Essays on the Dynamics of Order Books

DISSERTATION
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Finance

submitted by

Louis Müller

from

Zurich and Thayngen (Schaffhausen)

Approved on the application of

Prof. Dr. Karl Frauendorfer

and

Prof. Dr. Georg Pflug

Dissertation no. 5237

Difo-Druck GmbH, Untersiemau 2022

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, May 25, 2022

The President:

Prof. Dr. Bernhard Ehrenzeller

Acknowledgments

This dissertation would not have been possible without continued support from a number of remarkable people throughout my doctoral journey. In the following words, I would like to extend my special gratitude to all of them.

To begin with, I would like to thank Prof. Dr. Karl Frauendorfer for his valuable guidance and support a as a supervisor over the last few years. His targeted inputs have always been extremely helpful for all the research projects I was involved in throughout my doctoral journey. Moreover, I would like to acknowledge the helpful comments and suggestions of my co-supervisor Prof. Dr. Georg Pflug and Prof. Dr, Angelo Ranaldo. My particular gratitude also goes out to Alexandru Rif, having continuously supported and pushed me when things got increasingly difficult.

On the technical side, I would like to extend my gratitude to SIX Securities & Exchanges (SIX) for allowing me to analyze their vast data in the context for this dissertation. In particular, I would like to thank Marc Berthoud, Adam Matuszewski and Fabian Ochsner, who took their time to discuss findings and offered valuable insights from a practitioner's perspective. I would also like to thank Gido Haarbrücker and Claus Liebenberger for supporting me with answers and suggestions in the realm of information technology.

Finally, I would like to thank my biggest cheerleader and wife Shalini for her never-ending understanding and emotional support, my parents Christoph and Gabriela as well as all my friends. This would not have been possible without all of you.

Zurich, July 2022

Louis Müller

Contents

Summary

The three articles presented in this dissertation are addressing various topics in market microstructure in and around closing auctions in equity markets. Despite the growing market share of closing auctions relative to the continuous trading phase, academia has devoted relatively little research towards closing auctions. This dissertation in particular has been composed in cooperation with SIX Securities & Exchanges (SIX), using granular order-level data from their lit equities exchange of Swiss stocks.

The first chapter of this dissertation focuses on the volume traded in closing auctions. The analysis gauges the aggregated level of patience across all investors following continuous trading phases with varying underlying market conditions. The market conditions investigated contain both execution risk and liquidity. Moreover, the analysis explores how the marginal effect with respect to these market conditions varies when including expectations of how much volume will be traded in the subsequent closing auction.

The second chapter focuses on closing auctions on a more granular level, in that it investigates the process of price discovery throughout the 10-minute auctions. For this purpose, snapshots of the order book at various times during the auction are taken and uncrossed, in order to capture inflows and outflows of liquidity and the respective effect on hypothetical closing price. The paper thereby underlines the different roles investors using market- or limit orders assume during closing auctions.

The third and final chapter studies the compositions of closing order books at the end of the auction. It simulates the outflow of various percentages of liquidity from the top of the order book to visualize the distribution of price dislocations. Subsequently, the effects of the ratio between market- and limit orders on overnight returns are investigated. Finally, the chapter depicts a normalized version of the weighted price discovery contribution metric in order to compare the information content of closing returns to overnight returns.

Zusammenfassung

Die drei in dieser Dissertation vorgestellten Artikel befassen sich mit verschiedenen Themen der Marktmikrostruktur innerhalb von Schlussauktionen in Aktienmärkten. Trotz des wachsenden Marktanteils von Schlussauktionen im Vergleich zur kontinuierlichen Handelsphase hat sich die akademische Forschung bisher relativ wenig mit Schlussauktionen befasst. Diese Dissertation wurde in Zusammenarbeit mit der SIX Securities & Exchanges (SIX) verfasst, wobei granulare Daten auf Ebene von einzelnen Orders in Schweizer Aktien analysiert wurden.

Das erste Kapitel dieser Dissertation befasst sich genauer mit dem Handelsvolumen in Schlussauktionen. Die Analyse misst das aggregierte Niveau der Geduld über alle Marktteilnehmer nach der kontinuierlichen Handelsphase mit unterschiedlichen Marktbedingungen. Die untersuchten Marktbedingungen beinhalten sowohl das Ausführungsrisiko als auch die Liquidität. Darüber hinaus wird untersucht, wie der marginale Effekt in Bezug auf diese Marktbedingungen variiert, wenn Erwartungen über die Höhe des Handelsvolumens während der Schlussauktion berücksichtigt werden.

Das zweite Kapitel analysiert den Prozess der Preisfindung während der 10-minütigen Schlussauktionen auf einer detaillierteren Ebene. Zu diesem Zweck werden Momentaufnahmen des Orderbuchs in kurzen Intervallen über die ganze Auktion erfasst und aufgeschlüsselt. Dadurch können Liquiditätszuflüsse und -abflüsse und die jeweiligen Auswirkungen auf den hypothetischen Schlusskurs erfasst werden. Auf diese Weise wird deutlich, welche verschiedenen Rollen durch die Verwendung von Markt- oder Limitaufträge während der Schlussauktion von Marktteilnehmern eingenommen werden.

Im dritten und letzten Kapitel wird die Zusammensetzung der Orderbücher am Ende der Schlussauktion untersucht. Dabei wird der Abfluss verschiedener Prozentsätze von Liquidität aus dem oberen Teil des Orderbuchs simuliert zur Veranschaulichung der Verteilung von Preisabweichungen. Anschließend werden die Auswirkungen des Verhältnisses zwischen Markt- und Limit Orders auf die Übernachtrenditen untersucht. Schließlich wird eine normalisierte Version der gewichteten Preisentdeckungsbeitragsmetrik dargestellt, um den Informationsgehalt Preisschwankungen während Schlussauktionen und über Nacht zu vergleichen.

Chapter 1

Investor Participation in Closing Auctions

Investor Participation in Closing Auctions*

Louis Müller†

June, 2022

Working Paper

Abstract

In an environment where closing auctions are increasingly favoured over the continuous trading phase, policymakers may be tempted to regulate further outflow from the latter to the former. This study analyses what factors affect the choice between these two trading facilities and thus investor patience in some form. The results indicate that the degree of investor patience varies based on market conditions during the continuous trading phase. Increases in effective spreads and quoted depth enhance investor patience, whereas increases in intraday volatility and price impact have the opposite effect. For high-volume stocks, investors tend to be more patient when market quality deteriorates in comparison to low-volume stocks in the sample. Subsequently, a quantile regression framework is applied to investigate whether investors anticipate closing volume and thus adjust their sensitivities towards market conditions. There is evidence that investors become increasingly patient during periods of high intraday volatility, large intraday price movements and effective spreads on days when closing volume is large. These results imply that investors anticipate higher closing volume, which increases their propensity to trade in the closing auction.

Keywords: Closing auctions, intraday trading, market liquidity, trader patience
JEL Codes: G12, G14

*This work has been enabled by the cooperation of SIX Securities & Exchanges (SIX) and the Institute for Operations Research and Computational Finance (ior/cf-HSG) at the University of St. Gallen. I would hereby like to thank SIX for supplying detailed and granular order-level data for Swiss equities.

†University of St. Gallen, Switzerland: louis.mueller@unisg.ch

1 Introduction

The fact that closing auctions have become an integral part of daily trading cannot be denied anymore. Countless academic and non-academic studies show that market shares of closing auctions have steadily increased throughout the last decade. This tendency is being observed around the world and independent of regulatory environment, even through the extent of it varies substantially. In France, for instance, the market share of closing volume with respect to total volume has almost doubled from around 22% in late 2015 to over 40% in 2019 (Raillon, 2020).

One of the most plausible reasons for this development comes from the fact that there has been a shift from actively managed funds to passive exchange-traded funds (ETFs). ETFs are designed to track a predetermined index as closely as possible without any room for diverging investment strategies. In other words, their performance mirrors the market as closely as possible. In order to accurately reflect their underlying securities, ETFs require frequent rebalancing of positions. Closing auctions on major stock exchanges provide a reliable and representative end-of-day price, which is often used as a market-wide *reference price*. Such increased volumes during the close may also trigger a feedback loop. Execution algorithms of large institutional orders are usually trained to obtain the volume-weighted average price (VWAP) within a given trading day for their trades. With increased closing volume, end-of-day auctions gain more importance for such algorithms, consequently reinforcing the effect.

Concurrently, all other market participants have to make a choice of whether to participate during continuous trading hours or in the closing auction in order to execute their trades. With this decision in aggregation, investors reveal information about their (im)patience to trade. Unlike some theoretical models such as Foucault et al. (2005), Roşu (2009) this paper does not look at patience on a granular level, where impatient traders use market orders for instant execution as opposed to patient traders using limit orders. Nonetheless, both papers predict that liquidity demanders (using market orders) will act less aggressively if they are given an additional opportunity to trade. This work looks at this from a higher level with the choice of trading phase being the issue in focus here. This is more comparable to the work of Madhavan (1992), where investors see closing auctions as their last chance to trade. However, it would be an oversimplification to assume investor patience to be constant and independent of market conditions. More specifically, investors have the opportunity to observe order flow throughout the day and additionally decide when to execute. It is important to note, that postponing trades into the closing is associated with the loss of immediacy, i.e. there is a possibility that prices move against the investor in the meantime. For risk-averse investors, foregoing execution in the continuous phase introduces waiting costs. Under all these considerations, this paper attempts to shed light on the competitive dynamics between continuous trading and the closing auction.

This topic is highly relevant for both investors and regulators alike. Particularly with respect to the recent scrutiny about order flow concentration in the closing and the resulting impact on price efficiency. For policy makers, it is important to understand under which conditions investors choose to participate in closing auctions instead of continuous trading and vice-versa. This knowledge enables them to make more effective regulations to control the distribution of order flow throughout the trading day by setting the right incentives.

In order to address these issues, this paper aims to answer three main research questions. First, it explores how market conditions during the continuous trading phase affect participation in the closing auction. This paper sheds light on the impact of certain *market conditions* may have. *Market conditions* in the context of this study refers to the combination of both execution risk and liquidity. Execution risk in the context of this study is defined as risk for investors that the price may move against them if they don't execute immediately. This risk is mainly driven by realized volatility, which is measured through standard deviations over various time horizons [1]. For robustness purposes, execution risk is additionally also quantified through absolute order imbalances in the context of this study. In contrast to this, liquidity is a more elusive concept. In this context here, liquidity shall be defined following the definition of Kyle (1985, p. 1316) who pointed out three aspects of a liquid market. The first aspect is *tightness*, which refers to the bid-ask spread of a security. In liquid market, such spreads are low, which allows for cheap round-trip trades given no change in price. The second aspect is *depth*, quantifying the amount of one-sided order flow required in order to change the bid or ask against it. Markets with many available securities at the bid or ask are therefore considered deep. The third aspect is *resilience*, which quantifies the ability of a market to revert to the original price after a large one-sided uninformed shock. A shock is uninformed if it occurs out of randomness and does not predict the future course of the stock price. In summary, a liquid market is capable of absorbing large buy- or sell-order without high trading costs and persistent price adjustments. The second research question addressed in this paper is the exploration of whether market participants adjust their behavior based on the anticipation of closing volume. Raillon (2020) conceptually proposed that this anticipation could lead to the accumulation of investors using VWAP algorithms in order to optimize their execution.

[1] Volatility is measured over multiple time horizons in order to obtain more robust results as there are different kinds of volatility, such as very short-term quote flickering or more persistent price swings.

Even though the end-of-day closing of trading days is becoming more prominent in recent academic literature the topic of closing market share has not yet been thoroughly researched in current market conditions. There have been some studies two decades ago, scrutinizing the introduction of closing auctions by means of an event study, e.g. by focusing on measures of liquidity before and after the introduction of the new closing mechanism. Some existing empirical work that comes closest to this paper is by Pagano et al. (2013), who analyze the introduction of closing auctions on the NASDAQ in parallel to its existing continuous market. The new call auction mechanism (introduced in 2004) significantly reduced both spreads and volatility in the continuous market at the end of the day. In contrast, opening call auctions had much less effect on continuous hours. Moreover, the authors find that the event did not remarkably affect overall trading volume, i.e. volume from continuous trading hours merely shifted into closing auctions, but there was no *new* volume entering the aggregated market. In a comparable setting, Kandel et al. (2012) analyze the introduction of closing auctions in Italy and France. They yield very similar results, indicating a reduction of both volatility and spreads before the closing, but no impact on continuous trading otherwise. Moreover, they also observe a significant shift of volume from continuous trading to closing auctions, thus confirming the findings of Pagano et al. (2013). Importantly, Kandel et al. (2012) found that closing auctions attract significantly more volume, if their outcome represents the market-wide reference price. This endorses the fact that closing prices serve as an important market-wide reference beyond a single venue. Other noteworthy studies that find improved liquidity around the close and reason this with increased trader patience caused by the introduction of closing auctions include Aitken et al. (2005), Hagströmer and Nordén (2014), Inci and Ozenbas (2017), Kuo and Li (2011), and Pagano and Schwartz (2003). In summary, the evidence strongly for closing call auctions as a tool to enhance liquidity at the end of the day.

However, there are also critical voices in academia about the call auction mechanism in particular. While the previously mentioned papers found evidence in support of opening and closing using the call auction mechanism, this may depend heavily on the size of the stock. Call auctions are certainly a preferred solution if they are blessed with abundant liquidity. Importantly, call auctions have a chance of failure[2], in which case no shares are traded whatsoever. To underline this point, Ellul et al. (2005) analyze whether investors on the London Stock Exchange (LSE) preferred the call auction instead of an off-exchange dealership with guaranteed execution for their trades. The authors find that investors tend towards the dealership solution for smaller stocks, despite the inferior price efficiency. Similarly, Ibikunle (2015) find that small-cap stocks suffer from high failure rates at the

[2]Auction failure can occur in cases when the order book cannot be crossed, such that the lowest ask is higher than the highest bid.

opening due to lack of volume. Both of these studies imply that call auctions may not be the optimal solution for smaller stocks.

Another related strand of literature on auctions is more concerned with the accuracy of auction prices. From this angle, call auctions also seem to have a positive impact on market quality. For instance, Chang et al. (2008) analyze the introduction of call auctions on the Singaporean stock exchange in the year 2000. They find remarkable reductions in volatility, implying that prices with call auctions are more accurate and have less transitory noise incorporated. Moreover, they find that the introduction of call auctions significantly removed correlations between intraday- and overnight returns as well as a reduction of price reversals after the market re-opens in the morning. Similar results have been found after the tiered introduction of call auctions on the Paris stock exchange in 1996 and 1998 (Pagano & Schwartz, 2003), and additionally in an aggregated sample of 43 stock exchanges worldwide (Cordi et al., 2018).

While this study is not attempting to discredit any of the findings in the studies previously mentioned, markets (including closing auctions) have changed considerably in the last two decades, mostly through the migration to electronic trading and increasing removal of the human aspect from execution. More recent empirical work on closing auctions has been done by Bogousslavsky and Muravyev (2020). They find that the closing auction only has a small contribution to price discovery due to the influence large rebalancing activities by ETFs. This distorts closing prices with respect to the last observed midquote prior to the auction. However, this distortion is not based on new information on fair value but rather caused by large institutional order imbalances. Similarly, Hu and Murphy (2020) show that closing prices are much less efficient (i.e. revert over night) on so-called triple-witching days[3], when much uninformed volume is traded in the closing auction.

What the majority of these papers have in common, is that they find some (mostly positive) effect of call auctions on the remainder of the trading day. In contrast to those results, this paper focuses on the impact market conditions have on investors participating in these auctions. For this purpose, the participation in the closing auction for trade execution is interpreted as a choice for investors. Therefore, the propensity to wait until the closing auction for execution is a manifestation of *investor patience* (Kandel et al., 2012). In contrast to this, impatient investors prefer the immediacy of the continuous trading phase, which precedes the closing auction. Hence, investors have the choice between these two trading facilities based on their level of patience. However, each of these facilities has

[3]Triple-witching days occur on days when index futures, index options and stock options expire all on the same days. Delta neutral market participants who have exposure to any of these instruments will be required to offset their positions on that day to remain neutral.

different properties that may be preferred under varying market conditions.

The main contributions of this paper to the existing literature are two-fold. The first contribution is the presentation of evidence that investors tend to adjust their participation in closing auctions based on intraday market conditions. This speaks in favor of the hypothesis of Pagano (1989b), under which different trading facilities offer different features. It shows that investors actively monitor market conditions in order to decide which facility offers superior execution quality. The second contribution of this paper is related to the clustering of volume over the trading day. More specifically, it is shown that there exists a *pull-equilibrium* across trading facilities. In other words, there is evidence that the propensity of investors to trade in the closing auction increases when they expect the volume to be larger.

The paper takes advantage of comprehensive order-level data from SIX Securities & Exchanges (SIX), sampled over the entirety of 2018 and 2019, consisting of 498 trading days. The data comprises of all the 30 Blue Chip constituents of the Swiss Leadership Index (SLI). In order to evaluate investor patience, this study takes advantage of the fact that on SIX the continuous and closing auctions are strictly non-overlapping. This allows for the clear temporal separation of order flow across time. Therefore, the realized volume in the closing auction cannot have a direct effect on continuous market conditions. This convenient feature mitigates issues of endogeneity based on a Granger (1969) type argument. In addition to this, there are numerous factors that influence the closing volume, that are not in the scope of this paper. Therefore, a set of control variables is introduced, which includes metrics like dummy variables for the closing of derivative positions, expected- and unexpected continuous volume as well as continuous absolute returns. These metrics alone are able to explain around 41% of the variance of closing volumes. In contrast to this, market condition metrics of interest are divided into execution risk (intraday volatility and order imbalance) on the one hand and liquidity (tightness, depth, resilience (Kyle, 1985)) on the other hand.

The first set of results in this paper are based on a panel regression framework, which controls for all the out-of-scope metrics listed above. The results indicate that investors adjust their choice of participation in the closing auction based on continuous market conditions. More specifically, marginal changes in market quality during this phase have an impact on how much volume is executed in the closing auction. The evidence shows that a deterioration of execution quality has the opposite effect on closing participation as a deterioration in liquidity. On one hand, elevated levels of execution risk make investors less patient, in that they prefer immediate execution and therefore chose the continuous phase over the closing auction. On the other hand, deterioration of liquidity has the opposite effect, making investors more patient.

The second part of the results is extracted by modeling the anticipation of investors with respect to closing auction volume. For this reason, it extends the previous analysis by a quantile regression, which aims to explain the within variance of stocks. A quantile regression is an appropriate statistical technique, since it estimates one set of coefficients for each quantile of the dependent variable. If there are no expectations about the closing volume whatsoever, the coefficients (i.e. the sensitivities towards a change in market conditions) for each quantile would be identical. The results presented clearly reject this hypothesis. There is strong evidence that the propensity to choose the closing auctions upon change in market conditions increases when anticipated closing volume is high. This speaks in favour of a pull-equilibrium (similar to Pagano (1989a)) between the two trading facilities.

The remainder of this paper is structured as follows. Section 2 provides information about the market structure and institutional background. Section 3 will elaborate the data in detail and provide summary statistics. Section 4 will build a variety of testable hypotheses with respect to closing participation. Section 5 will build and explain the basic methodology used. Subsequently, section 6 will summarize all findings and either approve or reject the hypotheses stated prior. Finally, section 7 will conclude this work.

2 Institutional Background

The recent increase in closing volumes as opposed to the rest of the trading day can be explained by various factors. Raillon (2020) for instance, argues that this trend is facilitated mainly by four developments in the trading landscape:

(i) **Passive investing** has seen massive inflows through ETFs in recent years. Closing prices provide a precise benchmark on how to rebalance positions of passive funds. Moreover, index funds engage in huge block trades which require high liquidity. Participation in closing auctions provides ETFs with an opportunity to trade with superior liquidity and minimal tracking error.

(ii) **Best-execution** requirements, that were introduced under the Markets in Financial Instruments Directive (MiFID) II (effective early 2018), force brokers to trade in the best interest of their clients, i.e. with the lowest execution costs. Since closing prices are set in one single market only, brokers do not have to compare quotes across different venues. Instead, they can be certain to obtain the best execution price on behalf of their clients.

(iii) **Adverse selection** is an important consideration for traders during the day. In this context, it refers to the situation where some market participants have more information and are therefore making a profit at the expense of less informed traders. This is particularly problematic during the continuous trading phase with high market fragmentation. In this case, it is possible that the same security has different prices on other platforms for brief periods in time (Budish et al., 2015). Very fast trading strategies can thrive in those conditions (Biais et al., 2015).

(iv) **Execution algorithms** are learning that there is higher liquidity and potentially better execution quality during closing (according to (iii)). This in turn emphasizes end-of-day trading even stronger, eventually triggering a *positive feedback loop*. A similar argument is made in Pagano (1989b).

From these four points it becomes clear that the closing auction through its design has a structural advantage over continuous trading in several regards. The most important factor, however, is the *reference price* function of the closing price. This fact has been shown empirically by Kandel et al. (2012), where the closing auction attracts significantly more volume when it is equated to the market-wide *reference price*. This observation suggests that market participants are generally risk-averse by their willingness to settle for the reference price and foregoing potentially better prices. This logic can be extrapolated to the rest of the trading day. In this case, market participants would be more inclined to trade during the closing auction when intraday volatility is high. Since higher volatility entails potentially larger discrepancy between the obtained price and the closing price, investors would rather chose the same-day *reference price*. This is particularly true for brokers who must adhere to best-execution requirements, since finding the best price in fragmented markets on a volatile trading day may be extremely difficult.

