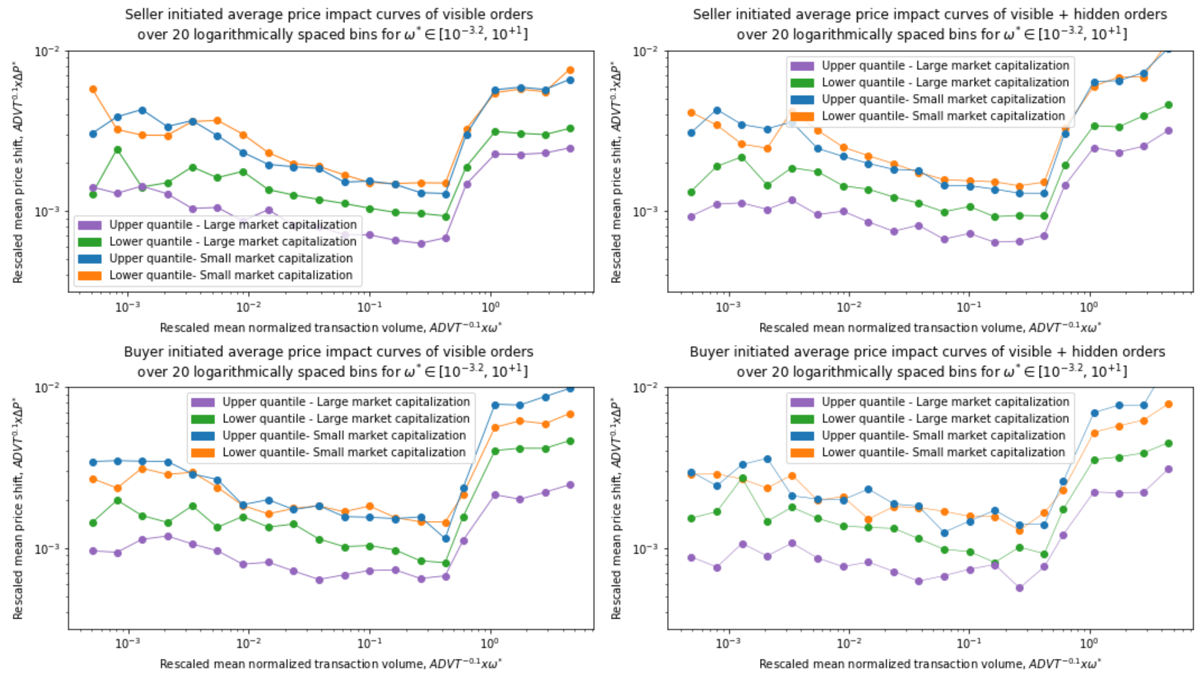


Figure 7. JSE. Re-scaled aggregated price impact curves of single trades of 64 liquid firms. The data composition consists of a total of 10, 977,280 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

The observations suggest that: (1) Hidden orders improve the relevance of the Lillo et al.(2003) average market impact function, (2) Hidden orders could be the important reason for the power-law dependence of the average market impact on the normalized volume, and (3) the market impact function of Lillo et al.(2003) is true depending on the condition that the trading history of orders is complete, at least with respect to facts at hand. In the event that the trading history of orders is dominated by hidden orders, the researcher may not necessarily observe the power-law distribution. Based on the observations that transpired over the results of the analysis, the Lillo et al.(2003) average market impact function is proven to be relevant to the Johannesburg Stock Exchange.

Figure 8 provides the results of buyer initiated and seller initiated trading of visible orders (left) and visible & hidden orders (right). The research held the standard deviation constant to allow the analysis to answer the research question of the impact of hidden orders on the average market impact curves. However, the advanced procedure would be to allow the standard deviation to vary to enable possible comments on the degree of confidence on the interpretation of the results. The research responds to how hidden orders affect the impact curves in a casual sense rather than to delve deeper into assessing the degree to which the interpretation of the results is true. Figure 8 shows that hidden orders effect a strict convex power-law far-right tail in contrast to visible orders which seem to effect a relatively concave power-law far-right tail. Moreover, the average market impact curves are volatile after the introduction of hidden orders especially towards the far-left of the impact curves. The observation suggests that hidden orders posses a material effect on the structure of the market impact of orders; which, translates into hidden orders, and

Table 1. JSE. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0019	2×10^{-04}	9.7	4×10^{-08}	0.0017	2×10^{-04}	9.5	8×10^{-08}
$\ln \omega$	8×10^{-04}	2×10^{-04}	5.0	1×10^{-04}	0.0013	2×10^{-04}	8.4	2×10^{-07}
R^2	0.5811				R^2	0.7973		
Adj.R^2	0.5579				Adj.R^2	0.7860		
MSE	8×10^{-04}				MSE	7×10^{-07}		
RMSE	8×10^{-07}				RMSE	8×10^{-04}		
Parameter:	'bias'	0.0019	λ	0.9981	'bias'	0.0017	λ	0.9983
	'slope'	8×10^{-04}	α	8×10^{-04}	'slope'	0.0013	α	0.0013
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0017	2×10^{-04}	8.9	7×10^{-07}	0.0016	2×10^{-04}	9.6	2×10^{-07}
$\ln \omega$	0.0011	2×10^{-04}	7.1	1×10^{-06}	0.0013	1×10^{-04}	9.6	7×10^{-07}
R^2	0.7361				R^2	0.8211		
Adj.R^2	0.7215				Adj.R^2	0.8112		
MSE	8×10^{-07}				MSE	6×10^{-07}		
RMSE	8×10^{-04}				RMSE	7×10^{-04}		
Parameter:	'bias'	0.0017	λ	0.9983	'bias'	0.0017	λ	0.9984
	'slope'	0.0011	α	0.0011	'slope'	0.0013	α	0.0013

have a significant influence over prices of stocks. Based on the observation of Figure 5 and Figure 8 the research comments that indeed hidden orders affect the structure of the average market impact curves.

5.1.1.2 Literature & Machine Learning Results (Johannesburg Stock Exchange)

Table 2 provides the machine learning results of all the stocks. Generalized Linear Models and Support Vector Machines produce high R^2 of 1 and 0.9871 respectively, and low $\hat{\text{MSE}}$ of 200.00 and 919.5449 respectively. The expectation was that Artificial Neural Networks would outperform the other machine learning techniques. The correlation metric, R^2,

Table 2. JSE. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.8348	0.8262	0.8527	0.7911	0.7530	0.8066	0.8649	0.8890	115.236
ω_{GLM}	0.0092	0.0093	0.0087	0.0113	0.0076	0.0059	0.0080	0.0104	2.6261
ω_{ANN}	1.1010	0.9940	0.9743	0.9787	0.9790	0.9785	0.9897	0.9764	1.1583
ω_{SVM}	0.6314	0.6568	0.6598	0.5946	0.5690	0.5544	0.6572	0.6472	171.533
ω_{RF}	1.1380	1.0027	0.9736	0.9697	0.9537	0.9462	1.0118	0.9681	8.5891
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$			R-Squared$_{SVM}$		R-Squared$_{RF}$		
1		0.1606			0.9871		0.7795		
Mean Squared Error									
MSE$_{GLM}$	MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$				
200.00	$2\times10^{+03}$		919.5449		$2\times10^{+03}$				

of the Artificial Neural Network was expected to be similar to the R^2 of the GLM. However, there are two main reasons why the results are not aligning to the expectation, the first being that the reported R^2 is not the R^2 of the training, rather it is calculated by training the machine learning techniques to make predictions. The estimated numbers are correlated with the exact numbers to measure the correlation measure, R^2. Consequently, the training must have been good; however, the test predictions are poor. Secondly, the package used to train the Artificial Neural Network randomly selects training, validation and testing sets. Therefore, in the event of not using seed selection, the subset used to train the Artificial Neural Network may be partly different from subsets that are used to train the other machine learning techniques. This is not to be understood as the data drawn from an alternative distribution. Bootstrapping was considered too computationally expensive and opened up a number of not in scope research questions around the most appropriate method to use.

Figure 9 presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which only hidden orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 10 presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which only hidden orders are estimated via Support Vector Machines and Random Forests. Figure 11 presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which all orders

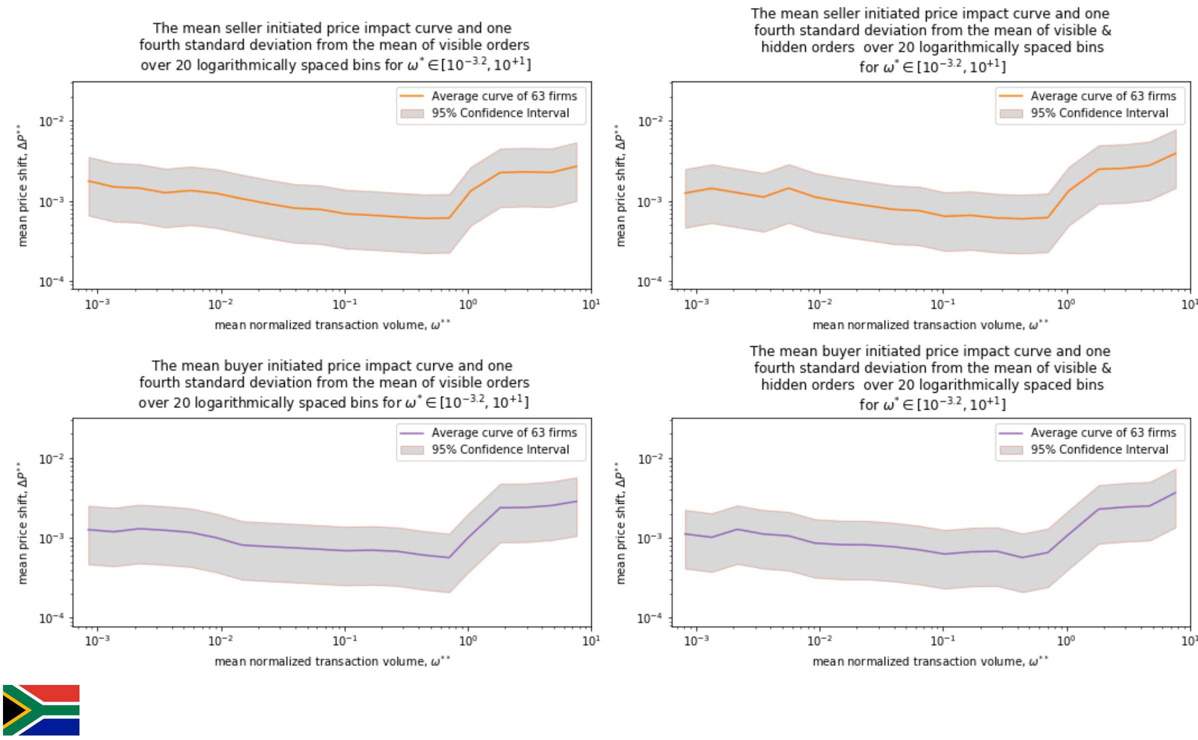


Figure 8. JSE. The mean price impact curves and the 95% confidence interval of a constant standard deviation from the mean of single trades of 63 liquid firms. The curves to the left are derived from the data of visible orders; meanwhile, the curves to the right are derived from the data of visible & hidden orders.

are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 12 presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which all orders are estimated via Support Vector Machines and Random Forests. It is unclear what really causes bad fits in Figures 11 - 12. However, hidden orders have additional characters in the data. It could be that the learning models have placed more emphasis on the parameters which measure the contribution of specific features that seem to be of importance to uncover the sizes of hidden orders; however, that can also cause poor performance over visible orders. The author remains uncertain on this issue, and therefore open to the reader's views.

Machine learning can indeed approximate the normalized sizes of transactions of both visible and hidden orders (Table 2). Furthermore, machine learning does well at the approximation of the aggregated average market impact of financial trades (Figure 13 to Figure 14). The most distinguished results that are observed are that when machine learning is employed to estimate only the sizes of hidden orders, the machine-generated aggregate average market impact curves tend to be smooth and fairly approximate the literature computed aggregated average market impact curves. However, the far-right tails, which are orders of normalized sizes greater than the mean normalized volume traded, are less representative of the reality.

When machine learning is employed to estimate the sizes of all orders and the aggregated average market impact thereof, the far left tails of the aggregated average market impact

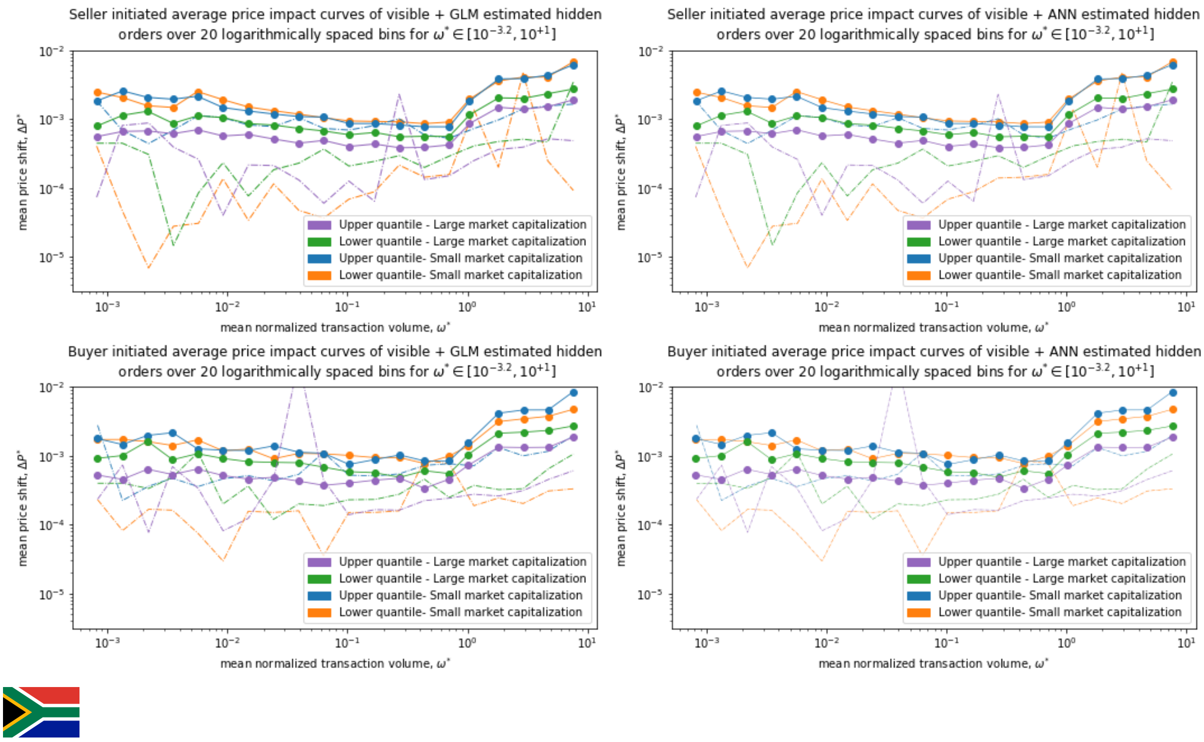


Figure 9. JSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated hidden orders, represented by dashed lines.

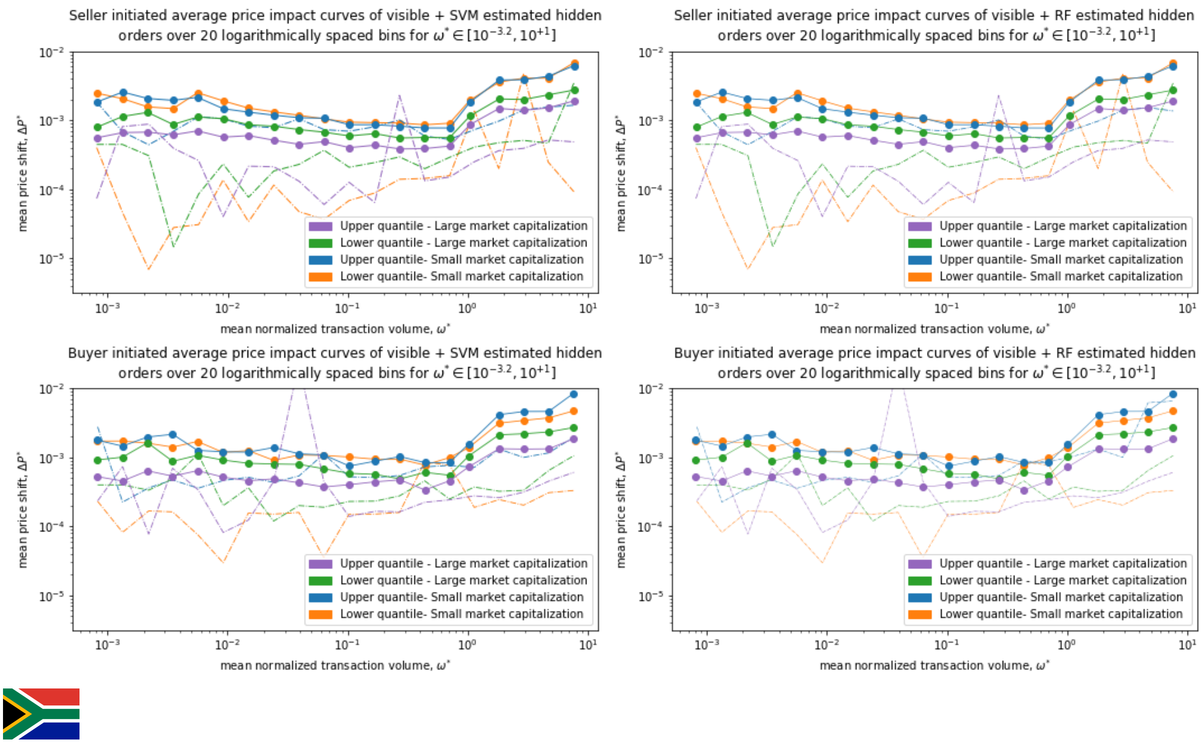


Figure 10. JSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated hidden orders, represented by dashed lines.

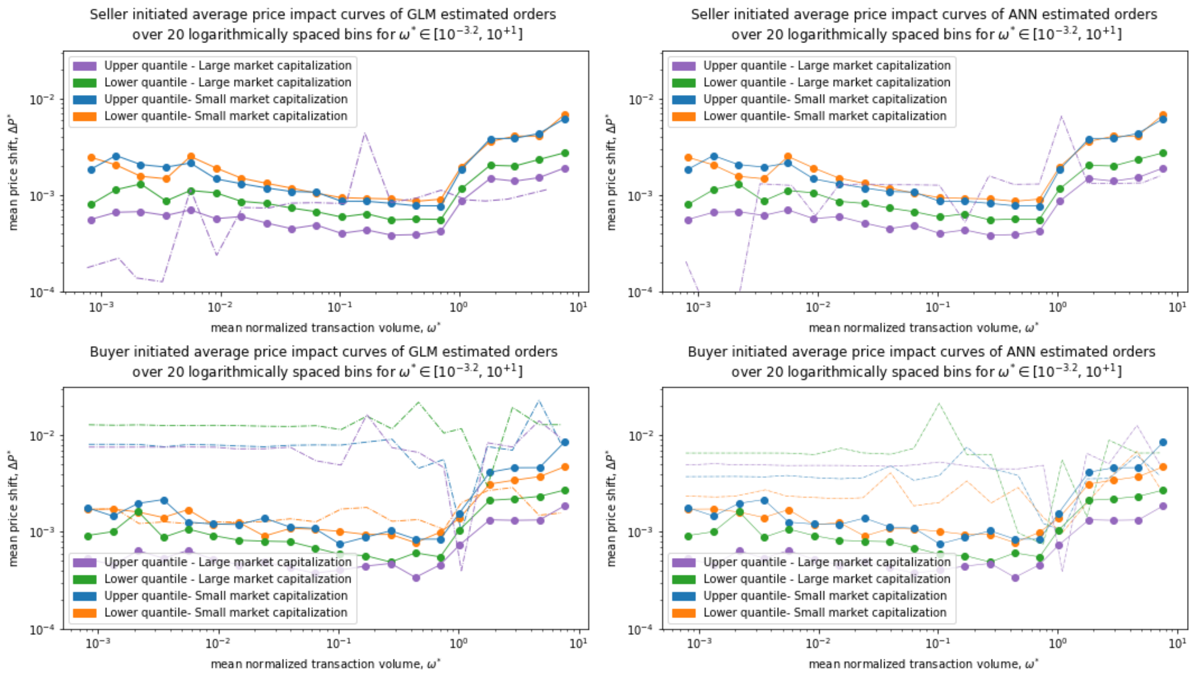


Figure 11. JSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

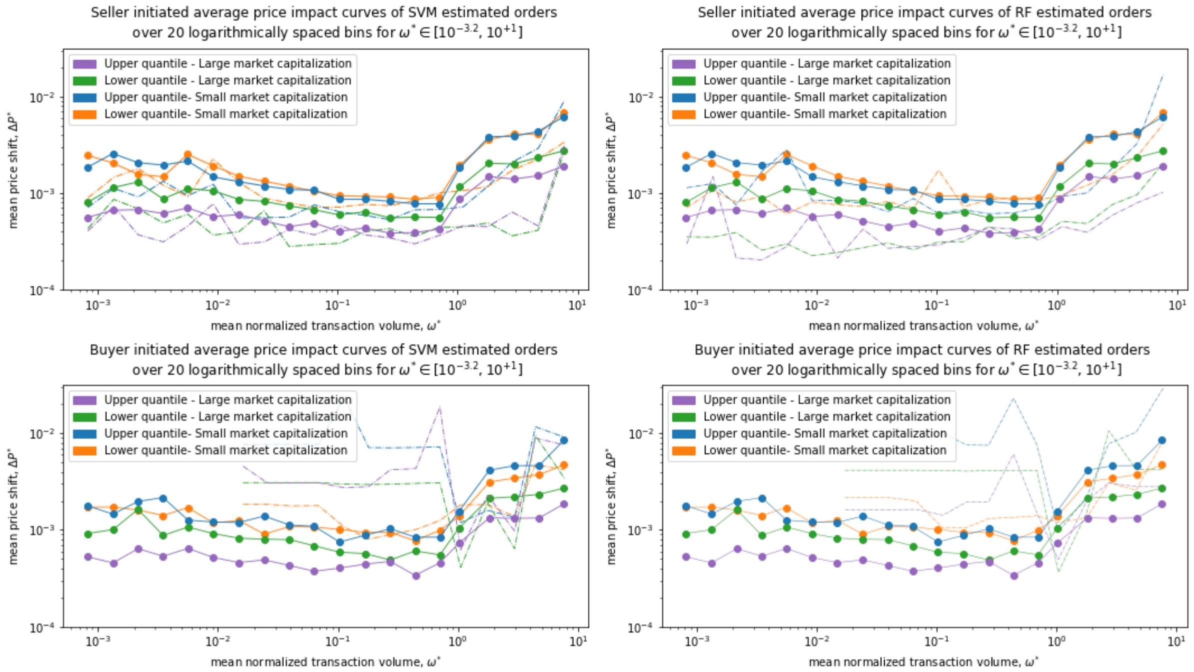


Figure 12. JSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

curves, which are orders of normalized sizes less than the mean normalized volume traded, are not representative of the reality. However, that occurs at a near-perfect approximation of the far-right tails. Thorough studies have been conducted in the context of the research to investigate the reasons for the observations [48, 49, 50]. The research discovered that when machine learning techniques are being used to estimate the sizes of hidden orders only, the techniques tend to keep in memory the estimation routine. The observations suggest that hidden orders execute within similar market conditions.

Machine learning techniques tend to perform extremely well when used to estimate all orders. However, missing information associated with certain orders such as visible orders can reduce the performance of the techniques. Since the techniques were designed to estimate hidden orders, and when the same techniques are being called to estimate all orders and the average market impact of all orders, only the far right tail of the average market impact curves are reproduced accurately, the research discovers empirically that hidden orders in the JSE are necessarily large. Moro et al.(2009) reported similar observations.

Akin to section 5.1.1.1 a separate analysis has been performed on the machine learning aspect of the research. Tables 3 - 7 indicate exceptional performance on average for machine learning on the estimation of the normalized sizes of hidden orders. Large R^2 and small $\hat{\text{MSE}}$ in the statistics are recorded and disclosed in the tables. Hence machine learning can, with limitations, accurately predict the sizes of hidden orders. Not only can hidden orders be uncovered, but the average market impact curves of visible orders and machine estimated hidden orders that are close to the average market impact curves of visible orders and true hidden orders could be calculated. Figures 13 -14 show that the mean of the average market impact curves of visible orders and machine estimated hidden orders approximates the mean of the average market impact curves of visible orders and true hidden orders well. Tables 3 - 7 and Figures 13 - 14 support the title of the research.

The research briefs the unfamiliar researcher on how the results of Figures 13 - 14 were derived. The average market impact curve of visible orders and true hidden orders was deduced by applying the approach of Lillo et al.(2003) on the data of visible orders and true hidden orders of individual firms. The average market impact curve of visible and estimated hidden orders was deduced by applying the approach of Lillo et al.(2003) on the data of visible orders and machine estimated orders of individual firms using the technique of section $4.4.3 - 4.4.6$. The individual average market impact curves have been averaged to determine the mean curve of the individual average market impact curves.

The confidence intervals were derived from calculating the varying standard deviation of the average price shifts of the individual firms in specified normalized volume bins. The research allowed the standard deviation to vary throughout. For a purpose of introducing some degree of uncertainty on the calculations such that a confidence interval could be derived about how the results are interpreted. The critical value of 90% has been multiplied with varying standard deviation figures and the product has been divided by the square root of the sample size. The results assist in turning casual comments into contextual comments that are substantiated. As the normalized volume increases over

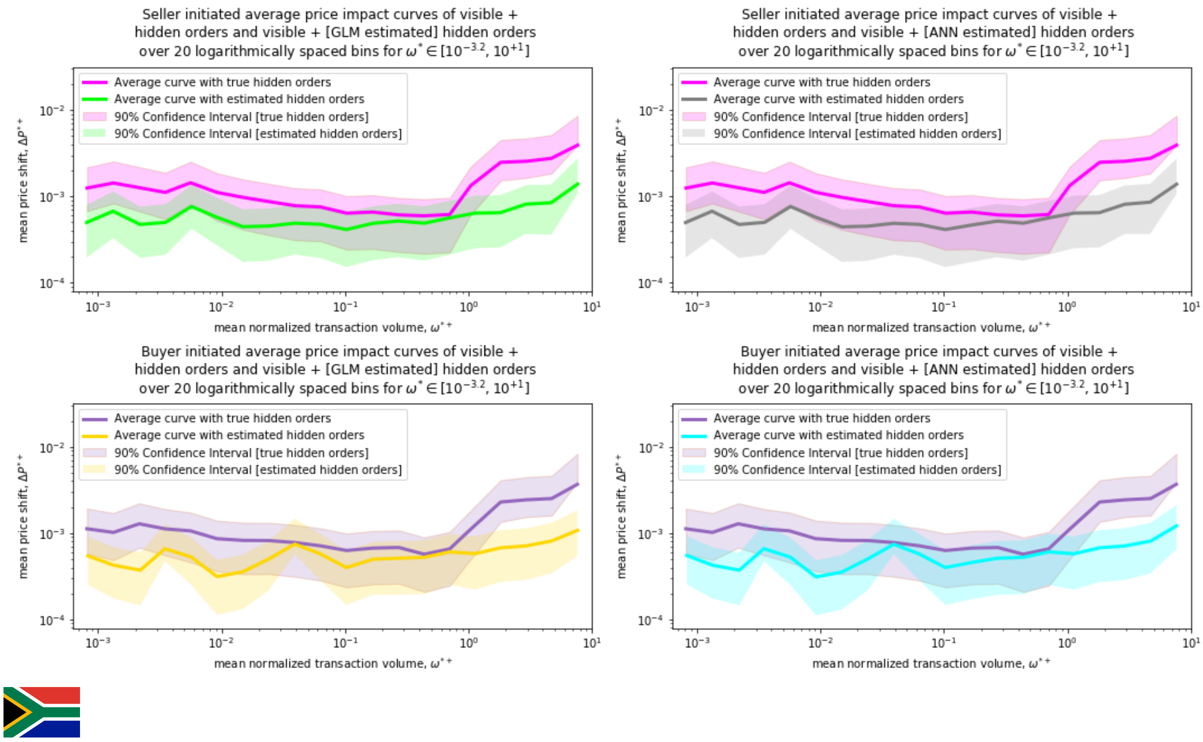


Figure 13. JSE. The mean price impact curves and the 90% confidence interval from the mean of single trades of 63 liquid firms. The sizes of hidden orders are estimated through generalized linear model (left plots), and artificial neural networks (right plots).

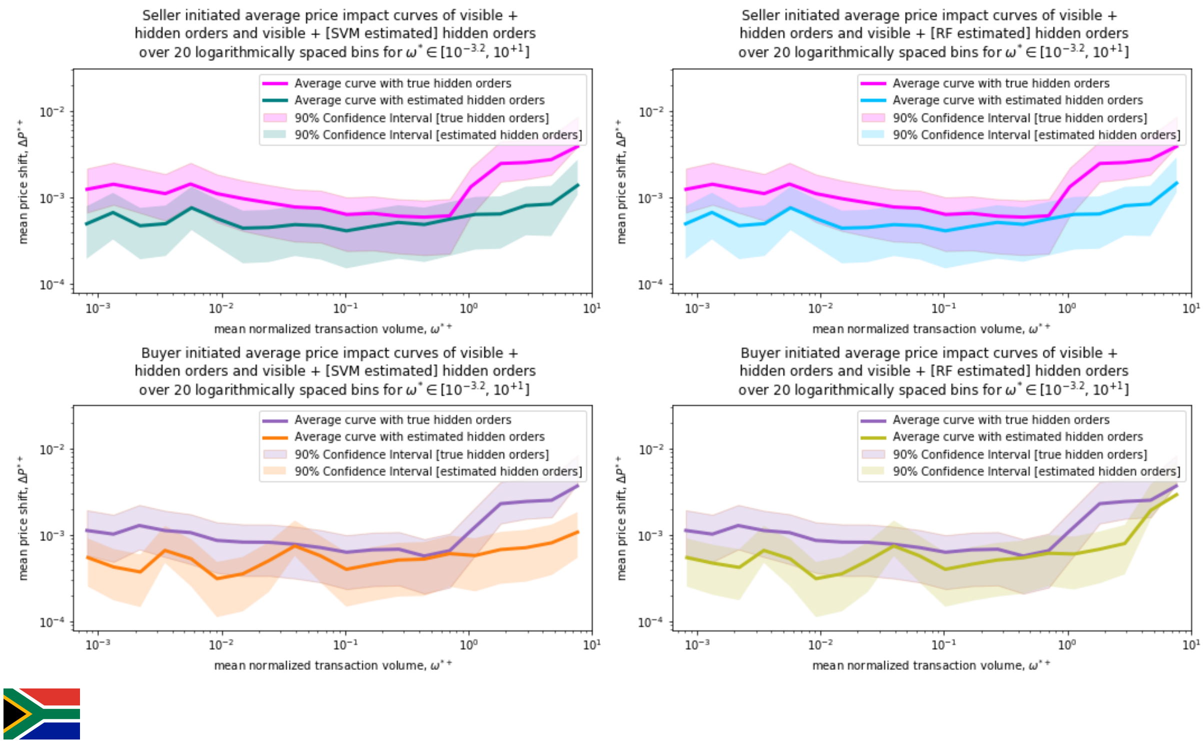


Figure 14. JSE. The mean price impact curves and the 90% confidence interval from the mean of single trades of 63 liquid firms. The sizes of hidden orders are estimated through support vector machines (left plots), and random forests (right plots).

Table 3. JSE. GLM parameters. Range and average across top 63 highly liquid stocks.