For the analysis in this paper the focus lies on the equity market of Switzerland's main stock exchange SIX Securities & Exchanges (SIX). On this specific exchange, the trading day is split into three main trading periods. In the morning, the day is started by an *opening auction* which takes place before 9:00am in the morning. The opening auction is followed by a *continuous trading phase* starting 9:00am and lasting until 5:20pm, lasting for 500 minutes. It therefore constitutes the longest phase and also the most significant in terms of trading volume (see table 1). Finally, the trading day is concluded with a *closing auction* from 5:20pm until 5:30pm. In contrast to many other exchanges in Europe, closing on SIX lasts for 10 minutes instead of the more common 5 minutes on other European exchanges. Despite trading being possible during all three periods, they employ two fundamentally different types of market structure as defined by Madhavan (1992):

1. **Order-driven market.** This type of market structure allows for continuous trading and is therefore implemented during the continuous trading phase. The market consists of limit orders signaling the willingness to either buy or sell at a predetermined price. Investors who want to trade at the prevailing prices have the opportunity to execute against the best limit orders of either side. This results in a dynamically changing limit order book. On the one hand, liquidity providers continuously adjust their limit orders to reflect the current state of information. On the other hand, liquidity demanders have the option to trade when they see prices dislocated from fundamental values.

2. **Call auctions.** In contrast to order-driven markets, call auctions aggregate orders over a period of time into a single price. The resulting price is determined through crossing the aggregated demand and supply of all orders that have been submitted during the period. Hence, this price is often referred to as the *uncrossing price*. Call auctions can therefore be understood as the consolidation of order flow over a given time period (Madhavan, 1992). This mechanism is particularly adequate for both *opening* and *closing* of the trading day. It has been shown empirically, that the introduction of call auction markets at either the beginning or ending of the trading day has positive effects on market quality. The beneficial effects mainly consist of a reduction in volatility and bid-ask spreads[4], improved price efficiency[5] and less price manipulation[6].

The closing auctions on the Swiss stock exchange are designed like call auctions. Over the course of the auction, investors receive several pieces of information that are continuously updated. Those include a hypothetical closing price and -volume, and the state of the order book on the main exchange. Despite most orders for the auction being submitted during the auction between 5:20pm and 5:30pm, the exchange allows investors to enter so-called ATC (at-close) orders during the continuous trading phase. These orders are invisible to other market participants and are activated once the auction starts[7].

[4]See Aitken et al. (2005), Hagströmer and Nordén (2014), Inci and Ozenbas (2017), Kuo and Li (2011), and Pagano et al. (2013).

[5]See Barclay et al. (2008), Bellia et al. (2017), Biais et al. (1999), Comerton-Forde et al. (2007), Cordi et al. (2018), and Pagano and Schwartz (2003).

[6]See Comerton-Forde and Putniņš (2011) and Hillion and Suominen (2004).

[7]Less than 1% of orders for closing auctions are submitted before the start of the auction. Therefore, the vast majority of orders are submitted within the closing auction

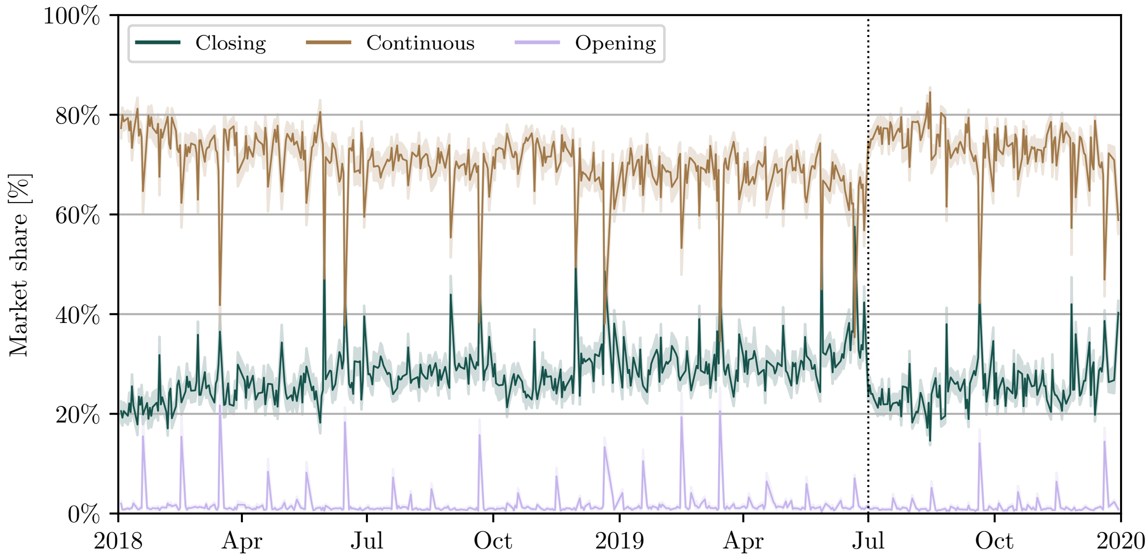


Figure 1: Daily average market share of different trading periods. The market shares by volume of each period were computed in daily frequency and divided by total volume for each individual stock in the sample. The plot includes the respective 99% confidence interval calculated daily using 500 bootstraps. The dotted vertical line represents the day when the European Union (EU) repealed the equivalence status of the Swiss stock exchange on July 1, 2019, which drove a lot of additional order flow into the local market. The top line, represents the continuous trading phase with the highest share of trading volume. The middle line represents the share of closing volume. The bottom line represents the share of trading volume during the opening auction.

3 Data

This study relies on level 2 order book data obtained directly from SIX. The sample period comprises the two full calendar years of 2018–2019 and only contains orders and trades in Swiss equities. The sample was deliberately terminated before the start of 2020 in order to avoid the extreme market conditions caused by the pandemic outbreak in the beginning of the subsequent year. The selected two-year time frame includes 116 million trades as well as 3.4 billion orders.

Figure 1 shows the evolution of market shares of trading volume throughout the sample. The volume is segmented by trading period and averaged across stocks in order to obtain the average market share in the cross-section of stocks. The chart reveals how the continuous phase dominates by market share, followed by the closing auction. Moreover, the chart visually reveals a steady increase of closing market share (approx. from 20% to 40%) at the expense of continuous market share (approx. from 80% to 60%) over the course of 1.5 years prior to the removal of *exchange-equivalence*[8]. After the structural break on July 1st 2019, this trend has resumed. This observation underlines the increasing importance of closing auctions throughout time, similarly stated in Raillon (2020) for the French market of CAC40 titles on Euronext Paris.

[8]On July 1st 2019, the EU commission decided to repeal the status of SIX which treated them as an equivalent trading platform to any other European venue. Consequently, European stocks could not be traded on SIX anymore and vice-versa.

In addition to this, the chart also shows frequent spikes of auction market shares in regular intervals at the expense of continuous trading. The reason for these irregularities are expiration dates of various derivatives contracts, such as options and futures. Market participants who still have open interest in such contracts at expiration may close their positions on the equity market. Therefore, events in the derivatives markets have clearly visible effects on the equity market depicted in fig. 1. On SIX there are two important events in this regard. First, option contracts can be bought with monthly expiration dates that are set to the third Friday of each month. Second, futures contracts expire quarterly on the third Friday of the months March, June, September and December.

In order to mitigate the problem of failing call auctions due to lack of limit orders (Ibikunle, 2015), the study only includes the 30 constituents of the SLI at the end of the sample period. Additions and removals from the index are not considered. Consequently, both ALC and SIKA have fewer observations due to its later addition to the SLI. All of the measures considered in this paper are directly extracted from the raw order book data, without reliance on any third-party vendors. The descriptive statistics in table 1 are presented on the basis of all stocks included in the sample. The table shows that there are substantial differences in terms of trading volumes in the cross-section of stocks, particularly in terms of the closing volume, ranging from an average of CHF 7mn. to CHF 132mn. In total, SIX facilitates around CHF 2.18bn. in continuous trading and almost CHF 900mn. in closing volume on an average trading day.

4 Hypothesis Development

It has been a long known fact that trading volume is clustered throughout the trading day. Admati and Pfleiderer (1988) were among the first to point out persistent patterns based on the existence of informed and uninformed traders. Their Kyle (1985) type model predicts that both types of traders will cluster around the same periods during the trading day which leads to pronounced patterns over time. What Admati and Pfleiderer (1988) call *discretionary liquidity traders*, i.e. market participants that are unconstrained in the timing of their execution, are at the core of this study. A similar argument is made by Pagano (1989a), who develops a theoretical prediction that the lack of volume can trigger a vicious cycle through increased volatility, which in turn will make trading even less attractive. Other notable studies include empirical evidence from McInish and Wood (1992, p. 760), who visually document a reverse J-shaped pattern, with the highest volume at the opening. The existence of such trading patterns indicates, that investors are flexible in terms of when they are willing to execute their orders throughout the trading

		Continuous trading volume							Closing auction volume						
symbol	obs.	μ_t	σ	min	25%	50%	75%	max	μ_t	σ	min	25%	50%	75%	max
ABBN	498	89.68	37.52	30.89	65.75	81.18	101.98	288.49	36.73	17.76	11.15	26.72	33.13	42.65	196.96
ADEN	498	37.53	16.85	12.36	27.77	34.34	43.02	186.24	16.17	7.11	5.07	11.72	14.59	18.96	55.55
ALC	181	73.69	79.35	17.94	44.56	55.06	75.23	910.22	25.79	29.85	6.47	15.59	19.94	26.14	364.80
AMS	498	44.23	33.29	9.67	24.21	34.70	54.11	319.09	8.05	6.16	0.00	4.17	6.59	10.43	83.69
BAER	498	32.26	14.46	9.98	22.72	29.37	38.40	124.03	14.84	14.50	4.46	10.52	13.18	16.54	311.20
CFR	498	89.22	37.78	29.34	66.60	81.15	102.14	395.51	37.10	16.00	12.71	27.66	34.17	42.75	160.38
CLN	498	25.68	19.89	7.23	16.19	20.72	29.30	272.77	7.68	5.06	2.06	5.23	6.61	8.86	62.13
CSGN	498	103.52	42.23	33.22	75.88	95.49	118.97	412.69	35.63	18.46	9.89	25.99	32.37	41.05	223.98
GEBN	498	40.85	22.21	9.61	28.94	36.20	46.29	249.62	17.84	9.15	5.31	12.42	15.86	20.84	76.07
GIVN	498	46.52	18.51	18.29	34.69	42.73	53.50	165.05	19.26	9.84	5.70	14.04	17.32	21.41	94.35
KNIN	498	22.48	11.79	6.65	14.98	20.06	26.74	103.79	9.84	6.90	2.59	6.99	8.87	10.86	116.36
LHN	498	75.42	31.02	27.73	54.59	69.61	88.66	301.97	31.66	15.85	9.18	22.49	28.75	36.68	184.02
LOGN	498	30.00	19.05	8.18	20.21	25.41	33.67	216.01	7.98	3.24	3.10	5.73	7.25	9.27	23.26
LONN	498	64.55	33.21	17.08	43.84	55.98	73.88	352.62	23.88	10.76	7.48	17.89	21.72	26.91	134.65
NESN	498	282.92	84.46	107.02	220.49	275.67	327.01	734.22	132.45	102.02	24.95	94.46	113.11	141.53	1141.42
NOVN	498	245.93	84.77	101.07	188.71	232.08	288.50	709.83	110.74	74.67	35.10	78.71	96.77	119.01	782.19
PGHN	498	33.77	15.16	12.38	24.16	30.64	38.92	182.38	13.24	7.62	4.39	9.33	11.78	14.79	106.34
ROG	498	235.00	84.16	96.83	177.49	218.88	275.35	733.06	112.82	65.57	29.42	83.19	100.42	125.19	740.19
SCHP	498	24.79	11.83	8.38	18.00	22.47	28.00	143.63	8.28	3.44	2.34	6.10	7.69	9.51	28.12
SCMN	498	46.36	20.10	18.33	34.73	42.53	51.60	261.67	21.15	11.14	7.55	15.38	19.03	23.68	104.32
SGSN	498	30.74	12.95	10.33	22.68	28.17	36.16	152.11	13.96	6.74	4.14	10.20	12.67	15.75	65.13
SIKA	388	50.30	23.46	16.11	34.43	45.92	58.06	194.87	17.69	9.95	6.86	13.06	15.74	19.27	112.00
SLHN	498	49.97	13.55	19.99	40.66	47.81	57.47	106.28	14.73	7.65	5.13	10.57	13.29	16.94	75.54
SOON	498	34.48	18.71	10.91	23.91	29.87	40.93	249.73	12.63	8.06	3.65	8.67	11.15	14.38	100.35
SREN	498	81.78	36.06	29.02	59.92	72.91	92.10	393.15	35.32	26.35	10.45	24.27	29.89	36.92	257.97
TEMN	498	31.00	26.93	7.20	17.92	24.48	36.21	433.03	11.01	27.32	3.25	6.77	8.80	11.29	607.47
UBSG	498	133.23	48.21	46.68	99.84	124.98	157.35	361.97	48.98	26.66	14.42	36.00	42.98	55.50	290.63
UHR	498	55.73	27.20	16.69	40.19	49.79	64.08	287.71	22.16	8.95	6.76	15.60	20.42	26.74	63.86
VIFN	498	25.04	14.58	7.60	16.00	21.09	28.63	141.25	6.73	4.08	1.93	4.69	5.97	7.35	53.64
ZURN	498	102.84	43.87	32.23	75.25	94.14	118.28	366.56	41.37	24.40	10.91	30.54	36.79	45.76	270.58

Table 1: Descriptive statistics. This table represents the descriptive statistics across all 30 SLI titles included in the analysis. It provides insights into the distribution of volumes across the continuous- and closing phase, which constitute the two most significant phases throughout the trading day by volume. All volumes are in million CHF and the measures were computed on a daily basis. The sample includes all trades and orders conducted in the years 2018 and 2019. The selection of titles represents all constituents of the SLI at the end of 2019 and thus the sample period. Some of the titles were included into the SLI at a later point in time, which means that there are fewer observations covered.

day.

When analyzing trading clusters, the consideration of *trader patience* plays an important role. Two of the most influential papers on this topic come from Foucault et al. (2005) and Roşu (2009), who both develop theoretical models with market participants, that are either *patient* or *impatient*. Both papers find that in equilibrium, patient(impatient) market participants prefer limit(market) orders. A different approach to quantify trader patience has been taken by Pascual and Veredas (2009), who propose a sequential ordered Probit model to disentangle patient from impatient traders. They find that the composition of the limit order book has a significant impact on the order choice investors make.

This concept of trader patience can also be extrapolated from a top-down perspective within each trading day. It has been shown that institutional investors use algorithms to execute large orders throughout one or several trading days in small increments in order to obtain optimal execution conditions and prevent other market participants from front-running their orders. Examples for this are well documented in Korajczyk and Murphy (2019) and Van Kervel and Menkveld (2019). These algorithms observe the same market signals and react to certain market conditions in potentially similar ways. These observations may in return trigger adjustments in trading strategies of each market participants in similar ways.

Consequently, trading patterns of market participants may be correlated during the trading day, which also includes the choice between the continuous trading phase and the closing auction, particularly with respect to the increasing importance of the closing auctions with respect to the remainder of the trading day. This trend has been observed throughout the last decades and particularly on European exchanges (Raillon, 2020). In the following sections, this study analyses two ways in which the allocation between these continuous trading phase and closing auction can be influenced, i.e. market conditions and expectations of closing volumes.

4.1 Continuous Market Conditions

The first set of hypotheses is designed to scrutinize the aggregated participation of market participants in the closing auction after observing market conditions during the continuous trading phase. Since there is no overlap between these two trading phases the issue of endogeneity at least is mitigated. Naturally, there are many factors during the continuous phase that may drive investor behavior. This set of hypotheses mainly focuses on two types of *market conditions* that are essential to investors in order to perceive a market to be of high quality. The term *market conditions* will later be used in the development of hypotheses and encompasses both execution risk and liquidity, as defined in the introduction above. It shall be noted that this study does not take price efficiency[9] into consideration as this is less important for the execution of trades unless there is speculative motivation.

In the past, various studies have been conducted on the reaction of investors with respect to market conditions. For instance, it can be observed that momentum strategies are commonly used by investors during the continuous trading hours. Empirical evidence of this can be found in Duffie (2010) and Gao et al. (2018), and a theoretical reasoning in Bogousslavsky (2016) who explains this by the infrequent rebalancing of *slow capital* from financial institutions. This supposes that some market participants react more quickly to changing market conditions than others. Therefore, some market participants may decide to execute their planned trades as soon as possible instead of waiting until the closing auction on days with increased execution risk. One reason for this behavior may be to prevent prices from further moving against them and hence revealing a certain degree of risk-aversion when volatility is high. For closing auctions, such behavior entails an outflow of volume from market participants who become impatient during the trading day. Another reason for this behavior may be founded in the argument of Foucault (1999), who predicts that investors are more likely to use more aggressive limit orders when volatility is low. This implies that investors are afraid of their limit orders being stale and having someone else taking advantage of this. Consequently, investors become less patient with their execution when volatility is high as immediacy becomes more important under the possibility of sub-optimal execution at close.

In addition to price-related measures such as price volatility and price movements, market conditions importantly also include liquidity-related measures. Both liquidity and volatility are closely related conditions in the market. Glosten and Milgrom (1985) and Kyle (1985) for instance explain the link between a deterioration of liquidity under increased price volatility. In particular, spreads become wider and depth becomes shallower when prices are more noisy. This can be explained by market makers or dealers adjusting their quotes based on the uncertainty and potential adverse selection that they are facing on a given trading day. Moreover, on volatile trading days, execution costs are expected to be higher, particularly for very large trades as liquidity becomes a decisive factor in order to minimize slippage[10]. It needs to be noted, however, that high volatility is not a necessary precondition for a deterioration in liquidity. In line with the previous argument, sub-par intraday liquidity may reduce investor patience and increase the inclination for

[9]Price efficiency is obtained if the security price in a market does not have any auto-correlation, such that it is martingale.

[10]Slippage is defined as the amount the price moves against the trader when executing a trade. For instance, if an institution wants to buy a large amount of stock, the price will tend to rise throughout execution.

immediate execution. On days with less liquidity during the continuous trading phase, investors may either expect the situation to not improve or even deteriorate until the end of the day. This makes such investors less likely to wait until the closing auction, consequently shifting volume into the continuous trading phase at the expense of the closing auction. Under this scenario, investors are becoming less patient when liquidity deteriorates.

Hypothesis 1 (H/1)**.** *Trading volume in closing auctions decreases when continuous market conditions deteriorate.*

This argument can also credibly be made from the opposite direction, where investors are becoming more patient when market conditions deteriorate. Under this inverse hypothesis, investors do not want to execute their trades in order to avoid potential unfavorable execution and minimize tracking error. A similar argument is made by Raillon (2020), who suggests that recently implemented best-execution requirements motivates the participation in closing auctions in order to obtain the market-wide *reference price*. The more strict best-execution requirements under MiFID II increasingly emphasizes the process of trade execution by brokers in terms of their client orders. In an increasingly fragmented market, obtaining optimal execution may be costly as quotes across venues may differ for short periods in time (Budish et al., 2015). Large intraday volatility makes this process only more complicated. Therefore, it can be expected that price swings or intraday noise may have an effect on closing participation. This mechanism would be even stronger on more volatile days, when efficient execution is more challenging. On those days, the closing auction offers safer execution with less price uncertainty. In addition to this, financial institutions have sophisticated algorithms nowadays in order to minimize their trading cost. For such algorithms, liquidity is a central concept required for optimal execution. Algorithms observing an illiquid intraday market (i.e. following the definition of Kyle (1985) through large spreads, low depth and/or low resilience) may be tempted to wait until the closing auction for more certainty in execution and less slippage. If many investors simultaneously make the same decision, i.e. through highly correlated trading algorithms, an increasingly liquid closing auction becomes a self-fulfilling prophecy. Meanwhile, volume is drained from the continuous trading phase as market participants become more patient in aggregation.

Hypothesis 2 (H/2)**.** *Trading volume in closing auctions increases when continuous market conditions deteriorate.*

4.2 Volume Clustering

The second contribution of this paper is concerned with the clustering of volume across the trading day. Academic literature has found this phenomenon in various studies within the continuous trading phase. However, it has not been explored across different trading phases or -facilities on the same trading platform. Examples of existing academic work most importantly include Admati and Pfleiderer (1988), who show theoretically that traders with private information will execute their trades when liquidity is highest, in order to minimize price impact and optimize their returns. This implies that they are patient enough not to require immediate execution. Consequently, volume will cluster across the continuous trading phase through a virtuous cycle. Similarly, Pagano (1989a) presents a theoretical model in which investors seek the highest liquidity[11] intraday in order to execute their trades without taking into consideration private information. This phenomenon has been well explored empirically, resulting in volume clustering after the opening and before the closing of the continuous trading phase. For instance, McInish and Wood (1992) observe a *reverse J-shaped* volume within each trading day. They attribute this to investors wanting to trade on new information that has surfaced overnight and to investors who have to reach their target allocation before the end of the day. The latter observation already indicates a reduction of investor patience towards the end of the trading day.

More applicable to closing auctions, Pagano (1989b) shows that competing trading platforms with identical properties cannot exist in equilibrium. The reasoning behind this is that investors would seek the platform with superior liquidity since there is no other measure to discriminate between the platforms. Consequently, all the volume will shift onto one platform making the other one obsolete. This argumentation can be extrapolated to the continuous trading phase and the closing auction. Both of these trading facilities compete for volume. However, in contrast to the assumptions from Pagano (1989b), these facilities have widely varying properties and are preferred by different investors. Therefore, the migration of all volume into one of these facilities is not being argued here, but rather that the set of market participants who is agnostic may choose in favor of where liquidity is superior[12]. This is the same logic behind trading algorithms who allocate their orders based on the expected volume, thus creating a positive feedback loop as argued by Raillon

[11] Liquidity in the context of Pagano (1989a) is defined in the sense of the *depth* market, i.e. how much passive volume is accessible to incoming market orders. Hence it captures periods when large orders can be executed without significantly affecting price.

[12] It also needs to be addressed that since the theoretical predictions of Pagano (1989b), capital markets have changed dramatically due to the introduction of electronic trading platforms and high-frequency traders (Carrion, 2013; Conrad et al., 2015; Hagströmer & Nordén, 2013; Hasbrouck, 2018). This led to significantly better price efficiency across platforms with diminishing mispricing (Budish et al., 2015).

(2020).

Hypothesis 3 (H/3)**.** *Investors are more willing to choose the closing auction if they expect high closing participation.*

This hypothesis implies a game-theoretic pull equilibrium among trading phases, where investors simultaneously decide on where to allocate their volume under the expectation of how other market participants will decide. Therefore, a clustering of volume in the closing auction should be observed to a larger extent when market conditions are more extreme.

5 Methodology

Quantifying investor patience is hard to achieve empirically. The disentangling of endogeneity issues arising with respect to other factors such as liquidity is particularly hard to obtain. One attempt to address this problem has been done by Pascual and Veredas (2009), who study the influence of liquidity on trader patience on a very granular basis on the Spanish Stock Exchange[13]. Similar to Foucault et al. (2005) and Roşu (2009), they assume that patient(impatient) traders have a preference for limit(market) orders. They find that in fact investors adjust their orders based on both the best quotes as well as the shape of the order book beyond the best quotes.

5.1 Regression Framework

Similar to Pascual and Veredas (2009), this study isolates trader patience from other market conditions. However, this study views patience from a much more top-down perspective over the entirety of the trading day and not on the level of the limit order book. For this purpose, the volume from the continuous trading phase is separated from volume of the closing auction, keeping in mind that the latter is determined strictly after the former. In other words, market conditions during the closing cannot impact the continuous auction, but the other way around is very well possible. Consequently, investors can decide to postpone their trades from the continuous trading phase to the closing based on unfavorable market conditions but not the other way around. This approach is somewhat similar to classic Granger (1969) causality, where causal relationships are expressed by shocks in one time-series leading another.

[13] Pascual and Veredas (2009) propose a sequential ordered Probit model that is separated into two steps. The first step determines the likelihood of an incoming trader to be either patient or impatient based on the state of the limit order book. In the second step, the arriving trader decides on the type of order to submit with respect to the current limit order book

Even after accounting for the temporal separation of both trading phases, it is important to account for potential correlation of volume shocks across securities. Such shocks can be caused by a variety of factors, e.g. macro news or release of company earnings. The challenge in this research design is mainly the distinction between expected and unexpected order flow. In econometric terms, the latter can be interpreted as the deviation from an expectation given in a pre-defined system. In order to account for unexpected shocks, a methodology similar to Barclay and Hendershott (2008) is applied. In their work, they attempt to determine the informativeness of opening prices segmented by expected and unexpected shocks, using autoregressive processes. In a similar manner, this paper replicates the expectation of trading volumes, based autoregressive processes estimated for the entire sample period. In order to keep the changes in volumes comparative for inferential reasons later on, all volumes are measured as logarithms of trading volumes with base two as

$$\log_2[V] \quad \forall \quad V \in \{V_{cont}, V_{close}\}$$

where V_{cont} represents the trading volume in currency terms during the continuous trading phase and V_{close} for the closing auctions. This procedure has the advantage of measuring the variances of volumes on a comparable scale for both large and small stocks, while constant differences can be accounted for by a simple constant term. Since this procedure is a strictly monotonous transformation of volumes, it preserves important statistical properties of each time series. One of these properties is non-stationarity in the time series. Since the aim of this paper is to determine how volume flows in and out of closing auctions, stationarity is a condition for meaningful inference and for the prevention of spurious regressions. For this purpose, a set of cross-sectionally augmented Dickey-Fuller (CADF) regressions is applied to volumes $y \in \{\log_2[V_{cont}], \log_2[V_{close}]\}$. They differ from ordinary augmented Dickey-Fuller tests, in that CADF allows for joint testing of non-stationarity within groups. For this purpose, the methodology described in Baltagi (2005, Section 12.3) is followed, which proposes individual regressions for each stock s, whereas d represents the trading day. The CADF regressions are estimated over $p = 5$ lags to represent the last five trading days (including the same weekday one week prior).

$$\Delta y_{s,d} = \alpha_s + \rho_s^* y_{s,d-1} + \delta_0 \overline{y}_{d-1} + \sum_{j=0}^{p} \delta_{j+1} \Delta \overline{y}_{d-j} + \sum_{k=1}^{p} \phi_k \Delta y_{s,d-k} + \varepsilon_{s,d} \tag{1}$$

In this equation, the cross-sectional average for each time period is included in the regression where $\overline{y}_d = N^{-1} \sum_{\hat{s}} y_{\hat{s},d}$. Moreover, the differences in the variable of question are defined as $\Delta y_{s,d} = y_{s,d} - y_{s,d-1}$ and $\Delta \overline{y}_d = \overline{y}_d - \overline{y}_{d-1}$. In order to derive the test statistic, Pesaran (2007) proposes to isolate all the t-statistics T_s of the lagged value ρ_s^* coefficients across regressions in (1). The t-statistics are then aggregated as an average to receive the test statistic CISP $= \mathbb{E}[T_s]$. This results in tests statistics of -5.892 for continuous- and -6.123 for closing volumes. Given the critical values[14] derived in Pesaran (2007), this result is highly significant for both time series and clearly rejects non-stationarity. Consequently, the problem of a spurious panel regression is ruled out for the volumes of both trading facilities.

In the further course of this paper, it is important to accurately control for the correlation between continuous- and closing volume as suggested in Barclay and Hendershott (2008). For that purpose, a set of $AR(p)$ models is estimated that are computed for each stock s individually

$$\log_2[V]_{s,d} = \alpha_s + \gamma_s \cdot d + \sum_{l=1}^{p} \varphi_{s,l} \cdot \log_2[V]_{s,d-l} + \varepsilon_{s,d} \qquad \forall \quad V \in \{V_{cont}, V_{close}\}. \tag{2}$$

All models were estimated using a constant α_s with a linear time trend γ_s. Figure 6 in the appendix shows the information criteria of different model specifications with $p \in \{0, 1, \ldots, 10\}$ to predict volumes for both continuous- and closing volumes (denoted as $V \in \{cont, close\}$) separately. For the continuous trading phase, the volume follows some autoregressive pattern with at least one lag in terms of all information criteria. Models with two lags explain around half the variance in continuous volume with an average R^2 of around 39% across stocks on a daily basis. For closing auctions, however, this is not the case. They are less predictable based on past realizations, with the BIC and HQIC (being the more strict information criteria) suggesting 0 lags (i.e. no autoregressive properties) as the optimal look-back period[15].

In a next step, the same specification of $p = 6$ is applied to compute the in-sample fitted values $\widehat{\log_2[V_{cont}]}$ as well as the error term $e(\log_2[V_{cont}])$. This estimation is only computed on the continuous trading volumes as the closing volumes do not have any autoregressive properties. Based on this procedure, the former will be referred to as expected volume and the latter as unexpected volume in the continuous trading phase. This decomposition is similarly executed by Barclay and Hendershott (2008) for opening price efficiency on the Nasdaq in order to account for variance in volumes.

$$\log_2[V_{cont}]_{s,d} = \underbrace{\widehat{\log_2[V_{cont}]}_{s,d}}_{\text{expected}} + \underbrace{e(\log_2[V_{cont}])_{s,d}}_{\text{unexpected}} \tag{3}$$

[14]Critical values depend on both the number of cross sections N and the number of time periods T. For the case of an intercept (see (1)) with $N = 30$ and $T = 500$, Pesaran (2007, p. 280) the critical values for CISPas follows: -3.84 (1%), -3.24 (5%) and -2.92 (10%).

[15]This decision of lags is based on visual evidence from fig. 6

Since the primary focus of this analysis is the prediction of closing volume at the end of the day, this setup requires a good environment which is less prone to issues of endogeneity. This comes from the fact that investors are given the chance to decide between continuous- and closing phase throughout the entirety continuous phase. Consequently, the outcome of the closing auction cannot influence the market conditions in the continuous phase. In order to isolate the effects of said market quality measures, the analysis is extended with three additional dummy variables. First, D^{month} equals 1 on days when options expire. Second, $D^{quarter}$ equals 1 on days when futures expire in March, June, September and December. Consequently, the days of future expiry dates is a strict subset of all option expiry dates. Third, $D^{equival}$ equals 1 after the repeal of the exchange equivalence on July 1st 2019.

In addition to dummy variables and (un)expected continuous volumes, this analysis also corrects for absolute intraday returns that are likely to influence the participation of the closing auction. Many financial institutions (particularly passively managed investment funds) have strict guidelines on their asset allocation. Large intraday price movements can skew this allocation and trigger correlated portfolio rebalancing across market participants. To take advantage of the high liquidity at the closing auction, an inflow of closing volume after large same-day continuous returns is to be expected. In this context, the continuous return is defined as

$$RET_{s,d}^{cont} = \ln\left[M_{s,d}^{5:20pm}\right] - \ln\left[M_{s,d}^{9:01am}\right] \tag{4}$$

where $M^{5:20pm}$ is the midquote just before the start of the closing auction and $M^{9:01am}$ is the midquote just after the settlement the opening auction. Therefore, $\left|RET_{s,d}^{cont}\right|$ represents the absolute log return of the continuous trading phase. For inferential purposes, this analysis is based on the following pooled OLS regression equation:

$$\begin{aligned} \log_2[V_{close}]_{s,d} = \alpha_s &+ \beta_1 \cdot D_{s,d}^{month} + \beta_2 \cdot D_{s,d}^{quarter} + \beta_3 \cdot D_{s,d}^{equival} \\ &+ \beta_4 \cdot \left|RET_{s,d}^{cont}\right| + \beta_5 \cdot \widehat{\log_2[V_{cont}]}_{s,d} + \beta_6 \cdot e(\log_2[V_{cont}])_{s,d} + \varepsilon_{s,d} \end{aligned} \tag{5}$$

In order to dissect the effects more precisely by trading volume in the continuous auction, the 30 SLI stocks are distributed into terciles, based on their trading volume during the continuous trading phase. These terciles are reassigned on a daily basis. The resulting quantiles reflect the most-traded (Q_1), neutral, (Q_2) and least-traded (Q_3) stocks in terms of trading volume. The descriptive statistics for these terciles can be found in table 2, which shows descriptive statistics within each of the terciles in million CHF. The table reveals that high-volume stocks in Q_1 have on average around three times the trading

Panel A: Continuous trading

	obs.	$\mu_{s,d}$	σ	min	25%	50%	75%	max
Q_1	4870	149.74	93.42	29.34	81.49	115.68	198.98	910.22
Q_2	4663	49.65	16.88	9.23	38.03	46.93	57.79	217.86
Q_3	4980	25.22	9.18	6.65	18.47	24.11	30.46	85.35

Panel B: Closing auctions

	obs.	$\mu_{s,d}$	σ	min	25%	50%	75%	max
Q_1	4870	63.44	62.45	6.30	30.57	42.38	84.89	1141.42
Q_2	4663	19.21	9.77	3.01	13.41	17.35	22.39	112.00
Q_3	4980	9.52	5.01	0.00	6.25	8.65	11.58	69.54

Table 2: Descriptive statistics by volume tercile. All the 30 Blue Chip constituents of the SLI are reassigned daily into volume terciles which are based on trading volume during the continuous phase. $Q_1(Q_3)$ represents stocks with the most(least) trading volume during the continuous phase. Q_2 is the neutral tercile.

volume in both continuous phase and closing auction compared to Q_2. Similarly, Q_2 has on average around double the trading volume compared to Q_3. Despite all 30 constituents counting as Blue Chips, these discrepancies in trading volume are substantial.

To determine adequate standard error correction for robust inference, a Durbin-Watson test for autocorrelation of the panel residuals $\hat{u}_{s,d}$ has been conducted. The test has been adjusted to account for the fact that the data is in panel format over a long period of time. The test statistic dw was therefore calculated as

$$dw = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \left(\hat{u}_{s,d} - \hat{u}_{i,t-1}\right)^2}{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{u}_{s,d}}$$

and resulted in a value of 1.613. By extrapolating the adjusted critical values from Bhargava et al. (1982) (as opposed to a simple one-dimensional Durbin-Watson test), this result is below the lower-bound critical value. Consequently, the panel is subject to positive autocorrelation in residuals. In an alternative test, an approach derived in Wooldridge (2010, Subsection 10.5.4) is applied. This requires a pooled OLS estimation of the equation $\hat{u}_{s,d} = \rho\hat{u}_{i,t-1} + \varepsilon_{s,d}$, using heteroskedasticity and autocorrelation (HAC) robust standard errors. The result of this tests indicates serial correlation with a p-value of 0.000 and $\hat{\rho} = 0.191$. In order to account for this feature of the analysis, HAC robust standard errors are computed, derived from the covariance matrix derived by Driscoll and Kraay (1998). These standard errors are robust to cross-sectional correlation as well as autocorrelation in residuals and are an extension of Newey and West (1987) to a panel

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	$\log_2[V_{close}]$	$\log_2[V_{close}]$	$\log_2[V_{close}]$	$\log_2[V_{close}]$	$\log_2[V_{close}]$	$\log_2[V_{close}]$
N	14362	14361	14361	14361	14361	14361
Adj. R^2	0.1574	0.2174	0.2698	0.3627	0.4206	0.4263
D^{month}	0.1043**	0.1136**	0.0799	0.1225***	0.0872*	0.0883*
	(2.1547)	(2.3969)	(1.6013)	(2.6986)	(1.8901)	(1.9432)
$D^{quarter}$	1.6999***	1.7071***	1.7489***	1.6071***	1.6491***	1.6562***
	(23.722)	(24.499)	(23.351)	(20.760)	(20.044)	(20.475)
$D^{equival}$	-0.0894**	-0.0810**	-0.1147***	-0.1148***	-0.1510***	-0.1493***
	(-2.5335)	(-2.4308)	(-3.6287)	(-3.6833)	(-4.7110)	(-4.7138)
$\lvert RET^{cont} \rvert$		0.2720***	0.2526***	0.1318***	0.1087***	
		(3.8996)	(3.6537)	(3.0597)	(2.6492)	
$\widehat{\log_2[V_{cont}]}$			0.4014***		0.4223***	0.3992***
			(13.215)		(14.098)	(13.087)
$e(\log_2[V_{cont}])$				0.5127***	0.5227***	0.4978***
				(22.849)	(23.850)	(23.044)
$Q_1 \times \lvert RET^{cont} \rvert$						0.2077***
						(5.5988)
$Q_2 \times \lvert RET^{cont} \rvert$						0.1265***
						(3.8324)
$Q_3 \times \lvert RET^{cont} \rvert$						0.0200
						(0.7387)
Fixed Effects	Entity	Entity	Entity	Entity	Entity	Entity

Table 3: Methodological framework results. This table regresses normalized closing auction volume $\log_2[V_{close}]$ onto a set of control variables derived from continuous trading without including any measures of market quality yet. $L_1(\log_2[V_{close}])$ is the lagged dependent variable by one day. Dummy variables equal 1 on option expiration dates (D^{month}), on future expiration dates ($D^{quarter}$) and after the repeal of exchange equivalence ($D^{equival}$). In terms of continuous volumes, $\widehat{\log_2[V_{cont}]}$ represents the expected component and $e(\log_2[V_{cont}])$ represents the unexpected component. All 30 SLI stocks are reassigned into volume terciles on a daily basis based on their continuous trading volume from highest (Q_1) to lowest (Q_3) volume stocks. The panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively.

setting using a Bartlett kernel function[16].

Table 3 reports how the derived control variables influence logarithmic closing volumes without taking into account any continuous market quality metrics yet. Lagged closing volumes are disregarded for two reasons. First, the fact that closing volumes do not behave like an autoregressive process[17] is questioning the predictive power of a lagged term. Second, the inclusion of a lagged term would violate the assumption of *strict*

[16]This approach is recommended by Hoechle (2007) for long panels where ($N \ll T$).

[17]This is also being indicated in Panel B of fig. 6.

exogeneity in fixed-effects panel settings, which would yield biased estimates.

The effect on volume flowing into closing auctions on derivative closing dates, however, are highly significant. This is particularly true on futures expiration dates which happen quarterly, when closing volume is $2^{1.6999} \approx 3.25$ times the volume compared to usual trading days. Days of option expiration have much less impact on closing volumes with only around 7% more volume than usual, but are still highly significant. Overall, table 3 suggests that the dummy variables alone are able to explain roughly 28% of the entire within variance. Moreover, investors have withdrawn approximately 6% of their volume from the closing auction after the repeal of exchange equivalence. All of these results can be visually confirmed in fig. 1. Another interesting observation can be made when looking at absolute returns throughout the continuous trading phase. These returns have a clearly positive impact on the volume in closing auction. Per percentage point continuous return, there is an inflow of approximately 20% into the closing auction. This speaks in favour of the hypothesis in Raillon (2020), that passively managed funds are forced to rebalance their portfolios after large intraday price movements shifts. This is true for both positive and negative returns. Interestingly, however, this effect is only pronounced for the more heavily traded stocks in Q_1 and Q_2. This implies that there is much less interest from passive funds in Q_3 stocks with low daily volume, as there is no significant effect on closing volume after large intraday price movements. In terms of expected and unexpected volumes throughout the continuous trading phase, table 3 shows clearly positive coefficients. This indicates that the volumes of the continuous phase and the closing auction are positively correlated. This observation indicates that overall, positive(negative) liquidity shocks in the continuous phase are not driven by outflows(inflows) at the closing auction. However, unexpected shocks during the continuous auction are able to explain more of the variance in closing auctions in terms of R^2 improvement than expected continuous volume.

5.2 Variables of Market Conditions

So far, a set of control variables has been derived, all of which have a significant impact on closing volume but are not the primary focus of this study. Section 4 lined out that the market conditions of interest are related to both execution risk and liquidity. Despite these two properties having a certain amount of overlap, they are approximated with a distinct set of variables that are all widely used in academic literature on market microstructure. To begin with, we will look at execution risk measures.

In order to capture this property, the first measure to be computed is *realized intraday volatility*. For this purpose, the continuous trading phase is subdivided into a number of sub-intervals, each lasting for h minutes. These equally-sized intervals consequently sum up to $K(h) = 500 \cdot h^{-1}$ as the total of intervals for each trading day, depending on the length of each interval. The return over each interval k is defined as the logarithmic difference of midquotes M at the beginning and at the end such that $RET_{k,i,t} = \log(M_{k,i,t}) - \log(M_{k-1,i,t})$. Consequently, $RET^{cont}_{s,d} = \sum_{k=1}^{K(h)} RET^{(h)}_{k,i,t}$ holds for each choice of h. In order to achieve this, the population standard deviation over the interval returns $RET^{(h)}$ are computed and subsequently normalized to one trading day to ensure comparability among interval sizes[18]. This is achieved using the equation

$$IVOLA^{(h)}_{s,d} = \sqrt{\sum_{k=1}^{K(h)} \left(RET^{(h)}_{k,i,t} - \left[RET^{cont}_{s,d} / K(h) \right] \right)^2} \tag{6}$$

where the measure realized volatility is estimated over several time horizons $h \in \{1, 5, 10, 20\}$ minutes. Despite the normalization to the whole trading day, the resulting value heavily depends on the chosen interval h. A small choice of h of 1 minute will capture much of the flickering of orders that can be caused by algorithmic market makers (Hasbrouck, 2018), whereas the choice of a larger interval (e.g. $h = 20$) does not account for transitory noise, but rather captures intraday price swings.

Another measurement that is related to volatility is order imbalance. The concept of order imbalance is a familiar concept in academic literature. One of the most well-known applications of it is by Chordia et al. (2008), who analyse the predictability of stock returns when looking at order imbalances over various time horizons. They use this predictive power as an inverse indicator of market efficiency. Order imbalance is a measure of whether the market shows buy- or sell pressure, when looking at the imbalance of market orders. For this reason, $VOL^{buy}(VOL^{sell})$ is defined as the volume initiated by buy(sell) market orders. The order imbalance is therefore defined as:

$$OIB_{s,d} = \log_2 \left[\frac{VOL^{buy}_{s,d}}{VOL^{sell}_{s,d}} \right]. \tag{7}$$

Hereby, order imbalance is defined somewhat differently from Chordia et al. (2008), in that it is not bounded[19]. Consequently, the measure defined in eq. (7) is able to capture more of the variance in case there is a large mismatch between buy- and sell-volume,

[18]This is the reason, why the sum of squares is not divided by the number of intervals $K(h)$ in the equation.

[19]Chordia et al. (2008) define order imbalance as follows:

$$ALTOIB_{s,d} = \frac{VOL^{buy}_{s,d} - VOL^{sell}_{s,d}}{VOL^{buy}_{s,d} + VOL^{sell}_{s,d}},$$

whereas in the other measure would merely converge to either -1 or 1.