Feature	Feature description	Metric	Coefficient value(s)	SE	tStat	pValue
x_{i0}	Bias	range	[-308.16, 147.82]	[4.40, 255.08]	[-6.64, 4.02]	[4.2E-11, 0.98]
		mean	-2.1223	30.6957	0.1792	0.3282
x_{i1}	Price	range	[-0.11 , 0.09]	[0.00 , 0.15]	[-5.45, 3.53]	[5.9E-08, 0.97]
		mean	-0.0056	0.0157	-0.1241	0.2963
x_{i2}	Bid price	range	[-1.69, 3.47]	[0, 1.64]	[-7.09, 32.04]	[6E-173,0.96]
		mean	0.0905	0.1423	1.7529	0.2322
x_{i3}	Bid size	range	[-0.00, 3.47]	[4.5E-07, 1.64]	[-4.44, 32.04]	[0 , 0.96]
		mean	0.0001	3.5E-05	4.5990	0.1748
x_{i4}	Ask price	range	[-9.12, 3.44]	[0 , 3.44]	[-35.71, 16.14]	[2E-208, 0.97]
		mean	-0.1049	0.3886	-0.4519	0.2599
x_{i5}	Ask size	range	[-0.00, 0.00]	[2.8E-07, 0.00]	[-9.26, 26.54]	[3E-133, 0.77]
		mean	7.1E-05	3.3E-05	3.9128	0.1440
x_{i6}	Turnover ratio	range	[-2E-07, 8E-07]	[6E-10, 4E-08]	[-23.0, 263.8]	[0 , 0.01184]
		mean	1.7E-07	6.2E-09	45.76073	0.0002
x_{i7}	Price change	range	[-38764,62354]	[0 , 32251]	[-36.25, 10.09]	[4E-213, 0.89]
		mean	519.76	3664.40	-0.5498	0.2564
x_{i8}	Spread prior to trade	range	[-1.5677, 4.5827]	[0 , 1.7189]	[-18.89, 35.57]	[4E-207, 0.92]
		mean	0.0706	0.1231	-0.3805	0.2517
x_{i9}	Spread thereafter trade	range	[-0.0367, 0.5368]	[0.0003, 0.0653]	[-5.97, 20.72]	[4E-85, 0.996]
		mean	0.0262	0.0095	2.4358	0.2077
x_{i10}	Mid quote price prior to trade	range	[0, 0]	[0, 0]	-	-
		mean	0	0	-	-
x_{i11}	Mid quote price subsequent to trade	range	[-6.8458, 9.0515]	[0.0005, 3.4355]	[-18.27, 35.76]	[8E-209, 0.95]
		mean	0.0198	0.5305	0.0759	0.2324
x_{i12}	Moving average value traded	range	[-3.7E-05, 0.00]	[6.8E-08, 9.8E-05]	[-5.91, 29.26]	[1E-155, 0.99]
		mean	1.3E-05	6.5E-06	1.2271	0.3484
x_{i13}	Volatility	range	[-300.7, 603.9]	[22.4769, 592.8]	[-4.42, 6.99]	[3.9E-12, 0.95]
		mean	56.0794	105.1373	0.5052	0.3016
x_{i14}	Momentum	range	[-148.677, 317.055]	[4.2484, 254.5]	[-4.4157, 6.83]	[1.2E-11, 0.99]
		mean	1.8005	30.3232	-0.2293	0.3023
x_{i15}	Order sign	range	[-0.6483, 0.7935]	[0.0172, 0.2491]	[-18.46, 14.51]	[2.0E-72, 0.89]
		mean	-0.0138	0.0585	-0.9273	0.1926
x_{i16}	Signed volume	range	[-0.9992, 0.9742]	[0.001, 0.05]	[-817.2, 132.8]	[0 , 0.7040]
		mean	0.0165	0.0180	-11.0807	0.0614
x_{i17}	Order imbalance	range	[-0.0069, 0.9744]	[5.3E-05, 0.0534]	[-5.35, 132.82]	[5.4E-10, 0.70]
		mean	-0.0005	0.0012	-0.3316	60.2522
x_{i18}	Liquidity	range	[0, 0]	[0, 0]	-	-
		mean	0	0	-	-
x_{i19}	Execution time	range	[-1.1814, 1.0729]	[0.0063, 1.8805]	[-3.8784, 7.87]	[63E-15, 0.99]
		mean	0.0017	0.1037	0.2718	0.5015
x_{i20}	PIN	range	[-254.8, 137.7]	[0 , 363.2]	[-0.97, 2.08]	[0.0378, 0.96]
		mean	0.4239	23.0793	0.1424	0.6601
x_{i21}	KL divergence counts	range	[-1.7634, 2.9993]	[0.3574, 3.6915]	[-1.4159, 3.27]	[0.0014, 0.99]
		mean	0.1674	1.1565	0.17662	0.5633
x_{i22}	KL divergence volume	range	[0, 0]	[0, 0]	-	-
		mean	0	0	-	-
	Ordinary R^2	Adjusted R^2	MSE	RMSE		
range	[0.2524, 0.9987]	[0.2431, 0.9987]	[0.14053, 17.3241]	[0.3749, 4.1622]		
mean	0.3716	0.3624	0.6572	0.8107		

the mean of the average normalized volumes, the market impact increases as a power. The confidence intervals are much wider for normalized volumes less than the mean of the average normalized volume, especially for orders of size strictly less than the mean

Table 4. JSE. GLM parameters of large capitalization CPI stock and small capitalization ADH stock.

Feature	Feature description	Coefficient	Capitalization	Coefficient value(s)	SE	tStat	pValue
x_{i0}	Bias	α_0	large	-77.3893	38.9286	-1.9880	0.0471
			small	64.9161	19.1075	3.3974	0.0007
x_{i1}	Price	α_1	large	-3.4E-05	0.0005	-0.0732	0.9417
			small	-0.0758	0.0222	-3.4165	0.0006
x_{i2}	Bid price	α_2	large	0	0	-	-
			small	0.4732	0.3313	1.4281	0.1534
x_{i3}	Bid size	α_3	large	0.0004	0.0002	1.5866	0.1130
			small	4.8E-05	1.2E-05	4.1349	3.7E-05
x_{i4}	Ask price	α_4	large	-0.0005	0.0006	-0.7579	0.4487
			small	0.3759	0.3369	1.1158	0.2646
x_{i5}	Ask size	α_5	large	0.0005	0.0002	2.3523	0.0189
			small	1.6E-05	1.2E-05	1.2818	0.2000
x_{i6}	Turnover ratio	α_6	large	7.7E-08	6.0E-09	12.8032	1.6E-34
			small	8.0E-08	1.1E-09	74.3792	0
x_{i7}	Price change	α_7	large	0	0	-	-
			small	1159.0493	997.1859	1.1623	0.2452
x_{i8}	Spread prior to trade	α_8	large	9.1E-05	0.0004	0.2075	0.8357
			small	0	0	-	-
x_{i9}	Spread subsequent to trade	α_9	large	-0.0003	0.0003	-0.9090	0.3636
			small	0.0669	0.0102	6.5724	6.0E-11
x_{i10}	Mid quote price prior to trade	α_{10}	large	0	0	-	-
			small	0	0	-	-
x_{i11}	Mid quote price subsequent to trade	α_{11}	large	0.0006	0.0005	1.1258	0.2606
			small	-0.7757	0.6718	-1.1548	0.2483
x_{i12}	Moving average value traded	α_{12}	large	3.6E-07	2.3E-07	1.6016	0.1096
			small	-7.7E-07	7.1E-06	-0.1089	0.9133
x_{i13}	Volatility	α_{13}	large	480.7735	218.5557	2.1998	0.0281
			small	82.0011	130.1168	0.6302	0.5286
x_{i14}	Momentum	α_{14}	large	66.8747	37.9810	1.7607	0.0786
			small	-59.3820	18.7701	-3.1636	0.0016
x_{i15}	Order sign	α_{15}	large	-0.1172	0.0435	-2.6952	0.0072
			small	-0.0887	0.0754	-1.1761	0.2397
x_{i16}	Signed volume	α_{16}	large	0.2910	0.0299	9.7457	2.2E-21
			small	0.1651	0.0106	15.5134	8.1E-52
x_{i17}	Order imbalance	α_{17}	large	0.0031	0.0026	1.2315	0.2185
			small	0.0015	0.0009	1.7245	0.0848
x_{i18}	Liquidity	α_{18}	large	0	0	-	-
			small	0	0	-	-
x_{i19}	Execution time	α_{19}	large	0.0580	0.3013	0.1924	0.8475
			small	-0.0163	0.0270	-0.6042	0.5458
x_{i20}	PIN	α_{20}	large	0	0	-	-
			small	-21.0066	64.8719	-0.3238	0.7461
x_{i21}	KL divergence counts	α_{21}	large	-0.4113	0.4101	-1.0028	0.3162
			small	-1.0880	2.4329	-0.4472	0.6548
x_{i22}	KL divergence volume	α_{22}	large	0	0	-	-
			small	0	0	-	-
	Ordinary R^2	Adjusted R^2	MSE	RMSE			
large	0.2828	0.2698	0.2914	0.5399			
small	0.7570	0.7552	8.4242	2.9024			

of the average normalized volume. The observation suggests that we could argue that market impact is a steady decreasing function of normalized volume for orders of size

Table 5. JSE. ANN parameters. Range, mean, small cap stock and large cap stock.

Metric	Range	Mean	Small Capitalization Stock	Large Capitalization Stock
Best Epoch	[1 , 1000]	154.7619	17	1000
Best Performance	[3.7527E-09, 623.70]	29.16	0.5381	4.21E-08
Best Validation Performance	[2.3542E-08, 2029.48]	97.44	1.4485	6.76E-08
Best Test Performance	[1.1619E-05, 2822.49]	269.38	0.5537	1.1619E-05

	R^2	MSE
small	0.9983	0.6771
large	1	1.7842E-06
range	[0.0313, 1]	[1.7842E-06, 549.93]
mean	0.8976	75.4800

strictly less than the mean of the average normalized volumes because the wider interval suggests less confidence in the interpretation of the results.

In addition, the research examined the performance of Generalized Linear Models (GLM) to determine if there could be common features that could be used to predict the sizes of hidden orders across a variety of firms. According to Figures 15, the mid quote price before, liquidity and KL divergence volume are not useful to the linear learning algorithms. Meanwhile, the turnover ratio contributes significantly. The observation suggests that in an ideal firm there exists a correlation between the volume of shares executed and the turnover ratio. Large - capitalization firm: CPI and small - capitalization firm: ADH have similar feature contribution effects. Figure 15 shows that although there are features that are exploited by the GLM to predict the sizes of hidden orders, there exists no set of market features that are consistently predictive of the sizes of hidden orders across different stocks. The research held the computing algorithms constant and proceeded to seek answers to research question from the BOVESPA, MOEX, NSE, SSE and many other markets.

In conclusion of the analysis on the Johannesburg Stock Exchange, the research shows that: Hidden orders play a significant role in the tails of the aggregated averaged market impact curves; Hidden orders drive the aggregated average market impact curves upward;

Table 6. JSE. SVM parameters. Range, mean, small cap stock and large cap stock.

Metric	Range	Mean	Small Capitalization Stock	Large Capitalization Stock
Gap	[0.0008, 0.0054]	0.0010	0.0010	0.0008
Delta Gradient	[0.0015, 0.1719]	0.0391	0.0264	0.0488
Largest KK Violation	[0.0010, 0.1288]	0.0292	0.0168	0.0392
Objective	[-1281.2, -5.19]	-228.8	-221.5	-290.1
Bias	[-0.320, -0.010]	-0.111	-0.0571	-0.3095
Epsilon	[0.0017, 0.0688]	0.0222	0.0091	0.0541
	R^2		MSE	
small	-		0.3416	
large	-		0.9213	
range	-		[0.0015, 1.3593]	
mean	-		0.4284	

Market capitalization separates the aggregated average market impact curves of small firms and large-capitalization firms, with large firms curves at the bottom and small-capitalization firms at the top; Market capitalization is responsible for the vertical evolution of the average market impact curves; Akin to the New York Stock Exchange the Lillo et al.(2003) average market impact function is of significant relevance to the Johannesburg Stock Exchange; Hidden orders significantly improve the relevance of the Lillo et al.(2003) average market impact function on the Johannesburg Stock Exchange; and Machine learning produces significantly accurate estimations of both the normalized volume and the average market impact curves.

In addition to the JSE the research investigated the answers of the research questions from a variety of markets. Discussion of the results are provided in subsection 5.1.2 to subsection 5.2.13. Supporting evidence is provided in the appendix.

Table 7. JSE. RF parameters. Range, mean, small cap stock and large cap stock.

Metric	Range	Mean	Small Capitalization Stock	Large Capitalization Stock
ResubLoss	[0.0293, 158.5]	0.1460	7.2578	0.0645
	R^2		MSE	
small	0.5012		16.1866	
large	0.0172		1023.4	
range	[0.0007, 0.8155]		[0.3505, 2755.4]	
mean	0.0897		734.6	

5.1.2 Brazil (Bolsa de Valores de Sao Paulo, BOVESPA)

5.1.2.1 Literature Results (Bolsa de Valores de Sao Paulo, BOVESPA)

Figure 20+ provides seller initiated and buyer initiated average market impact curves of the BOVESPA, both before and after the inclusion of hidden orders. Du Preez et al.(2015) reported similar results. Hidden orders seem to affect the center of the market impact curves. It is unclear what the role of market capitalization is. Table 10+ provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled averaged price shifts and re-scaled averaged normalized volumes, before and after the inclusion of hidden orders.

The inclusion of hidden orders lead to an improvement of the R^2 for seller initiated market impact curves by 1.6553% from 0.7793 to 0.7922. The $\mathrm{M\hat{S}E}$ has also improved by 1.1236% from 0.0178 to 0.0176. Although the inclusion of hidden orders has improved the R^2 of the buyer initiated market impact curves by 3.5558% from 0.7565 to 0.7834, the $\mathrm{M\hat{S}E}$ has worsened by 45.1613% from 0.0093 to 0.0135. The SE, tStat and pValue ranges are: $[0.0113, 0.0157]$, $[7.5, 8.3]$ and $[1 \times 10^{-07}, 3 \times 10^{-07}]$ respectively. Nevertheless, the general results imply the model of Lillo et al.(2003) to be relevant.

5.1.2.2 Literature & Machine Learning Results (Bolsa de Valores de Sao Paulo)

Table 27+ provides the machine learning results of all the stocks. It is unclear which technique yields accurate performance. Akin to the JSE much volume is traded towards the end. Figure 21+ presents the analytical and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 22+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Support Vector Machines and Random Forests.

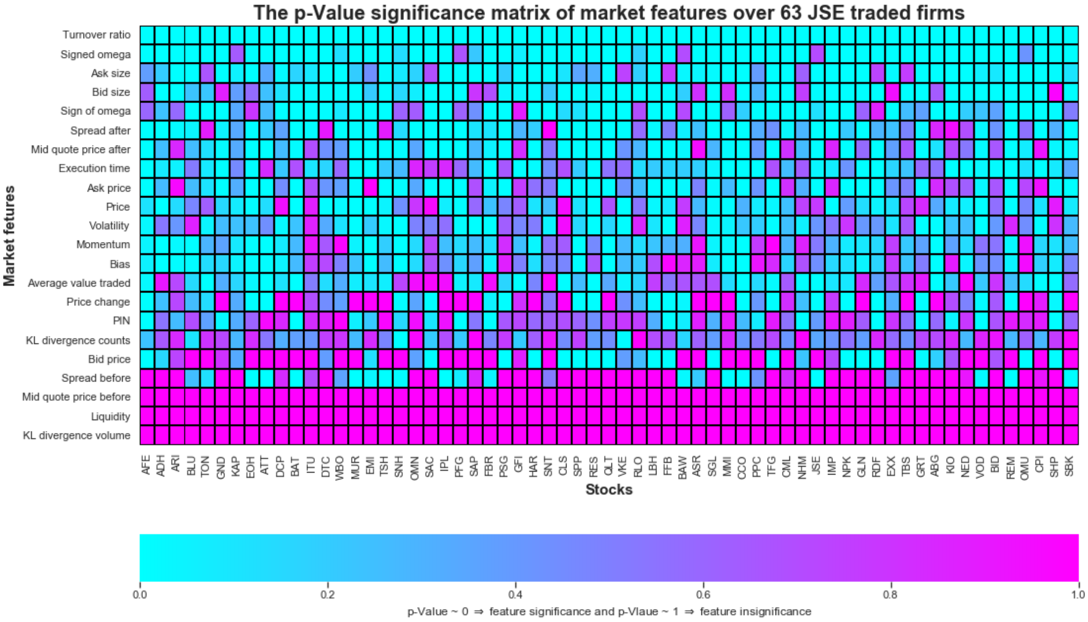


Figure 15. JSE. The p-value significance matrix of 63 JSE traded firms. The firms are ordered according to average daily value traded from small average daily value traded firms (left) to large average daily value traded firms (right). The p-Value reveals how significant a specified market feature is, with regard to the estimation of normalized transaction sizes using a generalized linear model. p-Value has a common scale of between 0 and 1, and that allows a cross examination of features across firms to establish patterns.

The unique observation of the results is that machine learning results fall below the literature results. A follow-up investigation was conducted. The research discovers that machine learning implies that many of the orders fall beyond the +1 edge. Consequently, the bins are deflated when transactions of the magnitude of normalized sizes in between the range $[-3.2, +1]$ are binned. The observation suggests that a significant number of transactions incur excessive market costs beyond the fair costs. One possible reason for this could be noise trading. Machine learning works partially for the determination of hidden orders and the average market impact.

5.1.3 Russia (Moscow Exchange, MOEX)

5.1.3.1 Literature Results (Moscow Exchange, MOEX)

Figure 23+ provides the results of the analysis of seller initiated and buyer initiated average market impact curves of the MOEX, both before and after the inclusion of hidden orders. Nonyane et al.(2018) reported different results. Perhaps differences in stock selection and grouping could be possible reasons for observed differences. Hidden orders seem to have no effect on the average market impact curves. It is unclear what

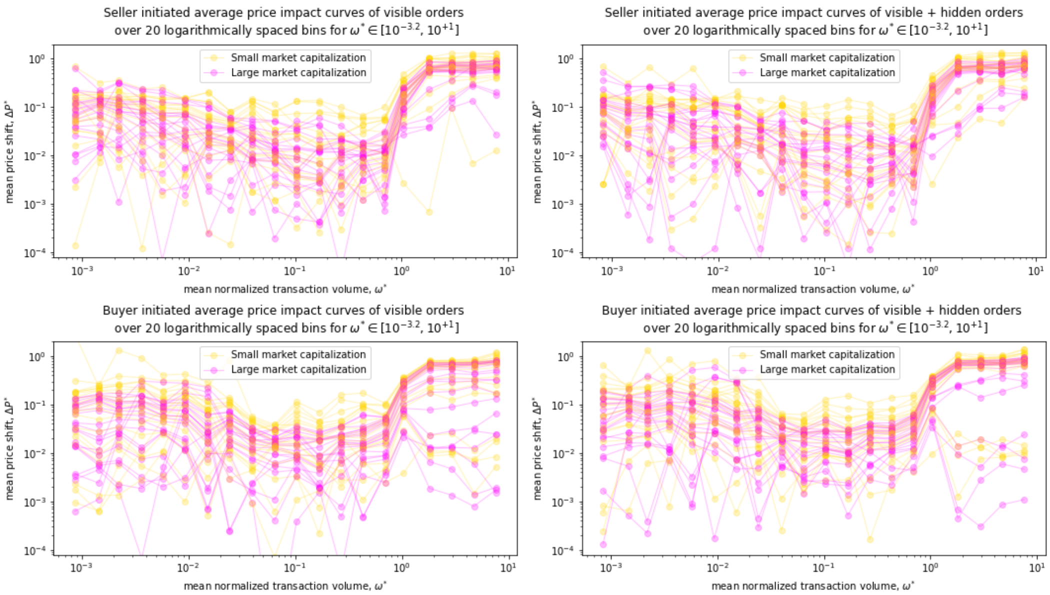


Figure 16. BOVESPA. Price impact curves of single trades of 42 relatively liquid firms. The curves are computed using the analytical literature approach. Lime curves represent firms with the average daily value traded larger than the median average daily value traded. Deeppink curves represent firms with the average daily value traded smaller than the median average daily value traded.The analysis period is from 28-09-2018 to 28-06-2019.

the role of market capitalization is. Table 11+ provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled averaged price shifts and re-scaled averaged normalized volumes, before and after the inclusion of hidden orders.

The R^2 for seller initiated market impact curves has improved by 2.1258% from 0.7324 to 0.7480 subsequent to the inclusion of hidden orders. The $\hat{\text{MSE}}$ has worsened by 1.4917% from 1.2804 to 1.2995. Meanwhile, the R^2 for buyer initiated market impact curves has improved by 0.5678% from 0.5107 to 0.5136 following the inclusion of hidden orders. The $\hat{\text{MSE}}$ has worsened by 1.3310% from 0.1127 to 0.1142. The SE, tStat and pValue ranges: $[0.0838, 0.5519]$, $[0.0836, 0.73125]$, and $[9 \times 10^{-07}, 0.6305]$. Other notable results are the values of α which are similar to the Harvey et al.(2016). The model of Lillo et al.(2003) seem to be relevant for seller initiated trading. The Lillo et al.(2003) model remains questionable for buyer initiated trading.

5.1.3.2 Literature & Machine Learning Results (Moscow Exchange)

Table 28+ provides the machine learning results of all the stocks. Much volume is exchanged at the start of the trading session. Figure 24+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 25+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Support Vector Machines

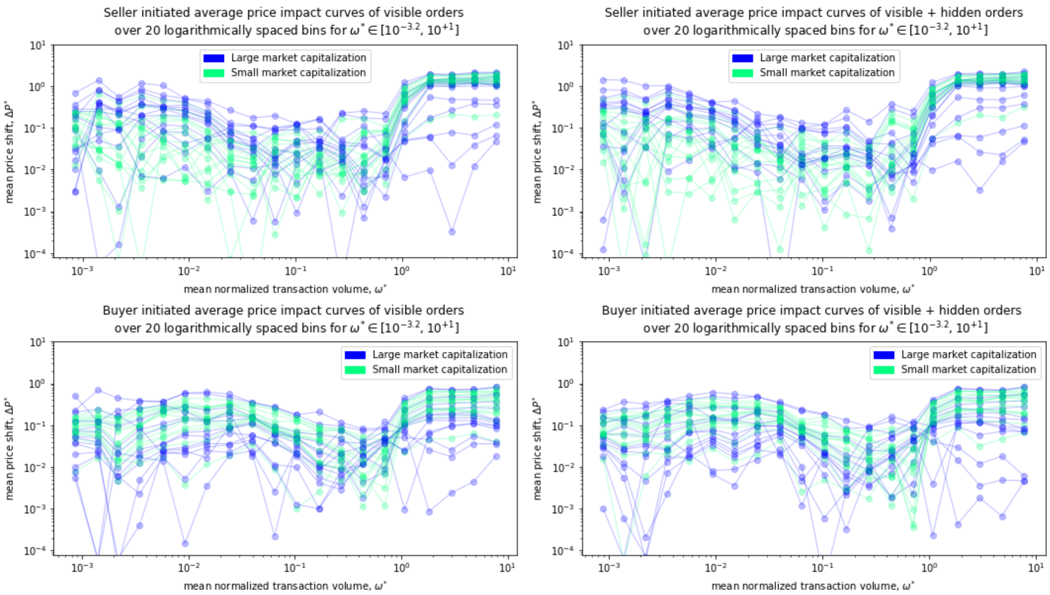


Figure 17. MOEX. Price impact curves of single trades of 30 relatively liquid firms. The curves are computed using the analytical literature approach. Teal curves represent firms with the average daily value traded larger than the median average daily value traded. Thistle curves represent firms with the average daily value traded smaller than the median average daily value traded. The analysis period is from 28-09-2018 to 28-06-2019.

and Random Forests. Artificial Neural Networks and Random Forest produce accurate results with the R^2 of 0.6495 and 0.6845 respectively, and the $\hat{\text{MSE}}$ of 2.6997 and 3.1195 respectively.

5.1.4 India (National Stock Exchange of India, NSE)

5.1.4.1 Literature Results (National Stock Exchange of India, NSE)

Figure 26+ provides the results of the analysis of seller initiated and buyer initiated aggregated average market impact curves, both before and after the inclusion of hidden orders. Nonyane et al.(2018) attempted to report on a similar analysis; however, insufficient data used in Nonyane et al.(2018) limits possible comparison to our research. Hidden orders cause the upward evolution of the impact curves. The role of market capitalization seem to be irrelevant.

Table 12+ provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes, before and after the inclusion of hidden orders. The R^2 for seller initiated market impact curves has improved by 7.4977% from 0.6442 to 0.6925 subsequent to the inclusion of hidden orders. The $\hat{\text{MSE}}$ has worsened by 50.6024% from 0.0083 to 0.0125. Meanwhile, the R^2 for buyer initiated market impact curves has improved by 0.3445% from 0.7547 to 0.7573 following the inclusion of hidden

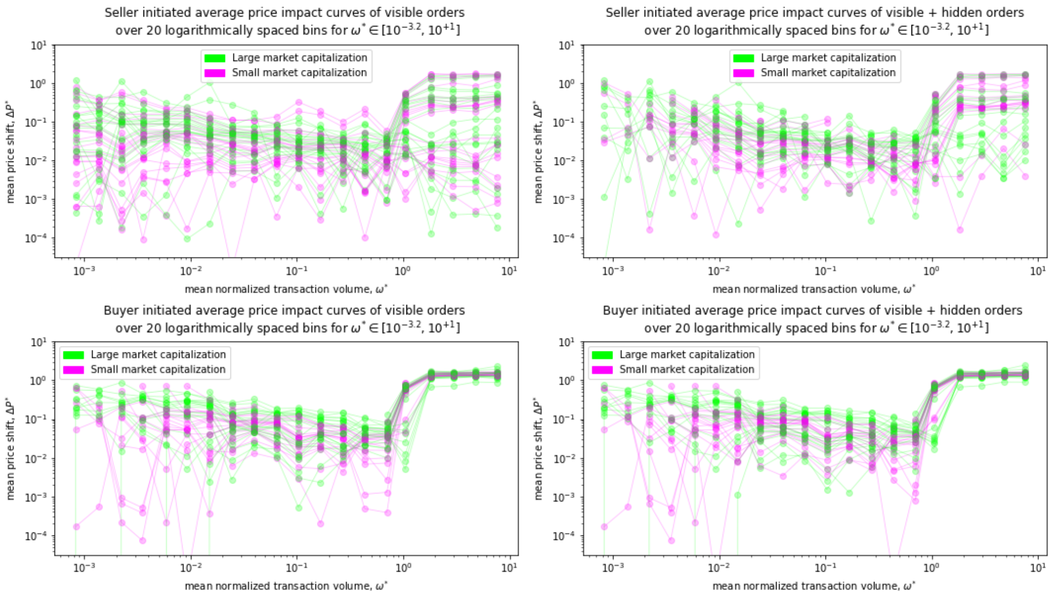


Figure 18. NSE. Price impact curves of single trades of 43 relatively liquid firms. The curves are computed using the analytical literature approach. Lime curves represent firms with the average daily value traded larger than the median average daily value traded. Magenta curves represent firms with the average daily value traded smaller than the median average daily value traded. The analysis period is from 28-09-2018 to 28-06-2019.

orders. The $\hat{\text{MSE}}$ has improved by 0.6309% from 0.0951 to 0.0945. The SE, tStat and pValue ranges: $[0.0214, 0.0777]$, $[0.1136, 1.8761]$, and $[7 \times 10^{-07}, 0.5886]$. Other notable results are the values of α which are similar to the Lillo et al.(2003). The model of Lillo et al.(2003) is relevant for NSE. Hidden orders produce marginal improvement on the performance of the Lillo et al.(2003) model on the NSE.

5.1.4.2 Literature & Machine Learning Results (National Stock Exchange of India)

Table 29+ provides the machine learning results of all the stocks. Figure 27+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 28+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Support Vector Machines and Random Forests. A similar average volume is traded throughout. Machine learning techniques are strongly affected by the presence of noise in the data. Many orders are implied to lie to the right beyond the +1 bound. Hence, machine estimated impact curves lie below the literature impact curves. The implication of the effect is that there could be a substantial amount of noise trading that results in many orders incurring excessive market impact costs.

5.1.5 China (Shanghai Stock Exchange, SSE)

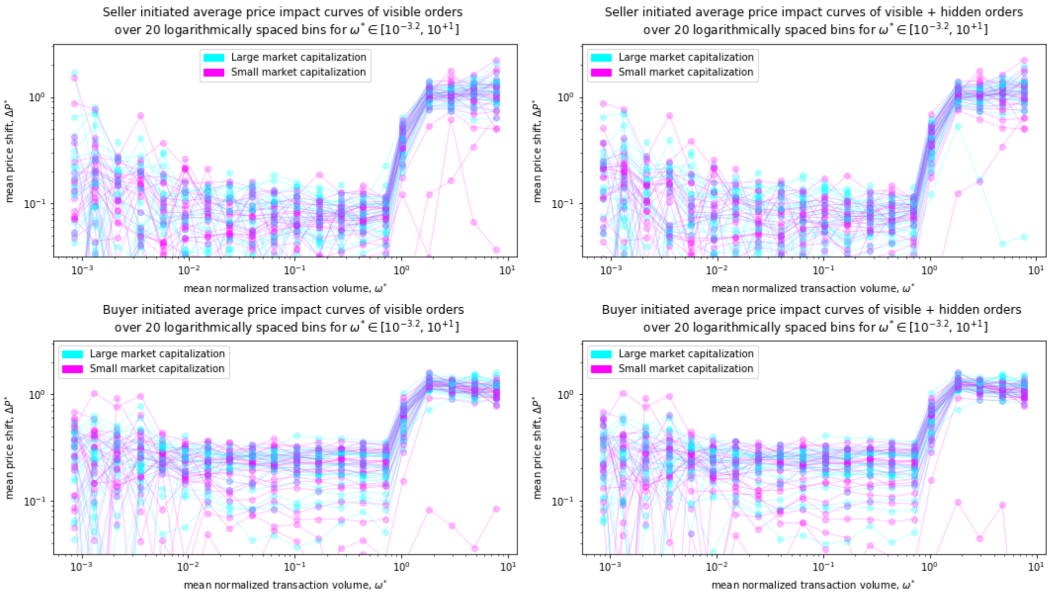


Figure 19. SSE. Price impact curves of single trades of 54 relatively liquid firms. The curves are computed using the analytical literature approach. Cyan curves represent firms with the average daily value traded larger than the median average daily value traded. Magenta curves represent firms with the average daily value traded smaller than the median average daily value traded. The analysis period is from 28-09-2018 to 28-06-2019.

5.1.5.1 Literature Results (Shanghai Stock Exchange, SSE)

Figure 29+ provides the results of the analysis of seller initiated and buyer initiated aggregated average market impact curves of the Shanghai Stock Exchange, both before and after the inclusion of hidden orders. Zhou et al.(2012) reported similar results. However, they considered data from the Shenzhen Stock Exchange, in contrast to our research which considered data from Shanghai Stock Exchange. Hidden orders have zero impact on the average market impact curves. Market capitalization has zero impact on the average market impact curves.

Table 13+ provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes, before and after the inclusion of hidden orders. The R^2 for seller initiated market impact curves has worsened by 0.1450% from 0.7596 to 0.7585 subsequent to the inclusion of hidden orders. The $\hat{\text{MSE}}$ has worsened by 2.5581% from 0.0430 to 0.0441. Meanwhile, the R^2 for buyer initiated market impact curves has worsened by 0.0152% from 0.6600 to 00.6599 following the inclusion of hidden orders. The $\hat{\text{MSE}}$ has improved by 0.6309% from 0.0951 to 0.0945. The SE, tStat and pValue ranges: $[0.0521, 0.0591]$, $[2.3339, 4.5753]$, and $[7 \times 10^{-07}, 0.0353]$. Other notable results are the values of α which are similar to the Lillo et al.(2003). Hence, the model of Lillo et al.(2003) is relevant. Hidden orders do not necessarily improve the performance of model of Lillo et al.(2003).

5.1.5.2 Literature & Machine Learning Results (Shanghai Stock Exchange)

Table 30+ provides the machine learning results of all the stocks. Much volume is exchanged towards the end of the trading session. Generalized Linear Models seem to be producing adequate predictions. Figure 30+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 31+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Support Vector Machines and Random Forests. The data for SSE is quite smooth. Conseque-
ntly, machine learning techniques produce quality results. The most notable effect is that contrary to the view that volume impacts price, price seems to be driving volume. The effect seems to be excessive market costs incurred by small volumes.

5.2 Developed Markets

Developed Markets are well-developed economies that have efficient capital markets that accommodate foreign ownership. The judgements of the results of sections 5.2.1 - 5.2.13 are based on the fact that the author has performed numerous data screening activities to compare actual normalized volume to estimated normalized volume. Unfortunately the author did not for see it as necessary to present such visualizations and statistics in the context of the current report. However, the author has attempted to scale down the large data into averages for purpose of sharing with readers the essence of actual numbers which were observed. Consider Tables 31+ - 43+ for empirical evidence.

5.2.1 United States of America (The New York Stock Exchange, NYSE)

5.2.1.1 Literature Results (New York Stock Exchange, NYSE)

Figure 32+ provides the results of the analysis of seller initiated and buyer initiated aggregated average market impact curves, both before and after the inclusion of hidden orders. Jiang et al.(2019) reported precisely similar results. Niu et Al.(2019) also reported similar results. Hidden orders affect the average market impact curves of small-capitalization firms. The curves change from what could be regarded as power-law to convex functions. Market capitalization separates small capitalization firms from large capitalization firms, with small capitalization firms appearing at the top and large capitalization firms appearing at the bottom.

Table 14+ provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes, before and after the inclusion of hidden orders. The R^2 for seller initiated market impact curves has worsened by 17.9006% from 0.4687 to 0.3848 subsequent to the inclusion of hidden orders. The $\hat{\text{MSE}}$ has improved by 50% from 2×10^{-07} to 10×10^{-08}. Meanwhile, the R^2 for buyer initiated market impact curves has improved by 3.9280% from 0.4557 to 0.4736 following the inclusion of hidden orders. The $\hat{\text{MSE}}$ has improved by 98.9655% from 0.0029 to 3×10^{-05}. The SE, tStat and pValue ranges: $[3.4 \times 10^{-05}, 0.0070]$, $[2.4178, 8.8745]$, and $[0.0011, 0.2102]$.

The Lillo et al.(2003) model is relevant to the average market impact in the NYSE. The only difference is the scale. Lillo et al.(2003) used the scale of between $[-2, +1]$ to bin transactions; meanwhile, the current research stretched the scale to between $[-3.2, +1]$ to bin transactions. Although, the $\hat{\text{MSE}}$ indicates the model of Lillo et al.(2003) to be relevant to the NYSE, the R^2 suggests that the average market impact of the NYSE may be gradually changing shape.

5.2.1.2 Literature & Machine Learning Results (New York Stock Exchange)

Table 31+ provides the machine learning results of all the stocks. Much volume is exchanged towards the end of the trading session. Support Vector Machines seem to be producing the desired results. Figure 33+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 34+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Support Vector Machines and Random Forests. The machine learning techniques seem to default (i.e do not yield satisfactory performance) for volumes less than the average volume traded; however, accurate performance is produced for volumes greater than the average volume traded. The observation suggests that hidden orders are essentially large in the NYSE.