Another aspect of market conditions analyzed in this paper is *liquidity*, which can be used quite loosely in the literature. In the context of this paper, the work of Kyle (1985) is referenced for the definition of liquidity, postulating three dimensions of liquidity: Tightness, depth, resiliency. All three of these dimensions are indispensable to the concept of liquidity and can also be measured numerically in markets.

The first of these dimensions, tightness, represents the cost of a round-trip trade. Since limit order books always have a spread between the best bid and the best ask, this spread represents the minimum cost that an investor must bear when using market orders. Thereby, liquidity is inversely related to the size of the bid-ask spread. In this paper, two types of bid-ask spreads are defined, which are also widely used in the literature. First, the *quoted spread* measures the quoted bid-ask spread as the difference between best bid and ask prices. The quoted spread is re-evaluated with each event at time t in the order book and expressed in basis points:

$$QS_{s,d,t} = \frac{P^{(ask)}_{s,d,t} - P^{(bid)}_{s,d,t}}{M_{s,d,t}} \cdot 10^4$$

where $P^{(bid)}_{s,d,t}(P^{(ask)}_{s,d,t})$ represents the best bid(ask) price in stock s on day d at time t. M_t is the midquote defined as the arithmetic mean of the best bid- and ask prices. Since we are operating in a continuous environment, the resulting spreads are time-weighted in order to account for sequences of rapid quote adjustment[20], which results in the *weighted quotes spread* (WQS) measure.

In a very similar manner, the *effective spread* is computed. In contrast to the quoted spread, the effective spread only takes into account an executed trade which is expressed as an event at time t in this context. By only taking into account trades instead of all quote adjustments, it implies that an investor deemed the spread small enough to submit a market order. It therefore does not only consider the willingness of the liquidity provider to post limit orders, but also the liquidity taker who executes against it. In market microstructure literature, this is often referred to as a more accurate measure of actual transaction costs. In practice, the *effective spread* is defined as

$$ES_{s,d,t} = q_{s,d,t} \cdot \frac{P_{s,d,t} - M_{s,d,t}}{M_{s,d,t}} \cdot 10^4$$

where $ALTOIB \in [-1, 1]$ is bounded on both sides. This property might lead to statistical issues for the panel regression approach presented in this paper.

[20]This phenomenon is also referred to as *quote flickering* (Hasbrouck, 2018). Without adjusting for this, periods of high volatility would receive oversized weight over more quiet periods.

where $q_{s,d,t}$ represents the direction of market order that initiated a trade in stock s on day d at time t, i.e. 1 for a buy and -1 for a sell. Given this indicator for direction, the effective spread always results in a positive value. Furthermore, the same time-weighted adjustment is done as with the quoted spreads above, resulting in the *weighted effective spread* (WES).

The second dimension, which is depth, refers to how much one-sided order flow is required to affect the price. In deep markets, even large orders will not affect the price substantially. Applying this concept to limit order books, one can measure how many securities are ready for purchase or sale on top of the book. For this purpose, *quoted depth* is measured at various levels of the order book g:

$$QD_{s,d}^{(g)} \quad \forall \quad g \in \{1,4,8\}.$$

The quoted depth is thereby the sum of all securities available up to the given level g. Consequently, the depth is monotonously increasing the farther out from the midquote it is measured ceteris paribus. Moreover, quoted depth is determined in currency terms, in order to account for different price levels of the securities analyzed.

The last liquidity dimension after Kyle (1985), resiliency, is more difficult to measure. It captures the ability of the market to quickly revert to the original price before a shock, e.g. caused by a large market order sweeping limit orders across multiple prices on the opposite side of the book. In the context of a limit order book, that means we are including a temporal dimension through the *price impact*, which measures how the price changes in a given time interval of h minutes after each trade executed at time t. Similar to the effective spread, q_t denotes the direction of the market order initiating the trade:

$$PI_{t,s,d}^{(h)} = q_{t,s,d} \cdot \frac{M_{t+h,s,d} - M_{t,s,d}}{M_{t,s,d}} \cdot 10^4 \quad \forall \quad h \in \{1,5,10\}. \tag{8}$$

This measure essentially yields a positive result if the midquote changes in the direction of the incoming market order after h minutes. Consequently, a resilient market exhibits a price impact of close to zero and a positive value otherwise. Price impact is a common measure in market microstructure literature and has also been applied in Carrion (2013), Conrad et al. (2015), and Hendershott et al. (2011). All of these paper extend this further by a full spread decomposition based on the framework of Glosten and Milgrom (1985).

6 Results

The methodology derived in section 5 will be used as a base for the presentation of all further empirical results. According to the hypotheses derived earlier, this section is subdivided into two subsections with analyses of intraday market conditions and volume clustering.

6.1 Continuous Market Conditions

The first set of results is based heavily on the regression framework derived in section 5.1, where the control variables were introduced. This section extends that approach by adding each variable considered in section 5.2 individually. However, the goal here is not to show the overall direct effect, but to discuss the effect in more granularity. More specifically, this section focuses on the changes in sensitivities with respect to the underlying market conditions

$$MarketCond \in \{IVOLA^{(h)}, OIB, WQS, WES, QD^{(g)}, PI^{(h)}\}.$$

In order to achieve this, the trading days are segmented into five quintiles with the respective dummy variable $M \in \{M_1, M_2, \ldots, M_5\}$ based on a given market condition. These quintiles are determined individually for each stock s in order remove structural differences between groups. For instance, if $MarketCond$ refers to quoted spreads, M_1 equals 1 for the 20% of trading days with the largest spreads and equals zero otherwise. These market quintiles are additionally interacted with the dummy variables for terciles $Q \in \{Q_1, Q_2, Q_3\}$ for intraday trading volume. All of these steps result in the following expression:

$$\begin{aligned}\log_2[V_{close}]_{s,d} = \alpha_s &+ \beta_1 \cdot D^{month} + \beta_2 \cdot D^{quarter} + \beta_3 \cdot D^{equival} \\ &+ \beta_4 \cdot e(\log_2[V_{cont}])_{s,d} + \beta_5 \cdot \widehat{\log_2[V_{cont}]}_{s,d} \\ &+ \sum_{k=1}^{3} Q_k \times \left[\beta_6^k \cdot \left|RET_{s,d}^{cont}\right| + \sum_{j=1}^{5} M_j \times \left(\gamma_j^k \cdot \log_2[MarketCond_{s,d}]\right)\right]\end{aligned} \tag{9}$$

Following this expression, we can see that all stockdays are separated into 15 buckets, based on a grid of intraday volume tercile as well as their market condition quintile. For each of these buckets, one coefficient γ_j^k is estimated respectively without any overlap, resulting in 15 coefficients. This approach enables the determination of varying sensitivities towards market conditions, i.e. second-order effects. In addition to this, all market conditions are transformed using logarithms with base 2, in order to retain comparability between the values. This step has to be done because there are inevitable differences in market conditions in different quintiles M. Logarithmization makes it possible to compare market conditions across buckets. Each of the 15 estimated coefficients γ_j^k thus represents the marginal effect of a doubling of $MarketCond$ from the average within each bucket. To test for differences within the γ_j^k coefficients standard Wald tests are applied following Baltagi (2011, Section 7.9).

The results in the following tables are organized to only focus on the estimates of the 15 unique γ_j^k coefficients per regression. The coefficients are organized by volume tercile (in rows) and market condition quintiles (in columns). The bottom row of each panel represents the difference between the most and least liquid stocks ($Q_1 - Q_3$). Similarly, the right-most column represents the difference of extreme quintiles with respect to market condition ($M_1 - M_5$).

The first set of results in table 4 are representative of analysis concerning measurements of execution quality. Specifically, the variables of interest here are intraday volatility as well as order imbalance. For this section, the results for intraday volatility are presented at the horizons over 1 and 20 minutes together with the results for market order imbalances. The results for other horizons for robustness purposes can be found in the appendix[21]. The results presented here show that intraday volatility has a clearly negative effect on closing auction volume. Therefore, higher intraday volatility coincides with an outflow of volume at the close. This is true for each of the two volatility horizons considered. Hence very short term quote flickering and intraday price swings have the same directional, albeit the latter has a larger impact on closing participation. It is important to remember that both the dependent variable as well as the independent variables of interest are measured in logarithmic terms with base 2. When looking at panel B for example, a coefficient of -0.1 can be interpreted as the outflow of 7% when volatility intraday doubles[22].

Another interesting observation concerning the volatility-related coefficients can also be made in terms of their extent. The coefficients in Panel B of table 4 are around 2–3 times as large as in Panel A, indicating that market participants react more sensitively to an increase in volatility measured over 20-minute horizons than the 1-minute counterpart. Keep in mind that by the definition in eq. (6), measurements of intraday-volatility across all horizons h are normalized to one trading day. Therefore, the volatility over the shorter time horizon does not need to be scaled in order to compare both measurements. Consequently, a doubling of 20-minute volatility makes market participants more

[21] See table 7 for further intraday volatility time horizons.

[22] This is given by the fact that the doubling of the volatility reduces the logarithmic trading volume by 0.1 leading to $2^{-0.1} \approx 0.93$, indicating a 7% reduction of volume in currency terms.

Panel A: $\log_2[IVOLA^{(1)}]$ — $R^2 = 0.4359$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0987***	-0.0640***	-0.0559***	-0.0446***	-0.0349***	-0.0638***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0000)
Q_2	-0.0470***	-0.0311**	-0.0164	-0.0238**	-0.0071	-0.0399***
	(0.0025)	(0.0128)	(0.1123)	(0.0161)	(0.3995)	(0.0008)
Q_3	0.0218	0.0258*	0.0215*	0.0278**	0.0324***	-0.0105
	(0.2541)	(0.0737)	(0.0794)	(0.0125)	(0.0017)	(0.4419)
$Q_1 - Q_3$	-0.1205***	-0.0897***	-0.0774***	-0.0724***	-0.0672***	-0.0533***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)

Panel B: $\log_2[IVOLA^{(20)}]$ — $R^2 = 0.4384$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.1607***	-0.1476***	-0.1255***	-0.1211***	-0.1028***	-0.0579***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Q_2	-0.1171***	-0.1009***	-0.0903***	-0.0891***	-0.0732***	-0.0439***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)
Q_3	-0.0599**	-0.0446**	-0.0503**	-0.0385**	-0.0332**	-0.0267*
	(0.0287)	(0.0448)	(0.0164)	(0.0460)	(0.0498)	(0.0624)
$Q_1 - Q_3$	-0.1007***	-0.1030***	-0.0752***	-0.0826***	-0.0696***	-0.0312***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0024)

Panel C: OIB — $R^2 = 0.4299$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	1.1072***	1.8122***	2.8768***	5.7020***	12.6303***	-11.5231***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Q_2	0.5735***	0.5277*	0.9497**	1.9184**	4.2829***	-3.7094**
	(0.0001)	(0.0622)	(0.0251)	(0.0102)	(0.0054)	(0.0110)
Q_3	-0.6040***	-0.9656***	-2.2199***	-3.5258***	-8.7803***	8.1763***
	(0.0002)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$Q_1 - Q_3$	1.7112***	2.7778***	5.0967***	9.2277***	21.4106***	-19.6994***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 4: Results of binning regressions for measures of execution risk. This table presents the binned γ coefficients from (9). Both the dependent and independent variables are measured in logarithms with base 2 to assure comparability across market quintiles. Each panel represents an independent estimation of coefficients. Panel A(B) shows the result when including the logs of intraday volatility measured over a time horizon of 1(20) minutes as defined in (6). Panel C exhibits the results based on order imbalance derived in (7) as an independent variable. Each panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively. Additionally, each panel tests for equality of the largest and smallest quantile using Wald tests.

impatient than a doubling of volatility over a 1-minute horizon, i.e. quote flickering. This observation is true across all volume terciles.

The results for order imbalance in table 4 are more ambiguous in that large stocks react positively in contrast to smaller stocks. Consequently, investors become more patient with their execution in large(small) stocks when order imbalances are large(small). In this regard, order imbalances have a different effect on closing participation than intraday volatility, since large stocks react more sensitively towards increases in volatility. In other words, investors become more impatient when volatility rises in large stocks, whereas they are calmer when volatility rises in smaller.

When analyzing the effects of execution risk across the market condition quintiles M, the patterns are more straightforward. When it comes to large- and medium-sized stocks, market participants are significantly less patient with higher execution risk during continuous trading hours, i.e. low intraday volatility or order imbalances. This implies that market participants prefer to execute their trades more immediately when volatility or order imbalances are large. Regarding small stocks, this phenomenon is observed to a much smaller extent for intraday volatility and in the opposite direction for order imbalance. Consequently, investors do not exhibit such impatient behavior on days with higher execution quality for small stocks. Overall, these results speak in favor of H/1 at the expense of H/2. Additionally, these findings particularly contradict the hypothesis of Raillon (2020) according to which investors tend to seek the closing auction as it is harder to otherwise achieve optimal execution. In this sample, the opposite behavior can be observed.

The results for the next type of market conditions, namely tightness, are exhibited in table 5 and are expressed in logarithms of quoted and effective spreads. Interestingly, the point estimates for the coefficients are estimated to be opposite. To begin with, quoted spreads exhibit slightly negative effects on closing participation, i.e. making investors less patient. In contrast to this, effective spreads increase investor patience in most cases (apart from small cap stocks on days with small effective spreads). One possible explanation for this divergence are potential correlations with other variables of market quality. Neither quoted spreads nor effective spreads are strictly exogenous, but depend on other market conditions. Nonetheless, reverse causality between the dependent and independent variables can be disregarded due to the temporal separation of the two measurements.

The second order implications approximated by the Wald tests for equality of the extreme quantiles in both dimensions are identical for both quoted and effective spreads. In both cases, investors tend to be more patient in larger stocks compared to smaller ones by a significant amount. The differences in coefficients of around 0.1 to 0.2 indicate a proportional volume difference of 7% respectively 15%. This result indicates that investors

Panel A: $\log_2[WQS]$ — $R^2 = 0.4410$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0675***	-0.0581*	-0.0575**	-0.0535*	-0.0463	-0.0212
	(0.0092)	(0.0532)	(0.0484)	(0.0969)	(0.2093)	(0.1778)
Q_2	-0.0972***	-0.1213***	-0.1215***	-0.1268***	-0.1204***	0.0233
	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.1215)
Q_3	-0.1648***	-0.1902***	-0.2097***	-0.2133***	-0.2075***	0.0427***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0028)
$Q_1 - Q_3$	0.0974***	0.1321***	0.1521***	0.1598***	0.1612***	-0.0639***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)

Panel B: $\log_2[WES]$ — $R^2 = 0.4817$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	0.1311***	0.0969***	0.0807***	0.0721***	0.1029***	0.0281
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1177)
Q_2	0.0929***	0.0651***	0.0312**	0.0012	0.0165	0.0764***
	(0.0000)	(0.0000)	(0.0183)	(0.9324)	(0.4681)	(0.0000)
Q_3	0.0309***	-0.0087	-0.0508***	-0.0841***	-0.0798***	0.1106***
	(0.0098)	(0.4948)	(0.0005)	(0.0000)	(0.0008)	(0.0000)
$Q_1 - Q_3$	0.1002***	0.1056***	0.1315***	0.1563***	0.1827***	-0.0825***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)

Table 5: Results of binning regressions for measures of tightness. This table presents the binned γ coefficients from (9). Both the dependent and independent variables are measured in logarithms with base 2 to assure comparability across market quintiles. Each panel represents an independent estimation of coefficients. Panel A shows the result when including the logs of the weighted quoted spreads. Panel B exhibits the results based on weighted effective spreads as an independent variable. Each panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively. Additionally, each panel tests for equality of the largest and smallest quantile using Wald tests.

expect execution during the auction to be cheaper in small stocks compared to larger stocks, and thus there is less need for immediate execution. In addition to this, investors are not affected by differences in spreads among large stocks. In the cross-section of small stocks, investors become increasingly more patient with increasing spreads, however. Conclusively, investors become more patient on days when measures of market tightness deteriorate.

Finally, table 6 presents the results based on measures of market depth as well as resilience. Further results on these measures can be found in the appendix[23]. To begin with, the results on market depth show that only the point estimates for the small stocks have a significant impact on closing participation. More specifically, increased depth in

[23]See table 8 for more results on depth, and table 9 for resilience.

Panel A: $\log_2(QD^{(8)})$ — $R^2 = 0.4367$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0214	-0.0759**	-0.0502*	-0.0665***	-0.0347*	0.0133
	(0.6307)	(0.0146)	(0.0698)	(0.0013)	(0.0778)	(0.7145)
Q_2	0.0322	0.0094	0.0005	-0.0139	-0.0060	0.0382**
	(0.3329)	(0.7339)	(0.9862)	(0.5489)	(0.7696)	(0.0421)
Q_3	0.1592***	0.1123***	0.0912***	0.0790***	0.0663***	0.0930***
	(0.0000)	(0.0002)	(0.0008)	(0.0010)	(0.0016)	(0.0000)
$Q_1 - Q_3$	-0.1807***	-0.1882***	-0.1414***	-0.1455***	-0.1010***	-0.0797**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0184)

Panel B: $\log_2(PI^{(1)})$ — $R^2 = 0.4404$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	0.0149	0.0877***	0.0824***	0.1335***	0.0828***	-0.0679**
	(0.3888)	(0.0022)	(0.0025)	(0.0005)	(0.0090)	(0.0182)
Q_2	-0.0071	-0.0019	0.0138	0.0403	0.0641**	-0.0711***
	(0.5816)	(0.9146)	(0.4955)	(0.1588)	(0.0458)	(0.0081)
Q_3	-0.1255***	-0.1220***	-0.1392***	-0.1213***	-0.1421***	0.0166
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.4891)
$Q_1 - Q_3$	0.1404***	0.2096***	0.2216***	0.2548***	0.2249***	-0.0845**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0153)

Table 6: Results of binning regressions for measures of depth and resilience. This table presents the binned γ coefficients from (9). Both the dependent and independent variables are measured in logarithms with base 2 to assure comparability across market quintiles. Each panel represents an independent estimation of coefficients. Panel A shows the result when including the logs of the quoted depth at the top eight levels of the order book. Panel B exhibits the results based on price impacts as defined in (8) determined over the course of one minute as an independent variable. Each panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively. Additionally, each panel tests for equality of the largest and smallest quantile using Wald tests.

those stocks makes investors significantly more patient, whereas a deterioration has the opposite effect. This indicates that investors do not feel the need to just *get the job done*. Even though this phenomenon can be observed across all market quintiles of the small stocks, investors become increasingly patient on days with deep order books. In contrast to this, investors in large stocks are more likely to take advantage of more convenient market conditions during the continuous trading phase, however, this effect is lesser in magnitude.

In terms of the third of the Kyle (1985) liquidity dimensions, resilience, most coefficients result in opposite signs. However, it is important to keep in mind that large depth indicates good liquidity, whereas large price impact has the opposite effect. Hence, it is desirable for a market to achieve price impact as low as possible in order to be viewed as liquid. The results on resilience overall show that particularly small stocks are very sensitive when price impact deteriorates. In that case, investors become much more impatient and execute their orders prior to the closing auction. In large stocks, the opposite effect can be observed, such that market participants tend to wait for the closing auction when they expect executions in the continuous phase to affect the price to a higher degree. When looking at the coefficient tests for equality across the five quintiles of market conditions, there are some negative second order effects for large and medium-sized stocks. In these stocks, investors are significantly more impatient in the quintile of days with the highest price impact.

After seeing all the results on the three liquidity dimensions following Kyle (1985) in tables 5 and 6, it becomes clear that market participants react differently in large and small stocks. Despite all the stocks in the sample being Blue Chip titles, there are significant differences. Overall, it can be said that investors in smaller stocks are much more sensitive towards deteriorating market conditions, particularly in terms of the liquidity measures. In these cases, investors tend to become more impatient and execute their trades earlier instead of during the closing auction. Therefore, H/2 seems to hold for smaller titles. In contrast to this, investors tend to become more patient when liquidity deteriorates in larger titles. This observation speaks more in favor of H/1. One reason for these opposite observations separated by market cap may lie in the expected volume in the closing auctions. Large stocks have an almost guaranteed abundance of liquidity in these auctions. Investors who experience sub-par liquidity conditions intraday are therefore inclined to rely on the closing auction. In smaller stocks, this abundance of liquidity is not guaranteed. Investors may thus expect either distortions of the closing price or even events where the auction does not cross (Ellul et al., 2005; Ibikunle, 2015). Conditions of illiquidity during the continuous trading phase may then be interpreted as a harbinger for an illiquid closing auction. In this case, investors may prefer to execute their trades during an illiquid continuous trading phase instead of a call auction with the same properties. This behavior could explain the differences between small and large stocks.

6.2 Volume Clustering

After having discussed the first half of the results on the impact of market conditions on closing volume, this section will shed light on the factors influencing the distribution of volume across the trading day in consideration of certain clusters. Previously, both theoretical work (Admati & Pfleiderer, 1988; Pagano, 1989a) as well as empirical evidence of volume clustering throughout the continuous phase of the trading day has been discussed in a previous section. However, one contribution of this paper is about the clustering across trading facilities. In order to achieve this, the methodology derived in section 5 is extended with the application of a quantile regression framework. Quantile Regression has not been used frequently in the literature of financial economics, but has rather been used in labor economics (Arias et al., 2002; Chamberlain, 1994; Fitzenberger & Franz, 1999) or micro-economics (Bassett et al., 2003; Eide & Showalter, 1998; Poterba & Rueben, 1995).