5.2.2 United Kingdom of Britain (The London Stock Exchange, LSE)

5.2.2.1 Literature Results (London Stock Exchange, LSE)

Figure 35+ provides the results of the analysis of seller initiated and buyer initiated aggregated average market impact curves of the London Stock Exchange, both before and after the inclusion of hidden orders. Willinski et al.(2014) cited similar graphics.

Table 15+ provides the results of a curve fitting of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volumes, before and after the inclusion of hidden orders. Seller initiated trading of small-capitalization firms may be cheap in the absence of hidden orders. Market capitalization seems to play no role. The poor statistics of Table 15+ suggests that the model of Lillo et al.(2003) is irrelevant to the LSE. The role of hidden orders remains unclear in general. However, the introduction of hidden orders seem to worsen the performance of the Lillo et al.(2003) model. Meaning hidden orders worsen the statistics of the fit of Lillo et al.(2003) model. Consider Table 15+ for evidence. The author makes similar remarks with the remaining analyzed markets from here on. The readers are advised to refer to the disclosed tables and / or graphs for clarity on comments which may not be clear.

5.2.2.2 Literature & Machine Learning Results (London Stock Exchange)

Table 32+ provides the machine learning results of all the stocks. A similar average volume is traded throughout. Support Vector Machines produce adequate results. Figure 36+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Generalized Linear Models and Artificial Neural Networks. Figure 37+ presents both the analytical results and machine learning results of the seller initiated and the buyer initiated aggregated average market impact curves in which orders are estimated via Support Vector Machines and Random Forests. All the machine learning techniques seem to be yielding adequate results.

5.2.3 United Arab Emirates (The Dubai Financial Market, DFM)

Hidden orders seem to have a major impact on the relevance of the Lillo et al.(2003) model on the DFM (Figure 38+ and Tables 16+ & 33+). Large capitalization firms incur large price changes, and small-capitalization firms incur small price changes, in contrast to the other markets such as the JSE and the NYSE. Machine learning techniques seem to pick up the tails well, in particular for seller initiated trading (Figures 39+ - 40+).

5.2.4 France (The Paris Stock Exchange, Euronext Paris)

The inclusion of hidden orders seem to worsen the performance of the model of Lillo et al.(2003) (Figure 41+ and Tables 17+ & 34+). Market capitalization determines the vertical location of the average market impact curves akin to the NYSE and the JSE. The only distinct observation is that the average value traded measure may not necessarily be an absolute measure since the curves' order is slightly different. The observation suggests that even large capitalization firms may have small average value traded. Machine learning performs well on both prediction of orders and the average market impact curves (Figures 42+ - 43+).

5.2.5 Singapore (The Singapore Exchange, SGX)

The model of Lillo et al.(2003) is irrelevant; however, the inclusion of hidden orders improves the performance of the model. Similar observations to the Euronext Paris are realized with regard to the role of market capitalization Lillo et al.(2003). Machine learning techniques, in particular GLM and ANN, yield satisfactory performance (Figures 44+ - 46+ and Tables 18+ & 35+).

5.2.6 Spain (The Bolsa de Madrid, BME)

The model of Lillo et al.(2003) seems to be relevant; however, the inclusion of hidden orders worsens the model performance. It is unclear what the role of market capitalization is. Machine learning yields fair performance (Figures 47+ - 49+ and Tables 19+ & 36+).

5.2.7 Australia (The Australia Securities Exchange, ASX)

The model of Lillo et al.(2003) seems to be relevant to buyer initiated trading before the inclusion of hidden orders. The inclusion of hidden orders worsens the performance of the model. Market capitalization implies that it is cheaper to trade small-capitalization firms in the absence of hidden orders; however, in the presence of hidden orders, it is cheaper to trade large capitalization firms (Figure 50+ and Table 20+). Lim et al.(2005) reported similar results to the results of Figure 50+. Machine learning produces a fair performance (Figures 51+ - 52+ and Table 37+).

5.2.8 Austria (The Vienna Stock Exchange, WBAG)

The model of Lillo et al.(2003) seem to be irrelevant. Hidden orders seem to play no role. Machine learning produces acceptable performance (Figures 54+ - 55+ and Tables 21+ & 38+).

5.2.9 Portugal (The Lisbon Stock Exchange, Euronext Lisbon)

The model of Lillo et al.(2003) does not seem to be relevant. Perhaps with much more data, the model would be relevant. Hidden orders make matters worse. Market capitalization plays no role. Machine learning does not yield satisfactory performance. Perhaps there is a need to increase the data (Figures 57+ - 58+ and Tables 22+ & 39+).

5.2.10 Belgium (The Brussels Stock Exchange, Euronext Brussels)

Large capitalization firms have large price changes, and small-capitalization firms have small price changes. The model of Lillo et al.(2003) is irrelevant. Hidden orders do not improve the performance of the model of Lillo et al.(2003). Machine learning provides outstanding performance (Figures 59+ - 61+ and Tables 23+ & 40+).

5.2.11 Norway (The Oslo Stock Exchange, OSE)

The model of Lillo et al.(2003) is irrelevant. Hidden orders cause the average impact curves to evolve upwards. It is unclear what the role of market capitalization is. Machine learning picks up the tails (Figures 62+- 64+ and Tables 24+ & 41+).

5.2.12 Finland (The Helsinki Stock Exchange, OMXH)

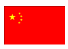
The model of Lillo et al.(2003) is irrelevant. The role of hidden orders is unclear. The role of market capitalization is unclear. Machine learning provides fairly acceptable performance (Figures 65+ - 67+ and Tables 25+ & 42+).

5.2.13 Germany (The Frankfurt Stock Exchange, FWB)

The model of Lillo et al.(2003) seem to be relevant. However, the presence of hidden orders seems to worsen the performance of the model. It seems to be cheaper to trade small-capitalization firms than to trade the large-capitalization firms. The role of market capitalization is identical to that of Euronext Paris. Machine learning reproduces the tails of the average impact curves. The observation suggests that hidden orders are essentially large (Figures 68+ - 70+ and Tables 26+ & 43+).

The research summarises the performance of the machine learning techniques in Table 8 and Table 9. Detailed statistics are provided in Table 2 and Tables 27+ - 43+. The results of Tables 8 - 9 were calculated by training machine learning techniques. Subsequent to the training, the machine learning techniques were used to estimate the sizes of traded orders. The $\widehat{\text{MSE}}$ was computed as the average of the square of differences between the estimated order sizes and the true order sizes.

Table 8. Summary statistics of the $\widehat{MSE}$ of emerging markets.

	Machine	Learning	Techniques	
	GLM	ANN	SVM	RF
	200.00	$2\times10^{+03}$	919.5449	$2\times10^{+03}$
	$3\times10^{+05}$	$3\times10^{+05}$	$3\times10^{+05}$	$3\times10^{+05}$
	3.8765	2.6997	3.6004	3.1195
	0.5979	0.0017	15.8741	0.0454
	$1\times10^{+03}$	$1\times10^{+04}$	$1\times10^{+04}$	$1\times10^{+03}$

Tables 8 - 9 suggest that Generalized Linear Models and Support Vector Machines are effective in estimating the sizes of orders, particularly for the Johannesburg Stock Exchange. Other markets suggest the contrary. Hence, each market suggest specified techniques to be compatible on the estimation of hidden orders.

Common observations among the results of subsection 5.2.3 to subsection 5.2.13 are the concave far-right tails of the average market impact curves. Farmer et al.(2008) argued that such observations result from increases in order flows predictability during the execution of large orders. Other notable observations are that the average market impact curves across markets do not necessarily take a similar shape to the market impact function of Lillo et al.(2003). Pazelt et al.(2017) suggested that a universal shape of the average market impact is non-linear, in which the scaling factor is determined as a Hurst exponent. Hence the master-curve for market impact would likely vary across markets.

Other related studies such as Hendricks et al.(2016) presented a feasible scheme for unsupervised state discovery, state detection and online learning for high-frequency quantitative trading agents faced with a multi-featured, asynchronous market data feed.

Correia et al.(2004) studied the cost of equity of companies listed on the Johannesburg Stock Exchange. Hendricks et al.(2014) improved post-trade implementation shortfall by up to 10.3% on average compared to the base model, based on a sample of stocks and trade sizes in the South African equity market.

Cui et al.(2012) modelled the price impact using an agent-based modeling approach. They investigated whether agent intelligence is a necessary condition when seeking to construct realistic price impact with an artificial market simulation. Prudent further investigations would include a broader literature, that has not yet been reviewed in the context of the research, with the aim of evaluating similarities across various reports to establish if there could be stylized facts that exist across markets which differ with operations and possibly many other factors.

5.3 Summary

Following the editor's advice the author wanted to stick to the proposed title of the research rather than to dilute the message for the reader by feeding much information on the analysis of the other markets. Thus the author decided to comment, in summarized form, on the analysis of the other markets. The concept which is proposed in the title of the research seem to be quite relevant for emerging markets as supported by Table 1 , Tables 10+ - 13+, Table 2, Tables 27+ - 30+ and Figures 20+ - 31+. Figures 13 -14 and Tables 3 - 7 emphasizes the claim on the JSE. However, the poor statistics of Tables 14+ - 26+ , Tables 31+ - 43+ and Figure 32+ - 70+ suggest shortcomings of the notion on developed markets. The author has a view that there may be other factors which remain unknown that could have prohibited acceptable performance of the techniques used as in general. In Tables 8 - 9, if the estimated normalized volumes are characterised by many outcomes that are sufficiently far from the true normalized volumes, then the mean squared error would likely to be large. Otherwise if a particular technique estimates individual normalized volumes that are sufficiently close to the true normalized volumes, then the mean squared error would be small. There may be other reasons that account for the vast differences of the results of Tables 8 - 9 which are unknown to the author at the moment. If one looks at a stock price process over a reasonably large time window, then the process looks like a wavelet. If one integrates many large time windows, then the whole process looks like a series of sines and cosines of a Fourier process. As future recommended research, the research seeks to understand stock price processes as functions of endogenous trading mechanisms including market impact. The utility of the project is to showcase the use of visible features and modern techniques to uncover the unknowns, for the purpose of achieving good decision making.

Table 9. Summary statistics of the $\widehat{MSE}$ of developed markets.

	Machine	Learning	Techniques	
	GLM	ANN	SVM	RF
	$3\times10^{+03}$	$3\times10^{+03}$	431.7919	$3\times10^{+03}$
	0.7026	16.2066	0.0347	0.1041
	5×10^{-05}	0.1103	0.0922	0.0215
	3.9254	3.4778	3.1706	2.7590
	4×10^{-06}	2.2279	0.7669	1.7554
	NaN	NaN	NaN	NaN
	3×10^{-05}	0.0420	0.0574	0.4283
	1×10^{-05}	12.4905	0.0801	0.0206
	0.0518	0.0093	0.0113	0.4024
	7×10^{-04}	0.9175	0.4878	0.2754
	0.0958	17.4278	13.8787	0.0277
	0.0105	0.0573	0.0821	0.1095
	3×10^{-33}	0.0060	0.0257	0.0466

Chapter 6

Conclusion

The average market impact model of Lillo et al.(2003) may not necessarily be the generic representation of the functional form of market impact across global financial markets. In some markets, the price seems to be the driving factor of volume traded as opposed to the norm of the volume being the driving factor of price. Large (small) market capitalization firms are cheap (expensive) to trade in some markets; meanwhile, the contrary occurs in other markets. The market capitalization of firms may not necessarily affect the average market impact curves. However, the market capitalization may separate the average market impact curves of small-capitalization stocks from the average market impact of curves of large-capitalization stocks in some markets. The role of hidden orders varies across markets. In some markets, hidden orders affect the tails of the average market impact curves; meanwhile, in some other markets, hidden orders may have zero effect on the average market impact curves. It would ideally be unreal to regard the average market impact of visible orders as reflective of the structure of the Lillo et al.(2003) average market impact model, in the absence of consideration of hidden orders, especially for the Johannesburg Stock Exchange equities market. Hence, the average market impact of Lillo et al.(2003) is particularly true depending on the condition that hidden orders are accounted for in the analysis.

Machine learning techniques suggest hidden orders to be executed within similar market conditions. For instance, if one quantifies market features according to the equations of section 4.4.1, then the techniques tend to assign more weight to the usual readings of the features for every normalized volume feed. If one attempts to test how well the used machine learning techniques perform, then one realizes that the prediction of normalized volume of visible orders is a volatile process relative to the outcomes of normalized volume of hidden orders. Such observations suggest that the readings on the market features are the same on average for estimated normalized volume of hidden orders. Machine learning techniques, in the context of the research, are sensitive to the unknowns. One possible future research question could be to reveal why our machine learning techniques are sensitive to unknowns. Machine learning can with limitations and varying degrees of exploitation of features, predict the sizes and the average market impact of orders. Machine learning could potentially impact a variety of research fields in financial markets. Future research proposes to study equities price processes as wavelets for reasons that are stated under section 5.3. The clearing process is applied to evaluate the effect of momentum that is introduced by the reactions of agents. The equities price processes are uncovered as a combination of sines and cosines of Fourier transform processes. The degree to which the equities price processes are driven by economic factors rather than the overreactions of agents is revealed. Future equities prices are predicted via machine learning.

The machine learning techniques yielded varying degrees of performances over a range of markets that were analysed. Perhaps, this gives an indication that different techniques adopt differently to different data patterns. This provokes a thought that there could be much more sophisticated techniques that could yield better performances. The author speculates that financial research would potentially call for tailor-made machine learning techniques. All these observations yield the overall effect that there could be a change in the machine learning research landscape which is triggered by the characteristics of financial data. The impact of the used machine learning techniques is such that traditional methods of doing analysis on financial research could soon be history.

Bibliography

[1] Almgren, R., Chriss, N. 2000. Optimal execution of portfolio transactions, J. Risk, vol. 3, no. 2, pp. 5–39.

[2] Amihud, Y., Mendelson, H. 1986. Asset pricing and the bid-ask spread, Journal Financial Economics, (17): 223 - 249.

[3] BARRA, Market impact model handbook (Berkeley, California, Barra, 1997).

[4] Berkowitz, S. A., Logue, D. E., Noser, E. A. J. 1988. The Total Cost of Transactions on the NYSE, Journal of Finance, American Finance Association. 43 (1): 97–112.

[5] Bertsimas, D., Lo, A. W. 1998. Optimal control of execution costs, Journal of Financial Markets, Elsevier, vol. 1(1), pages 1-50, April.

[6] Bouchaud, J.P., Farmer, J.D., Lillo, F. 2009. How markets slowly digest changes in supply and demand, in Handbook of Financial Markets: Dynamics and Evolution, North-Holland, Elsevier.

[7] Bouchaud, J.P. 2017. Price Impact, Capital Fund Management, Encyclopedia of Quantitative Finance, R. Cont Edt.; [q-fin.TR], arXiv: 0903.24.281. John Wiley Sons Ltd.

[8] Breiman, L., Friedman, J., Olshen, R., Stone, C. 1984. Classification and Regression Trees, Boca Raton, FL: CRC Press.

[9] Breiman, L. 1996. Bagging Predictors, Machine Learning 26, pp. 123–140.

[10] Breiman, L. 2001. Random Forests, Machine learning, Volume. 45, page. 5–32.

[11] Brooks, J. 1968. The Fluctuation: The Little Crash in '62, in "Business Adventures: Twelve Classic Tales from the World of Wall Street". (New York: Weybright Talley, 1968).

[12] Caudill, M. 1989. Neural Networks Primer, San Francisco, CA: Miller Freeman Publications.

[13] Correia, C., Uliana, E. 2004. Market segmentation and the cost of equity of companies listed on the Johannesburg Stock Exchange, South African Journal of Accounting Research 18, 65–86. doi:10.1080/10291954. 2004.11435109.

[14] Cui W., Brabazon A. 2012. An agent-based modeling approach to study price impact, 2012 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), pages 1-8.

[15] DARPA Neural Network Study, Lexington, MA: M.I.T. Lincoln Laboratory, 1988.

[16] Du Preez B.S. Market Micro-Structure. MSc Thesis (2015).

[17] Easely, D., Keifer, N.M., O'Hara, M. 2002. Is information risk a determinant of asset returns? Journal of Finance, 57(5):2185-2221.

[18] Efron, B., Tibshirani, R. J. 1993. An Introduction to the Bootstrap, New York: Chapman Hall.

[19] Evans, M., Lyons, R. 2002. Order flow and exchange rate dynamics, J. Political Economy, vol. 110, no. 1, pp. 170–180.

[20] Fan, R.E., Chen P.H., Lin C.J. 2005. Working Set Selection Using Second Order Information for Training Support Vector Machines, The Journal of Machine Learning Research, No: 6: 1871–1918.

[21] Farmer, J.D., Lillo, F. 2004. On the origin of power-law tails in price fluctuations, Quantitative Finance, 4: C7–C11. [Taylor Francis Online].

[22] Farmer J.D., Gerig A., Lillo F., Waelbroeck., H. 2008. The market im of large trading orders: Correlated order flow, asymmetric liquidity, and efficient prices.

[23] Francioni, R., Schwartz, R. A. 2017. Equity Markets in Transition: the Value Chain, Price Discovery, Regulation, and Beyond. https://doi.org/10.1007/978-3-319-45848-9.

[24] Gabaix, X., Gopikrishnan, P., Plerou, V Stanley, H. 2006 Institutional investors and stock market volatility, Quarterly Journal of Economics, 121:461-504.

[25] Gatheral, J., Schied, A., Slynko, A. 2010. Transient linear price impact and Fredholm integral equations.

[26] Gatheral, J. 2011. No-Dynamic-Arbitrage and Market Impact, To appear in Quantitative Finance.

[27] Glosten, L.R., Milgrom, P.R. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, J. Financ. Econ., 14: 71–100.

[28] Hagan, M.T., Menhaj, M. 1994. Training feedforward networks with the Marquardt algorithm, IEEE Transactions on Neural Networks, Vol. 5, No. 6, pp. 989–993.

[29] Hagan, M.T., Demuth, H.B., Beale, MH. Neural Network Design, Boston, MA: PWS Publishing, 1996, Chapters 11 and 12.

[30] Hasbrouck, J., Seppi, D. 2001. Common Factors in Prices, Order Flows and Liquidity, Journal of Financial Economics, 59, 388-411.

[31] Hastie, T., Tibshirani, R., Friedman, J. 2001. The Elements of Statistical Learning. New York: Springer.

[32] Harvey, M. (2014), Master Curve Honours Project, Program in Advanced Mathematics of Finance.

[33] Harvey, M., Hendricks, D., Gebbie, T., Wilcox, D. 2016. Deviations in expected price impact for small transaction volumes under fee restructuring, Elsevier B.V. Physica A, 471, pp 416-426.1602.049.

[34] Hendricks D. An online adaptive learning algorithm for optimal trade execution in high-frequency markets. PhD Thesis (2016).

[35] Hendricks, D., Wilcox, D. 2014. A reinforcement learning extension to the Almgren-Chriss framework for optimal trade execution,in: 2014IEEE Conference on Computational Intelligence for Financial Engineering and Economics (CIFEr), IEEE. pp. 457–464. doi:10.1109/CIFEr.2014. 6924109.

[36] Huang, T.M., Kecman, V., Kopriva, I. 2006. Kernel-Based Algorithms for Mining Huge Data Sets: Supervised, Semi-Supervised, and Unsupervised Learning. Springer, New York.97835403168.

[37] Huberman, G., Stanzl, W. 2004. Price Manipulation and Quasi-Arbitrage, Econometrica, Vol.74, pp 1247-1276.

[38] Jegadeesh, N., Sheridan T. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency, Journal of Finance 48, 65-91.

[39] Jiang Z.Q., Xie W.J., Zhou W.X., Sornette D. 2019. Multifractal analysis of financial markets: a review Reports on Progress in Physics 82:12, pages 125901.

[40] Jones, Charles M., Gautam Kaul, and Marc L. Lipson. 1994. Transactions, volume, and volatility, Review of Financial Studies 7 (Winter): 631–51.

[41] Kl¨ock, F., Schied, A., Sun, Y. 2011. Existence and absence of price manipulation in a market impact model with a dark pool. Available at SSRN 1785409.

[42] Kyle, AS. 1985. Continuous auctions and insider trading. Econometrica, 53(6): 1315–1336, [Crossref], [Web of Science ®], [Google Scholar].

[43] Kullback, S., Leibler, R.A., 1951. On information and sufficiency, Annals of Mathematical Statistics. 22 (1): 79–86. doi:10.121.

[44] Lee, C.M.C., Ready, M.J. 1991. Inferring Trades Direction from Intraday Data, The Journal of Finance, Vol.46, No.2, pp 733-746.

[45] Li, T., Almgren, R. 2016. Option Hedging with Smooth Market Impact, Market Microstructure and Liquidity. 2: 1650002.

[46] Lillo, F., Mantegna, R., Farmer, J.D. 2003. Master Curve for Price-Impact Function, Nature, Vol 421.

[47] Lim, M., Coggins, R. 2005. The immediate price impact of trades on the Australian Stock Exchange, Quantitative Finance, 5:4, 365-377, DOI: 10.1080/14697680500151400.

[48] Maake. W., (2014), AMF Honours Research Proposal, Price Impact and Optimal Execution on the JSE (sup. Prof. D. Wilcox). Program in Advanced Mathematics of Finance

[49] Maake, W., (2014), AMF Project Report, (sup. Prof. D. Wilcox), Program in Advanced Mathematics of Finance.

[50] MAAKE, Witness; VAN ZYL, Terence. Applications of Machine Learning to Estimating the Sizes and Market Impact of Hidden Orders in the BRICS Financial Markets. Journal of Advanced Studies in Finance, [S.l.], v. 11, n. 1, p. 28-64, aug. 2020. ISSN 2068-8393. Available at: <https://journals.aserspublishing.eu/jasf/article/view/5398>. Date accessed: 29 aug. 2020. doi: https://doi.org/10.14505//jasf.v11.1(21).04.

[51] Marquardt, D. 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters, SIAM Journal on Applied Mathematics, Vol. 11, No. pp. 431–441.

[52] Meinshausen, N. 2006. Quantile Regression Forests, Journal of Machine Learning Research, Vol. 7, pp. 983–999.

[53] Mutswari. P., (2014), AMF Honours Research Proposal, Price Impact and Optimal Trading Strategy and Supply / Demand Dynamics, (sup. Prof. D. Wilcox). Program in Advanced Mathematics of Finance

[54] Miccichè, S., Bonanno, G., Lillo, F., and Mantegna, R.N. 2002. Physica A 314 756.

[55] Moro, E., Vicente, J., Moyano, L. G., Gerig, A., Farmer, J. D., Vaglica, G., Lillo, F., and Mantegna, R. N. 2009. Market impact and trading profile of hidden orders in stock markets, Physicalreview. E, Statistical, nonlinear, and soft matter physics 80:066102.

[56] Neter J., Kutner M.H., Nachtstein C.J., Wasserman. 1996. Applied Linear Statistical Models, IRWN.

[57] Nguyen, D., Widrow, B. 1990. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights, Proc. of the Int. J. Conf. on Neural Networks, (3), pp. 21–26.

[58] Niu H., Wang W., Zhang, J. 2019. Recurrence duration statistics and time-dependent intrinsic correlation analysis of trading volumes: A study of Chinese stock indices, Physica A: Statistical Mechanics and its Applications 514, pages 838-854.

[59] Nonyane E., Gebbie T. 2018. Calculating the price response of stocks in emerging markets. MSc dissertation.

[60] Obizhaeva, A., and Jiang W. 2009. Optimal trading strategy and supply/demand dynamics, Working paper, University of Maryland and MIT.

[61] O'Hara, M. 1995. Market Microstructure Theory, Blackwell, Oxford, ISBN 1-55786-443-8. The Review of Financial Studies, Vol. 8, Is. 4, pg 1235.

[62] Patzelt F., Bouchaud J.P. 2017. Universal scaling and nonlinearity of aggregate price impact in financial markets, Capital Fund Management.

[63] Potters, M., Bouchaud, J.P. 2003. More Statistical Properties of order books and price impact, Physica, A,324, 133. Statistical Mechanics and its Applications.

[64] Plerou, V., Gopikrishnan, P., Gabaix, X., Stanley, H.E. 2002. Quantifying Stock Price Response to Demand Fluctuations, Physical Review E, 66, 027104.

[65] Seber, G.A.F. 1977. Linear Regression Analysis, New York: Wiley. 2nd edition.

[66] Vapnik, V. 1995. The Nature of Statistical Learning Theory. Springer, New York.

[67] Wilinski, M., Cui, W., Brabazon, A. 2014. An analysis of price impact functions of individual trades on the London Stock Exchange, 2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), 1-8.

[68] Zhou W.X. 2012. Universal price impact functions of individual trades in an order-driven market, Quantitative Finance 12:8, pages 1253-1263.

Table 10. BOVESPA. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0370	0.0335	1.1	0.3244	0.0300	0.0333	0.9	0.4148
$\ln \omega$	0.1253	0.0157	8.0	3×10^{-07}	0.1294	0.0156	8.3	2×10^{-07}
R^2	0.7793				R^2	0.7922		
Adj.R^2	0.7671				Adj.R^2	0.7807		
MSE	0.0178				MSE	0.0176		
RMSE	0.1331				RMSE	0.1331		
Parameter:	'bias'	0.0370	λ	0.9637	'bias'	0.0300	λ	0.9704
	'slope'	0.1253	α	0.1253	'slope'	0.1294	α	0.1294
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0526	0.0235	2.2	0.0531	0.0662	0.0292	2.3	0.0572
$\ln \omega$	0.0839	0.0113	7.5	1×10^{-07}	0.1114	0.0138	8.1	2×10^{-07}
R^2	0.7565				R^2	0.7834		
Adj.R^2	0.7430				Adj.R^2	0.7713		
MSE	0.0093				MSE	0.0135		
RMSE	0.0947				RMSE	0.1161		
Parameter:	'bias'	0.0526	λ	0.9488	'bias'	0.0662	λ	0.9359
	'slope'	0.0839	α	0.0839	'slope'	0.1114	α	0.1114

Table 11. MOEX. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0680	0.2835	0.1901	0.6305	0.0393	0.2856	0.0836	0.6274
$\ln \omega$	3.9859	0.5480	7.2825	9×10^{-07}	4.0296	0.5519	7.3125	9×10^{-07}
R^2	0.73243				R^2	0.7480		
Adj.R^2	0.7324				Adj.R^2	0.7340		
MSE	1.2804				MSE	1.2995		
RMSE	1.1256				RMSE	1.1256		
Parameter:	'bias'	0.0680	λ	0.9343	'bias'	0.0393	λ	0.9615
	'slope'	3.9859	α	3.9859	'slope'	4.0296	α	4.0296
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.4296	0.0838	5.0638	2×10^{-04}	0.4208	0.0845	4.9204	3×10^{-04}
$\ln \omega$	0.6928	0.1604	4.3376	4×10^{-04}	0.7049	0.1621	4.3613	4×10^{-04}
R^2	0.5107				R^2	0.5136		
Adj.R^2	0.4835				Adj.R^2	0.4866		
MSE	0.1127				MSE	0.1142		
RMSE	0.3331				RMSE	0.3355		
Parameter:	'bias'	0.4296	λ	0.6508	'bias'	0.4208	λ	0.6565
	'slope'	0.6928	α	0.6928	'slope'	0.7049	α	0.7049

Table 12. NSE. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0403	0.0214	1.5269	0.0406	0.0388	0.0269	1.8761	0.1421
$\ln \omega$	0.0619	0.0099	5.9521	2×10^{-04}	0.0798	0.0124	6.3813	7×10^{-06}
R^2	0.6442				R^2	0.6925		
Adj.R^2	0.6244				Adj.R^2	0.6755		
MSE	0.0083				MSE	0.0125		
RMSE	0.0848				RMSE	0.0848		
Parameter:	'bias'	0.0403	λ	0.9605	'bias'	0.0388	λ	0.9619
	'slope'	0.0619	α	0.0619	'slope'	0.0798	α	0.0798
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0086	0.0777	0.1136	0.5886	0.0089	0.0775	0.1171	0.5859
$\ln \omega$	0.2679	0.0359	7.4488	7×10^{-07}	0.2688	0.0358	7.5031	7×10^{-07}
R^2	0.7547				R^2	0.7573		
Adj.R^2	0.7411				Adj.R^2	0.7438		
MSE	0.0951				MSE	0.0945		
RMSE	0.3082				RMSE	0.3074		
Parameter:	'bias'	0.0086	λ	0.9914	'bias'	0.0089	λ	0.9911
	'slope'	0.2679	α	0.2679	'slope'	0.2688	α	0.2688

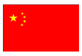


Table 13. SSE. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$									
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS		
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue	
(Int.)	0.1217	0.0521	2.3339	0.0353	0.1229	0.0528	2.3362	0.0342	
$\ln \omega$	0.1833	0.0243	7.5616	7×10^{-07}	0.1848	0.0246	7.5400	8×10^{-07}	
R^2	0.7596				R^2	0.7585			
Adj.R^2	0.7462				Adj.R^2	0.7451			
MSE	0.0430				MSE	0.0441			
RMSE	0.2071				RMSE	0.2071			
Parameter:	'bias'	0.1217	λ	0.8854	'bias'	0.1229	λ	0.8844	
	'slope'	0.1833	α	0.1833	'slope'	0.1848	α	0.1848	
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS		
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue	
(Int.)	0.2696	0.0591	4.5673	3×10^{-04}	0.2695	0.0590	4.5753	3×10^{-04}	
$\ln \omega$	0.1647	0.0277	5.9411	2×10^{-05}	0.1644	0.0277	5.9407	2×10^{-05}	
R^2	0.6600				R^2	0.6599			
Adj.R^2	0.6411				Adj.R^2	0.6410			
MSE	0.0552				MSE	0.0549			
RMSE	0.2347				RMSE	0.2342			
Parameter:	'bias'	0.2669	λ	0.7637	'bias'	0.2695	λ	0.7638	
	'slope'	0.1647	α	0.1647	'slope'	0.1644	α	0.1644	

Table 14. NYSE. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	$5.9{\times}10^{-04}$	$1{\times}10^{-04}$	4.0981	0.0033	$4.6{\times}10^{-04}$	$10{\times}10^{-05}$	4.7138	0.0011
$\ln \omega$	$2{\times}10^{-04}$	$5{\times}10^{-05}$	2.9267	0.0381	$9.7{\times}10^{-05}$	$3.4{\times}10^{-05}$	2.4178	0.0611
R^2	0.4687				R^2	0.3848		
Adj.R^2	0.4097				Adj.R^2	0.3165		
MSE	$2{\times}10^{-07}$				MSE	$10{\times}10^{-08}$		
RMSE	$4{\times}10^{-04}$				RMSE	$4{\times}10^{-04}$		
Parameter:	'bias'	$5.9{\times}10^{-04}$	λ	0.9994	'bias'	$4.6{\times}10^{-04}$	λ	0.9995
	'slope'	$2{\times}10^{-04}$	α	$2{\times}10^{-04}$	'slope'	$9.7{\times}10^{-05}$	α	$9.7{\times}10^{-05}$
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0085	0.0070	4.3114	0.0726	0.0015	$7{\times}10^{-04}$	8.8745	0.0253
$\ln \omega$	0.0080	0.0033	4.1297	0.0092	$2{\times}10^{-04}$	$3{\times}10^{-04}$	4.7595	0.2102
R^2	0.4557				R^2	0.4736		
Adj.R^2	0.4255				Adj.R^2	0.4443		
MSE	0.0029				MSE	$3{\times}10^{-05}$		
RMSE	0.0279				RMSE	0.0029		
Parameter:	'bias'	0.0085	λ	0.9915	'bias'	0.0015	λ	0.9985
	'slope'	0.0080	α	0.0080	'slope'	$2{\times}10^{-04}$	α	$2{\times}10^{-04}$

Table 15. LSE. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$									
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS		
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue	
(Int.)	0.0034	0.0011	3.5506	0.0070	0.0031	0.0011	2.8129	0.0253	
$\ln \omega$	-1$\times10^{-04}$	8$\times10^{-04}$	0.2217	0.6755	3$\times10^{-04}$	0.0011	0.2378	0.8172	
R^2	0.0200				R^2	0.0060			
Adj.R^2	-0.0617				Adj.R^2	-0.0934			
MSE	2$\times10^{-05}$				MSE	1$\times10^{-05}$			
RMSE	0.0035				RMSE	0.0035			
Parameter:	'bias'	0.0034	λ	0.9966	'bias'	0.0031	λ	0.9969	
	'slope'	-1$\times10^{-04}$	α	-1$\times10^{-04}$	'slope'	3$\times10^{-04}$	α	3$\times10^{-04}$	
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS		
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue	
(Int.)	0.0922	0.0913	1.0286	0.3236	0.0651	0.0339	3.7406	0.0629	
$\ln \omega$	0.1884	0.0415	3.8431	0.1142	0.0503	0.0153	2.6831	0.3123	
R^2	0.4273				R^2	0.2758			
Adj.R^2	0.3955				Adj.R^2	0.2356			
MSE	0.1528				MSE	0.0336			
RMSE	0.3637				RMSE	0.1350			
Parameter:	'bias'	0.0922	λ	0.9119	'bias'	0.0651	λ	0.9370	
	'slope'	0.1884	α	0.1884	'slope'	0.0503	α	0.0503	

Table 16. DFM. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0055	5×10^{-04}	10.927	1×10^{-08}	0.1386	0.0755	1.7535	0.1040
$\ln \omega$	4×10^{-04}	2×10^{-04}	1.8552	0.1012	0.1092	0.0369	3.2023	0.0170
R^2	0.1631				R^2	0.3527		
Adj.R^2	0.1166				Adj.R^2	0.3168		
MSE	4×10^{-06}				MSE	0.1027		
RMSE	0.0020				RMSE	0.0020		
Parameter:	'bias'	0.0055	λ	0.9945	'bias'	0.1386	λ	0.8706
	'slope'	4×10^{-04}	α	4×10^{-04}	'slope'	0.1092	α	0.1092
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0054	5×10^{-04}	11.099	2×10^{-08}	0.1422	0.0756	1.8015	0.0955
$\ln \omega$	4×10^{-04}	2×10^{-04}	1.8541	0.1064	0.1080	0.0370	3.1398	0.0184
R^2	0.1636				R^2	0.3435		
Adj.R^2	0.1171				Adj.R^2	0.3070		
MSE	4×10^{-06}				MSE	0.1027		
RMSE	0.0019				RMSE	0.2996		
Parameter:	'bias'	0.0054	λ	0.9946	'bias'	0.1422	λ	0.8674
	'slope'	4×10^{-04}	α	4×10^{-04}	'slope'	0.1080	α	0.1080

Table 17. Euronext Paris. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1232	0.0126	8.1015	2×10^{-04}	0.1247	0.0106	11.3108	6×10^{-07}
$\ln \omega$	0.0123	0.0058	1.4423	0.3892	0.0015	0.0049	0.2675	0.6333
R^2	0.1525				R^2	0.0183		
Adj.R^2	0.1055				Adj.R^2	-0.0363		
MSE	0.0036				MSE	0.0019		
RMSE	0.0502				RMSE	0.0502		
Parameter:	'bias'	0.1232	λ	0.8841	'bias'	0.1247	λ	0.8828
	'slope'	0.0123	α	0.0123	'slope'	0.0015	α	0.0015
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1442	0.0188	7.4450	1×10^{-04}	0.1415	0.0157	8.8577	7×10^{-06}
$\ln \omega$	0.0159	0.0086	1.7530	0.2484	0.0113	0.0072	1.5671	0.1373
R^2	0.1656				R^2	0.1204		
Adj.R^2	0.1192				Adj.R^2	0.0716		
MSE	0.0060				MSE	0.0040		
RMSE	0.0749				RMSE	0.0623		
Parameter:	'bias'	0.1442	λ	0.8657	'bias'	0.1415	λ	0.8681
	'slope'	0.0159	α	0.0159	'slope'	0.0113	α	0.0113

Table 18. SGX. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln\lambda + \alpha \ln\omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0251	0.0037	7.3351	2×10^{-06}	0.0889	0.0461	1.9833	0.0647
$\ln\omega$	0.0011	0.0018	0.9561	0.4054	0.0426	0.0223	1.8587	0.0843
R^2	0.0632				R^2	0.1616		
Adj.R^2	0.0112				Adj.R^2	0.1150		
MSE	3×10^{-04}				MSE	0.0426		
RMSE	0.0146				RMSE	0.0146		
Parameter:	'bias'	0.0251	λ	0.9752	'bias'	0.0889	λ	0.9149
	'slope'	0.0011	α	0.0011	'slope'	0.0426	α	0.0426
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0267	0.0058	5.7533	4×10^{-04}	0.0552	0.0347	1.8050	0.1085
$\ln\omega$	0.0024	0.0028	1.2477	0.2941	0.0353	0.0165	2.0491	0.0695
R^2	0.0898				R^2	0.1904		
Adj.R^2	0.0393				Adj.R^2	0.1454		
MSE	9×10^{-04}				MSE	0.0248		
RMSE	0.0231				RMSE	0.13773		
Parameter:	'bias'	0.0267	λ	0.9737	'bias'	0.0552	λ	0.9463
	'slope'	0.0024	α	0.0024	'slope'	0.0353	α	0.0353

Table 19. BME. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.0701	0.0078	9.3778	3×10^{-07}	0.1138	0.0158	6.1603	6×10^{-05}
$\ln \omega$	0.0167	0.0038	3.2578	0.0266	-0.0024	0.0075	0.6451	0.4694
R^2	0.3593				R^2	0.0643		
Adj.R^2	0.3237				Adj.R^2	0.0123		
MSE	0.0034				MSE	0.0090		
RMSE	0.0311				RMSE	0.0311		
Parameter:	'bias'	0.0701	λ	0.9323	'bias'	0.1138	λ	0.8924
	'slope'	0.0167	α	0.0167	'slope'	-0.0024	α	-0.0024
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1087	0.0163	8.7057	2×10^{-05}	0.1479	0.0244	5.5087	0.0023
$\ln \omega$	0.0498	0.0076	4.8591	0.0034	0.0096	0.0114	0.7802	0.5978
R^2	0.5311				R^2	0.0898		
Adj.R^2	0.5050				Adj.R^2	0.0392		
MSE	0.0101				MSE	0.0116		
RMSE	0.0647				RMSE	0.0972		
Parameter:	'bias'	0.1087	λ	0.8970	'bias'	0.1479	λ	0.8625
	'slope'	0.0498	α	0.0498	'slope'	0.0096	α	0.0096

Table 20. ASX. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.3222	0.0529	6.2474	1×10^{-05}	0.2735	0.0251	11.1602	5×10^{-09}
$\ln \omega$	0.0525	0.0266	1.1616	0.5491	0.0106	0.0100	0.4389	0.3699
R^2	0.1133				R^2	0.1336		
Adj.R^2	0.0640				Adj.R^2	0.0854		
MSE	0.05964				MSE	0.0127		
RMSE	0.2102				RMSE	0.2102		
Parameter:	'bias'	0.3222	λ	0.7246	'bias'	0.2735	λ	0.7607
	'slope'	0.0525	α	0.0525	'slope'	0.0106	α	0.0106
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.6322	0.0712	8.9358	1×10^{-06}	0.6774	0.0486	14.3202	4.9227
$\ln \omega$	0.1715	0.0336	5.112	3×10^{-04}	0.0524	0.0234	2.1440	0.0786
R^2	0.5853				R^2	0.2056		
Adj.R^2	0.5622				Adj.R^2	0.1615		
MSE	0.0809				MSE	0.0434		
RMSE	0.2834				RMSE	0.1929		
Parameter:	'bias'	0.6322	λ	0.5314	'bias'	0.6774	λ	0.5079
	'slope'	0.1715	α	0.1715	'slope'	0.0524	α	0.0524

Table 21. WBAG. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1169	0.00172	5.0213	6×10^{-04}	0.1436	0.0440	2.9934	0.1077
$\ln \omega$	0.0080	0.0076	0.3923	0.4728	0.0174	0.0208	0.8987	0.4128
R^2	0.0409				R^2	0.0489		
Adj.R^2	-0.0124				Adj.R^2	-0.0040		
MSE	0.0165				MSE	0.0416		
RMSE	0.0689				RMSE	0.0689		
Parameter:	'bias'	0.1169	λ	0.8897	'bias'	0.1436	λ	0.8662
	'slope'	0.0080	α	0.0080	'slope'	0.0174	α	0.0174
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1181	0.0176	5.3787	2×10^{-04}	0.1409	0.0444	2.9997	0.1258
$\ln \omega$	0.0085	0.0081	0.4754	0.4690	0.0148	0.0220	0.6520	0.5258
R^2	0.0406				R^2	0.0238		
Adj.R^2	-0.0127				Adj.R^2	-0.0305		
MSE	0.0169				MSE	0.0415		
RMSE	0.0701				RMSE	0.1752		
Parameter:	'bias'	0.1181	λ	0.8886	'bias'	0.1409	λ	0.8686
	'slope'	0.0085	α	0.0085	'slope'	0.0148	α	0.0148

Table 22. Euronext Lisbon. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.6979	0.0844	9.2029	5×10^{-06}	0.4832	0.0651	7.4000	5×10^{-06}
$\ln \omega$	0.0739	0.0391	2.2305	0.01670	0.0319	0.0291	1.0900	0.2999
R^2	0.2180				R^2	0.0635		
Adj.R^2	0.1745				Adj.R^2	0.0115		
MSE	0.1589				MSE	0.0903		
RMSE	0.3357				RMSE	0.3357		
Parameter:	'bias'	0.6979	λ	0.4976	'bias'	0.4832	λ	0.6168
	'slope'	0.0739	α	0.0739	'slope'	0.0319	α	0.0319
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.7839	0.1021	8.0701	4×10^{-06}	0.5348	0.0787	7.2873	4×10^{-04}
$\ln \omega$	0.1133	0.0488	2.4131	0.0701	0.0515	0.0378	1.3617	0.1901
R^2	0.2472				R^2	0.0934		
Adj.R^2	0.2054				Adj.R^2	0.0430		
MSE	0.2246				MSE	0.1316		
RMSE	0.4041				RMSE	0.3120		
Parameter:	'bias'	0.7839	λ	0.4566	'bias'	0.5348	λ	0.5858
	'slope'	0.1133	α	0.1133	'slope'	0.0515	α	0.0515

Table 23. Euronext Brussels. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.3427	0.0479	6.9866	0.0593	0.3604	0.0317	10.2461	0.0730
$\ln \omega$	0.0570	0.0226	2.6451	0.0753	0.0183	0.0149	1.7091	0.2154
R^2	0.2770				R^2	0.1558		
Adj.R^2	0.2368				Adj.R^2	0.1089		
MSE	0.0398				MSE	0.0237		
RMSE	0.1898				RMSE	0.1898		
Parameter:	'bias'	0.3427	λ	0.7099	'bias'	0.3604	λ	0.6974
	'slope'	0.0570	α	0.0570	'slope'	0.0183	α	0.0183
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.2851	0.0348	6.5703	0.0275	0.3259	0.0228	15.8390	0.0187
$\ln \omega$	0.0520	0.0162	2.9809	0.0171	0.0206	0.0112	1.7786	0.1275
R^2	0.3274				R^2	0.1541		
Adj.R^2	0.2900				Adj.R^2	0.1071		
MSE	0.0221				MSE	0.0.0124		
RMSE	0.1380				RMSE	0.0.0903		
Parameter:	'bias'	0.2851	λ	0.7519	'bias'	0.3259	λ	0.7219
	'slope'	0.0520	α	0.0520	'slope'	0.0206	α	0.0206

Table 24. OSE. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.3031	0.1994	1.2246	0.3255	0.0499	0.0284	1.3108	0.2886
$\ln \omega$	0.0599	0.2433	0.2925	0.7877	0.1653	0.0218	2.6756	0.5715
R^2	0.0266				R^2	0.3204		
Adj.R^2	-.2168				Adj.R^2	0.1505		
MSE	0.2010				MSE	0.0115		
RMSE	0.3015				RMSE	0.3015		
Parameter:	'bias'	0.3031	λ	0.7385	'bias'	0.0499	λ	0.9513
	'slope'	0.0080	α	0.0080	'slope'	0.0174	α	0.0174
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.4637	0.0595	8.1198	1×10^{-05}	0.6255	0.0719	8.3655	3×10^{-06}
$\ln \omega$	0.1012	0.0278	3.7484	0.0268	0.0313	0.0358	0.5564	0.3259
R^2	0.4248				R^2	0.0687		
Adj.R^2	0.3929				Adj.R^2	0.0169		
MSE	0.0585				MSE	0.0928		
RMSE	0.2363				RMSE	0.2836		
Parameter:	'bias'	0.4637	λ	0.6290	'bias'	0.6255	λ	0.5350
	'slope'	0.1012	α	0.1012	'slope'	0.0313	α	0.0313

Table 25. OMXH. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln\Delta P \sim -\ln\lambda + \alpha\ln\omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1497	0.0221	6.4892	2×10^{-04}	0.1671	0.0207	8.0980	3×10^{-06}
$\ln\omega$	0.0166	0.0104	1.3827	0.2058	-0.0065	0.0096	-0.2277	0.5800
R^2	0.0991				R^2	0.0208		
Adj.R^2	0.0491				Adj.R^2	-0.0336		
MSE	0.0117				MSE	0.0129		
RMSE	0.0873				RMSE	0.0873		
Parameter:	'bias'	0.1497	λ	0.8610	'bias'	0.1671	λ	0.8461
	'slope'	0.0166	α	0.0166	'slope'	-0.0065	α	-0.0065
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1278	0.0116	9.3543	7×10^{-07}	0.1268	0.0200	7.5959	7×10^{-06}
$\ln\omega$	0.0180	0.0051	2.0254	0.1914	-0.0014	0.0092	0.5136	0.5457
R^2	0.2422				R^2	0.0529		
Adj.R^2	0.2001				Adj.R^2	3×10^{-04}		
MSE	0.0033				MSE	0.0109		
RMSE	0.0463				RMSE	0.0794		
Parameter:	'bias'	0.1278	λ	0.8800	'bias'	0.1268	λ	0.8809
	'slope'	0.0180	α	0.0180	'slope'	-0.0014	α	-0.0014

Table 26. FWB. Curve fitting results of the Lillo et al.(2003) average market impact function on the seller and buyer initiated re-scaled aggregated averaged price shifts and re-scaled aggregated averaged normalized volume, before and after the inclusion of hidden orders.

Linear Regression Model: $\ln \Delta P \sim -\ln \lambda + \alpha \ln \omega$								
SELLER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1744	0.0426	4.2820	0.0011	0.2524	0.0233	10.7262	1×10^{-08}
$\ln \omega$	0.0727	0.0203	3.8099	0.0038	-4×10^{-04}	0.0109	-0.05367	0.7852
R^2	0.4409				R^2	0.0055		
Adj.R^2	0.4098				Adj.R^2	-0.0497		
MSE	0.0349				MSE	0.0098		
RMSE	0.1690				RMSE	0.1690		
Parameter:	'bias'	0.1744	λ	0.8400	'bias'	0.2524	λ	0.7769
	'slope'	0.0727	α	0.0727	'slope'	-4×10^{-04}	α	-4×10^{-04}
BUYER	VISIBLE	ORDERS			VISIBLE	&HIDDEN	ORDERS	
Coefficient	Estimate	SE	tStat	pValue	Estimate	SE	tStat	pValue
(Int.)	0.1166	0.0212	5.5100	4×10^{-05}	0.2364	0.0210	10.7801	1×10^{-08}
$\ln \omega$	0.0265	0.0103	2.4878	0.0330	-0.0117	0.0098	-1.1133	0.3094
R^2	0.2552				R^2	0.0690		
Adj.R^2	0.2138				Adj.R^2	0.0173		
MSE	0.0083				MSE	0.0077		
RMSE	0.0840				RMSE	0.0836		
Parameter:	'bias'	0.1166	λ	0.8899	'bias'	0.1409	λ	0.7895
	'slope'	0.0265	α	0.0265	'slope'	-0.0117	α	-0.0117

Table 27. BOVESPA. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	NaN	0.6359	0.5868	0.6325	0.6358	0.7654	0.7527	0.7162	$1{\times}10^{+03}$
ω_{GLM}	NaN	0.0703	0.0611	0.0603	0.0555	0.0640	0.0581	0.0584	18.0428
ω_{ANN}	NaN	1.0739	1.1450	1.1327	1.0670	1.0127	0.9890	0.9618	0.8904
ω_{SVM}	NaN	0.9431	0.8427	0.8742	0.8679	0.8891	0.9755	0.9175	27.9336
ω_{RF}	NaN	1.6793	1.1072	0.9475	0.8878	0.9954	1.0279	0.0263	39.1540
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
$3{\times}10^{+05}$		$3{\times}10^{+05}$		$3{\times}10^{+05}$		$3{\times}10^{+05}$			

Table 28. MOEX. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	3.7667	1.0523	1.3113	1.4694	1.1473	1.8118	1.0801	1.2188	1.4412
ω_{GLM}	0	0.1692	0.1662	0.1540	0.1603	0.6326	0.1668	0.1753	0.1888
ω_{ANN}	1	1.1484	1.0245	0.9981	0.9960	1.05376	1.0083	0.9968	0.9956
ω_{SVM}	1.2608	1.3002	1.4621	1.4175	1.4862	1.5313	1.3117	1.3908	1.4865
ω_{RF}	1.1306	1.0974	1.1105	1.0670	1.0520	1.1464	1.0107	1.0135	1.0347
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		0.6495		0.6225		0.6845			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
3.8765		2.6997		3.6004		3.1195			

Table 29. NSE. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	1.0421	1.0437	1.0118	0.9878	0.9767	1.0615	1.5309	NaN	NaN
ω_{GLM}	0	0	0	0	0	0	0	NaN	NaN
ω_{ANN}	1.0584	1.0392	1.0156	1.0608	1.0309	1.0715	1.3950	NaN	NaN
ω_{SVM}	2.6949	5.1757	3.6874	5.22348	3.1673	3.5226	3.2400	NaN	NaN
ω_{RF}	0.9990	1.0264	1.0116	1.0136	1.0274	1.0449	1.2511	NaN	NaN
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
0.5979		0.0017		15.8741		0.0454			

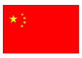


Table 30. SSE. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.7825	0.8758	0.9306	NaN	0.9076	1.0402	123.5415	NaN	NaN
ω_{GLM}	0.8942	0.9310	0.9356	NaN	0.9512	1.0198	96.2742	NaN	NaN
ω_{ANN}	1.0370	1.0355	1.0400	NaN	1.0455	1.0408	2.3583	NaN	NaN
ω_{SVM}	0.8236	0.9653	1.0061	NaN	1.0331	1.1582	14.8148	NaN	NaN
ω_{RF}	0.8787	0.9466	0.9727	NaN	0.9510	1.0568	22.0267	NaN	NaN
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
$1\times10^{+03}$		$1\times10^{+04}$		$1\times10^{+04}$		$10\times10^{+03}$			

Table 31. NYSE. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	1.1975	0.8837	0.9218	0.8532	0.8410	0.8724	0.7754	23.9505	15.9322
ω_{GLM}	0.1072	0.0841	0.0628	0.0756	0.0659	0.1055	0.1371	3.2578	3.0166
ω_{ANN}	1.0746	1.0059	1.0332	1.0032	0.9951	0.9925	1.0405	1.4458	1.6812
ω_{SVM}	1.4452	0.9318	0.9517	0.8898	0.8677	0.9118	0.7870	25.8093	28.0583
ω_{RF}	1.1453	1.1607	1.0851	1.0855	1.0918	1.1167	1.1852	2.9330	2.1734
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		0.7311		0.9166		0.6801			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
$3\times10^{+03}$		$3\times10^{+03}$		431.7919		$3\times10^{+03}$			

Table 32. LSE. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	1.0381	1.0357	1.0236	0.9406	0.9274	0.9856	1.0215	1.7441	2.6483
ω_{GLM}	0.8749	0.7480	0.7548	0.7528	0.7318	0.6991	0.8277	0.9642	0.4904
ω_{ANN}	4.9865	5.0059	4.9690	5.0249	5.0078	5.0998	5.0132	5.0365	5.2856
ω_{SVM}	0.9253	0.9072	0.9477	0.8321	0.8156	0.8498	0.9703	1.6134	0.6420
ω_{RF}	1.1560	1.1608	1.1478	1.1318	1.1322	1.1237	1.1355	1.1602	1.2231
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		0.7446		0.9249		0.7579			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
0.7026		16.2066		0.0347		0.1041			

Table 33. DFM. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	NaN	0.8669	0.8048	0.8732	1.0975	1.3405	NaN	NaN	NaN
ω_{GLM}	NaN	0.9297	0.9398	1.0073	1.1338	1.4340	NaN	NaN	NaN
ω_{ANN}	NaN	1.4719	1.3018	1.3611	1.5033	1.1890	NaN	NaN	NaN
ω_{SVM}	NaN	0.9084	0.9162	0.8466	0.9478	0.9041	NaN	NaN	NaN
ω_{RF}	NaN	0.7893	0.7980	0.7991	0.9117	1.3951	NaN	NaN	NaN
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
5×10^{-05}		0.1103		0.0922		0.0215			

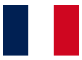


Table 34. Euronext Paris. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.9711	0.8879	0.9254	0.8246	0.9119	0.9819	0.9522	0.9857	1.4307
ω_{GLM}	0.8105	0.7948	0.8994	0.7597	0.8414	0.8762	0.7807	0.8780	0.9728
ω_{ANN}	1.0367	1.0203	1.0238	1.0323	1.0115	1.0500	0.9987	0.9939	1.0086
ω_{SVM}	1.4610	1.8383	1.4386	1.3021	1.3279	1.7068	1.4809	2.7955	1.3204
ω_{RF}	1.0905	0.8336	0.9226	0.8013	0.8759	0.9140	0.8706	8774	1.1621
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
3.9254		3.4778		3.1706		2.7590			

Table 35. SGX. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.8004	0.9718	0.9190	2.6227	1.0015	0.9727	0.9514	0.9022	1.2545
ω_{GLM}	0.8855	0.9011	0.9174	2.6122	0.9306	1.0454	1.0595	0.8486	1.2742
ω_{ANN}	1.8813	1.7515	1.9368	2.0797	1.5133	1.7326	1.7932	1.8547	1.9370
ω_{SVM}	0.8546	1.0024	1.1005	2.0385	0.9843	1.0032	1.0746	0.9604	0.9364
ω_{RF}	0.8088	0.7837	0.8246	1.2866	0.8110	8673	0.8525	0.7909	1.0080
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
1		0.1337		0.9326		0.3442			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
4×10^{-06}		2.2279		0.7669		1.7554			

Table 36. BME. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.6600	0.7573	0.8123	0.8375	0.7164	0.7370	0.6565	0.8579	1.4960
ω_{GLM}	0.7357	0.8418	0.8762	0.8735	0.7017	0.7667	0.7077	0.9770	1.5306
ω_{ANN}	0.8688	1.0381	0.8335	1.0529	0.9254	0.9568	0.9244	0.8749	1.0657
ω_{SVM}	0.9515	0.8989	1.0051	0.8874	0.8241	1.0154	1.2297	0.6932	1.3317
ω_{RF}	1.1418	1.0434	1.0301	1.0482	1.0401	1.1230	1.1243	1.2814	1.3564
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
0.0947		0.0073		0.0097		2×10^{-04}			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
NaN		NaN		NaN		NaN			

Table 37. ASX. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	1.7862	0.8213	0.7144	0.8090	0.9490	0.9228	0.9007	82.4236	0.5926
ω_{GLM}	1.0697	0.6197	0.5079	0.4929	0.4287	0.4972	0.3372	20.5444	0.5721
ω_{ANN}	1.1466	1.0080	0.9945	1.2254	1.0264	0.9958	0.9994	1.1225	1.0178
ω_{SVM}	2.1154	0.7473	0.8350	0.8129	1.0569	1.0347	1.0488	11.1213	0.2407
ω_{RF}	2.0934	1.5767	1.5829	1.5093	1.5508	1.5419	1.4492	3.1994	1.6401
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
3×10^{-05}		0.0420		0.0574		0.4283			

Table 38. WBAG. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.9750	1.0937	0.8088	0.8305	0.8096	1.8963	0.9374	0.9522	0.8649
ω_{GLM}	0.7489	0.7068	0.7436	0.8403	0.7841	1.8089	0.9142	0.9166	0.8962
ω_{ANN}	5.2773	5.2773	5.4895	5.4858	5.5387	5.2321	4.8819	5.0642	5.0128
ω_{SVM}	0.8977	0.9077	0.9362	0.7508	0.8329	1.6162	0.6804	0.7417	0.5574
ω_{RF}	1.0344	1.0913	1.0068	0.9945	1.0121	1.1576	0.9563	1.0011	0.9250
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
1		0.9579		0.9016		0.9653			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
1×10^{-05}		12.4905		0.0801		0.0206			

Table 39. Euronext Lisbon. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.7322	1.3275	NaN	NaN	NaN	NaN	NaN	21.1138	NaN
ω_{GLM}	0.3970	0.1170	NaN	NaN	NaN	NaN	NaN	12.7292	NaN
ω_{ANN}	0.9727	0.9905	NaN	NaN	NaN	NaN	NaN	1	NaN
ω_{SVM}	0.9958	0.9646	NaN	NaN	NaN	NaN	NaN	3.4542	NaN
ω_{RF}	1.5957	1.8306	NaN	NaN	NaN	NaN	NaN	3.5920	NaN
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
0.0518		0.0093		0.0113		0.4024			

Table 40. Euronext Brussels. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	1.0499	1.0785	1.1399	0.8126	0.9511	0.7488	0.7991	0.8309	4.3099
ω_{GLM}	0.7292	0.7969	0.6137	0.6925	0.7455	0.7324	0.6691	0.6854	0.8568
ω_{ANN}	1.2339	1.28021	1.2359	1.1092	1.1608	1.1722	1.1209	1.0240	1.3828
ω_{SVM}	1.2930	1.6101	0.9840	1.3962	1.3627	1.1936	1.3346	1.2936	1.1956
ω_{RF}	0.9867	1.0306	0.9816	0.8623	0.9918	0.8712	0.8995	0.8853	0.8970
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
1		0.7288		0.8943		0.6146			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
7×10^{-04}		0.9175		0.4878		0.2754			

Table 41. OSE. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.8829	0.7998	.9043	0.7580	1.2921	1.0081	0.8638	1.5866	NaN
ω_{GLM}	0.6634	0.7113	0.6009	0.6597	0.5656	0.6179	0.5952	0.8069	NaN
ω_{ANN}	2.6566	2.5134	2.2011	2.5734	2.2842	2.1707	2.4475	2.6340	NaN
ω_{SVM}	0.9081	1.1969	1.1336	1.5802	1.3994	1.8552	5.3012	1.0936	NaN
ω_{RF}	0.9443	0.9704	1.0214	0.9388	1.1308	1.0728	1.0201	1.0551	NaN
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
NaN		NaN		NaN		NaN			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
0.0958		17.4258		13.8787		0.0277			

Table 42. OMXH. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	1.3104	0.8851	0.8657	0.9656	0.9221	0.8638	0.7740	0.9333	0.6942
ω_{GLM}	0.7074	0.9140	0.8029	0.9629	0.8108	0.7636	0.7299	0.7587	0.6182
ω_{ANN}	1.2095	1.1231	1.0773	1.1028	1.0782	1.0196	1.0734	1.0778	1.0318
ω_{SVM}	0.6004	1.0006	0.9344	1.2090	1.0225	0.7965	0.7808	0.9227	0.9615
ω_{RF}	1.2957	1.2979	1.2843	1.3847	1.3876	1.3548	1.3037	1.4338	1.3339
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
1		0.9513		0.7377		0.9741			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
0.0105		0.0573		0.0821		0.1095			

Table 43. FWB. Hourly aggregated averaged true transaction sizes, and GLM, ANN, SVM and RF estimated normalized transaction sizes of all the stocks.

All firms									
Time (h)	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
ω_{TRUE}	0.9963	0.9723	0.9700	1.0346	0.9555	0.9564	0.9234	0.9684	0.9766
ω_{GLM}	0.6876	0.7433	0.7080	0.7046	0.6993	0.7033	0.6968	0.7216	0.8135
ω_{ANN}	1.0020	1.0111	1.0203	1.0384	1.0188	1.0171	1.0374	1.0190	1.0236
ω_{SVM}	0.9549	1.0873	0.9201	0.9507	1.0521	0.9544	0.9468	1.0056	0.8759
ω_{RF}	1.0556	1.0975	1.0464	1.0692	1.0410	1.0271	1.0403	1.0250	1.0788
R-Squared									
R-Squared$_{GLM}$		R-Squared$_{ANN}$		R-Squared$_{SVM}$		R-Squared$_{RF}$			
1		0.9728		0.8428		0.9930			
Mean Squared Error									
MSE$_{GLM}$		MSE$_{ANN}$		MSE$_{SVM}$		MSE$_{RF}$			
3×10^{-33}		0.0060		0.0257		0.0466			

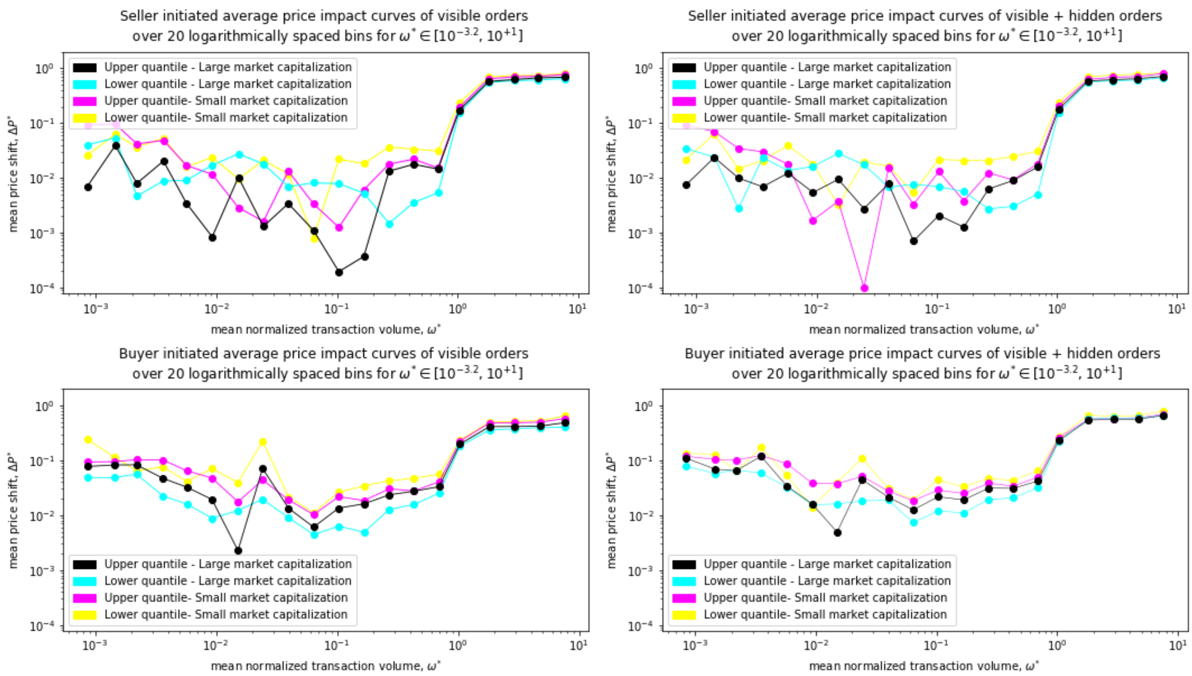


Figure 20. BOVESPA Aggregated price impact curves of single trades of 42 liquid firms. The data composition consists of a total of 6, 572, 183 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