The concept of quantile regression was first introduced by Koenker and Bassett (1978) and subsequently extended by Koenker and Hallock (2001). The idea behind quantile regression comes from the observation that different quantiles of the dependent variable may be influenced in a different way by the independent variables. For instance, Yu et al. (2003, p. 335) provide an example of the weight of children by age. They argue that in order to make accurate predictions about the expected weight profile over the course of their lives, it is important to not infer measures based on averages. More specifically, under- or overweight children may behave differently from children with median weight given the same conditions otherwise. Another example presented in Koenker and Hallock (2001, p. 146) is concerned with the relationship between income and food expenditure on a household-level. Intuitively, households with lower income have larger marginal food expenditure. This non-linearity in the data can be well captured by estimating different sets of parameters for each quantile of the dependent variable. Consequently, quantile regressions can provide a more granular insight into the data as opposed to mean-based methods of estimation.

In contrast to panel regressions, quantile regressions allow the inclusion of one additional dimension, which is the quantile of the dependent variable. Instead of minimizing the squared errors as in the case for ordinary least squares (OLS), quantile regressions optimize a different set of parameters for each τth quantile of the dependent variable, where $\tau \in (0, 1)$. For this purpose, let $\rho_\tau(\,\cdot\,)$ be a *tilted absolute value function* (Koenker & Hallock, 2001), which represents the loss function of the form

$$\rho_\tau(u) = u \cdot \left(\tau - \mathbb{I}_{[u<0]}\right) \qquad \forall \qquad u \in \mathbb{R}, \tag{10}$$

where $\mathbb{I}_{[u<0]}$ is an indicator function which results in 1 when u is negative and 0 otherwise. Due to the indicator function in (10), the output of the loss function is strictly non-negative. Additionally, the loss function is linear in u. Following Koenker and Hallock (2001), this loss function is applied to the residuals of the regression $Q^\tau\left(Y \mid X\right) = X\beta^\tau$

for each quantile individually, such that

$$
\begin{aligned}
\hat{\beta}^{\tau} &= \underset{\beta \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i\beta\right) \\
&= \underset{\beta \in \mathbb{R}^k}{\arg\min} \left[(\tau - 1) \sum_{y_i < x_i\beta} (y_i - x_i\beta) + \tau \sum_{y_i \geq x_i\beta} (y_i - x_i\beta) \right].
\end{aligned}
$$

Consequently, quantile regression minimizes the absolute error with respect to the τ quantile instead of the squared error in the case of OLS. This makes the procedure more robust with respect to outliers, that are known to distort OLS parameter estimates. Koenker and Hallock (2001) refer to the optimization of a quantile regression as minimization of the sum of *asymmetrically weighted absolute residuals*. In addition to this, we get a distinct set of optimal parameters $\hat{\beta}^{\tau}$ for each quantile τ of the dependent variable y. In contrast to this, OLS solely provides one set of parameters which is derived over the whole sample average.

In this application of quantile regression, the goal is to determine whether there is a certain expectations channel active when it comes to predicting closing volume. In other words, it aims to determine whether market participants adjust their sensitivities towards market conditions based on the expected closing volume. If there are any systematic differences between the reactions of investors, this would imply that investors indeed take into account the volume which is expected to be traded at the close. For this purpose, the dependent variable of choice remains the closing volume $\log_2[V_{close}]$. In the first set of results, the quantile regression will only consider the absolute returns of the continuous trading phase $\left|RET_{s,d}^{cont}\right|$ as defined in (4). This results in the following regression equation:

$$
\begin{aligned}
Q^{\tau}\left(\log_2[V_{close}]^{*}_{s,d}\right) = \alpha &+ \beta_1^{\tau} \cdot D^{month} + \beta_2^{\tau} \cdot D^{quarter} + \beta_3^{\tau} \cdot D^{equival} \\
&+ \beta_4^{\tau} \cdot e\left(\log_2[V_{cont}]\right)^{*}_{s,d} + \beta_5^{\tau} \cdot \widehat{\log_2[V_{cont}]}^{*}_{s,d} + \gamma^{\tau} \cdot \left|RET_{s,d}^{cont}\right|
\end{aligned}
\tag{11}
$$

This regression equation is estimated individually for each $\tau \in \{0.05, 0.1, \ldots, 0.95\}$ and for each size tercile, resulting in a total of 57 estimations per variable of interest. For this reason, all volume-related variables of the regression are de-meaned within each stock, which is denoted by an asterisk ($*$). This methodology is similar to the within-transformation for panel regressions in that it disregards the variance between groups. Since we are only interested in the variance within each size tercile and not between them, there are no fixed-effects or stock-based dummy variables added back to the equation. Confidence intervals were derived based on the asymptotic statistics laid out in Koenker (1994).

The results of the quantile regression with continuous returns are presented in fig. 2.

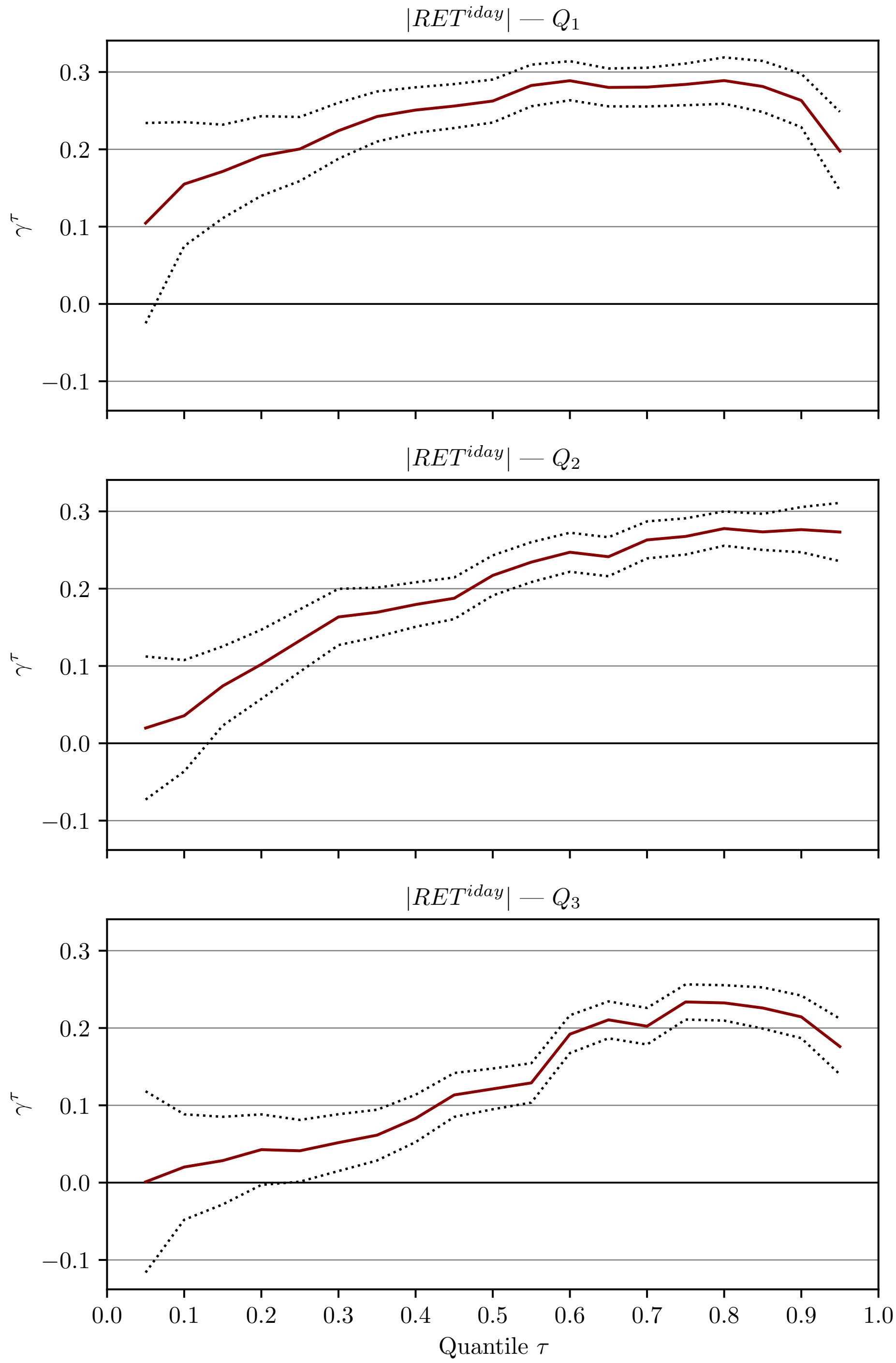


Figure 2: Results of the quantile regression on intraday returns. This figure presents the results from the quantile regression on the returns from the continuous trading phase $\left|RET^{cont}_{s,d}\right|$. The regression equation is defined in (11). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

The horizontal axes of the panels shown represent the quantile of the dependent variable, i.e. the de-meaned logarithmic closing volume. Therefore, the plotted line can be interested as the marginal effect of an independent variable depending on the quantile τ. The figure shows a consistent pattern across all three size terciles, in that the results are upward sloping. This indicates that investors become increasingly more patient with respect to each percentage point of absolute continuous return. Small and medium stocks in particular exhibit the most significant increases from the lowest to the highest τ quantile under consideration of the presented confidence intervals.

The results in table 3 already implied a positive impact of continuous returns on closing volumes[24], the reason being large passive investors being required to rebalance their portfolios after intraday price swings. The novelty of the results in fig. 2 lies in the observation, that investors are more willing to postpone their trades into the closing auction when they expect liquidity to be higher. Particularly on days with very low closing volumes, an additional percent in continuous return has barely any effect on the closing volume. This increase in sensitivity towards continuous returns is indicative of investor awareness of optimal execution. On days where investors expect large closing volumes, the closing auction automatically becomes a more viable option to execute trades.

In a next step, this quantile regression approach is extended by the variables of market conditions introduced in section 5.2, represented as $MarketCond$. Similar to to the results on market conditions, the basic regression equation is extended by each of the variables for market conditions individually. This results in the following equation:

$$\begin{aligned} Q^{\tau}\left(\log_2[V_{close}]^{*}_{s,d}\right) &= \alpha + \beta^{\tau}_1 \cdot D^{month} + \beta^{\tau}_2 \cdot D^{quarter} + \beta^{\tau}_3 \cdot D^{equival} + \beta^{\tau}_4 \cdot \left|RET^{cont}_{s,d}\right| \\ &+ \beta^{\tau}_5 \cdot e\left(\log_2[V_{cont}]\right)^{*}_{s,d} + \beta^{\tau}_6 \cdot \widehat{\log_2[V_{cont}]}^{*}_{s,d} + \gamma^{\tau} MarketCond^{*}_{s,d} \end{aligned} \tag{12}$$

This equation represents an extension to (11), in that it also includes the market condition with the relevant γ^{τ} coefficient, which will be the subject of the following plots.

The results for measures of execution risk are presented in fig. 3, including intraday volatility computed over the time horizons of 1 and 20 minutes[25], and order imbalance on a given day. Akin to the results for absolute continuous returns above, the curves for intraday volatility are upward sloping, particularly for the volatility over the 20-minute horizon. This implies that investors become more patient and willing to wait for the closing auction if they expect the volume to be larger. Under the expectation of small

[24]This effect has been found to hold for large and medium stocks. Small stocks have not shown any effect in the panel setting.

[25]For robustness purposes, further time horizons of intraday volatility are presented in fig. 7 in the appendix.

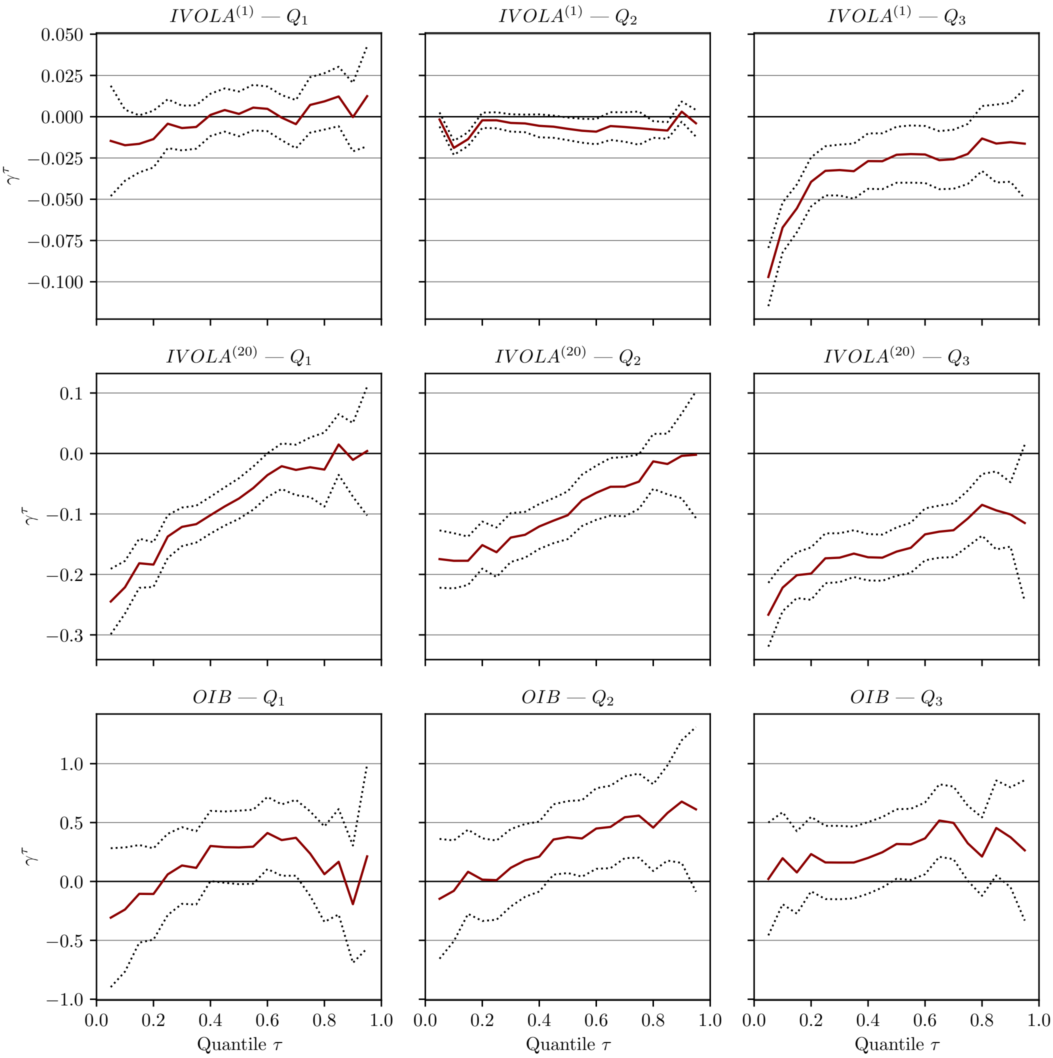


Figure 3: Results of the quantile regression on measures of execution risk. This figure presents the results from the quantile regression on measures of execution risk, namely intraday volatility at the 1- and 20-minute time horizons (defined in (6)) as well as order imbalance (defined in (7)). The regression equation is defined in (12). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

closing volumes, however, investors become much more impatient with each additional percentage point of intraday volatility, as described in table 4. The results of order imbalances also tend to have positive slopes, but due to the wider confidence bands this cannot be underlined with statistical significance. Overall, however, investors tend to become more patient with respect to their reaction to measures of execution risk when they expect higher volume during the closing auction, which is in line with the pull-equilibrium hypothesis H/3.

Transitioning into the three measures of liquidity by Kyle (1985), fig. 4 presents the results for tightness, namely quoted and effective spreads. Both of these measures show an upward sloping trend with comparatively small confidence intervals, particularly for the

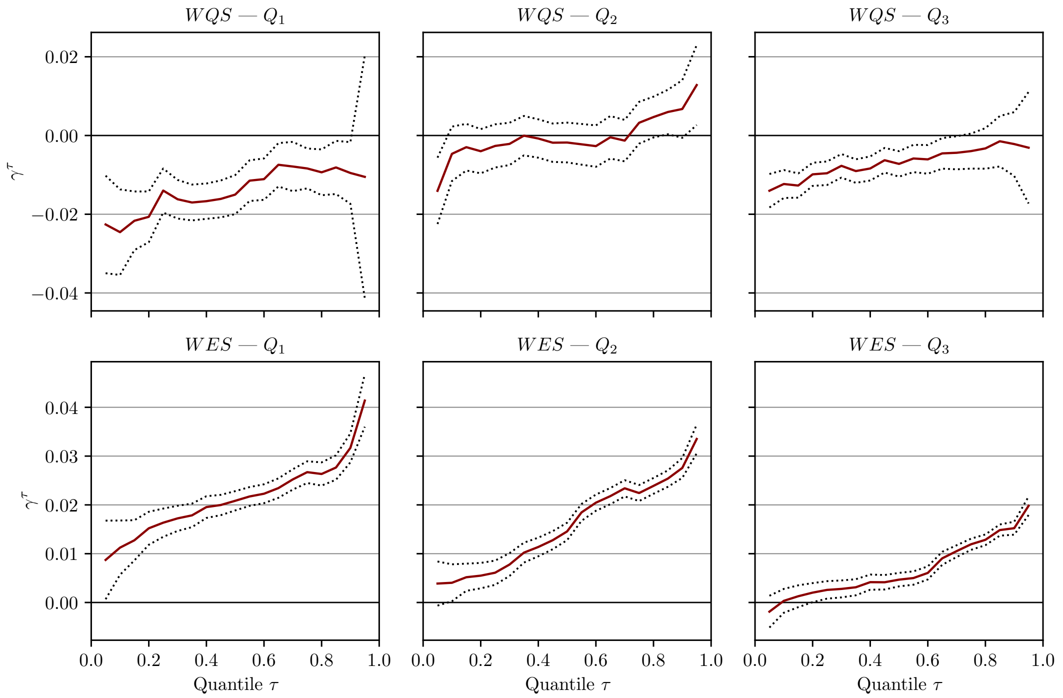


Figure 4: Results of the quantile regression on measures of tightness. This figure presents the results from the quantile regression on measures of execution risk, namely the weighted quoted spreads WQS as well as weighted effective spreads WES. The regression equation is defined in (12). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

weighted effective spread measure. This is in line with the previous results on quantile regression, indicating higher inclination to wait for the closing auction when the expectation for larger closing volume is high.

Finally, the results for depth and resilience are presented in fig. 5. More specifically, this figure presents the results for depth of the top 8 levels of the book[26] as well as the price impact over the 5-minute time horizon[27]. The estimated γ coefficients for depth are negatively sloping. It is important to note that depth has an inverse relationship with liquidity. Therefore, the results on quoted depth imply that under the expectation of large closing volume, investors become more impatient when depth increases. However, this also entails increased patience if depth decreases, i.e. liquidity deteriorates, which is in line with the previous findings. For price impact, the figure shows similar results, albeit less pronounced and significant, particularly for large stocks. Nonetheless, the positive slope in terms of price impact remains.

In conclusion, the approach of quantile regression revealed a consistent pattern throughout all measures of execution risk as well as liquidity. On all accounts, the marginal propensity to choose the closing auction under an incremental deterioration of market

[26] For robustness purposes, fig. 8 shows more detailed results on market depth.
[27] For robustness purposes, fig. 9 shows more detailed results on market resilience.

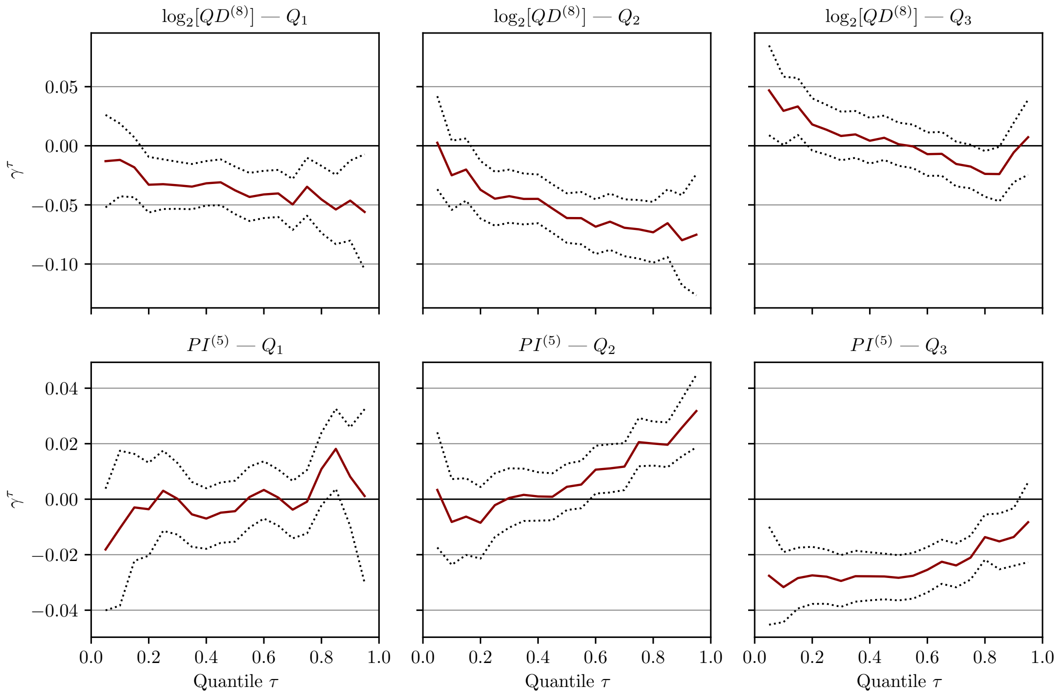


Figure 5: Results of the quantile regression on measures of depth and resilience. This figure presents the results from the quantile regression on measures of execution risk, namely the logarithm of quoted depth at the top 8 levels of the order book as well as the price impact over a time horizon of 5 minutes (defined in (8)). The regression equation is defined in (12). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

conditions is larger when the expected closing volume is high. In more simplified terms, this means that investors adjust their sensitivities towards market conditions based on their expectations. Given the same deterioration of market conditions, investors become more patient when they expect the closing auction will offer higher liquidity. This evidence speaks in favour of H/3, which has already been shown during the continuous trading phase (Admati & Pfleiderer, 1988; Pagano, 1989a). However, this work shows that these forces are also in place across subsequent trading facilities with varying market structures.