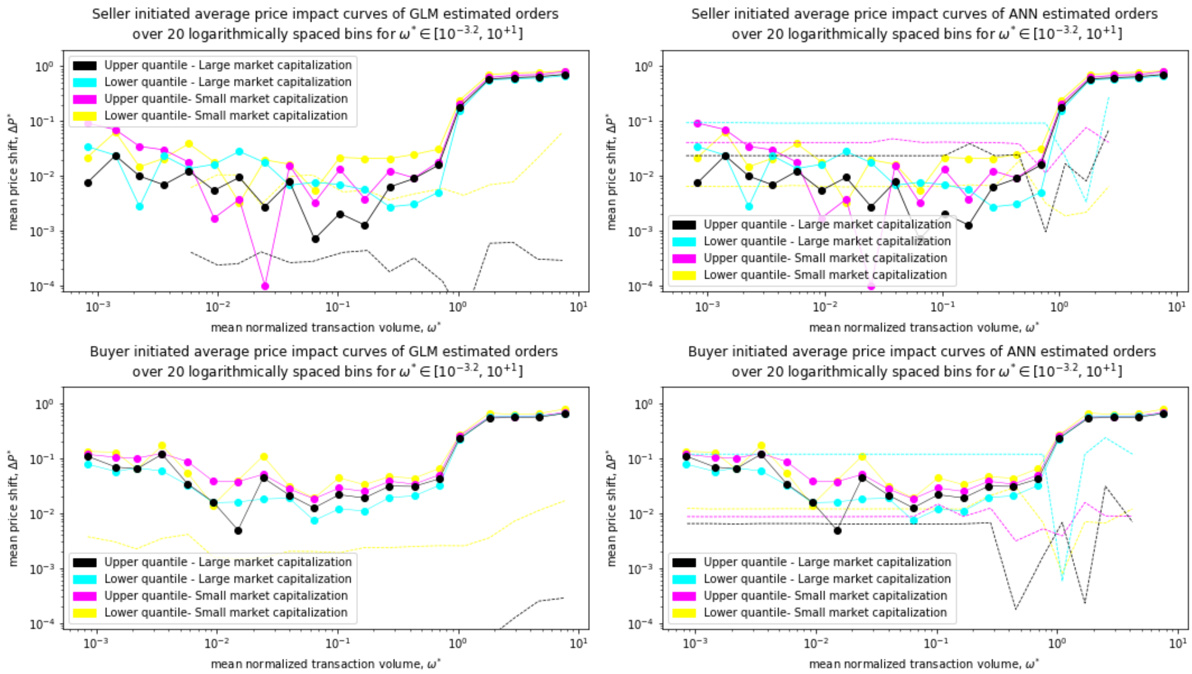


Figure 21. BOVESPA. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

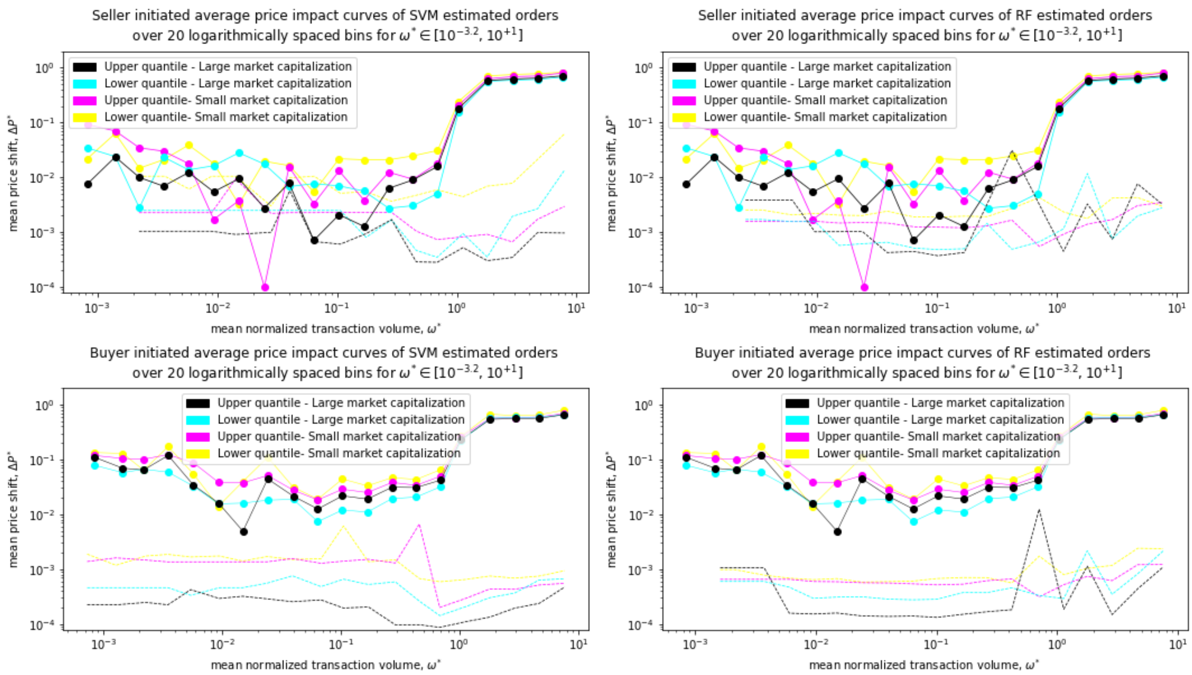


Figure 22. BOVESPA. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

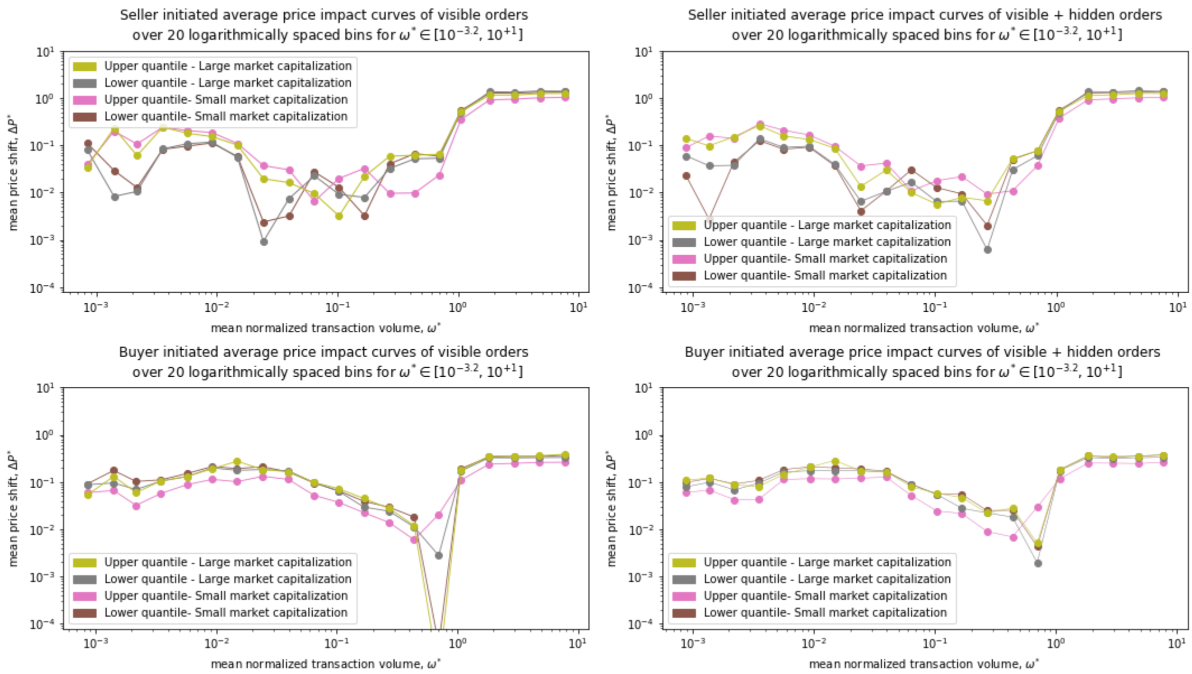


Figure 23. MOEX. Aggregated price impact curves of single trades of 30 liquid firms. The data composition consists of a total of 3, 091, 547 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

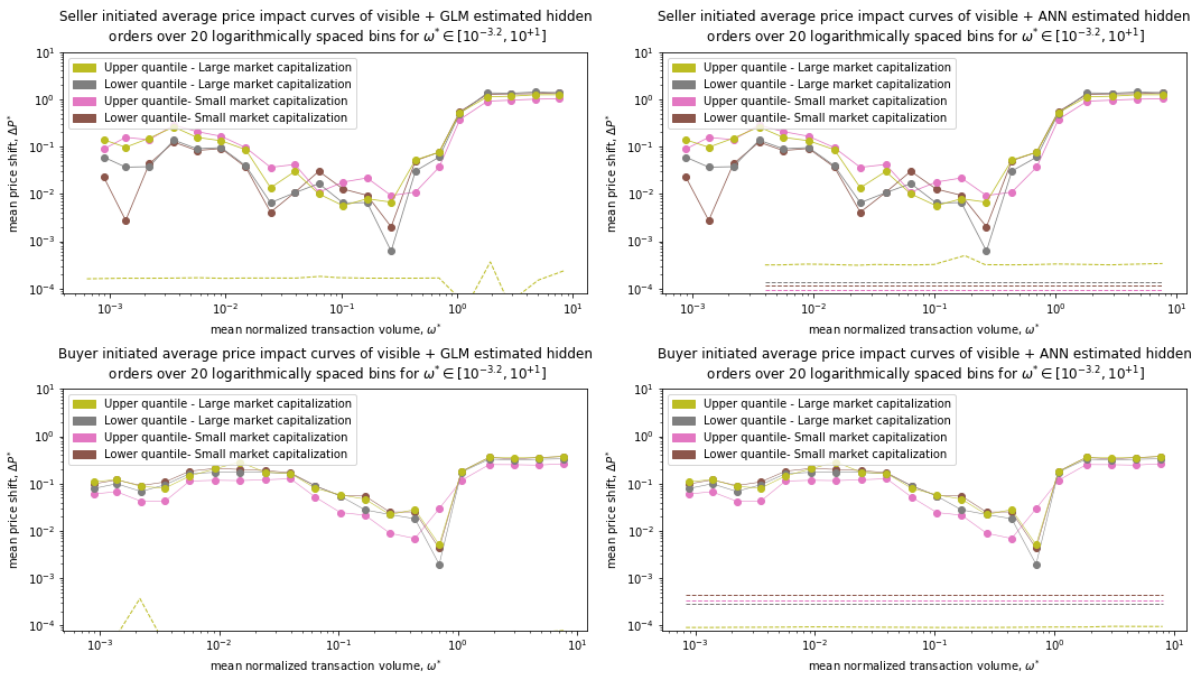


Figure 24. MOEX. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

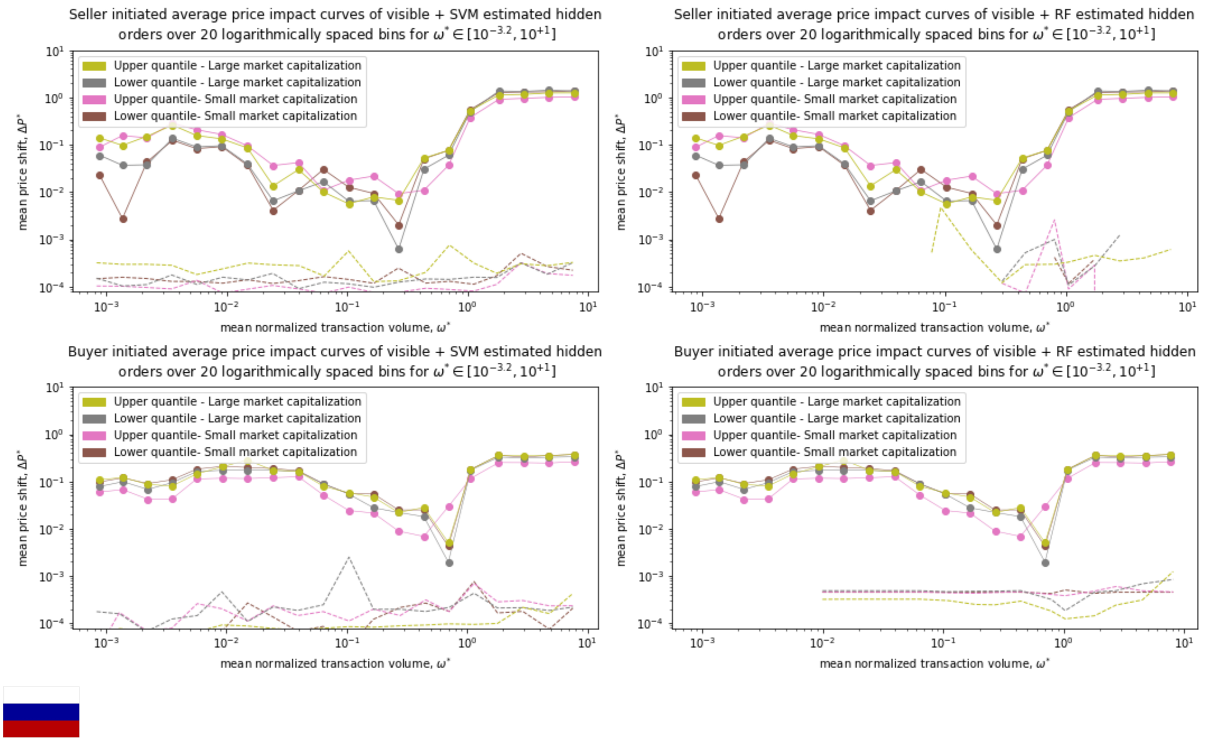


Figure 25. MOEX. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

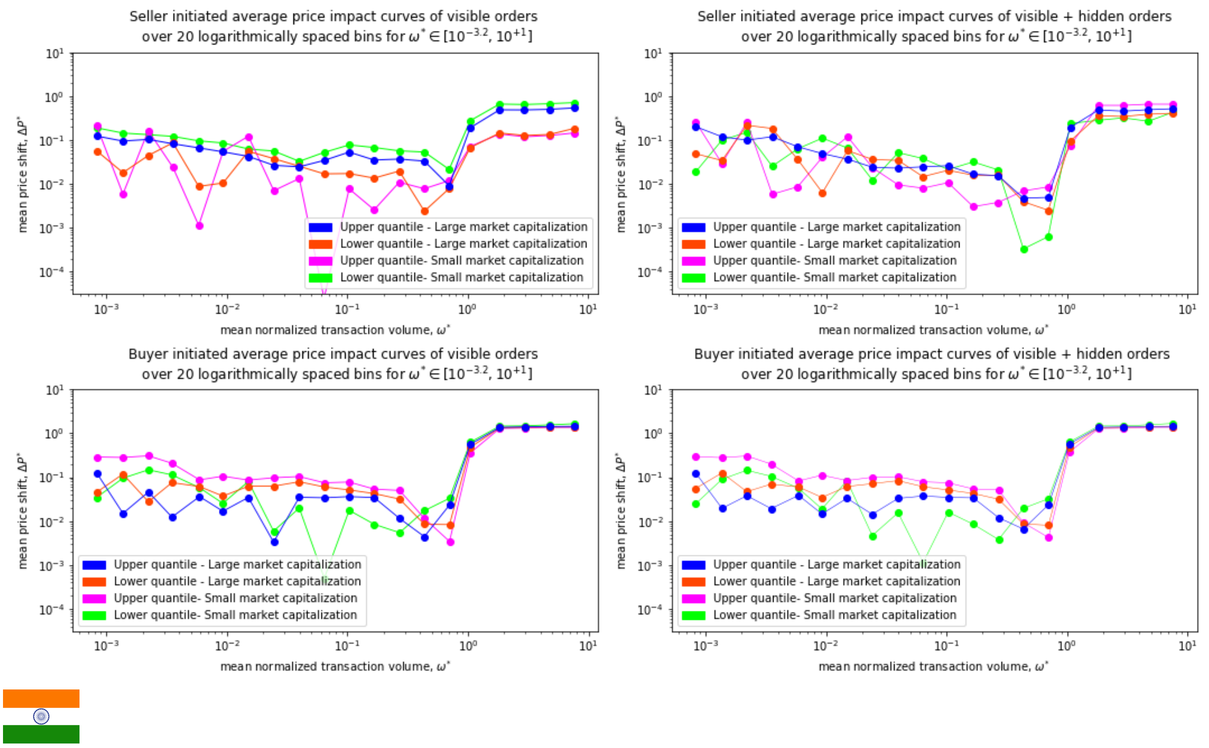


Figure 26. NSE. Aggregated price impact curves of single trades of 43 liquid firms. The data composition consists of a total of 4, 095, 817 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

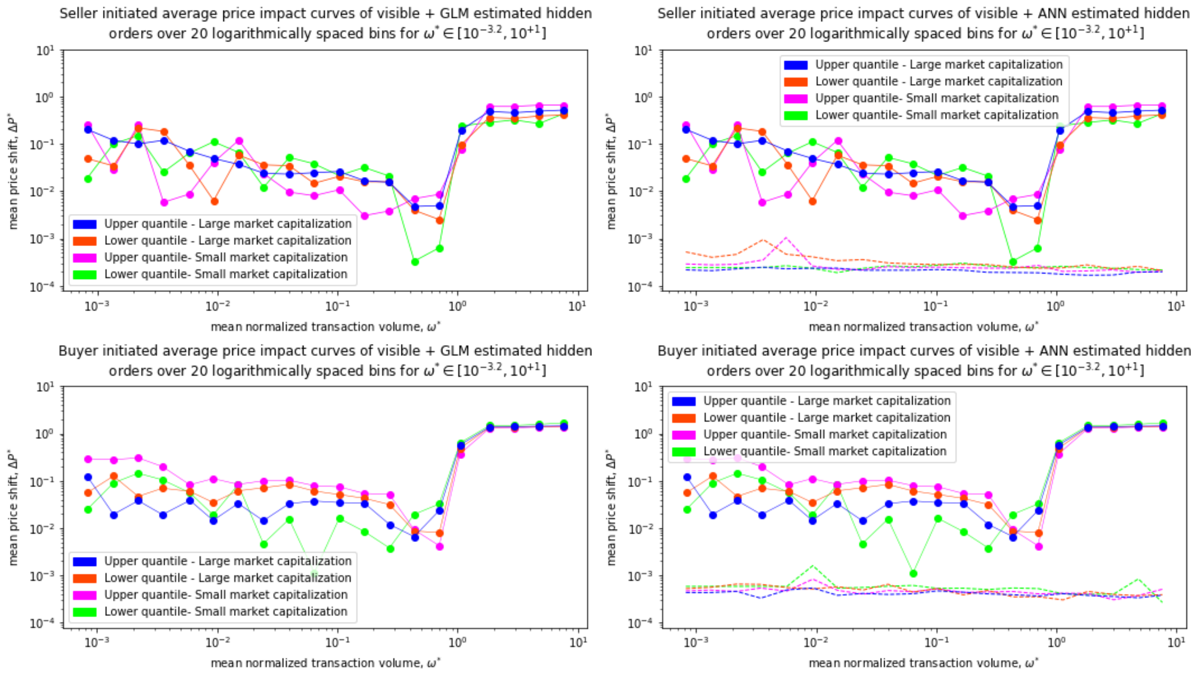


Figure 27. NSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

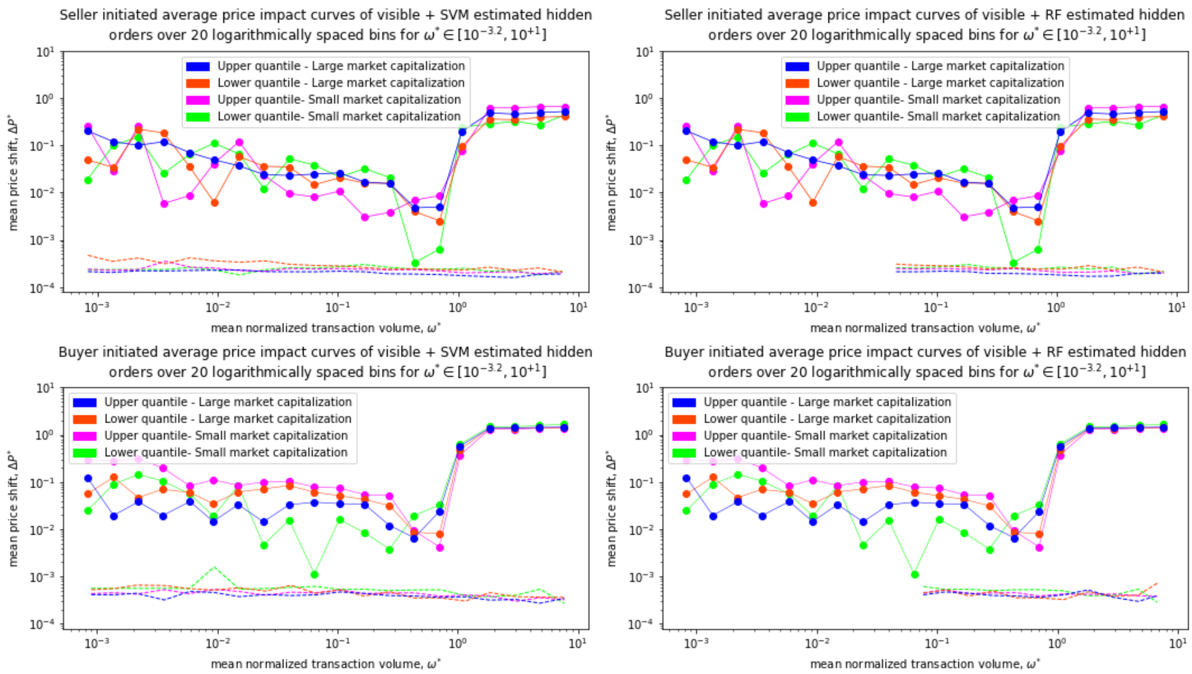


Figure 28. NSE Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

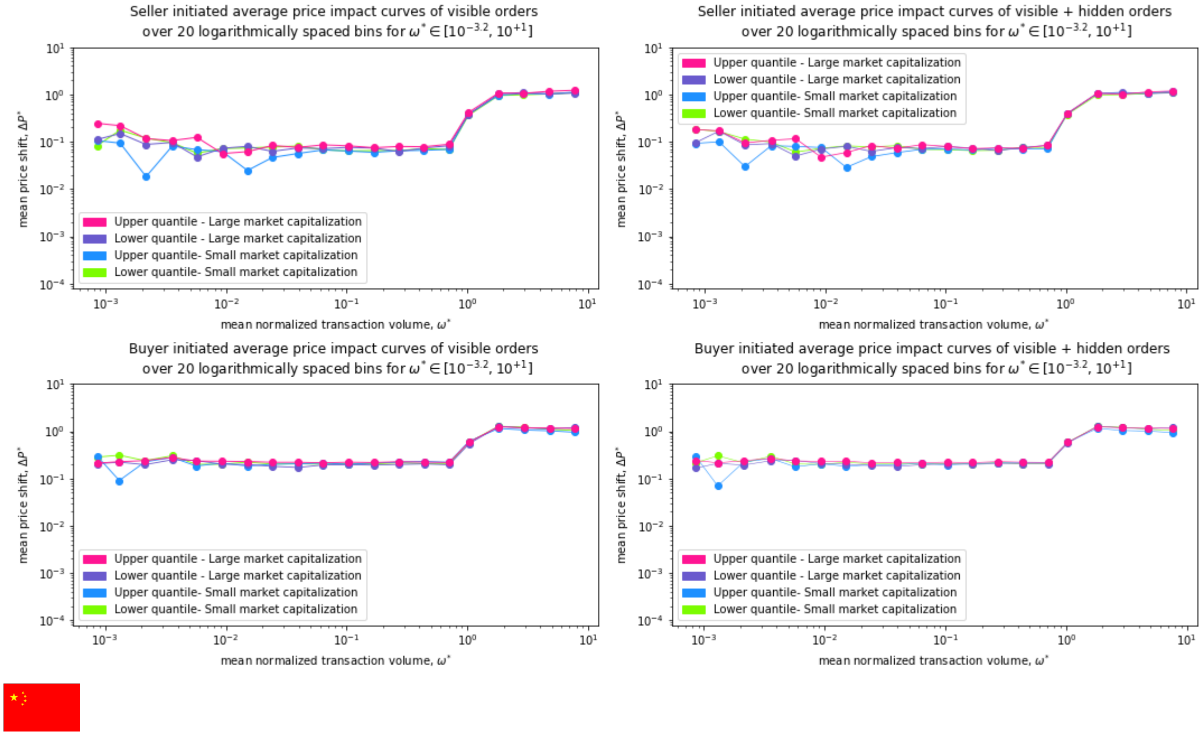


Figure 29. SSE. Aggregated price impact curves of single trades of 54 liquid firms. The data composition consists of a total of 5, 226, 211 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

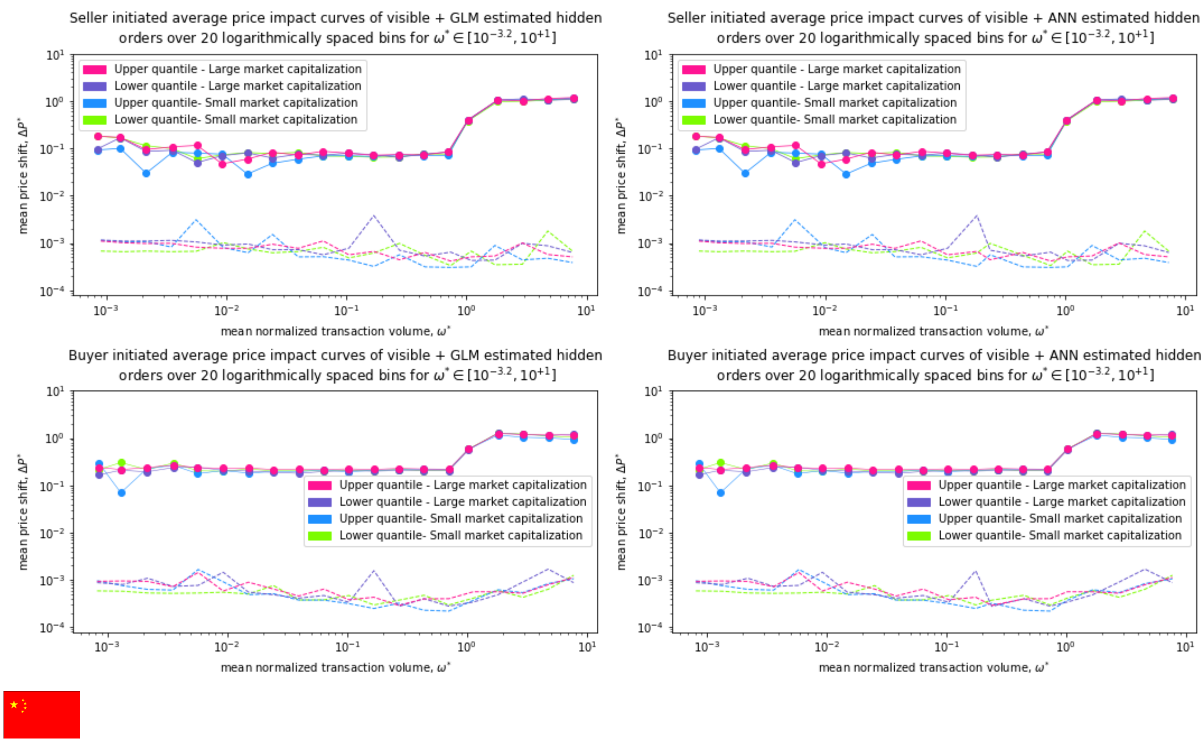


Figure 30. SSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

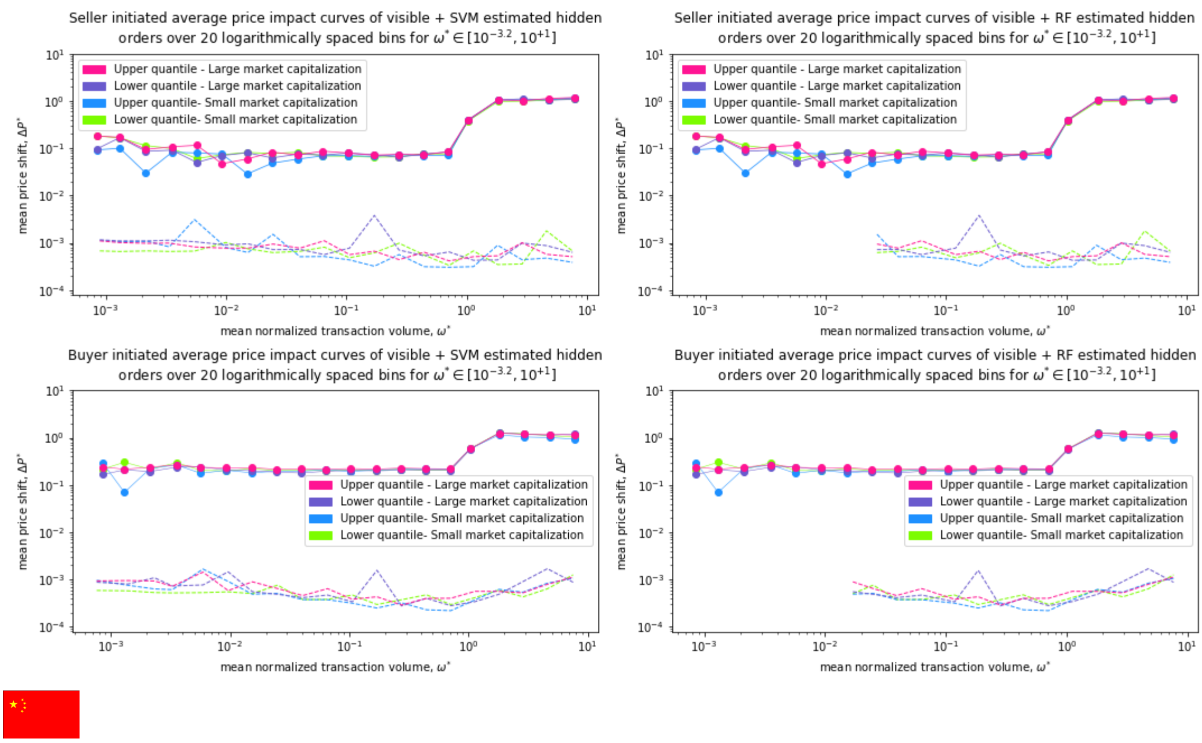


Figure 31. SSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

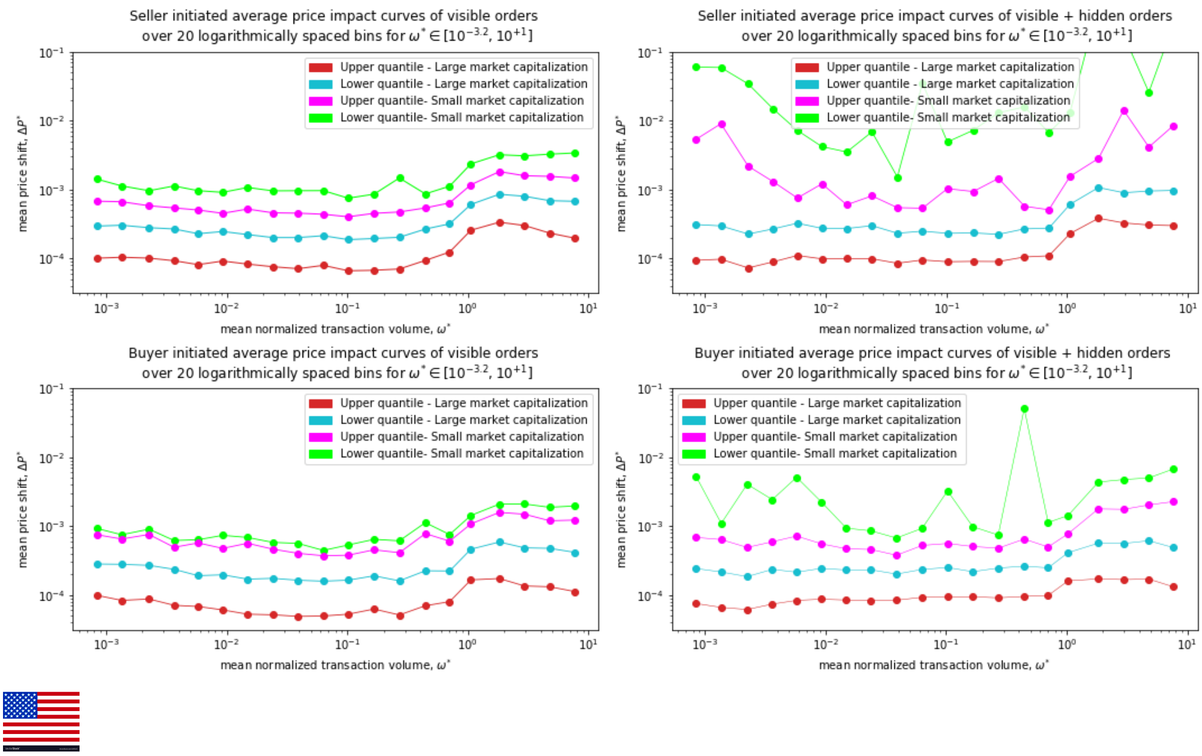


Figure 32. NYSE. Aggregated price impact curves of single trades of 442 liquid firms. The data composition consists of a total of 29, 404, 122 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

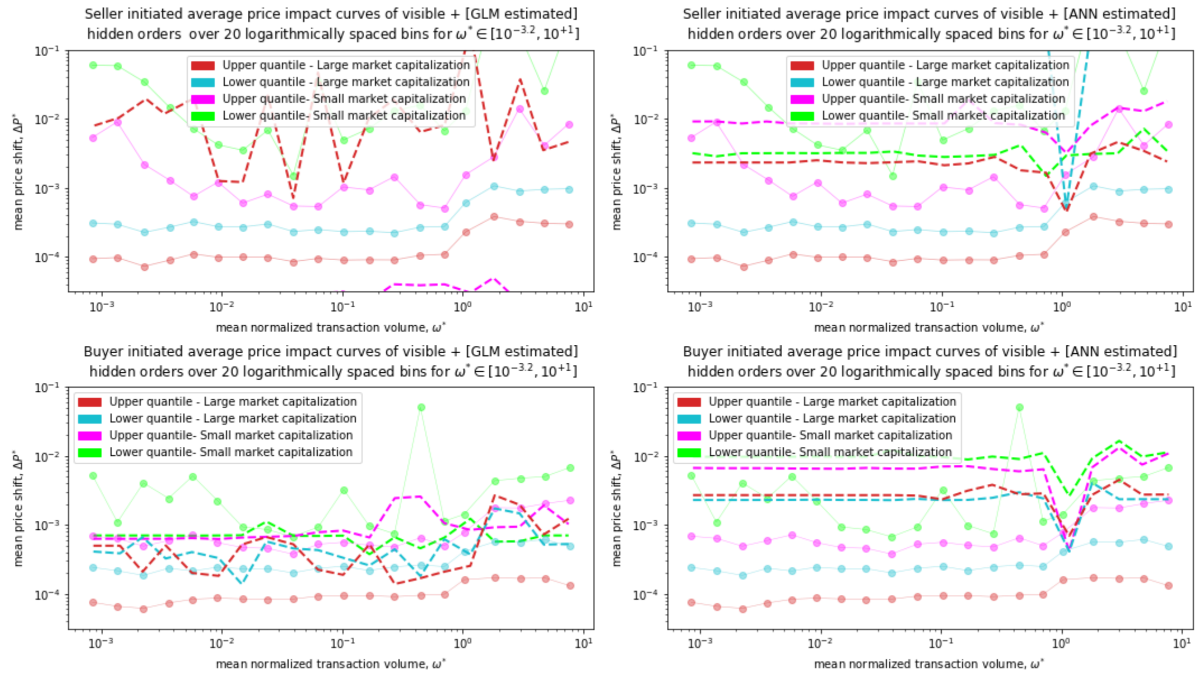


Figure 33. NYSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

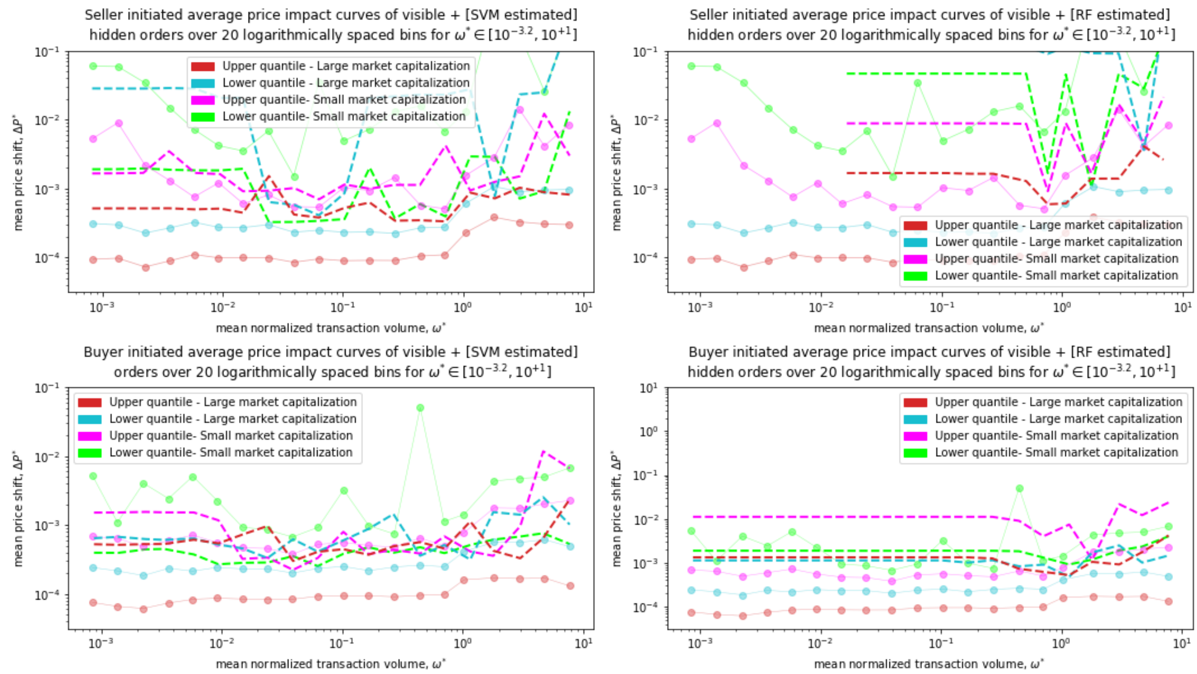


Figure 34. NYSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

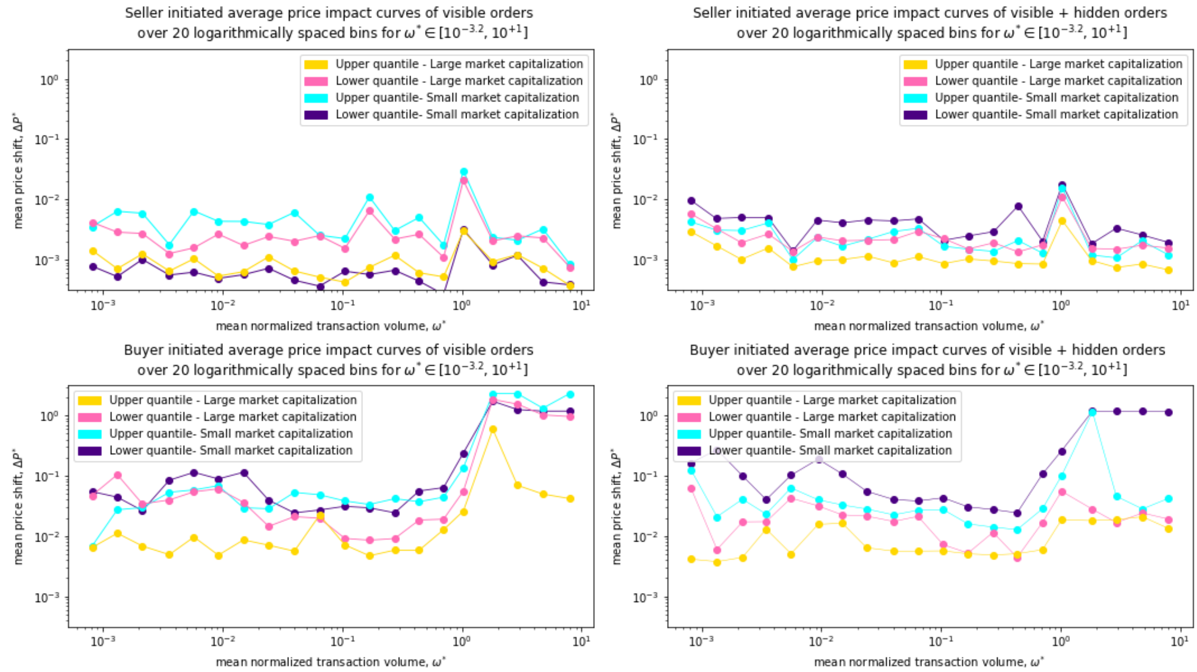


Figure 35. LSE. Aggregated price impact curves of single trades of 21 liquid firms. The data composition consists of a total of 577, 464 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

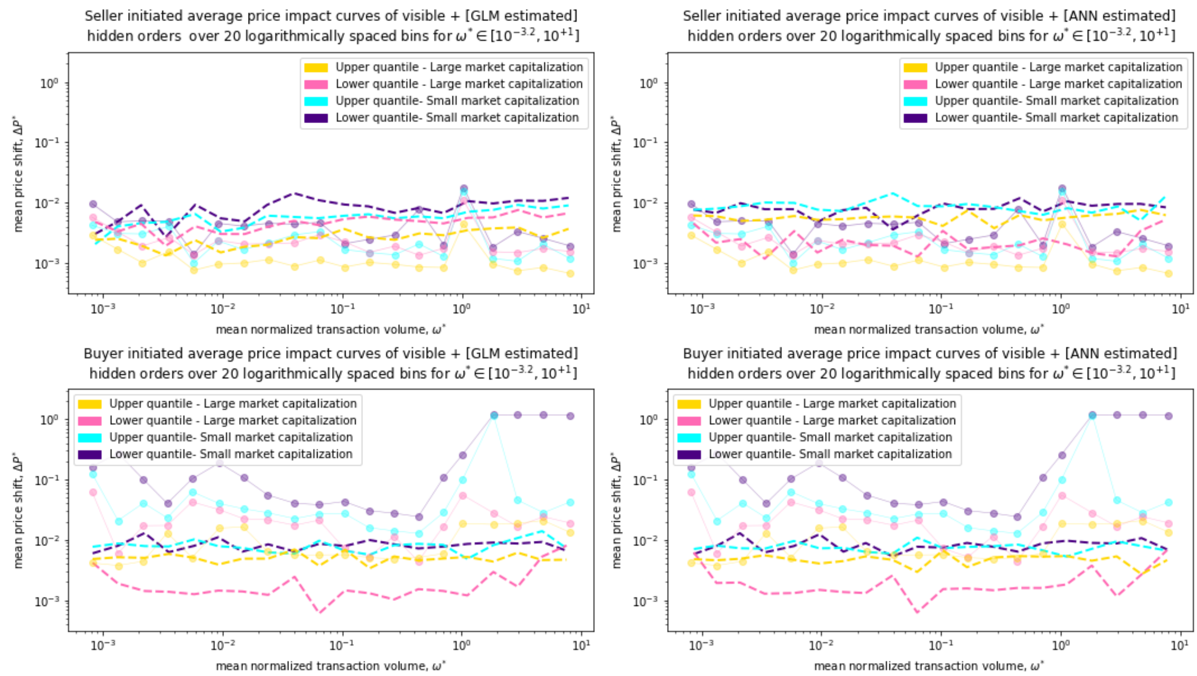


Figure 36. LSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

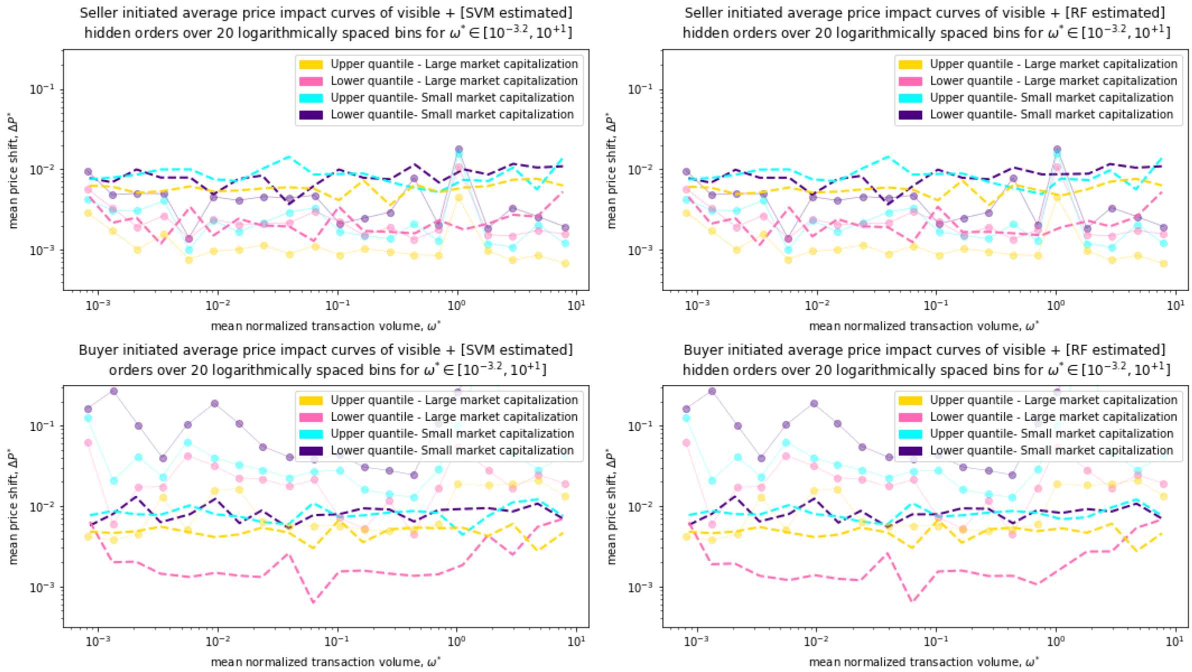


Figure 37. LSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

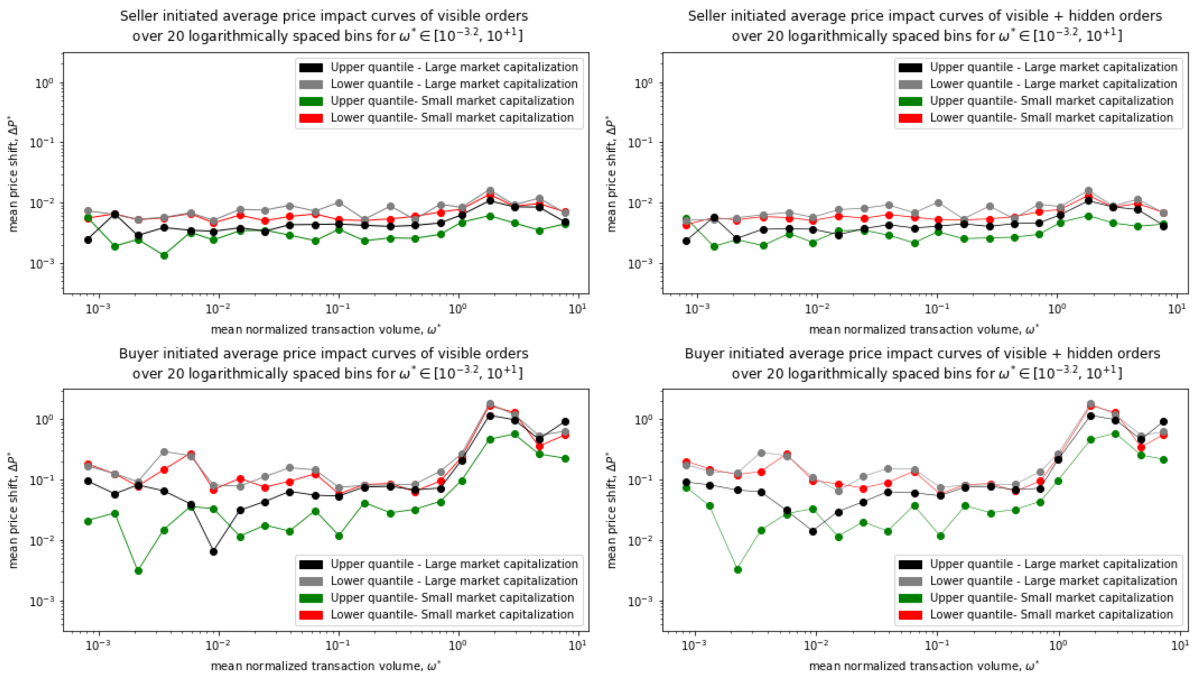


Figure 38. DFM. Aggregated price impact curves of single trades of 23 liquid firms. The data composition consists of a total of 342, 698 transactions. The period of analysis is from 28 - 09 - 2018 to 28 - 06 - 2019.

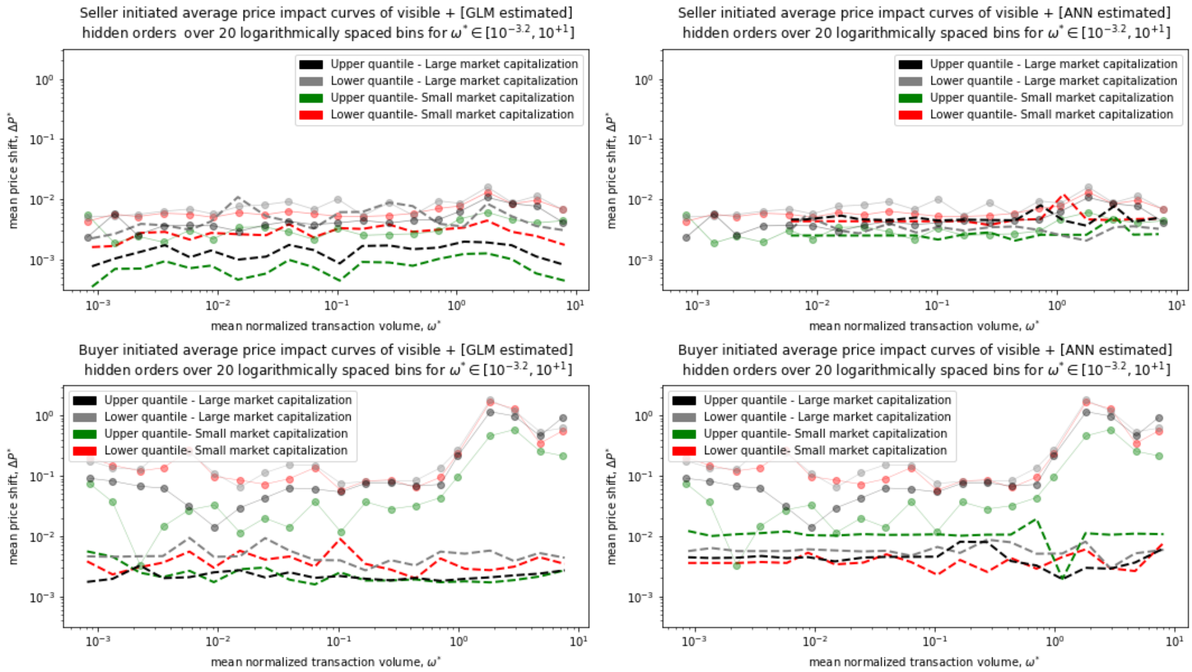


Figure 39. DFM. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

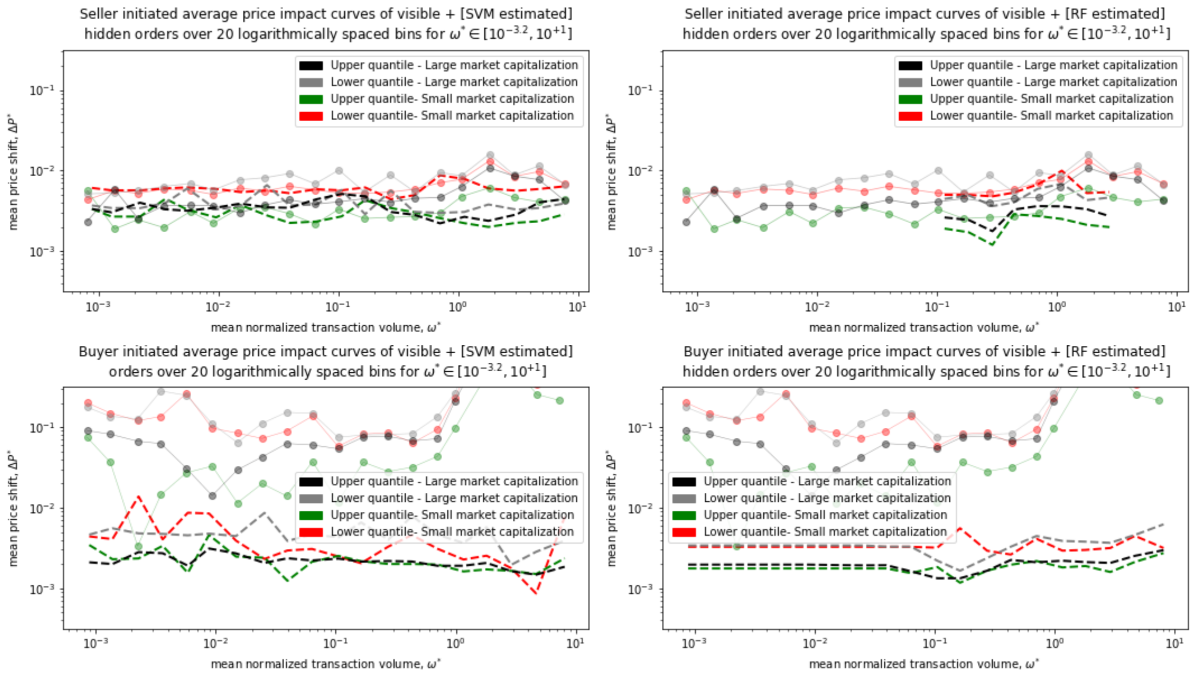


Figure 40. DFM. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

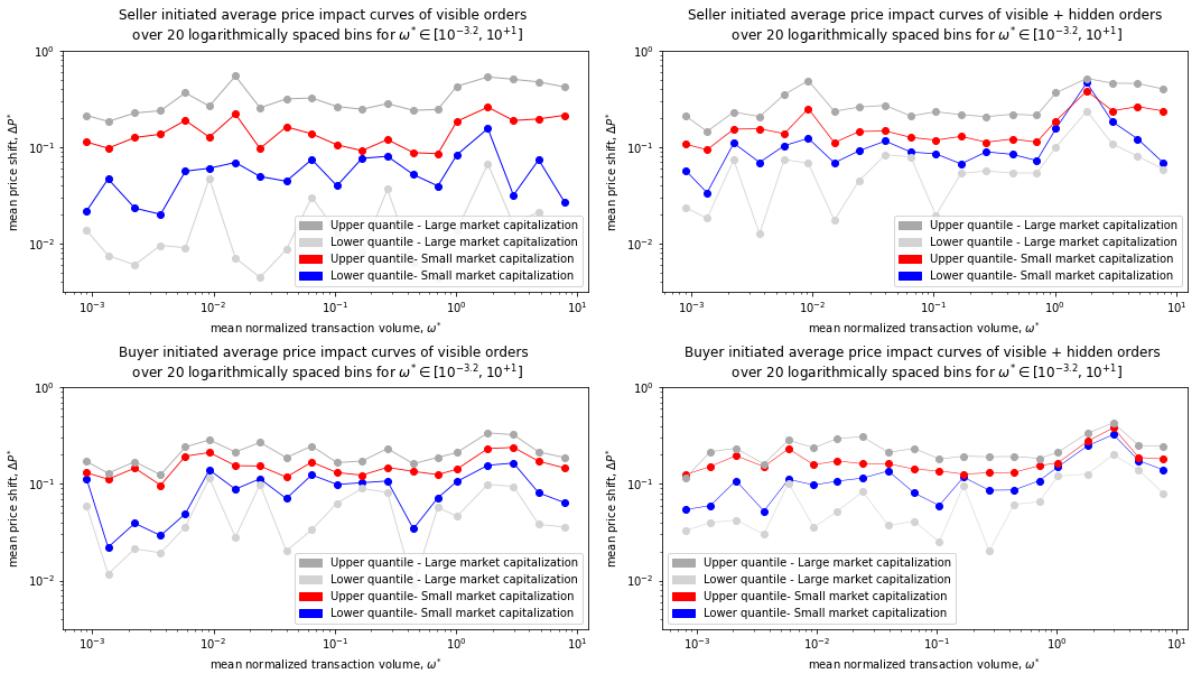


Figure 41. Euronext Paris. Aggregated price impact curves of single trades of 40 liquid firms. The data composition consists of a total of 551, 029 transactions. The period of analysis is from 24 - 10 - 2019 to 22 - 01 - 2020.

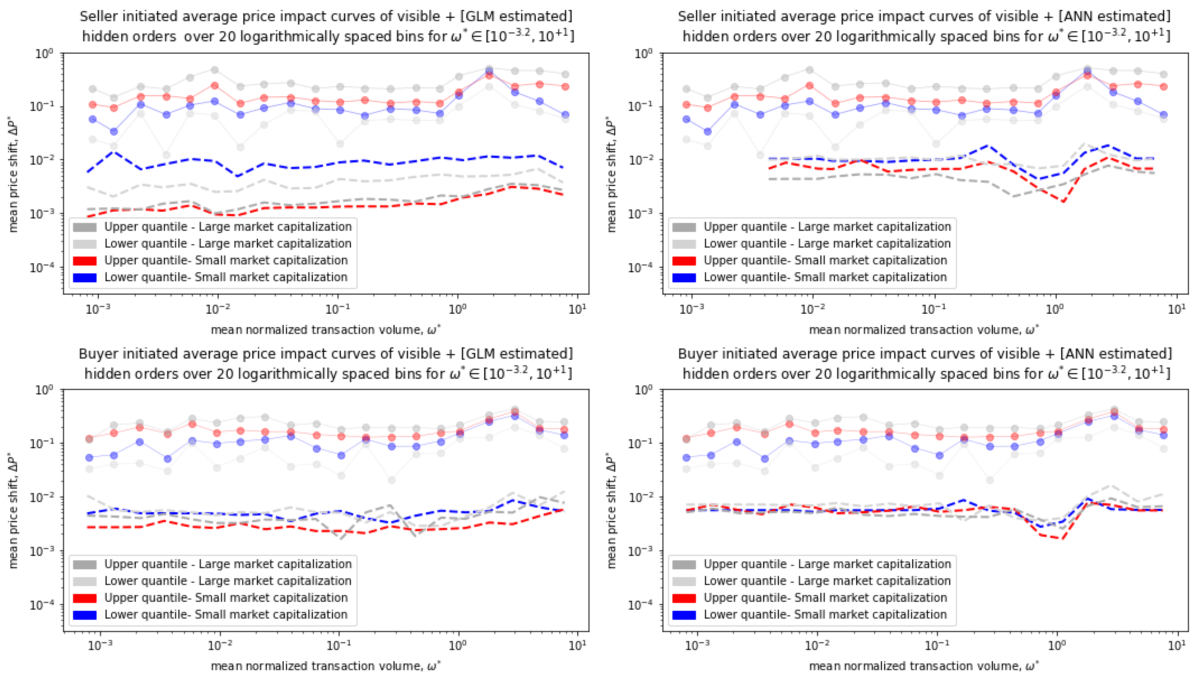


Figure 42. Euronext Paris. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

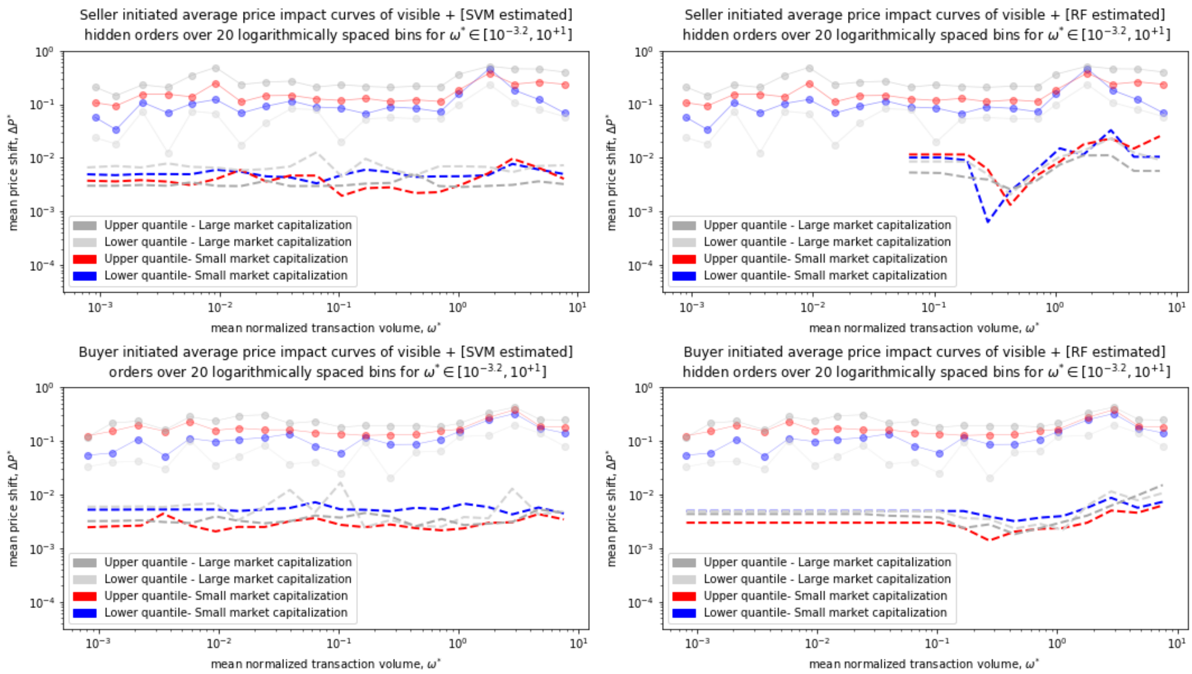


Figure 43. Euronext Paris. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

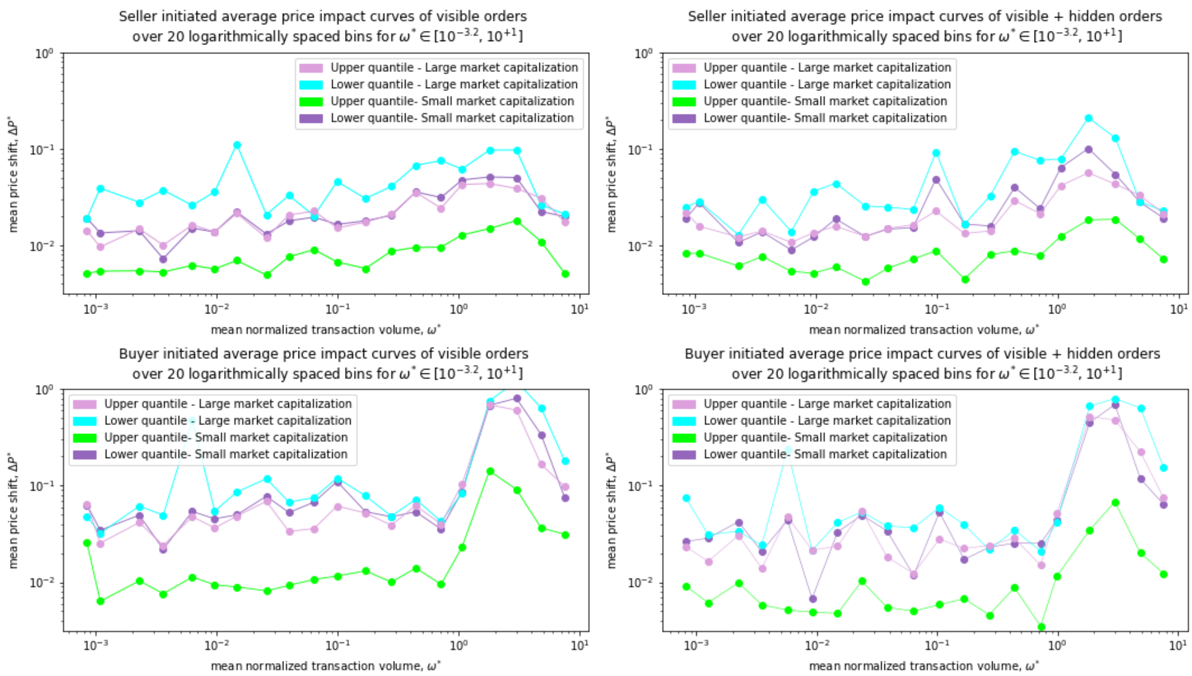


Figure 44. SGX. Aggregated price impact curves of single trades of 20 liquid firms. The data composition consists of a total of 179, 392 transactions. The period of analysis is from 26 - 10 - 2019 to 24 - 01 - 2020.

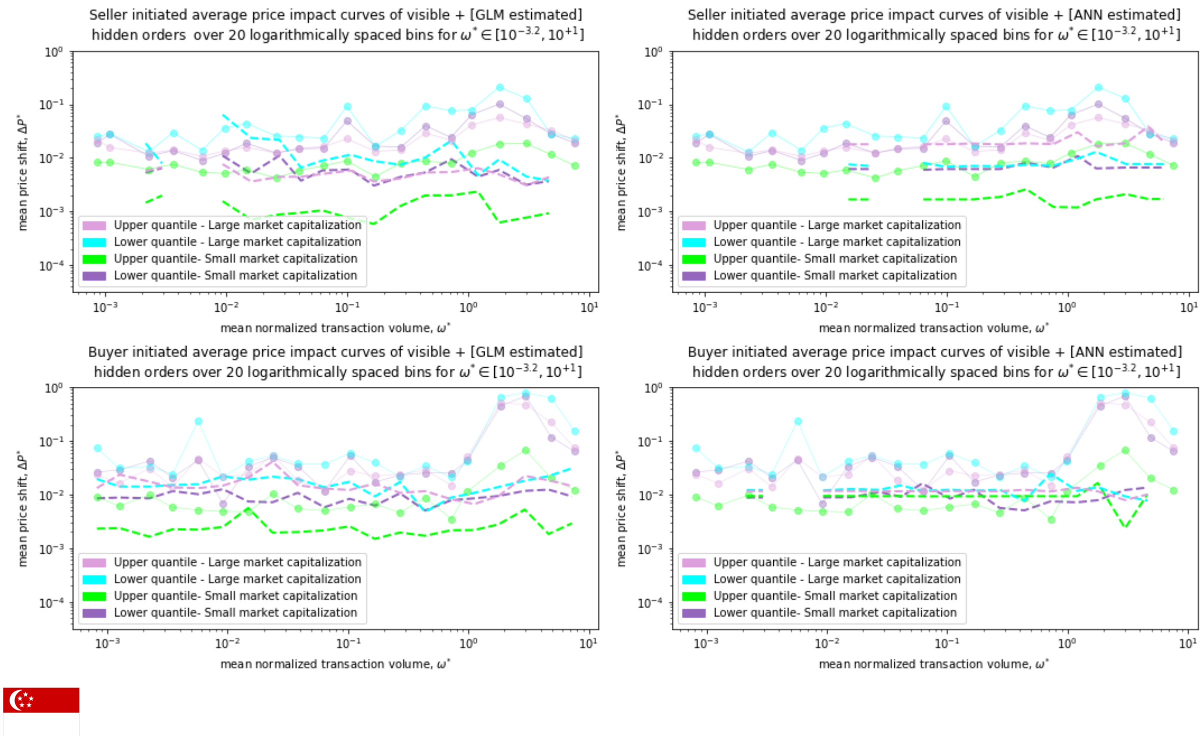


Figure 45. SGX. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

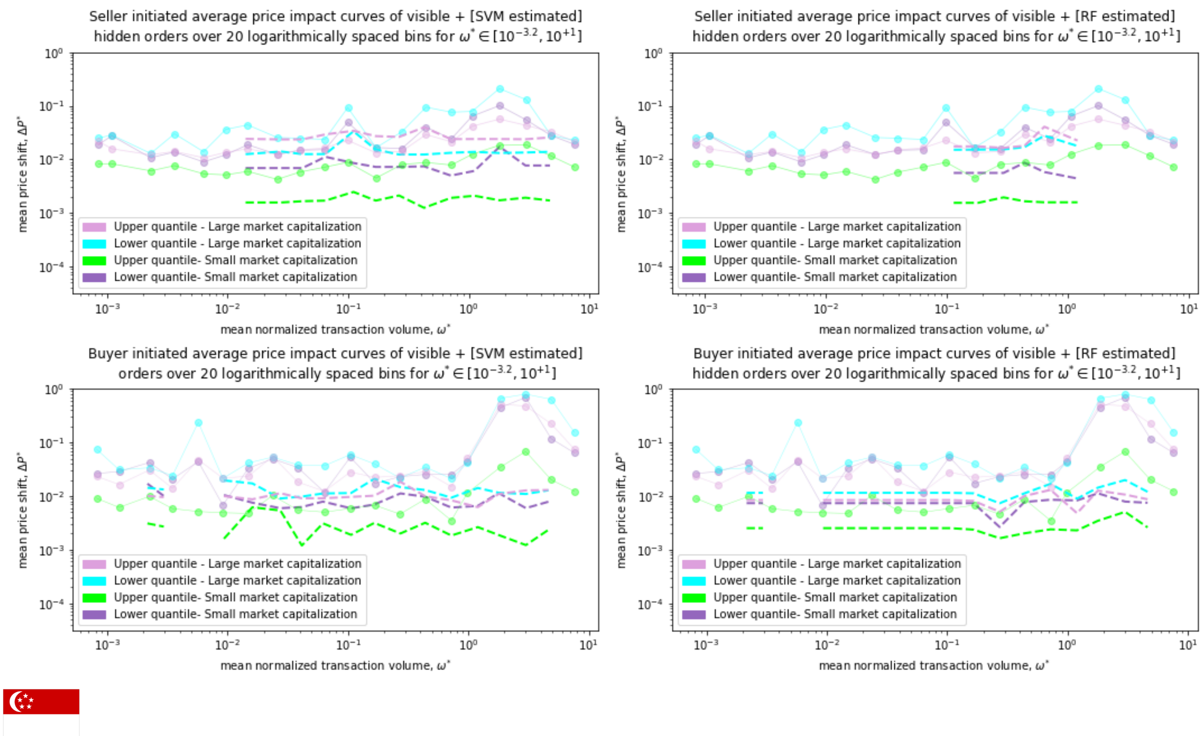


Figure 46. SGX. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

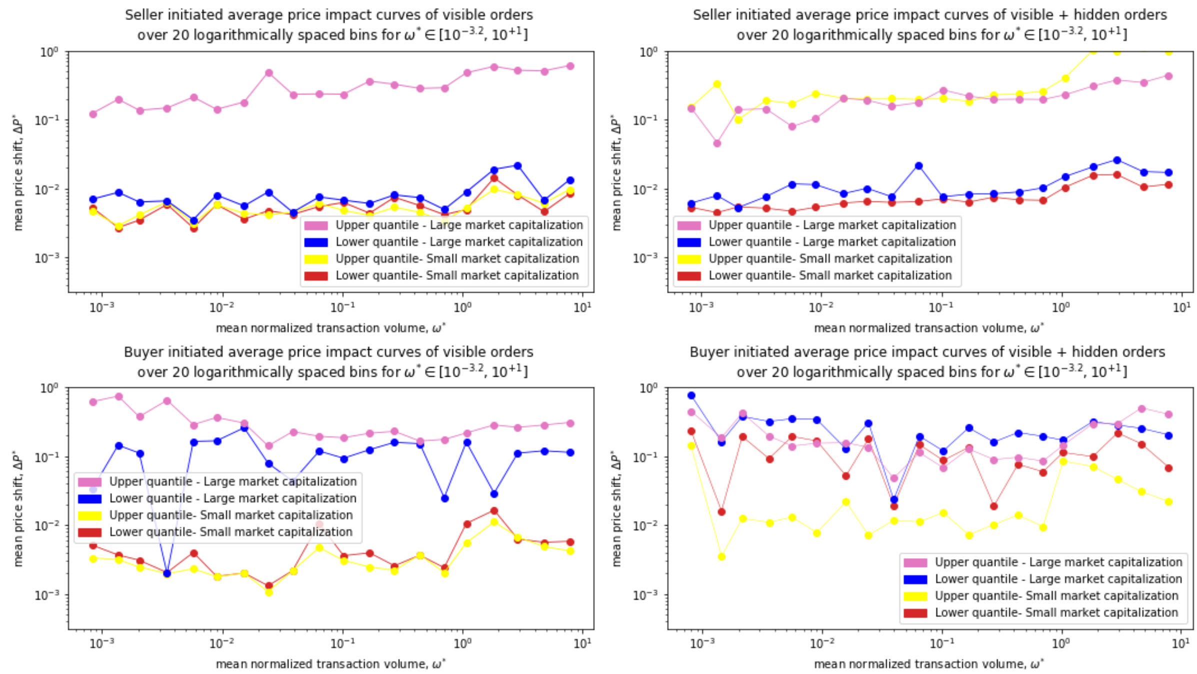


Figure 47. BME. Aggregated price impact curves of single trades of 19 liquid firms. The data composition consists of a total of 453, 360 transactions. The period of analysis is from 24 - 10 - 2019 to 22 - 01 - 2020.