7 Conclusion

In recent years, the closing auction has become an increasingly important part of the trading day, in that it attracts a growing amount of volume at the expense of the continuous trading phase. This phenomenon has been observed in equity markets around the world, but it is particularly pronounced in European markets. Regulators are aware of this trend and fear that such outflows from the continuous trading phase may lead to a deterioration in market quality. Therefore, research into the drivers of participation is of timely interest.

Regulators may be particularly interested in finding on the driving factors for the choice of participation in the closing auction over the continuous trading phase. Ultimately, this is an important piece of information when introducing new regulations aiming to limit the outflow into the closing auction by setting the correct incentive structures.

Using order-level data from SIX Securities & Exchanges (SIX) over a two-year horizon, this study finds evidence that investors indeed carefully monitor market conditions during the continuous trading phase and subsequently decide whether to enter the closing auction. To reach this conclusion, the fact that the continuous trading phase is strictly preceding the closing auction without overlap is taken advantage of. In a first step, a panel regression reveals a set of control variables that must be accounted for in order to better isolate the impact of market conditions. For the scope of this paper, the term market conditions includes variables with respect to execution risk and liquidity.

In terms of results, this study offers two main contributions to the existing literature. First, it finds that investors react differently to changes in execution risk and liquidity. With respect to execution risk, which comprises measures of volatility and order imbalance, investors become increasingly impatient upon deterioration, leading to outflows of closing volume. However, as liquidity during the continuous phase gets worse, investors become more patient, leading to an increase in closing volume. In addition to this, this study finds significant differences in how investors react to stocks of varying sizes despite all of stocks in the entire sample being Blue Chips. More specifically, investors in the largest stocks become significantly more patient than small stocks under the same incremental increase of volatility. For measures of liquidity, the opposite can be observed. Investors in high-volume stocks are more patient and rely on the liquidity of closing auctions. Finally, when market conditions are already unfavorable for investors, the sensitivity with respect to patience increases further.

The second contribution of this paper lies in the evidence for a *pull equilibrium* with respect to closing volume. To show this, a quantile regression approach is implemented, thereby modeling an expectations channel for investors. The results show that investors become increasingly patient with their order submission if they expect higher liquidity in the closing auction overall. For instance, an absolute intraday return of one percent makes investors relatively more patient if they expect larger closing volume than if they don't. For deterioration in other market conditions, this finding is replicated, particularly for measures of liquidity such as effective spreads.

Based on these results, it can be concluded that closing auctions are a very important phase of the trading day, as investors can chose them for their executions in order to obtain superior execution on particularly illiquid days. Regulations intended to limit consolidation of volume in the closing auction should take this into account. However, more research is needed to fully understand the choices of different trading facilities that market participants face today. This must be further investigated with particular regard to the increasing fragmentation of equity markets nowadays.

References

Admati, A. R., & Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The review of financial studies*, *1*(1), 3–40.

Aitken, M., Comerton-Forde, C., & Frino, A. (2005). Closing call auctions and liquidity. *Accounting & Finance*, *45*(4), 501–518.

Arias, O., Hallock, K. F., & Sosa-Escudero, W. (2002). Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins data. *Economic applications of quantile regression* (pp. 7–40). Springer.

Baltagi, B. H. (2005). *Econometric analysis of panel data 3rd edition england* (3rd ed.). John Wiley & Sons.

Baltagi, B. H. (2011). *Econometrics* (5th ed.). Springer.

Barclay, M. J., & Hendershott, T. (2008). A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance*, *15*(5), 839–849.

Barclay, M. J., Hendershott, T., & Jones, C. M. (2008). Order consolidation, price efficiency, and extreme liquidity shocks. *Journal of Financial and Quantitative Analysis*, 93–121.

Bassett, G. W., Tam, M.-Y., & Knight, K. (2003). Quantile models and estimators for data analysis. *Developments in robust statistics* (pp. 77–87). Springer.

Bellia, M., Pelizzon, L., Subrahmanyam, M. G., & Yuferova, D. (2017). *Coming Early to the Party*, Safe Working Papers.

Bhargava, A., Franzini, L., & Narendranathan, W. (1982). Serial correlation and the fixed effects model. *The Review of Economic Studies*, *49*(4), 533–549.

Biais, B., Foucault, T., & Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial economics*, *116*(2), 292–313.

Biais, B., Hillion, P., & Spatt, C. (1999). Price discovery and learning during the preopening period in the paris bourse. *Journal of Political Economy*, *107*(6), 1218–1248.

Bogousslavsky, V. (2016). Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance*, *71*(6), 2967–3006.

Bogousslavsky, V., & Muravyev, D. (2020). *Should we use closing prices? institutional price pressure at the close* [Working Paper].

Budish, E., Cramton, P., & Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, *130*(4), 1547–1621.

Carrion, A. (2013). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, *16*(4), 680–711.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. *Advances in econometrics: Sixth world congress* (pp. 171–209).

Chang, R. P., Rhee, S. G., Stone, G. R., & Tang, N. (2008). How does the call market method affect price efficiency? evidence from the singapore stock market. *Journal of Banking & Finance*, *32*(10), 2205–2219.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, *87*(2), 249–268.

Comerton-Forde, C., Lau, S. T., & McInish, T. (2007). Opening and closing behavior following the introduction of call auctions in singapore. *Pacific-Basin Finance Journal*, *15*(1), 18–35.

Comerton-Forde, C., & Putniņš, T. J. (2011). Measuring closing price manipulation. *Journal of Financial Intermediation*, *20*(2), 135–158.

Conrad, J., Wahal, S., & Xiang, J. (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics*, *116*(2), 271–291.

Cordi, N., Félez-Viñas, E., Foley, S., & Putninš, T. (2018). *Closing time: The effects of closing mechanism design on market quality* (tech. rep.). Working paper.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, *80*(4), 549–560.

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *The Journal of finance*, *65*(4), 1237–1267.

Eide, E., & Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics letters*, *58*(3), 345–350.

Ellul, A., Shin, H. S., & Tonks, I. (2005). Opening and closing the market: Evidence from the london stock exchange. *Journal of Financial and Quantitative Analysis*, *40*(4), 779–801.

Fitzenberger, B., & Franz, W. (1999). Industry-level wage bargaining: A partial rehabilitation—the german experience. *Scottish journal of political economy*, *46*(4), 437–457.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial markets*, *2*(2), 99–134.

Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *The review of financial studies*, *18*(4), 1171–1217.

Gao, L., Han, Y., Li, S. Z., & Zhou, G. (2018). Market intraday momentum. *Journal of Financial Economics*, *129*(2), 394–414.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, *14*(1), 71–100.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Hagströmer, B., & Nordén, L. (2013). The diversity of high-frequency traders. *Journal of Financial Markets*, *16*(4), 741–770.

Hagströmer, B., & Nordén, L. (2014). Closing call auctions at the index futures market. *Journal of Futures Markets*, *34*(4), 299–319.

Hasbrouck, J. (2018). High frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis*.

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

Hillion, P., & Suominen, M. (2004). The manipulation of closing prices. *Journal of Financial Markets*, *7*(4), 351–375.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal*, *7*(3), 281–312.

Hu, E., & Murphy, D. (2020). *Vestigial tails? floor brokers at the close in modern electronic markets* [Working Paper].

Ibikunle, G. (2015). Opening and closing price efficiency: Do financial markets need the call auction? *Journal of International Financial Markets, Institutions and Money*, *34*, 208–227.

Inci, A. C., & Ozenbas, D. (2017). Intraday volatility and the implementation of a closing call auction at borsa istanbul. *Emerging Markets Review*, *33*, 79–89.

Kandel, E., Rindi, B., & Bosetti, L. (2012). The effect of a closing call auction on market quality and trading strategies. *Journal of Financial Intermediation*, *21*(1), 23–49.

Koenker, R. (1994). Confidence intervals for regression quantiles. *Asymptotic statistics* (pp. 349–359). Springer.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, *15*(4), 143–156.

Korajczyk, R. A., & Murphy, D. (2019). High-frequency market making to large institutional trades. *The Review of Financial Studies*, *32*(3), 1034–1067.

Kuo, W., & Li, Y.-C. (2011). Trading mechanisms and market quality: Call markets versus continuous auction markets. *International review of Finance*, *11*(4), 417–444.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 1315–1335.

Madhavan, A. (1992). Trading mechanisms in securities markets. *The Journal of Finance*, *47*(2), 607–641.

McInish, T. H., & Wood, R. A. (1992). An analysis of intraday patterns in bid/ask spreads for nyse stocks. *the Journal of Finance*, *47*(2), 753–764.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation. *Econometrica*, *55*(3), 703–708.

Pagano, M. (1989a). Endogenous market thinness and stock price volatility. *The Review of Economic Studies*, *56*(2), 269–287.

Pagano, M. (1989b). Trading volume and asset liquidity. *The Quarterly Journal of Economics*, *104*(2), 255–274.

Pagano, M., Peng, L., & Schwartz, R. A. (2013). A call auction's impact on price formation and order routing: Evidence from the nasdaq stock market. *Journal of Financial Markets*, *16*(2), 331–361.

Pagano, M., & Schwartz, R. A. (2003). A closing call's impact on market quality at euronext paris. *Journal of Financial Economics*, *68*(3), 439–484.

Pascual, R., & Veredas, D. (2009). What pieces of limit order book information matter in explaining order choice by patient and impatient traders? *Quantitative Finance*, *9*(5), 527–545.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of applied econometrics*, *22*(2), 265–312.

Poterba, J. M., & Rueben, K. S. (1995). The effect of property-tax limits on wages and employment in the local public sector. *The American Economic Review*, *85*(2), 384–389.

Raillon, F. (2020). The growing importance of the closing auction in share trading volumes. *Journal of Securities Operations & Custody*, *12*(2), 135–152.

Roşu, I. (2009). A dynamic model of the limit order book. *The Review of Financial Studies*, *22*(11), 4601–4641.

Van Kervel, V., & Menkveld, A. J. (2019). High-frequency trading around large institutional orders. *The Journal of Finance*, *74*(3), 1091–1137.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *52*(3), 331–350.

Appendix

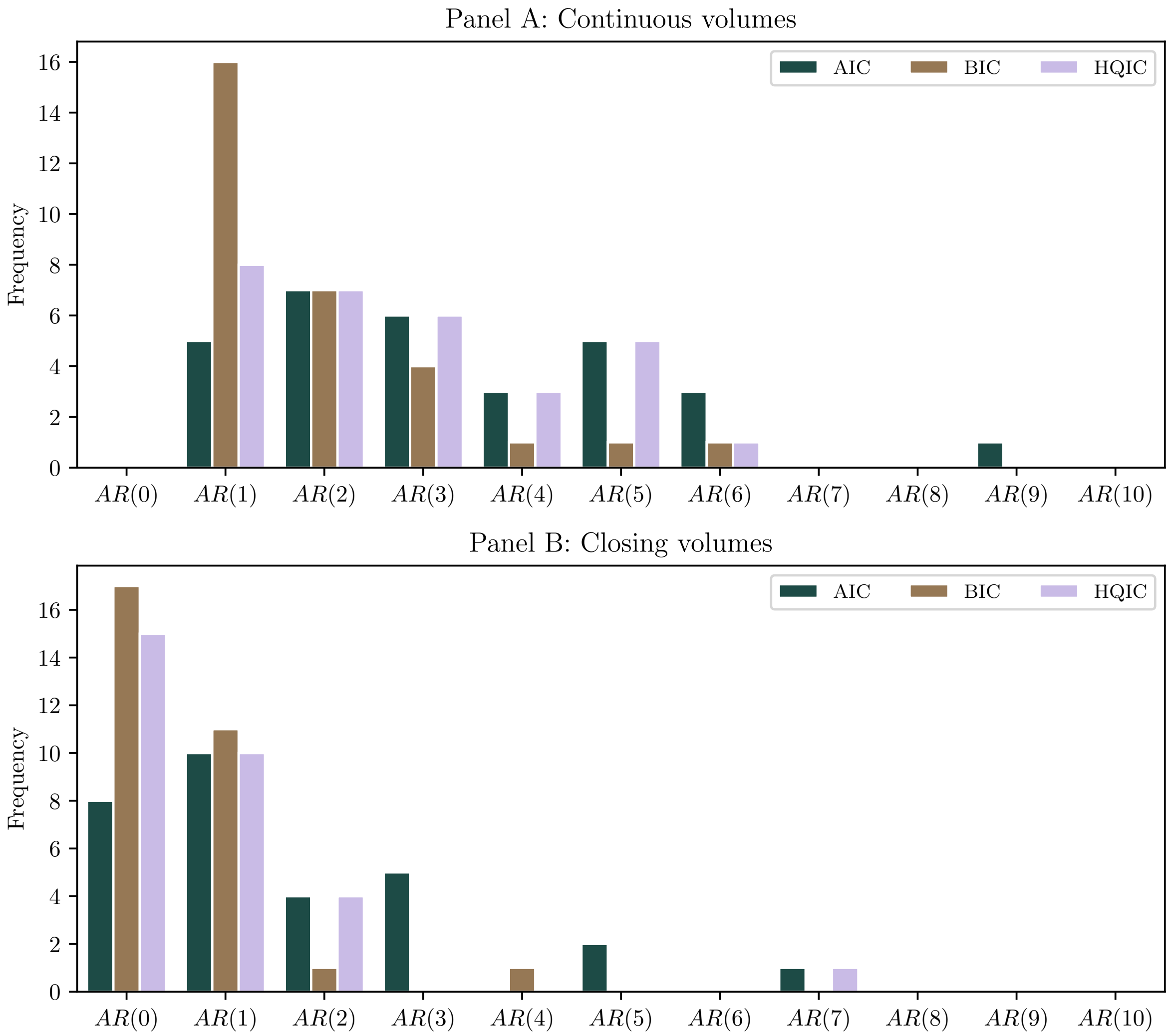


Figure 6: Distribution of optimal lag by information criterion. The underlying AR(p) models to this figure are estimated using a constant and linear time trend. Panel A reports information criteria for volumes of the continuous trading phase. Panel B reports information criteria for volumes of the closing auction. The analysis includes estimations of three common information criteria: Akaike (AIC), Bayesian (BIC) and Hannan-Quinn (HQIC). The information criteria were calculated for all $AR(p)$ models for $p \in \{0, 1, \ldots, 10\}$. All the autoregressive models are estimated with a constant and a linear trend, in order to account for changes in levels. The models were fitted for each individual stock of the sample. Out of these three criteria, the BIC offers the most restrictive properties and is therefore assigned the highest importance for the further analysis.

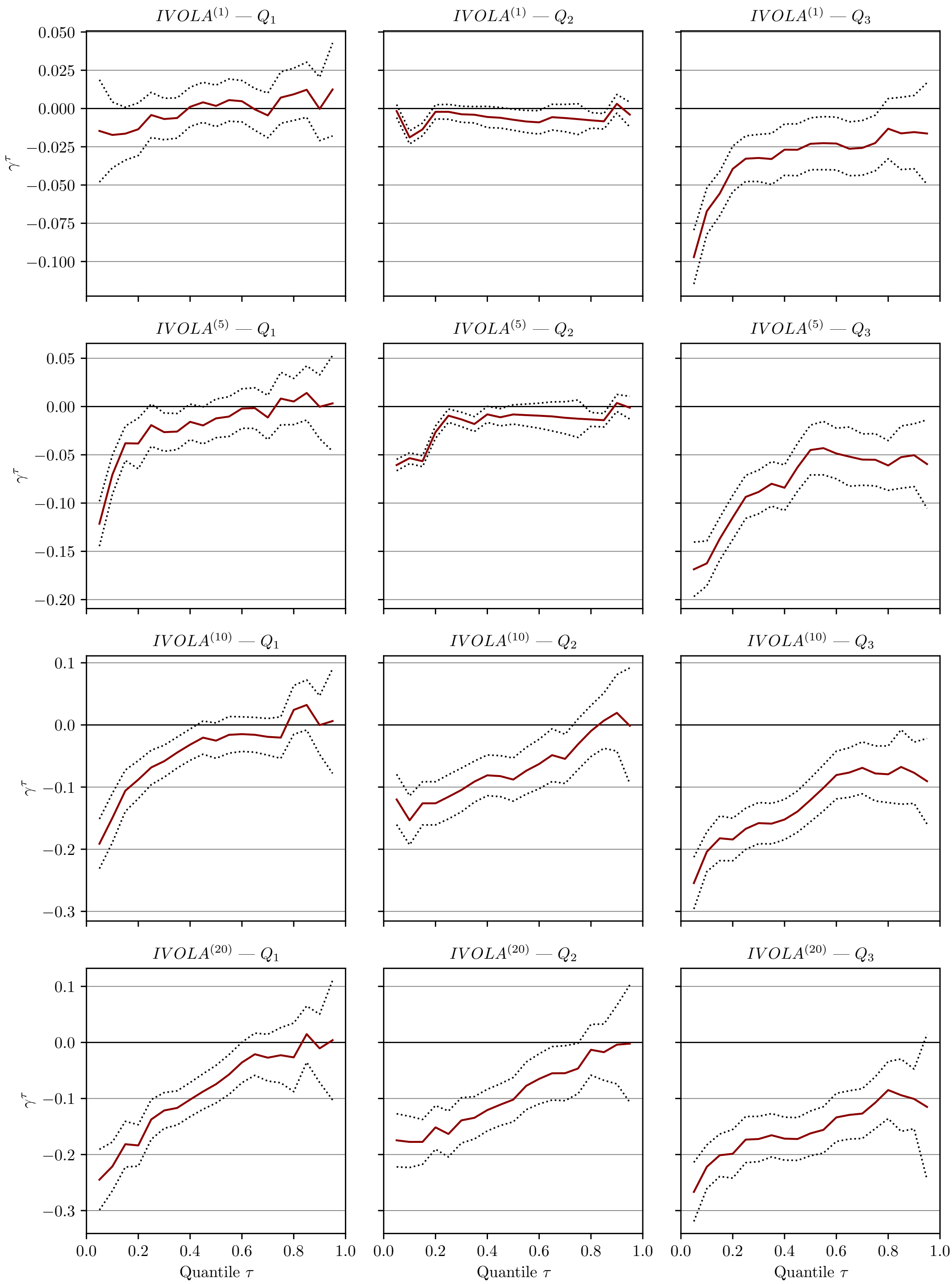


Figure 7: Robustness results of the quantile regression on intraday volatility. This figure presents the results from the quantile regression on measures of intraday volatility computed over a time horizon of 1,5,10,20 minutes (defined in (8)). The regression equation is defined in (12). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

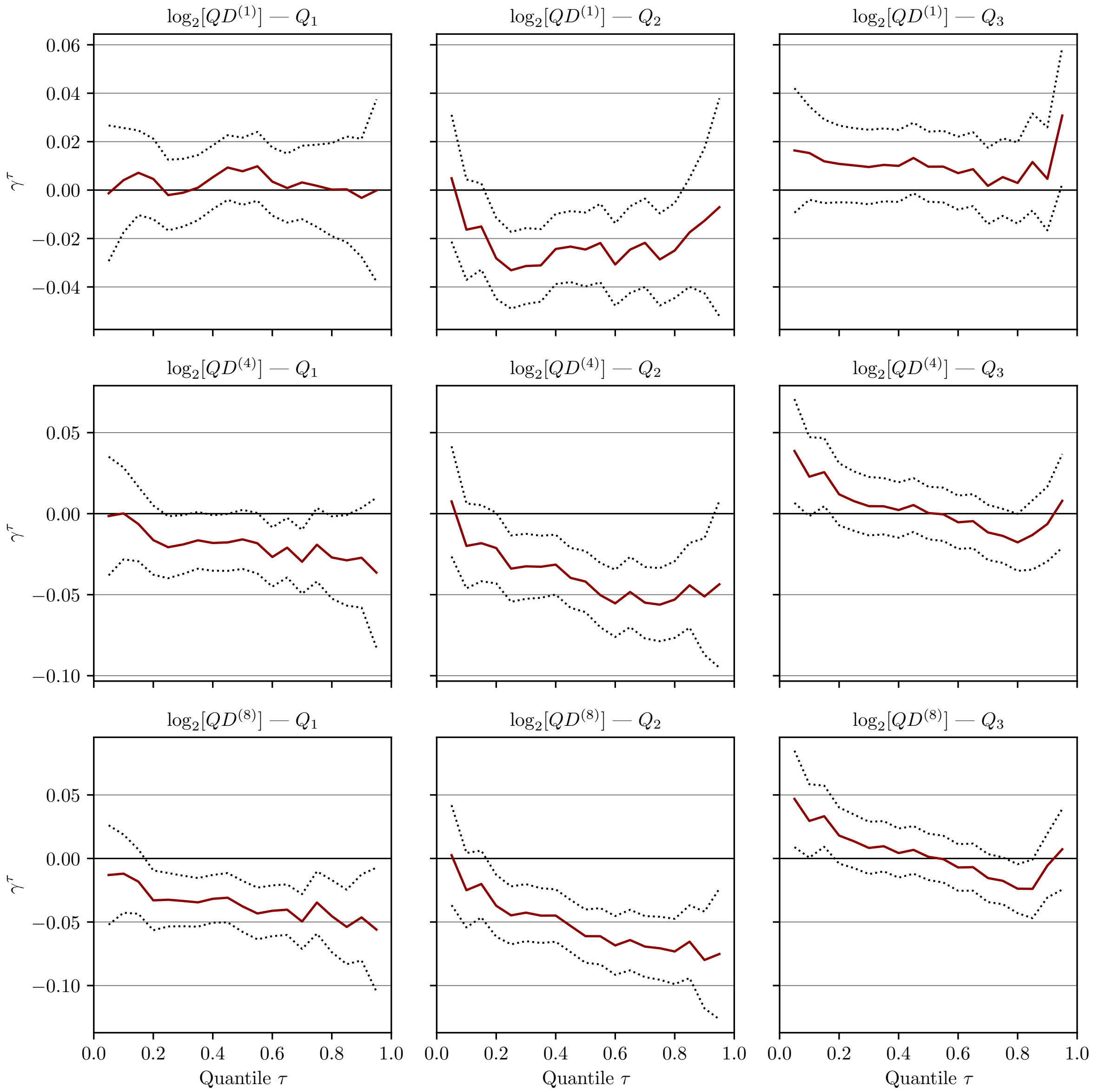


Figure 8: Robustness results of the quantile regression on market depth. This figure presents the results from the quantile regression on measures of market depth, i.e. the quoted depth over the top 1,4,8 levels of the order book. The regression equation is defined in (12). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