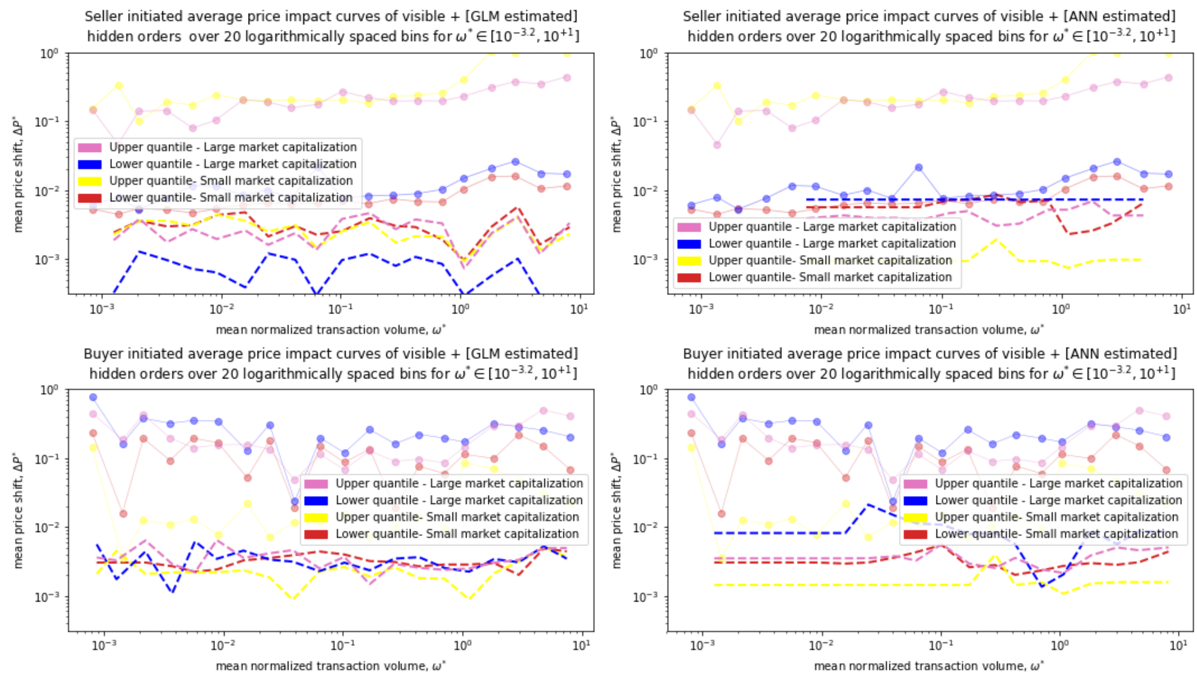


Figure 48. BME. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

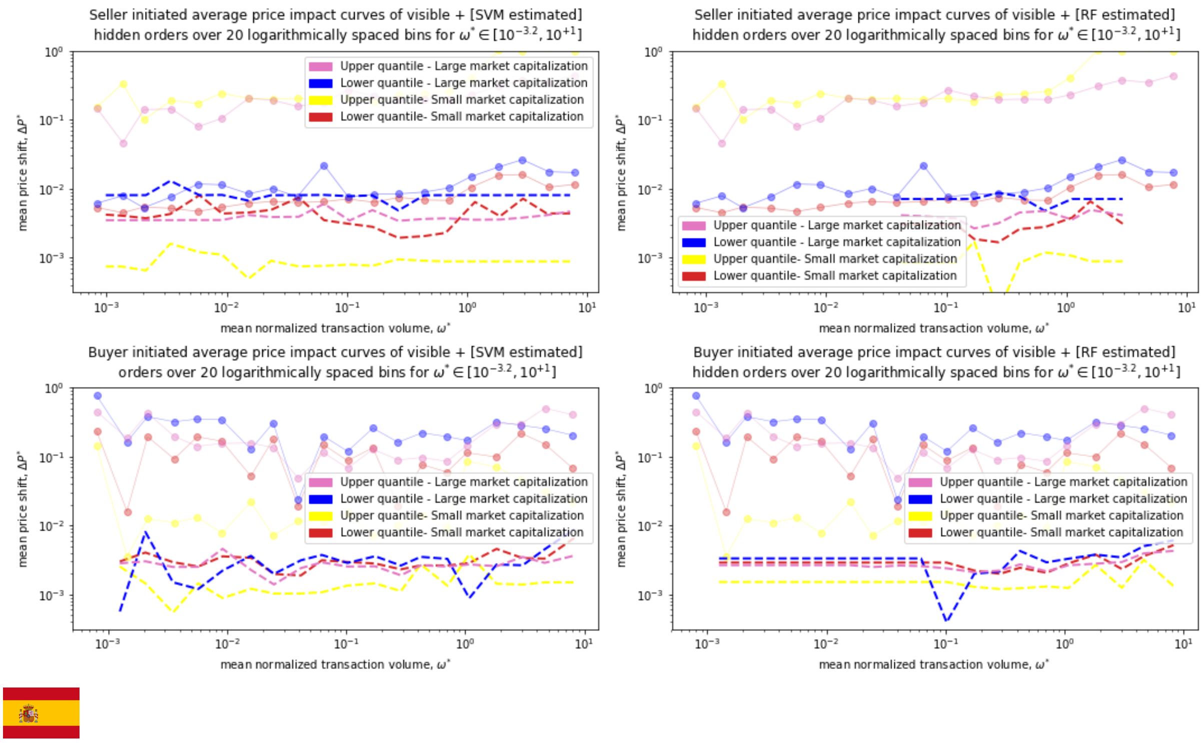


Figure 49. BME. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

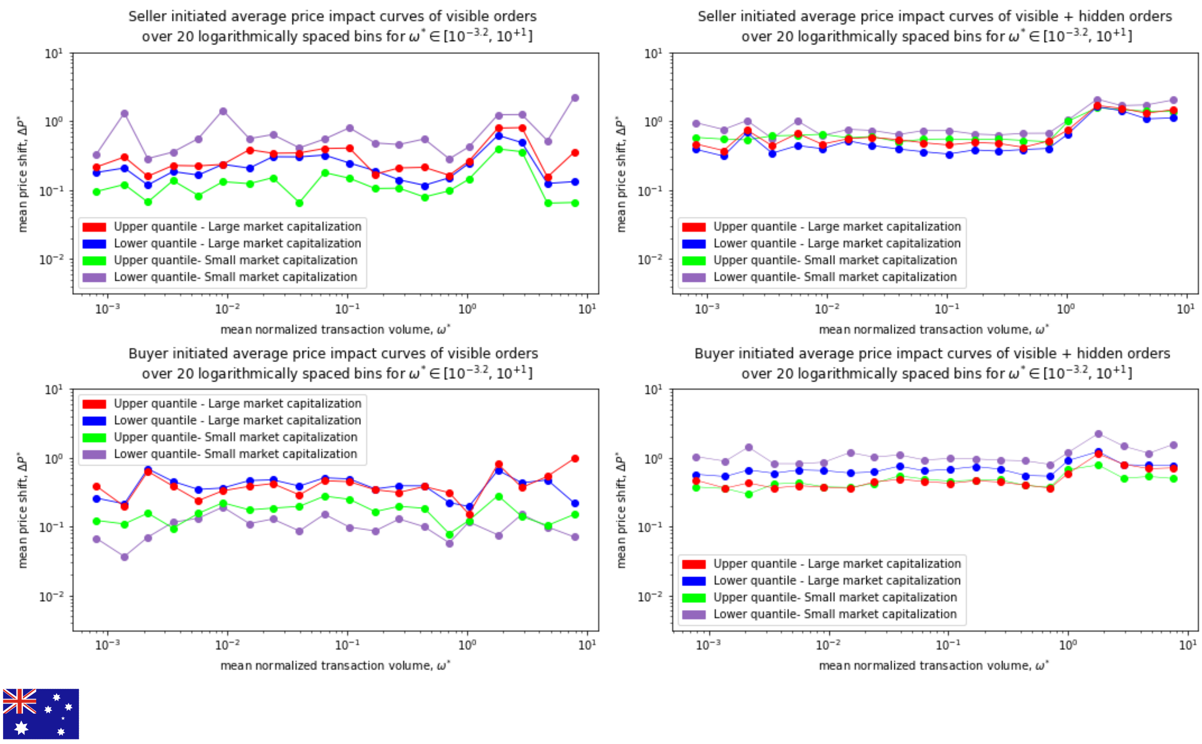


Figure 50. ASX. Aggregated price impact curves of single trades of 25 liquid firms. The data composition consists of a total of 778, 458 transactions. The period of analysis is from 25 - 10 - 2019 to 23 - 01 - 2020.

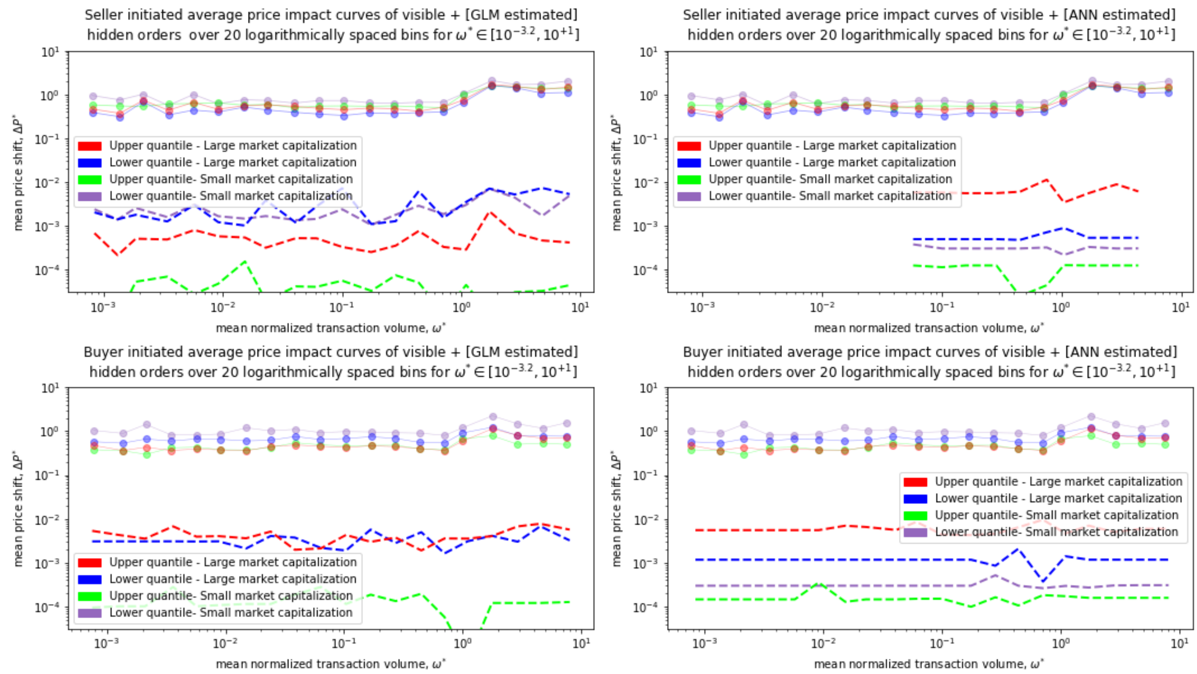


Figure 51. ASX. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

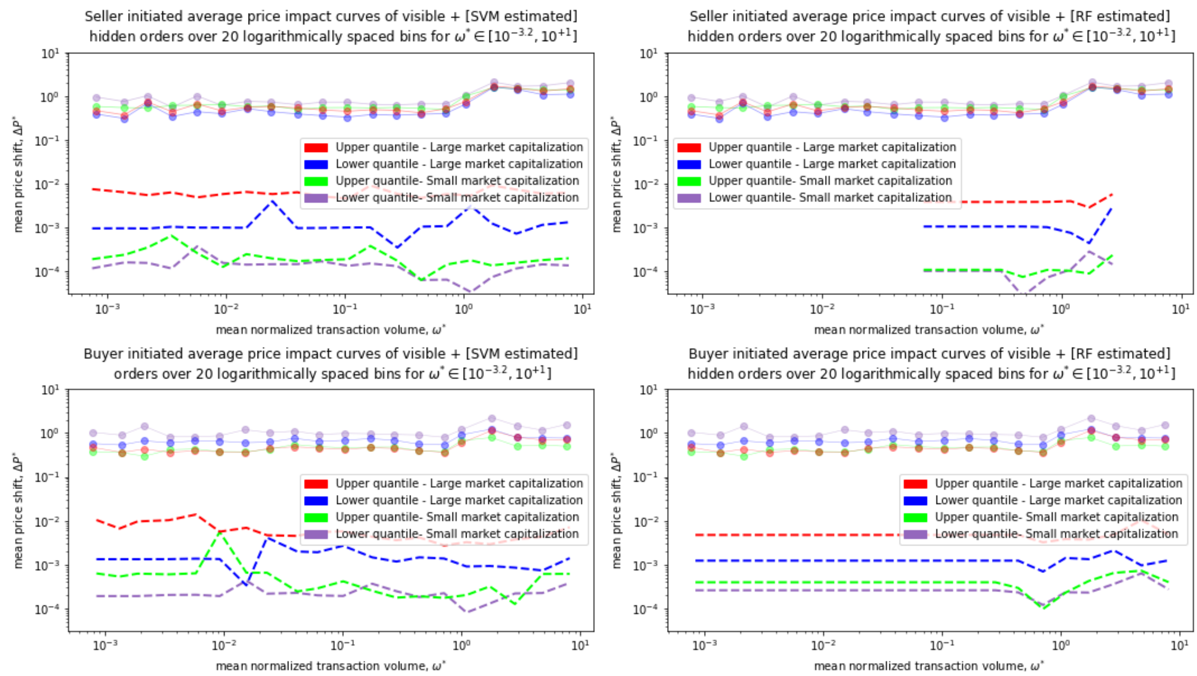


Figure 52. ASX. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

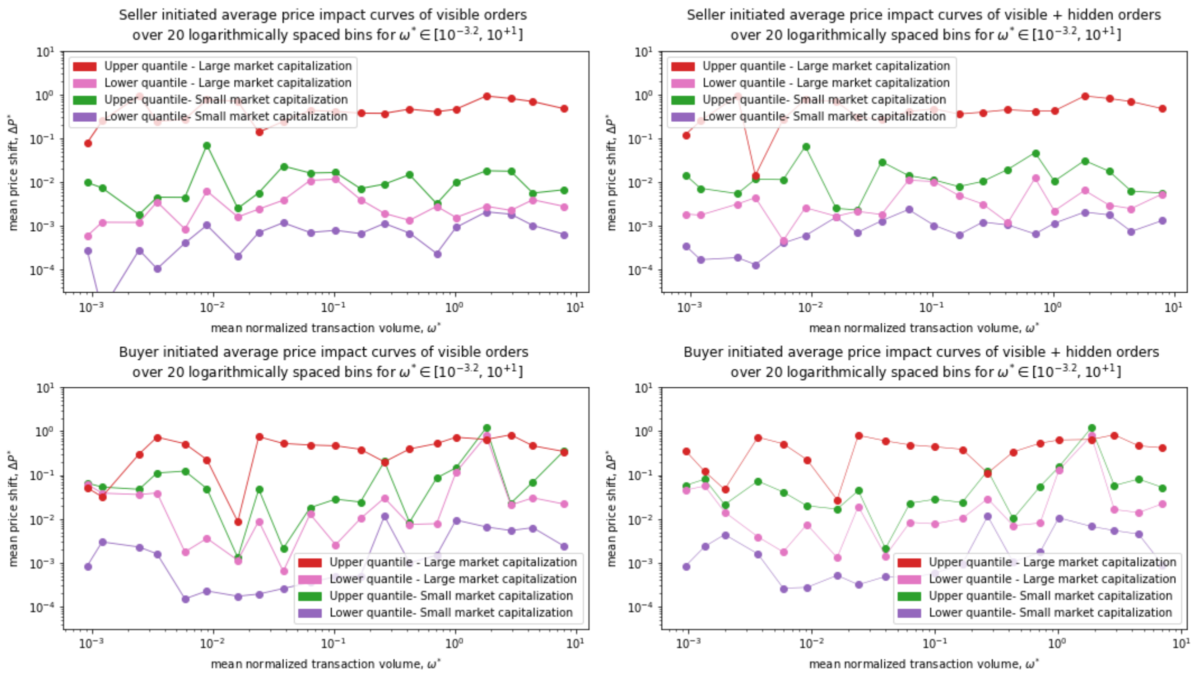


Figure 53. WBAG. Aggregated price impact curves of single trades of 5 liquid firms. The data composition consists of a total of 70, 194 transactions. The period of analysis is from 23 - 10 - 2019 to 21 - 01 - 2020.

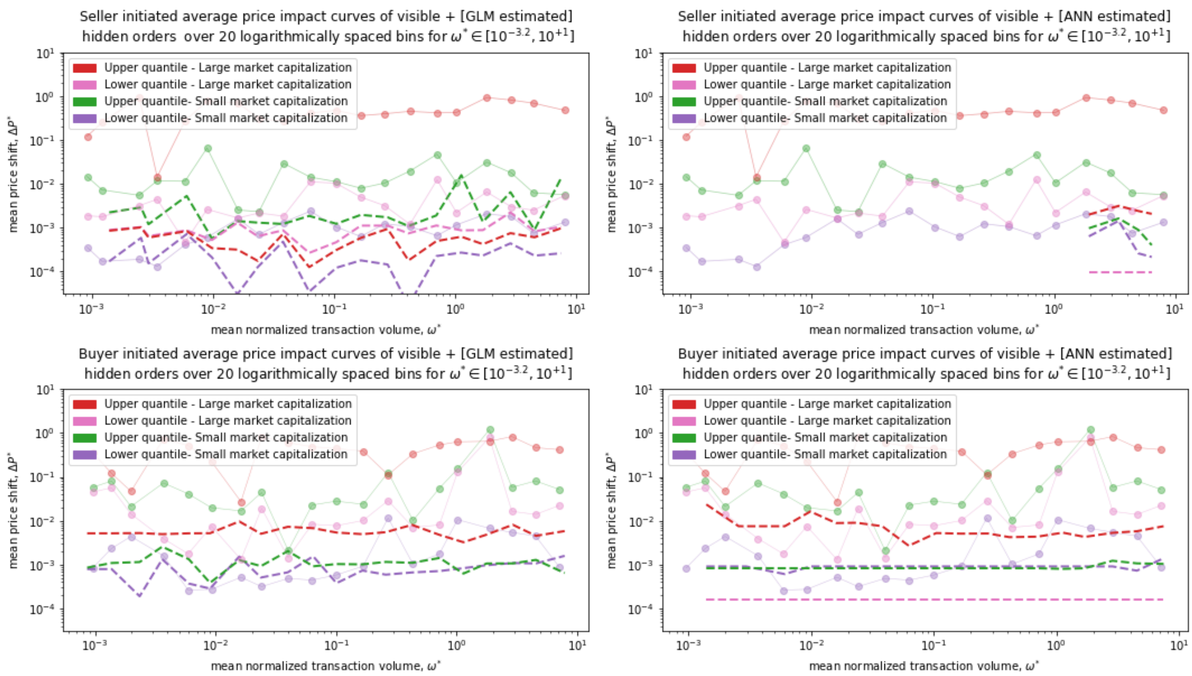


Figure 54. WBAG. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

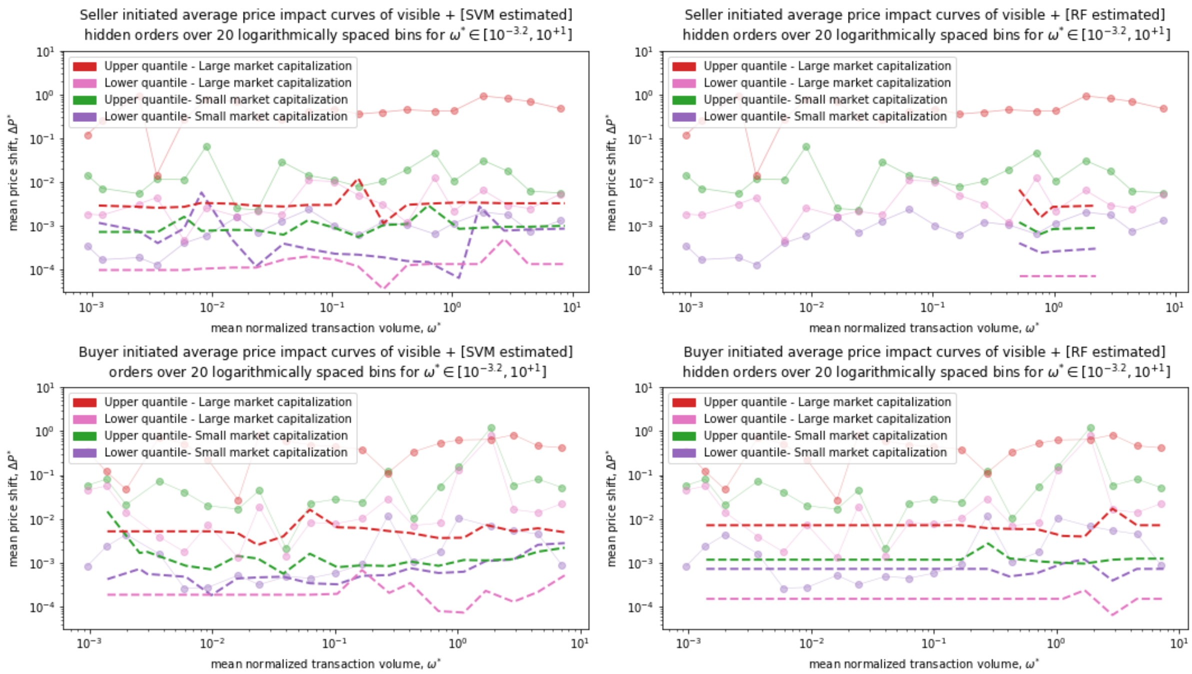


Figure 55. WBAG. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

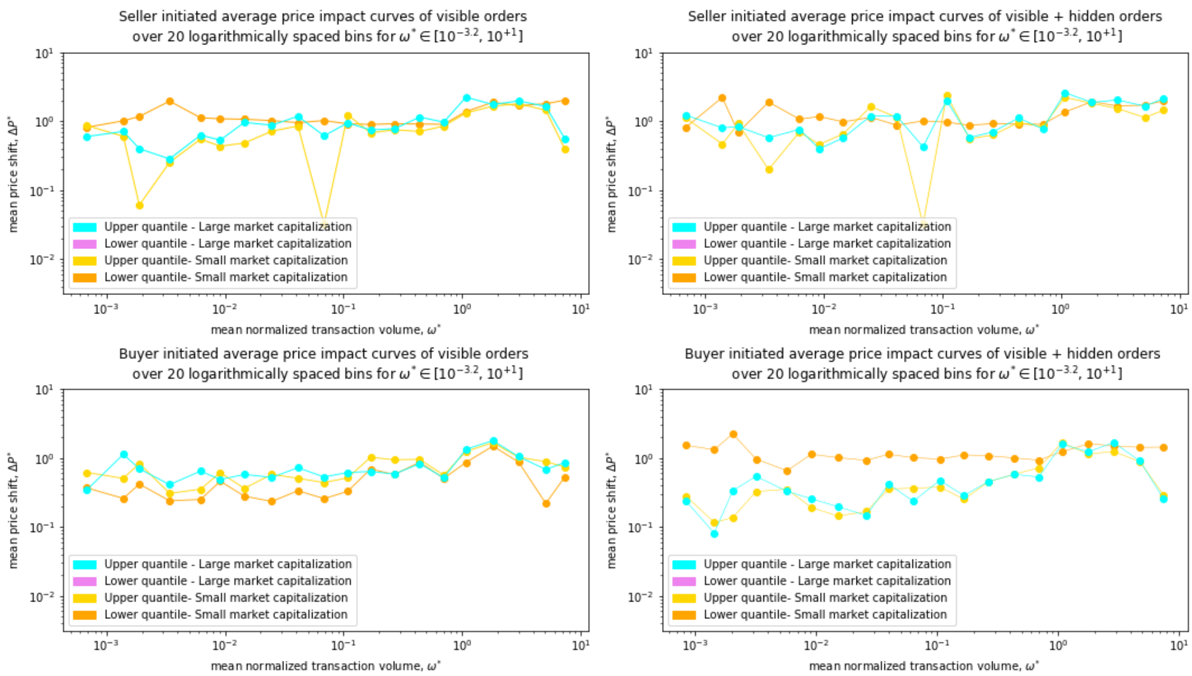


Figure 56. Euronext Lisbon. Aggregated price impact curves of single trades of 4 liquid firms. The data composition consists of a total of 56, 202 transactions. The period of analysis is from 23 - 10 - 2019 to 21 - 01 - 2020.

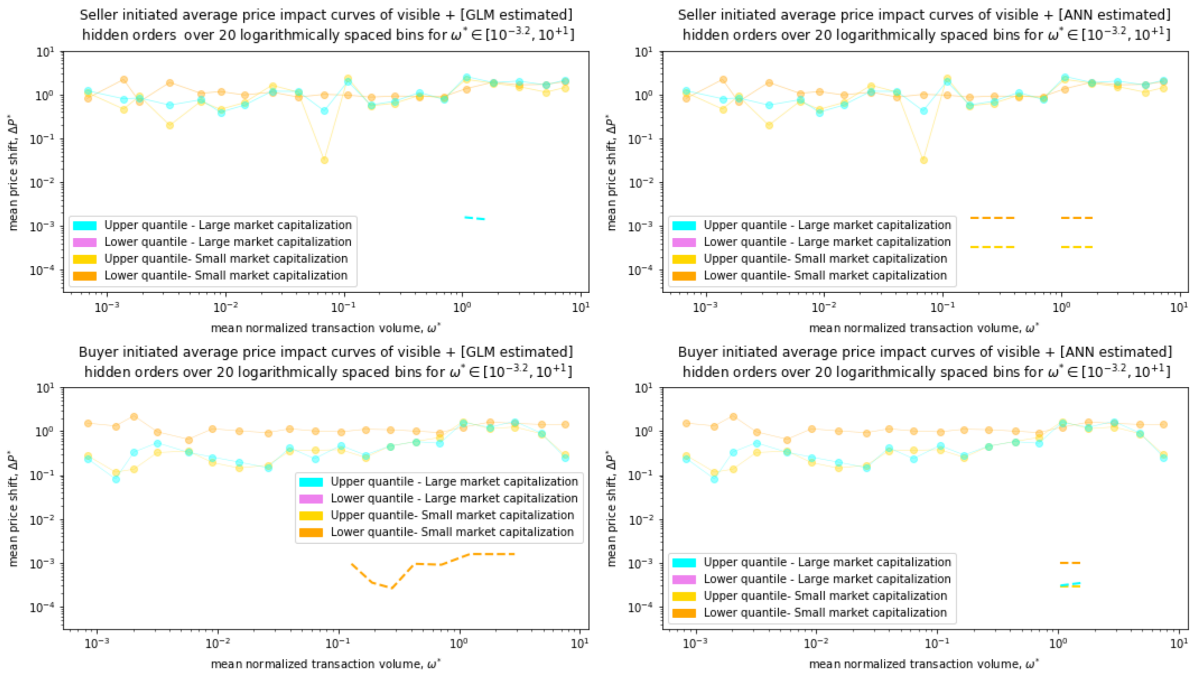


Figure 57. Euronext Lisbon. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

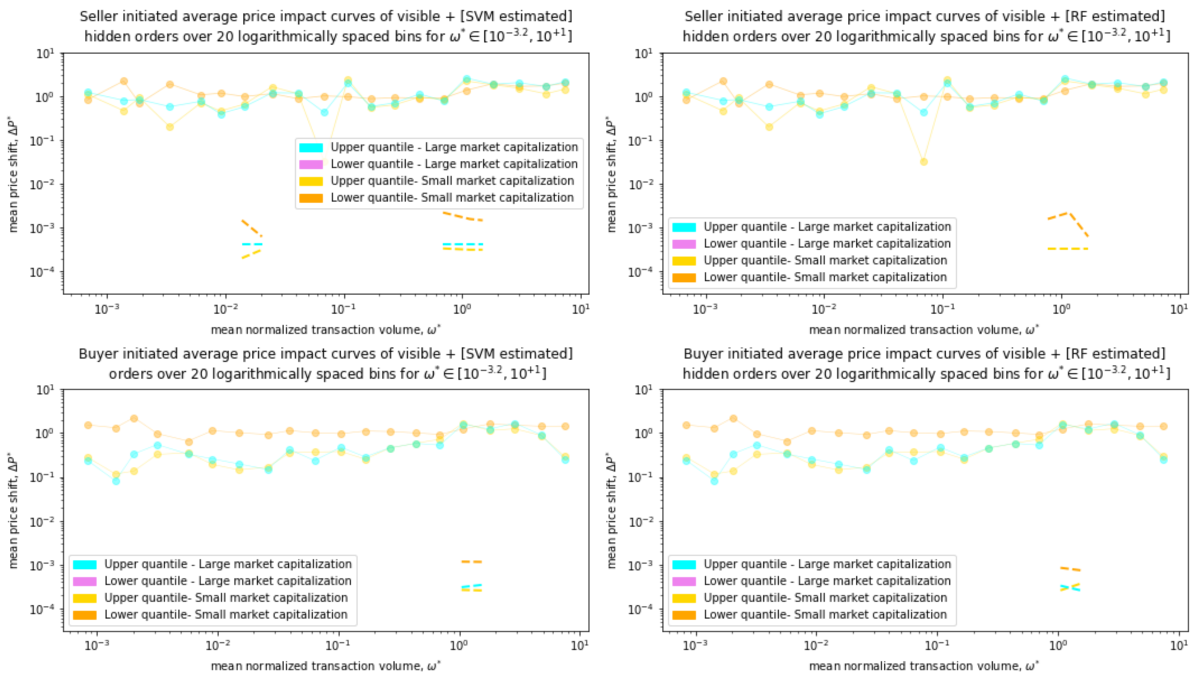


Figure 58. Euronext Lisbon. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

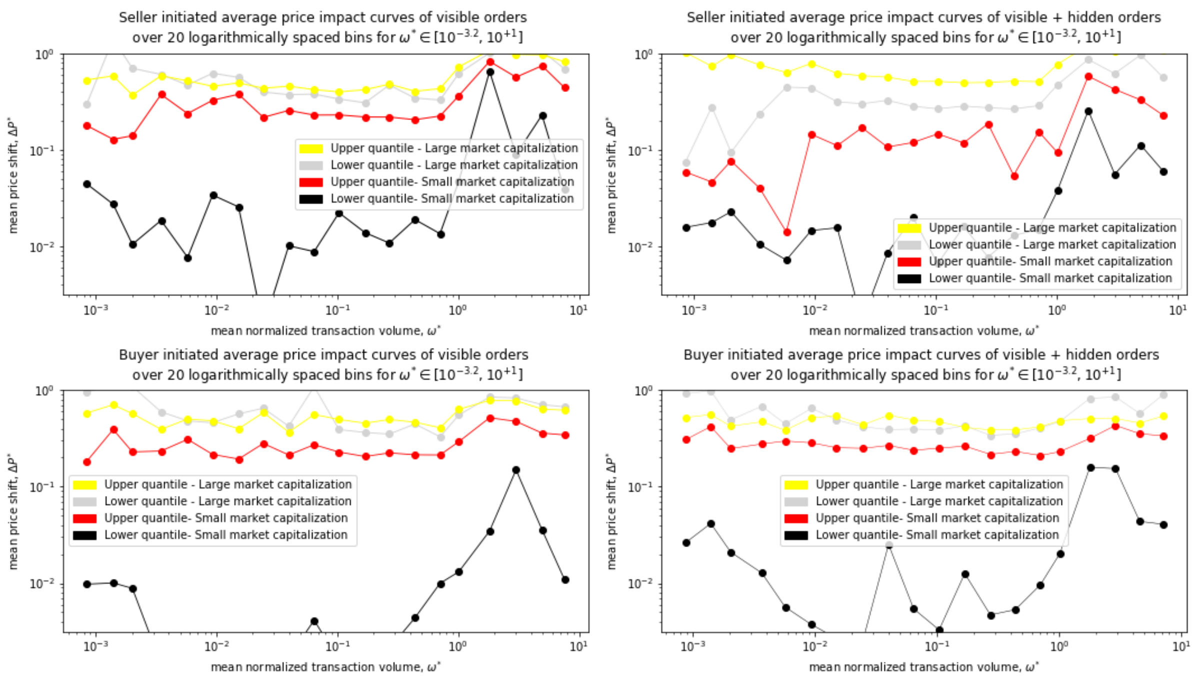


Figure 59. Euronext Brussels. Aggregated price impact curves of single trades of 17 liquid firms. The data composition consists of a total of 413, 227 transactions. The period of analysis is from 24 - 10 - 2019 to 22 - 01 - 2020.