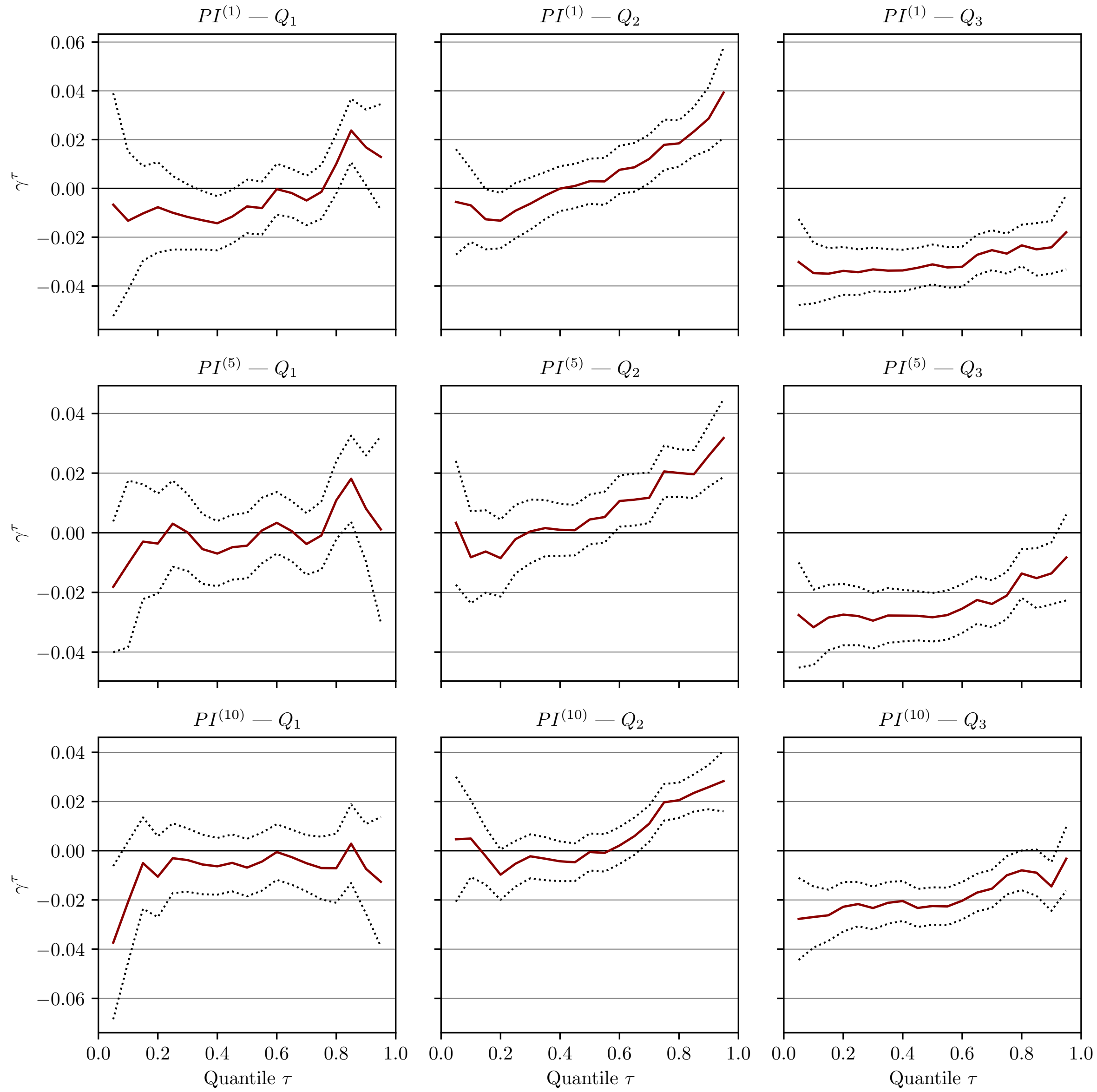


Figure 9: Robustness results of the quantile regression on market resilience. This figure presents the results from the quantile regression on measures of market resilience, i.e. the price impact over a time horizon of 1,5,10 minutes (defined in (8)). The regression equation is defined in (12). Each panel represents an independent estimation of the regression equation. The horizontal axis represents the quantile of the dependent variable, i.e. the de-meaned logarithm of the closing volume $\log_2[V_{close}]^*_{s,d}$. The regression is estimated for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$. The vertical axis exhibits the coefficient of the independent variable of interest. The dotted lines represent the confidence intervals at the level of 95% significance.

Panel A: $\log_2[IVOLA^{(1)}]$ — $R^2 = 0.4359$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0987***	-0.0640***	-0.0559***	-0.0446***	-0.0349***	-0.0638***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0000)
Q_2	-0.0470***	-0.0311**	-0.0164	-0.0238**	-0.0071	-0.0399***
	(0.0025)	(0.0128)	(0.1123)	(0.0161)	(0.3995)	(0.0008)
Q_3	0.0218	0.0258*	0.0215*	0.0278**	0.0324***	-0.0105
	(0.2541)	(0.0737)	(0.0794)	(0.0125)	(0.0017)	(0.4419)
$Q_1 - Q_3$	-0.1205***	-0.0897***	-0.0774***	-0.0724***	-0.0672***	-0.0533***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)

Panel B: $\log_2[IVOLA^{(5)}]$ — $R^2 = 0.4370$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.1089***	-0.0865***	-0.0827***	-0.0711***	-0.0549***	-0.0540***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Q_2	-0.0661***	-0.0516***	-0.0386***	-0.0430***	-0.0280***	-0.0382***
	(0.0003)	(0.0006)	(0.0028)	(0.0002)	(0.0083)	(0.0012)
Q_3	-0.0025	0.0009	0.0054	0.0089	0.0128	-0.0153
	(0.9063)	(0.9555)	(0.7228)	(0.5246)	(0.3012)	(0.2292)
$Q_1 - Q_3$	-0.1064***	-0.0874***	-0.0881***	-0.0800***	-0.0677***	-0.0387***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0008)

Panel C: $\log_2[IVOLA^{(10)}]$ — $R^2 = 0.4376$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.1288***	-0.1027***	-0.1035***	-0.0806***	-0.0763***	-0.0526***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Q_2	-0.0733***	-0.0710***	-0.0664***	-0.0505***	-0.0452***	-0.0282**
	(0.0010)	(0.0001)	(0.0001)	(0.0006)	(0.0005)	(0.0210)
Q_3	-0.0122	-0.0147	-0.0152	-0.0088	-0.0036	-0.0086
	(0.6239)	(0.4594)	(0.4099)	(0.6050)	(0.8047)	(0.5373)
$Q_1 - Q_3$	-0.1166***	-0.0881***	-0.0883***	-0.0718***	-0.0727***	-0.0440***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)

Panel D: $\log_2[IVOLA^{(20)}]$ — $R^2 = 0.4384$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.1607***	-0.1476***	-0.1255***	-0.1211***	-0.1028***	-0.0579***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Q_2	-0.1171***	-0.1009***	-0.0903***	-0.0891***	-0.0732***	-0.0439***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)
Q_3	-0.0599**	-0.0446**	-0.0503**	-0.0385**	-0.0332**	-0.0267*
	(0.0287)	(0.0448)	(0.0164)	(0.0460)	(0.0498)	(0.0624)
$Q_1 - Q_3$	-0.1007***	-0.1030***	-0.0752***	-0.0826***	-0.0696***	-0.0312***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0024)

Table 7: Robustness results of binning regressions for measures of execution risk. This table presents the binned γ coefficients from (9). Both the dependent and independent variables are measured in logarithms with base 2 to assure comparability across market quintiles. Each panel represents an independent estimation of coefficients. Panels A,B,C and D show the result when including the logs of intraday volatility measured over a time horizon of 1,5,10,20 minutes respectively, as defined in (6). Each panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively. Additionally, each panel tests for equality of the largest and smallest quantile using Wald tests.

Panel A: $\log_2(QD^{(1)})$ — $R^2 = 0.4384$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0065	0.0053	0.0057	0.0014	-0.0033	-0.0033
	(0.5822)	(0.6102)	(0.5675)	(0.8836)	(0.6993)	(0.6184)
Q_2	0.0309***	0.0361***	0.0282***	0.0202**	0.0122	0.0187***
	(0.0051)	(0.0009)	(0.0040)	(0.0320)	(0.1754)	(0.0001)
Q_3	0.0664***	0.0664***	0.0623***	0.0538***	0.0481***	0.0182***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0009)
$Q_1 - Q_3$	-0.0729***	-0.0611***	-0.0566***	-0.0524***	-0.0514***	-0.0215***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0058)

Panel B: $\log_2(QD^{(4)})$ — $R^2 = 0.4369$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0253	-0.0374	-0.0318	-0.0336*	-0.0163	-0.0090
	(0.3802)	(0.1324)	(0.1426)	(0.0650)	(0.3701)	(0.6078)
Q_2	0.0303	0.0295	0.0018	0.0067	0.0067	0.0236**
	(0.2408)	(0.2104)	(0.9335)	(0.7458)	(0.7185)	(0.0444)
Q_3	0.1159***	0.0880***	0.0756***	0.0682***	0.0564***	0.0595***
	(0.0000)	(0.0002)	(0.0006)	(0.0005)	(0.0019)	(0.0000)
$Q_1 - Q_3$	-0.1412***	-0.1254***	-0.1073***	-0.1018***	-0.0727***	-0.0685***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Panel C: $\log_2(QD^{(8)})$ — $R^2 = 0.4367$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0214	-0.0759**	-0.0502*	-0.0665***	-0.0347*	0.0133
	(0.6307)	(0.0146)	(0.0698)	(0.0013)	(0.0778)	(0.7145)
Q_2	0.0322	0.0094	0.0005	-0.0139	-0.0060	0.0382**
	(0.3329)	(0.7339)	(0.9862)	(0.5489)	(0.7696)	(0.0421)
Q_3	0.1592***	0.1123***	0.0912***	0.0790***	0.0663***	0.0930***
	(0.0000)	(0.0002)	(0.0008)	(0.0010)	(0.0016)	(0.0000)
$Q_1 - Q_3$	-0.1807***	-0.1882***	-0.1414***	-0.1455***	-0.1010***	-0.0797**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0184)

Table 8: Robustness results of binning regressions for measures of depth. This table presents the binned γ coefficients from (9). Both the dependent and independent variables are measured in logarithms with base 2 to assure comparability across market quintiles. Each panel represents an independent estimation of coefficients. Panels A,B and C show the results when including the logs of quoted depth in currency terms at the top 1,4,8 levels of the book respectively. Each panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively. Additionally, each panel tests for equality of the largest and smallest quantile using Wald tests.

Panel A: $\log_2(PI^{(1)})$ — $R^2 = 0.4404$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	0.0149	0.0877***	0.0824***	0.1335***	0.0828***	-0.0679**
	(0.3888)	(0.0022)	(0.0025)	(0.0005)	(0.0090)	(0.0182)
Q_2	-0.0071	-0.0019	0.0138	0.0403	0.0641**	-0.0711***
	(0.5816)	(0.9146)	(0.4955)	(0.1588)	(0.0458)	(0.0081)
Q_3	-0.1255***	-0.1220***	-0.1392***	-0.1213***	-0.1421***	0.0166
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.4891)
$Q_1 - Q_3$	0.1404***	0.2096***	0.2216***	0.2548***	0.2249***	-0.0845**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0153)

Panel B: $\log_2(PI^{(5)})$ — $R^2 = 0.4364$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	-0.0009	0.0425	0.1209***	0.1121**	0.0063	-0.0072
	(0.9507)	(0.1194)	(0.0041)	(0.0422)	(0.6758)	(0.7475)
Q_2	-0.0009	0.0164	0.0073	0.0655***	0.0684***	-0.0694***
	(0.9287)	(0.2415)	(0.6963)	(0.0079)	(0.0005)	(0.0016)
Q_3	-0.1035***	-0.1087***	-0.0897***	-0.1277***	-0.1181***	0.0145
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.5571)
$Q_1 - Q_3$	0.1026***	0.1512***	0.2106***	0.2398***	0.1244***	-0.0218
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0003)	(0.5160)

Panel C: $\log_2(PI^{(10)})$ — $R^2 = 0.4341$

	M_1	M_2	M_3	M_4	M_5	$M_1 - M_5$
Q_1	0.0054	0.0522**	0.0838**	0.1807***	-0.0132	0.0186
	(0.7259)	(0.0416)	(0.0328)	(0.0092)	(0.3232)	(0.4271)
Q_2	0.0075	0.0192	0.0572***	0.0742***	0.0281*	-0.0206
	(0.5257)	(0.2271)	(0.0036)	(0.0030)	(0.0568)	(0.2744)
Q_3	-0.0713***	-0.0757***	-0.0687***	-0.0824***	-0.0263	-0.0450*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2422)	(0.0504)
$Q_1 - Q_3$	0.0766***	0.1279***	0.1525***	0.2631***	0.0131	0.0635*
	(0.0001)	(0.0000)	(0.0005)	(0.0002)	(0.6224)	(0.0650)

Table 9: Results of binning regressions for measures of depth and resilience. This table presents the binned γ coefficients from (9). Both the dependent and independent variables are measured in logarithms with base 2 to assure comparability across market quintiles. Each panel represents an independent estimation of coefficients. Panels A,B and C show the results when including the logs of the price impacts over time horizons of 1,5,10 minutes respectively as defined in (8). Each panel was estimated using entity-fixed effects. Reported standard errors are derived using the Driscoll-Kraay covariance matrix. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively. Additionally, each panel tests for equality of the largest and smallest quantile using Wald tests.

Chapter 2

Price Discovery Within Closing Auctions

Price Discovery Within Closing Auctions*

Louis Müller†

June, 2022

Working Paper

Abstract

At a time when financial markets rely increasingly on accurate closing prices it is paramount to understand the discovery process that leads to these prices. This paper analyses order flow patterns throughout the closing call auctions on the Swiss stock exchange. Price dislocations in the first eight minutes of the auction are on average reverted by more than 90%. Furthermore, the results indicate that aggressive limit orders are submitted in the opposite direction with respect to the previous price dislocation and thus exerting a counteracting force. In contrast, market orders are submitted independently regarding the price path. Moreover, market order imbalances are found to contain some amount of information particularly in the beginning of the auction, which is ultimately reversed. Finally, closing returns are less efficient following auctions with negative price dislocation and upward reversion, indicating overcompensation of informative sell order flow. Overall, the findings suggest that closing auctions are so effective at absorbing liquidity shocks that they may in fact conceal informative order flow.

Keywords: Closing auctions, price discovery, price efficiency, order imbalances
JEL Codes: G12, G14

*This work has been enabled by the cooperation of SIX Securities & Exchanges (SIX) and the Institute for Operations Research and Computational Finance (ior/cf-HSG) at the University of St. Gallen. I would hereby like to thank SIX for supplying detailed and granular order-level data from their equities exchange.

†University of St. Gallen, Switzerland: louis.mueller@unisg.ch

1 Introduction

The rising importance of closing auctions as daily trading facility has been increasingly scrutinized by academic literature as well as by practitioners. Among other trends, shifts towards passive investment strategies and increasingly strict best-execution requirements for brokers have lead to closing auctions capturing market share of around 40% on Euronext Paris, at the expense of the continuous trading phase (Raillon, 2020). The reason for this lies in the importance of closing prices as a tool for benchmarking entire baskets of stocks. This development is remarkable despite the relatively short duration of these auctions versus the rest of the trading day. However, it can be explained by the favorable properties of call auctions compared to quote driven continuous markets as soon as there is large demand for liquidity. These advantages of closing auctions as a new trading facility were first addressed in Madhavan (1992) and have been well documented ever since.

This increase in demand for execution at the close is what mainly motivates this paper. With the increasing amount of capital that is transacted during this time of the day, obtaining efficient prices becomes particularly crucial. It is therefore paramount to better understand the process under which the price is formed incrementally over the course of the auction. There are already numerous papers on the the process of price discovery in the continuous trading phase, which was driven by the fact that most volume throughout a trading day is (still) transacted during this time. However, the closing auction has not been analyzed with the same rigour.

Among other parties, this research may be particularly interesting for financial regulators as well as operators of exchanges and other trading platforms. Due to the increasing volumes at close as opposed to the rest of the trading day, many alternative trading platforms have made efforts to to capture some of this volume at the expense of traditional exchanges. In addition to this, trading venues tend to charge higher fees for trades executed during the closing auction as opposed to the continuous trading phase. This development threatens to disrupt traditionally highly centralized closing auctions by fragmenting the order flow over multiple venues. In contrast to continuous trading, closing auctions must result in one single price that is universally accepted (Kandel et al., 2012), which is determined on the listing exchange. However, the fragmentation of the order flow may have adverse effects on price discovery and ultimately leads to less efficient closing prices.

For all these reasons, it is important to understand the process in more detail by analyzing closing auctions on a granular level, with particular focus on timing and effect of incoming order flow. Besson and Fernandez (2021) recently stated that closing auctions on various Euronext exchanges in Europe usually show large inflows of orders both at the beginning and the end of the auction, whereas it is much more quiet in between. This clustering of volume indicates that investors act strategic, particularly the ones submitting their orders at the end of the auction. This pattern is reminding of observations made by McInish and Wood (1992), whereas spreads are highest in the beginning and the end of the trading day.

The order flow behind this pattern is the main focus of this paper. Particularly, this paper distinguishes between market- and aggressive limit order flow. The academic literature has shown frequently, that these types of orders are used by different types of traders. Some relevant examples for this are Foucault et al. (2005), Goettler et al. (2005), Kaniel and Liu (2006), and Roşu (2009). The literature finds that during continuous trading, market orders are predominantly used by impatient investors who value immediacy of execution over the optimal execution price. However, in during call auctions, these order types have fundamentally different implications since the execution of all orders falls on the same timestamp. Therefore, investors do not have the opportunity to take advantage of temporary mispricings before the auction clears. Contrary to the continuous trading phase, market orders in call auctions can only only guarantee *certain* execution but not *immediate* execution and thus, price uncertainty remains until final clearing.

There are multiple research questions that this paper addresses. First, the question of whether investors adjust their order submission strategy based on the past price path is addressed. The answer to this question allows an assessment of whether order flow is occurring in reaction to the unrealized return into the auction or whether it is independent thereof. Second, it looks into the question of how well the order flow predicts the future return within the scope of the auction auction. The answer to this question allows for the distinction between informed and uninformed order flow. While there is disagreement in academic literature about whether informed investors use limit or market orders for the continuous auction, this paper provides an answer to this question in the context of closing auctions. The third question is about whether the price path has an impact on the efficiency of the final closing price. This particularly concerns the occurrence of large price dislocations and the subsequent correction before clearance. For instance, it may be plausible that large initial spikes may go uncorrected until the end of the auction and only revert overnight or that informative spikes are overlooked and reverted.

The methodology used to answer these questions differs in several aspects from previous analyzes have not been done before in academic literature of closing auctions. One important point is that closing auctions are sliced into 61 intervals of 10 seconds duration each in order to obtain a reasonably granular picture of order flow. In addition to this, there is a strict focus on aggressive orders, particularly in the context of limit order flow[1]. More specifically, aggressive limit buy(sell) orders are strictly above(below) the current clearing price and therefore affect that price going forward. This calculation is highly dependent on the current state of the order book and therefore allows an accurate picture of price discovery. To the best of my knowledge, there is no other study that isolates aggressive order flow with comparable rigour. Finally, the analysis applies Multinomial Logit models that focus on the sign of order imbalances. This paper concentrates on directional order imbalances due to elevated levels of noise in order flow data or the complete absence thereof.