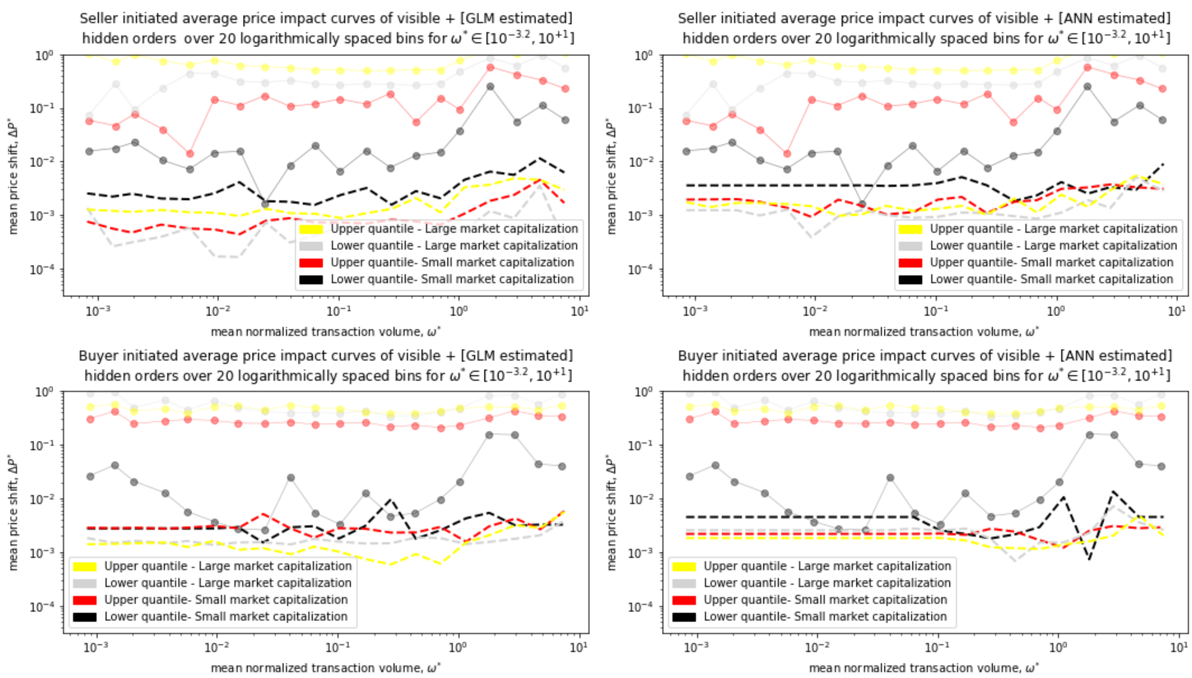


Figure 60. Euronext Brussels. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

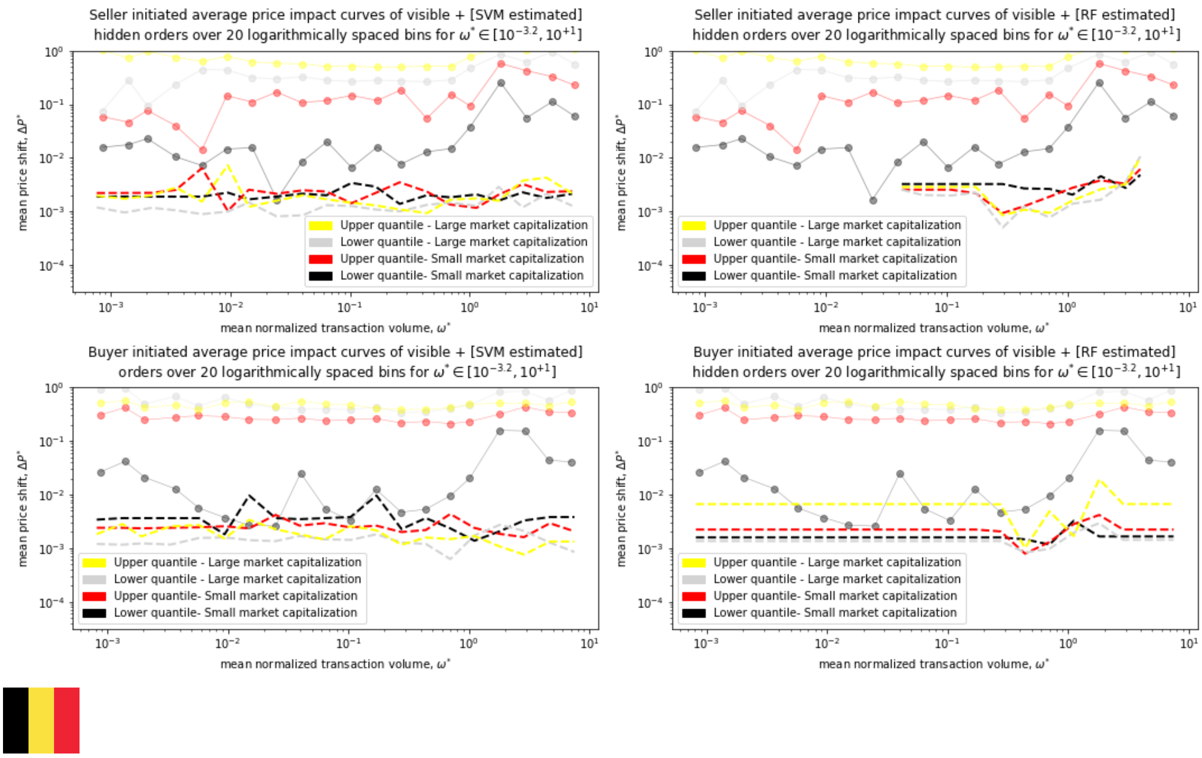


Figure 61. Euronext Brussels. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

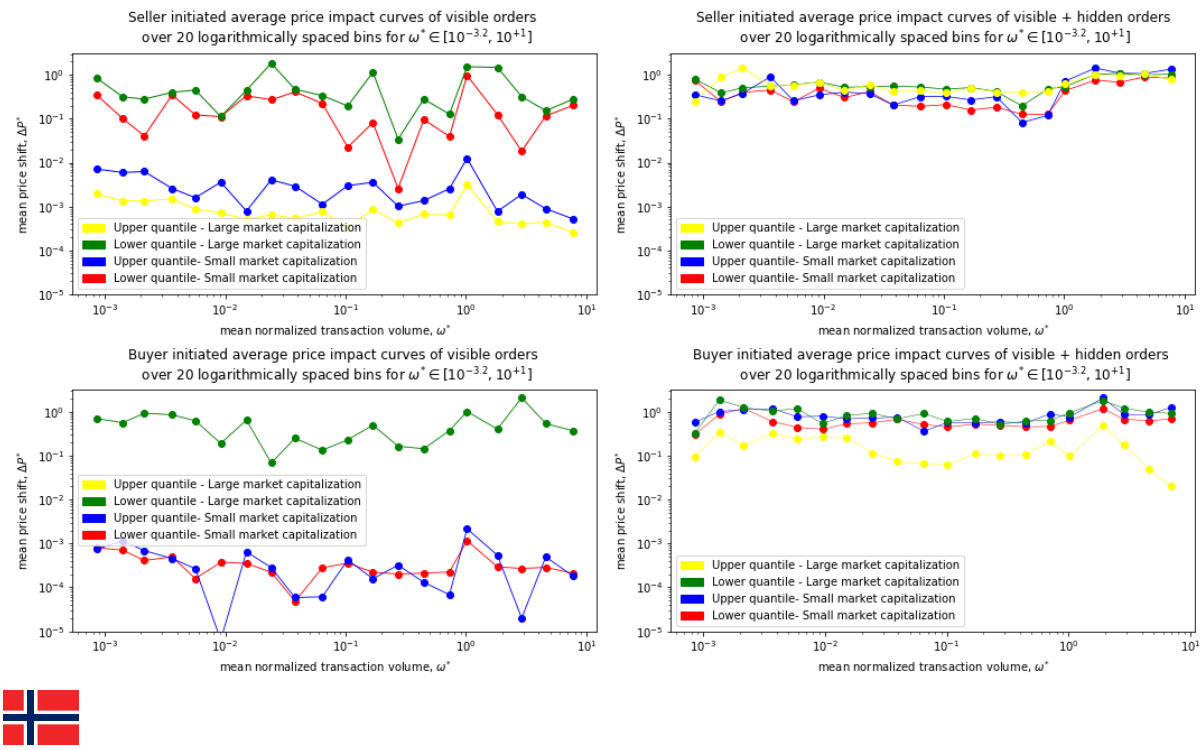


Figure 62. OSE. Aggregated price impact curves of single trades of 8 liquid firms. The data composition consists of a total of 207, 182 transactions. The period of analysis is from 26 - 10 - 2019 to 24 - 01 - 2020.

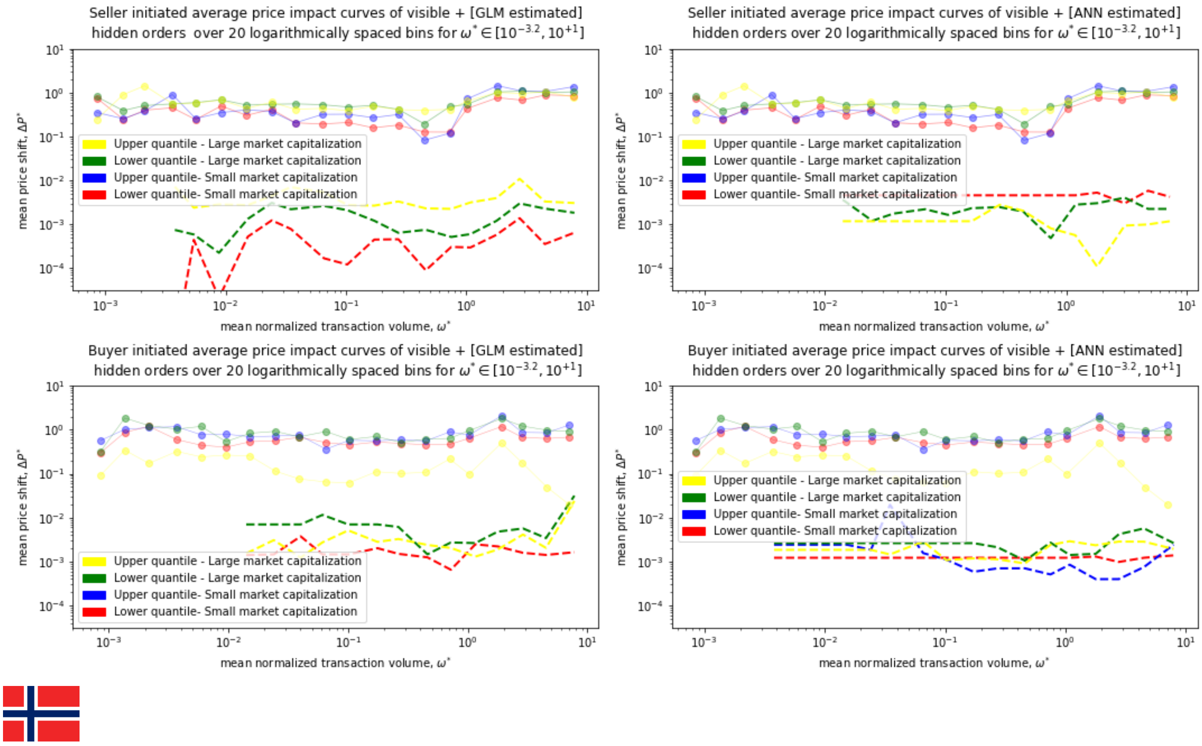


Figure 63. OSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

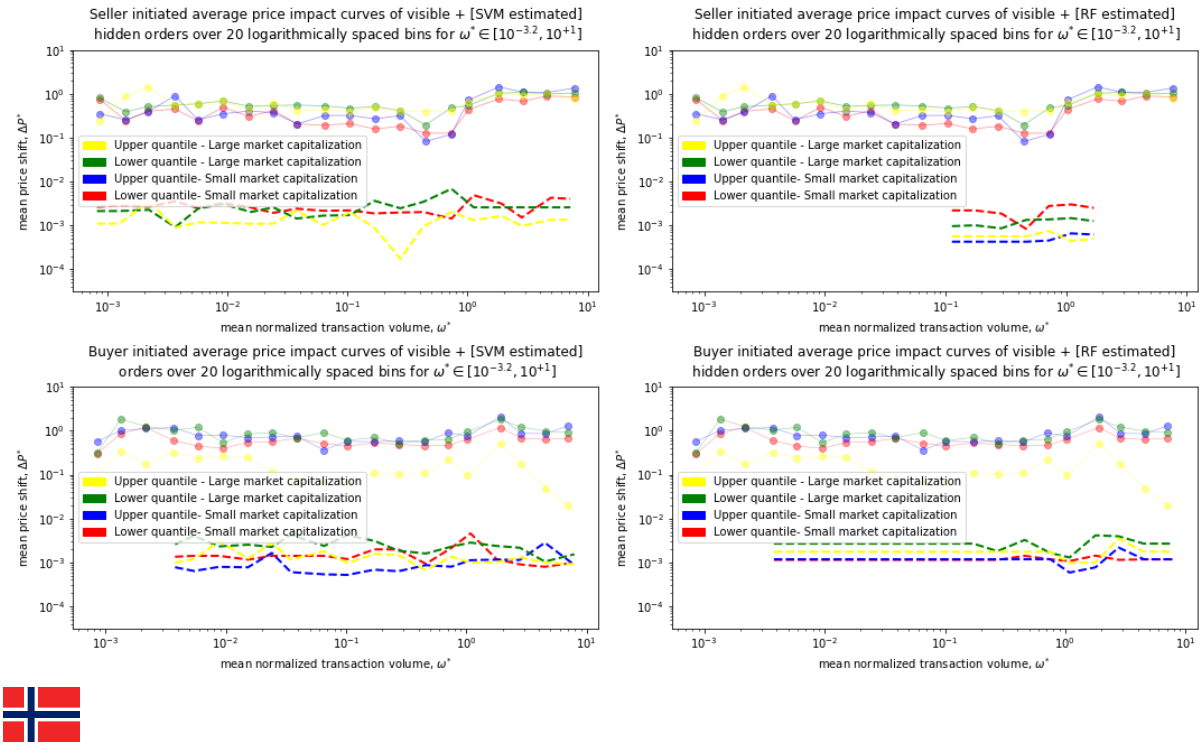


Figure 64. OSE. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

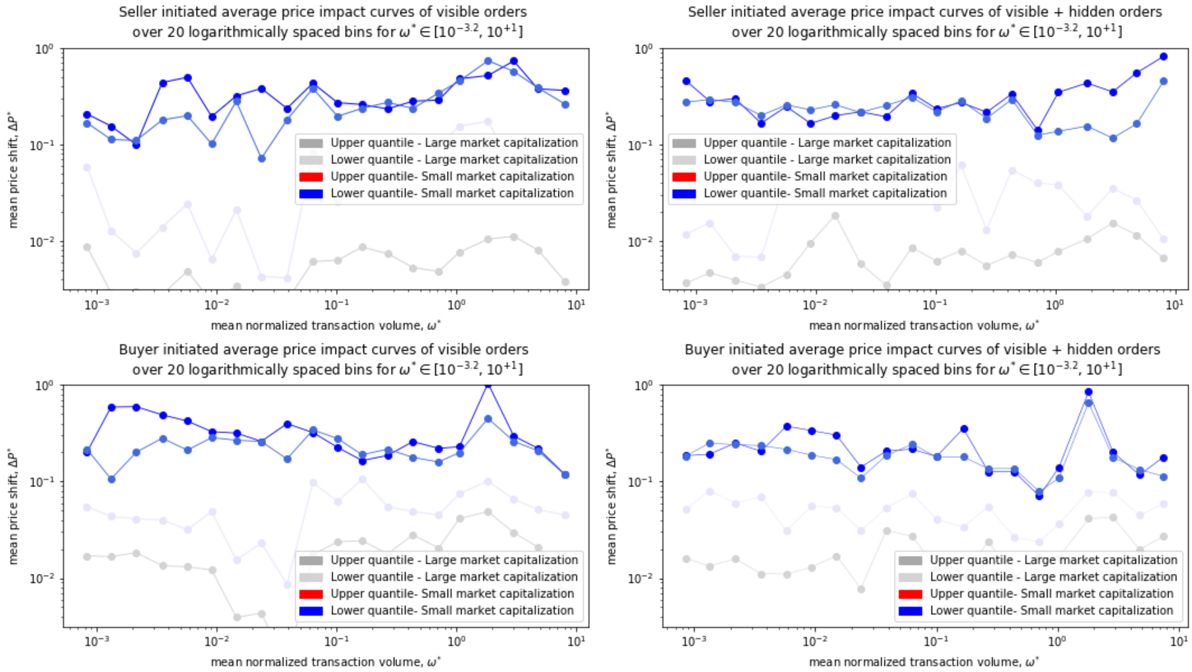


Figure 65. OMXH. Aggregated price impact curves of single trades of 33 liquid firms. The data composition consists of a total of 326, 721 transactions. The period of analysis is from 23 - 10 - 2019 to 21 - 01 - 2020.

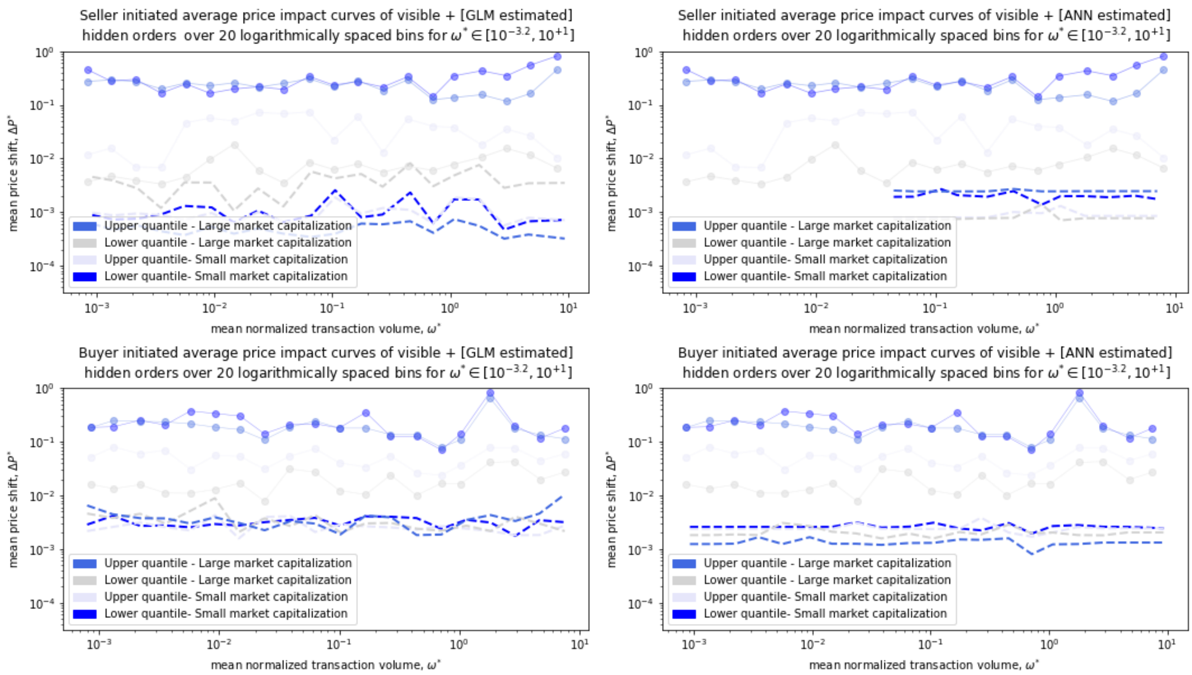


Figure 66. OMXH. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

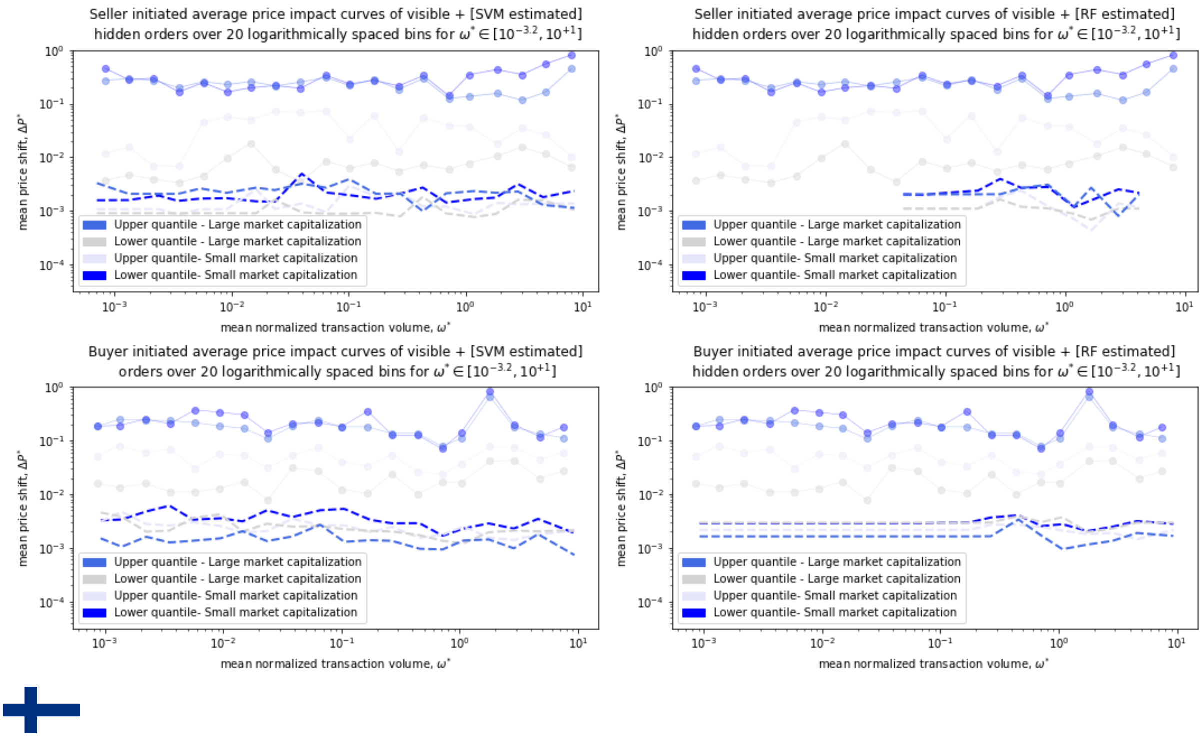


Figure 67. OMXH. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.

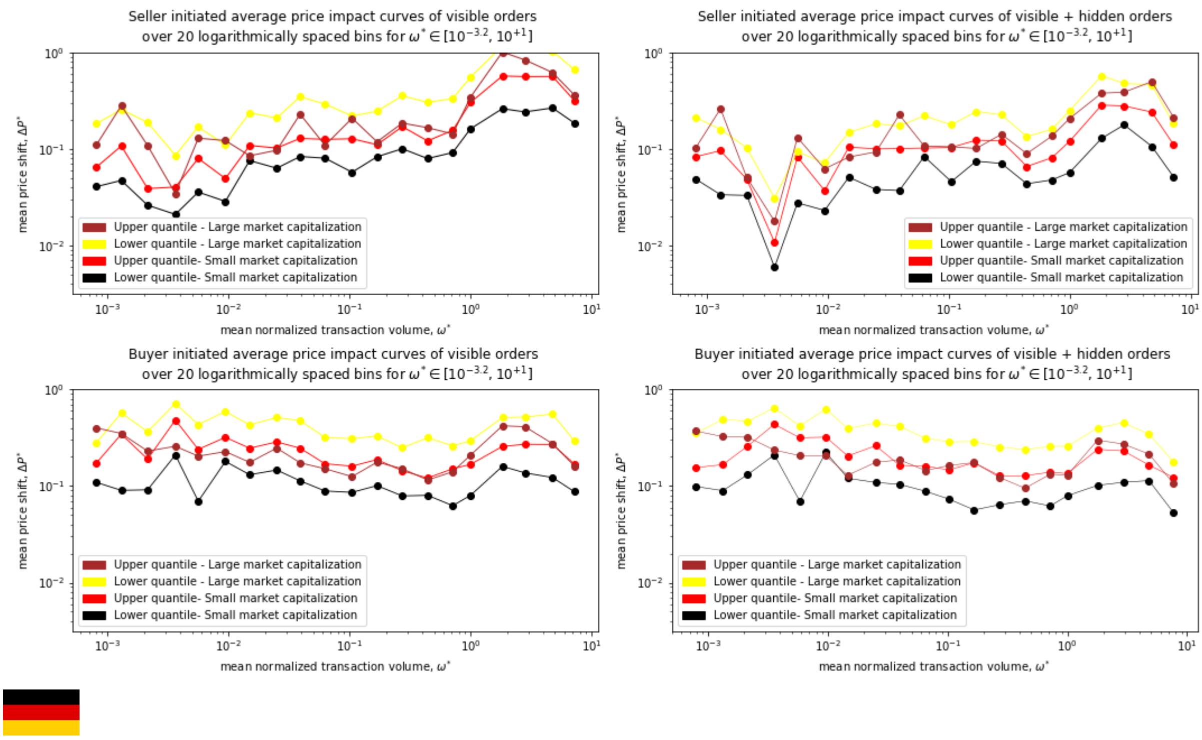


Figure 68. FWB. Aggregated price impact curves of single trades of 43 liquid firms. The data composition consists of a total of 1, 313, 429 transactions. The period of analysis is from 24 - 10 - 2019 to 22 - 01 - 2020.

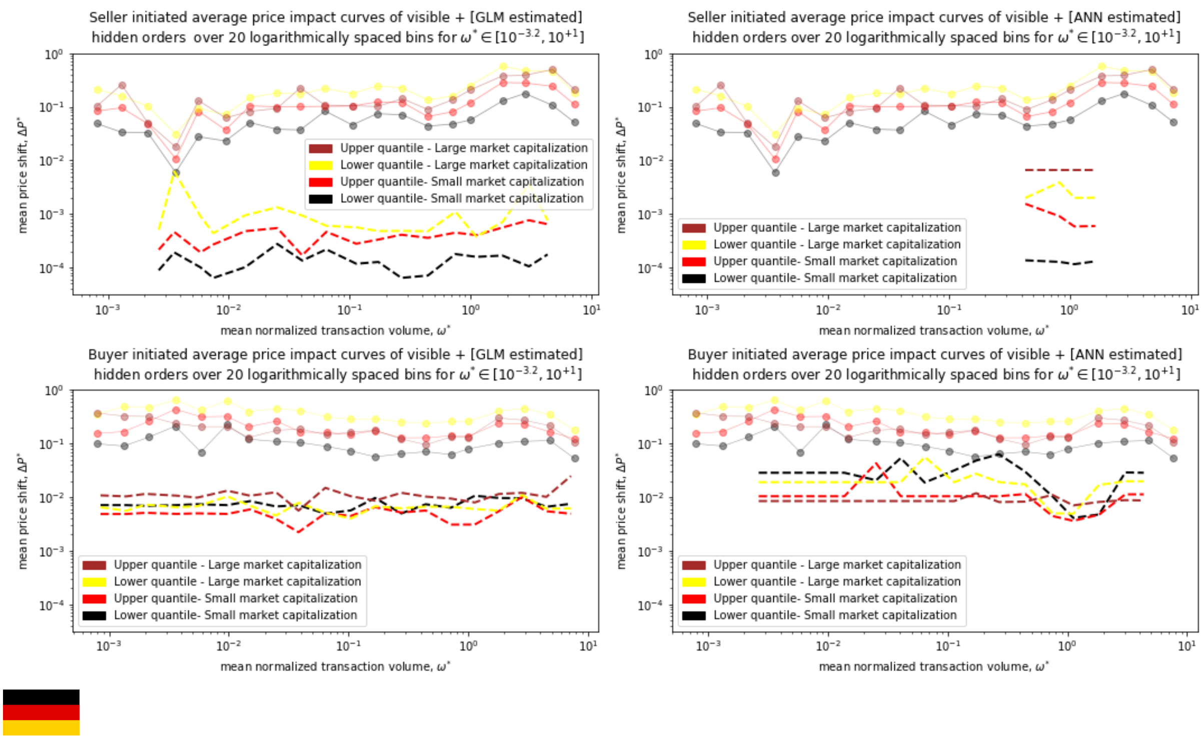


Figure 69. FWB. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of GLM and ANN machine estimated orders, represented by dashed lines.

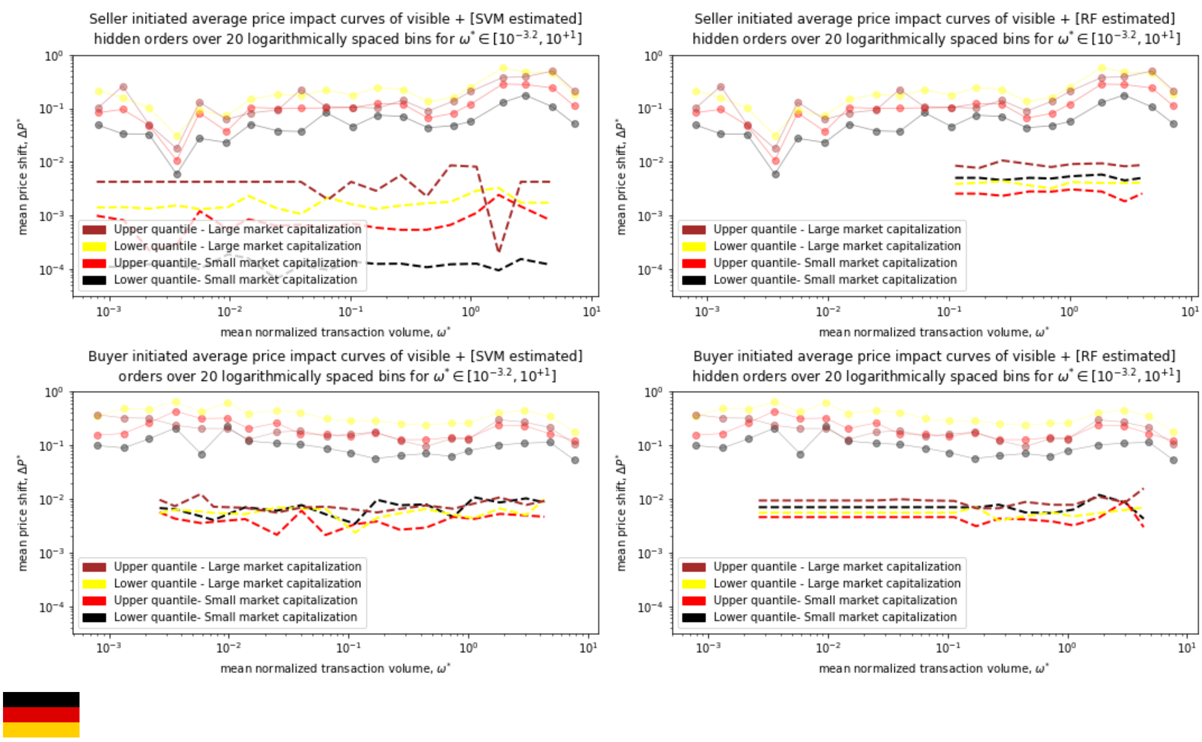


Figure 70. FWB. Aggregated average market impact curves of true orders, represented by solid lines with circular nodes, and aggregated average market impact curves of SVM and RF machine estimated orders, represented by dashed lines.