This analysis yields several main findings. To begin with, price dislocations in the first 8 minutes of the auction are reversed by an extent of over 90%, indicating that there is substantial amounts of noise during this period. Only in the last two minutes of the auction the price starts to approach the ultimate closing price. Another finding states that aggressive limit order imbalances are pointing against the return since the beginning of the auction and therefore exert a counteracting force onto the price. In contrast to this, market orders arrive independently without regards to the previous price path. However, imbalances of market orders negatively predict the return until the end of the auction when accounting for the past price path. The same phenomenon does not apply to imbalances of aggressive limit orders, which don't reveal any additional information. Lastly, the analysis shows that closing returns are partly reverted to over several time periods overnight. Closing prices are collectively rejected to be martingale by means of statistical tests. The inefficiency is particularly pronounced after auctions with the largest negative price dislocations that are reverted upwards before the auction clearing.

The interpretation of these findings leads to the main contributions of this paper with respect to the existing body of literature. First, the reversion of initial price shocks is indicative of closing auctions being successful in what they were designed to achieve, that is absorbing large and correlated shocks of liquidity. A liquidity shock in the paper refers to a situation when a large amount of orders enters or leaves the order book within a short period of time, which can significantly change the order book and ultimately change the uncrossing price. However, this raises the question of whether this consistent price reversion may also curtail new information which should be incorporated into prices for the sake of price efficiency. Using this observation, informed investors who wish to keep their advantage from the market until having fully entered their position may consider

[1] In call auctions, limit orders that are passive (i.e., non-aggressive) have no impact on the closing price, because they would not be executed under the current order book. In contrast to this, aggressive limit orders have an impact on the uncrossing price.

to submit market orders in the beginning of the closing auction to reach their objective. Second, market order and aggressive limit orders are submitted by two distinct types of investors. On the one hand, market orders are used for what academic literature calls *liquidity traders*[2], who do not have a view on the price but require liquidity to rebalance their portfolios. On the other hand, aggressive limit orders are submitted by investors in reaction to the past price path. Examples for this type may be investors who perform a market making function. Third, the fact that closing returns are not collectively efficient implies that the dissemination of new information is limited, in particular due to the partial reversion of such returns overnight. However, the observation that closing returns are particularly inefficient following downward spikes that are reversed upward may indicate that market makers may falsely compensate valid sell signals, which in fact were informative.

This paper is structured as follows. Section 2 briefly surveys relevant literature to this topic and how this paper complements the existing body. Section 3 elaborates the data extraction process and introduces the call auction mechanics on the Swiss stock exchange. Section 4 presents the first set of results relating to the interaction of price discovery and order flow within the closing auction. Section 5 takes a step back and assesses the efficiency of closing prices by means of unbiasedness regressions with respect to longer time horizons. Finally, section 6 summarizes the findings and provides an outlook for potentially useful follow-on research.

2 Related Literature

The focus of this paper lies on the interaction of price discovery and order flow during the closing auction. There exists an extensive body of literature on price discovery during the continuous trading phase. Particularly during the last two decades, researchers in market microstructure were heavily scrutinizing the impacts of high-frequency onto market conditions for all remaining investors (e.g. Chaboud et al. (2014) and Hendershott et al. (2011)). In parallel to this, the focus has recently also shifted more to towards closing auctions due to rising relative volumes, constituting an increasingly important liquidity event. In this niche, most of the literature is concerned with the introduction of call auctions at the end of the day, which happened around 20 years ago around the world. What has not been properly addressed in the literature is order flow throughout the closing auction. The reason for this may be twofold. First, data for auctions are not easy to obtain and a large amount of programming is required to reconstruct order books throughout the

[2] For further information refer to Anand et al. (2005), Bloomfield et al. (2005), Kalay and Wohl (2009), and Rindi (2008) as examples.

entirety of the auction. Second, due to the nature of call auctions no trades are occurring until the uncrossing of the auction. This is in contrast to the continuous trading phase, during which investors have the opportunity to exploit potential mispricings at any time. Consequently, paper attempts to fill exactly this gap in the literature. This section will provide a brief survey of the most relevant literature with respect to both price discovery and closing auctions as well as their currently limited overlap.

To begin with, the literature on price discovery during the continuous phase is discussed in brief. For this paper, Foucault et al. (2005) and Roşu (2009) are among the most relevant published works. Both constitute dynamic theoretical models for the explanation of order flow types, with similar results. Foucault et al. (2005) finds that patient traders tend to use limit orders for their trades, as they value price certainty over immediacy of execution. Roşu (2009) shows that market orders lead to a temporary price overshoot but are likely corrected over time and that higher trading activity causes lower price impact. Similarly, Kaniel and Liu (2006) propose a model derived from the Glosten and Milgrom (1985) equilibrium which indicates that informed traders prefer to use limit orders instead of market orders, therefore indicating that limit order imbalances are more informative about future price movements, whereas market orders are used more by investors with requirements to rebalance their portfolio (so called liquidity traders). In contrast to this, Anand et al. (2005) find that informed investors tend to prefer the usage of market orders in order to take advantage of temporary mispricings in the market. Moreover, Griffiths et al. (2000) also find that investors make a conscious decision about order choice, weighing opportunity cost versus price impact. Additionally, they find that market order flow is often positively correlated and thus revealing new information to other market participants.

Moreover, there are also relevant works with respect to the endogeneity of limit order books. One of the most important papers in this regard is Parlour (1998), who develops a theoretical model in which all investors are exactly aware of the state of the order book as well as of the effect of their own order placement strategies and thus act in a highly strategic manner. Consequently, all investors optimize for their optimal outcome considering the subsequent reactions of other market participants. As an extension to this, Pascual and Veredas (2009) apply a sequential model empirically and show that patient investors adjust their order submissions based on the state of the order book on their side of the book, whereas impatient traders mostly care about the opposite side of the book. Importantly, Pascual and Veredas (2009) look at continuous order books, where best bid- and ask- orders cannot overlap unlike during call auctions.

Due to their design, call auctions have significantly different properties as opposed to the continuous trading phase. The most important two of which are the inability to transact immediately before the final clearing and the overlap of bid and ask order books. The early literature with respect to call auctions, which were first proposed in the seminal paper by Madhavan (1992), were concerned with the introduction to such markets as a complement to the existing continuous trading. Most papers concluded that the introduction of call auctions for both opening and closing had positive effects for investors due to these auctions being designed to absorb large simultaneous liquidity shocks. The beneficial effects observed after the introduction of such auctions mainly consist of a reduction in volatility and bid-ask spreads (Aitken et al., 2005; Hagströmer & Nordén, 2014; Inci & Ozenbas, 2017; Kandel et al., 2012; Kuo & Li, 2011; Pagano et al., 2013), improved price efficiency (Barclay et al., 2008; Bellia et al., 2017; Biais et al., 1999; Comerton-Forde et al., 2007; Cordi et al., 2018; Pagano & Schwartz, 2003) and less price manipulation (Comerton-Forde & Putniņš, 2011; Hillion & Suominen, 2004).

More recent contributions to the literature go beyond this event study approach and analyze the market mechanism in more detail. For instance, Barclay and Hendershott (2003) find that after-hours trading in the US is has an impact on price discovery, despite the fact that the resulting prices are mostly inefficient. Moreover, they find that the average trade after-hours contains more information as opposed to during normal trading hours, however, overall most information is still revealed during normal trading hours. In addition to this, Barclay et al. (2008) find that a centralized call auction outperforms a decentralized mechanisms, where there are different order books existing simultaneously. This is particularly true on days with high demand for liquidity such as triple witching days[3] without substantial new information in terms of pricing but solely for rebalancing purposes. Therefore, consolidation of auctions (e.g. on the main exchange) is potentially the most desirable outcome with respect to price discovery. Barclay and Hendershott (2008) also find that since the introduction of pre-open trading some of the price discovery has shifted out of the opening auction into the new facility.

Some more recent works on closing auctions include Bogousslavsky and Muravyev (2020) who look at how closing prices can be dislocated due to liquidity shocks caused by large institutional investors. They find that closing prices are quite inefficient and are often reversed over night and sometimes even open short-lived arbitrage opportunities by means of a violation of the put-call parity. The authors are cautious of potential over-reliance on closing prices for benchmarking and rebalancing purposes due to such non-informative distortions. In addition to this, Hu and Murphy (2020) show that the New York Stock Exchange (NYSE) has bad price discovery properties, as physically present floor brokers have near-exclusive access to the book in the last few minutes of the auction. This allows

[3]Triple witching days occur on a quarterly basis when future and options expire on the same days. Many investors with open interest are required to equalize their positions to remain market neutral.

for superior information versus the rest of the market. This effect has been shown to be mitigated during the 2020 pandemic when the mechanism was halted due to stay-at-home orders. Moreover, they also find that order-driven NASDAQ call auctions outperform in terms of price efficiency.

Finally, the literature on detailed order flow during the auction is relatively slim. To begin with, Smith et al. (2003) develop a model for call auctions and the subsequent distribution of clearing prices given varying order flow during the continuous trading phase. Importantly, they make the limiting assumption of independent and identically distributed (IID) random order flow, which disregards any possibility for persistent order flow patterns inside the closing auction. The same model is extended by Derksen et al. (2020), who find that pricing preferences of actors during the trading day are realized during the closing auction. However, they only consider trading days during which the closing price lies within the intraday price range, which may constitute a sampling bias. Theissen and Westheide (2020) analyze the activity of designated market makers during the call auction on German equity markets and find that they are most active in stocks with relatively low liquidity as well as during periods of elevated volatility. This indicates that these agents are fulfilling their purpose of providing liquidity to the market when needed most. The last work discussed here is Besson and Fernandez (2021) which is a descriptive practitioner's report and most similar to this paper in terms of content. More specifically, they look at several closing auctions throughout Europe through a lens of incremental intervals. However, their study is different from this paper in three key points. First, the authors of the report do not single out aggressive limit order flow given by the simultaneous state of the order book but rather the final clearing. Second, they do not distinguish between inflows and outflows of orders, despite both being important for the determination of the ultimate call auction order book. Third, they do not capture the causal relationship of order flow depending on both the price path as well as previous order flow. Their main findings conclude that price impacts of market orders are smaller during call auctions as opposed to the continuous trading phase and that most activity as measured by volume happens in the beginning and the end of the auction. Overall, this paper does not find any conflicting evidence to their work.

3 Data

The data used in this paper was obtained directly from SIX, which is the national stock exchange of Switzerland. The data on Swiss equities offers very high granularity and includes every order and trade submitted to the exchange including the hidden ones[4]. The analyzed time period includes three calendar years from January 2018 until December 2020, comprising 743 unique trading days.

During this time period, stocks that are listed on SIX and therefore trade on their primary exchange are analyzed exclusively. Several filters are applied to the stock universe to narrow down the selection further. First, stocks with an average closing volume below CHF 5 million per day are disregarded. This mitigates the risk of including stocks that may not be liquid enough to result in a reliable auction crossing which may be an issue for many investors, as shown in Ellul et al. (2005) and Ibikunle (2015). Second, only stocks that have been in the sample for at least 250 trading days are considered. This step accounts undesirable effects for natural listings and de-listings. After the application of these filters, the stock-universe is ultimately reduced to 69 stocks. Taking into account the number of trading days for each stock results in a sample of 40,011 stock-days.

The closing auction on SIX is designed like a normal call auction. It is initiated at 17:20 after the continuous trading phase is halted, such that both trading phases are strictly non-overlapping. Investors are allowed to submit two types of orders. The first type are market orders that are executed at the ultimate closing price. Limit orders on the other hand are submitted with a limit price. Limit buy(sell) orders are only executed if the ultimate closing price is equal or below(above) the limit price. The auction lasts for at least 10 minutes. Thereafter, the closing occurs randomly within a two-minute interval, in order to prevent manipulation of the closing price[5]. At the close, all market- and limit orders are gathered into aggregated demand- and supply. The closing price is determined where crossed volume is maximized.

The way the data was sampled allows us to go much more granular than this. For data collection, a recursive algorithm was run for each stock-day in the sample to reconstruct the limit order book (LOB) at every point in time throughout the trading day[6], which also applies to the closing auction. For the purpose of this study, instantaneous snapshots of the LOB were sampled in 10-second intervals, starting at the beginning of the auction. The intervals were sampled even without any update in the LOB. The first interval of the auction at timestamp 0 includes two types of orders. First, orders that were in the order book during the continuous trading phase and did not expire. Second, orders that were placed before the beginning of the closing auction but have only been activated for

[4]For instance iceberg orders that are not visible to any other market participants apart from the submitting party and the exchange.

[5]Some relevant studies with respect to random endings of auctions are Comerton-Forde and Putniņš (2011), Cordi et al. (2018), and Hillion and Suominen (2004).

[6]This is done because all past orders within the same trading day may have an impact on the current LOB.

the auction[7]. In practice, the closing auction on SIX is lasting for at least 600 seconds, with an unpredictable closing thereafter in order to prevent manipulation of the closing price. In order to keep the panel of auctions well balanced, only the first 600 seconds of the closing call auction are considered for the most part of this analysis. This results in a total of 61 intervals per closing auction on each stock-day plus one observation of the last prevailing order book before the auction starts. This process of data-gathering results in around 2.7 million data points.

Each of these data points represents one full LOB at a given point in time, which is subsequently uncrossed in order to receive an indication of the current uncrossing price ($ival\_price$) as well as uncrossing volume, which is broadcasted to investors in real-time and hence public information. Based on the current uncrossing prices, the return between two intervals in percentage points is computed as

$$ival\_rets_{s,t,l} = \ln\left(\frac{ival\_price_{s,t,l}}{ival\_price_{s,t,l-1}}\right) \cdot 100 \tag{1}$$

where s and t stand for stock and trading day respectively and l represents the closing auction interval. This interval return measure will be calculated for each available interval starting with $ival\_rets_{s,t,0}$. In this setting, $ival\_price_{s,t,-1}$ represent the last observed midquote before the beginning of the closing auction. In contrast to this, $ival\_price_{s,t,0}$ represents observed uncrossing price at the very start of the auction, at which time the order book only contains orders that have been submitted before the closing auction[8]. Similarly, $ival\_price_{s,t,60}$ represents the order book after the 600-second timestamp into the auction, which is the last order book before the random closing window starts. Thus, we receive a total of 61 interval returns. The next step consists of the calculation of cumulative returns across intervals:

$$cumul\_rets_{s,t,l} = \sum_{m=0}^{l} ival\_rets_{s,t,m} \tag{2}$$

This measure represents the return between the last continuous LOB midquote and a given interval l during the closing auction. The reason for this additivity is the logarithmic nature of interval returns.

[7] A good example for this type of orders are market on close (MOC) orders, that are submitted by institutional investors such as ETFs who are benchmarked against closing prices.

[8] Despite these two timestamp occurring in the same seconds, the LOBs are fundamentally different. This has two main reasons. First, some orders are only valid during the continuous trading phase and expire automatically before the auction. Second, investors have the option to enter orders during the trading day that remain dormant during the continuous phase and only enter the book at the start of the auction. One example for this are market-on-close orders that allow investors to purchase a stock at the prevailing closing price.

One important contribution of this paper lies in the fact, that it exclusively considers *aggressive* order flow. An aggressive order thereby constitutes an order that would be executed at the time of submission given the simultaneous state of the LOB[9]. Clearly, market orders are always aggressive, as they induce execution at any price. For limit orders, aggressiveness is more nuanced as it depends on the simultaneous hypothetical uncrossing price. For instance, a limit buy(sell) order is only aggressive if its submission price is above(below) the current uncrossing price. The granular data provided for this study allows for the phasing out non-aggressive limit orders that do not have an immediate impact on the uncrossing. This is an improvement on the approach taken in Besson and Fernandez (2021), who classify aggressive limit orders with respect to the *final* closing price instead of the *current* one. By ignoring limit orders that are beyond the current closing price, the approach laid out in this paper more accurately depicts at what point during the auction limit orders have an impact on the uncrossing price.

As this study is analyzing the interaction between price discovery and order flow, it is crucial to define a measure of order imbalance. Empirically, there are several approaches to capture this measure. By definition, an imbalance measure aims to express the interaction between two flows (f_1 and f_2) which are both expressed in positive currency terms. This study follows a new approach to define such imbalances, which has not been used in academic literature so far to the best of my knowledge. More specifically, it uses natural logarithms in order to capture the degree of the order imbalance in both directions. The resulting metric is calculated as:

$$IMB(f_1, f_2) = \begin{cases} \ln(f_1 - f_2 + 1), & \text{if } f_1 \geq f_2 \\ -\ln(f_2 - f_1 + 1), & \text{otherwise} \end{cases} \qquad \forall \qquad f_1 \geq 0, f_2 \geq 0. \tag{3}$$

This definition comes with one particularly relevant property for the following statistical analysis. As opposed to the other approaches used in the literature, this measure has no lower and upper bounds. This enables the metric to capture the extent of imbalances more accurately and therefore allowing for more precise statements about counteracting- or reinforcing imbalances during the closing auction. One disadvantage of this measure, however, is the propensity to generate outliers, which may potentially affect consistency of coefficients under quadratic loss functions, such as with OLS estimation. For robustness purposes, all the results presented later have additionally been replicated using the

[9]An example for a paper looking into aggressiveness of order flow in the continuous trading phase is Degryse et al. (2005), who assign different levels of aggressiveness to obtain a more detailed view.

alternative order imbalance metric which is commonly used in academic literature[10]. The results are evidently found to be robust with respect to the choice of order imbalance metric. Nonetheless, there are some clear differences between the two measures, as depicted in fig. 9. The logarithmic measure defined in eq. (3) results in a differentiable plane at each combination of positive flows. The traditional academic measure shows a discreet jump around the point where both points are zero. As order flow during closing auctions can be rather sparse, due to its absence in many time intervals, the former option has more desirable qualities.

Based on eq. (3), several measures of order imbalance are computed in order to capture order flow in elaborate detail. In order to achieve this, types of order flows are to be defined. To begin with, $mkt\_buys\_in_{s,t,l}(mkt\_buys\_out_{s,t,l})$ represents the logarithmic volume of inflowing(outflowing) market buy orders in stock s, on day t and during interval l since the start of the auction. Similarly, $mkt\_sells\_in(mkt\_sells\_out)$ represents inflows(outflows) of market sell orders. For aggressive limit orders, the exact same logic applies, but the prefix in each variable is replaced with $lim\_$. All of these variables are used to define the first set of order flows:

$$
\begin{aligned}
mkt\_in\_imb_{s,t,l} &= IMB\left(mkt\_buys\_in_{s,t,l},\ mkt\_sells\_in_{s,t,l}\right)\\
mkt\_out\_imb_{s,t,l} &= IMB\left(mkt\_buys\_out_{s,t,l},\ mkt\_sells\_out_{s,t,l}\right)\\
lim\_in\_imb_{s,t,l} &= IMB\left(lim\_buys\_in_{s,t,l},\ lim\_sells\_in_{s,t,l}\right)\\
lim\_out\_imb_{s,t,l} &= IMB\left(lim\_buys\_out_{s,t,l},\ lim\_sells\_out_{s,t,l}\right)
\end{aligned}
$$

These measures capture the imbalances of in- and outflows for both market and limit orders. Call auction markets are quite unique in the sense that a one-sided outflow of orders[11] may have a significant impact on the uncrossing price. This is particularly true for the removal of market order from the book, which is not possible during the continuous

[10] In contrast to the more common order imbalance measure in eq. (3), imbalances may also be defined as a limited metric. In the academic literature, the metric most commonly used and defined in Chordia et al. (2002, 2008), Chordia and Subrahmanyam (2004), and Holden and Jacobsen (2014) is the following:

$$
ALTIMB(f_1, f_2) = \begin{cases} (f_1 - f_2)(f_1 + f_2)^{-1}, & \text{if } f_1 + f_2 > 0 \\ 0, & \text{otherwise} \end{cases} \qquad \forall \quad f_1 \geq 0, f_2 \geq 0.
$$

According to this definition, the imbalance value is bound by $IMB \in [-1, 1]$. The measure can take the value 1(-1) under the condition that $f_1(f_2)$ is positive while $f_2(f_1)$ is zero. Therefore, the measure takes into account the magnitude of order imbalance to a lesser extent, which mitigates the issue of outliers frameworks based on OLS. The disadvantage of this measure, however, is the inability to capture the extent of order imbalances when one side does not show any activity. This is crucial in this analysis as most intervals are quiet from at least one side of the book most of the time.

[11] An outflow of orders means that existing orders are cancelled, which in turn affects the ultimate order book

	μ	σ	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
mkt_in_imb	0.04	7.21	-13.54	-12.69	-11.39	0.00	0.00	0.00	11.40	12.70	13.55
mkt_out_imb	-0.03	3.82	-11.89	-8.43	0.00	0.00	0.00	0.00	0.00	7.79	11.75
lim_in_imb	-0.18	7.92	-13.74	-12.94	-11.84	0.00	0.00	0.00	11.68	12.83	13.64
lim_out_imb	-0.11	4.98	-12.53	-11.31	-7.02	0.00	0.00	0.00	0.00	11.02	12.37
mkt_imb	0.01	7.85	-13.71	-12.93	-11.83	0.00	0.00	0.00	11.80	12.92	13.70
lim_imb	-0.22	8.04	-13.69	-12.90	-11.85	-7.52	0.00	0.00	11.67	12.78	13.60
ival_rets	0.00	0.55	-0.80	-0.38	-0.13	0.00	0.00	0.00	0.13	0.38	0.82
cumul_rets	0.04	1.81	-4.08	-2.94	-1.58	-0.45	0.05	0.54	1.71	2.85	4.03

Table 1: Descriptive statistics of granular closing auction data. This table presents the descriptive statistics over all 10-second intervals over the closing auctions, including distributional information. *ival_rets* represents the return between two interval prices as defined in eq. (1), *mkt_in_imb*(*mkt_out_imb*) represents the imbalance of market order inflows(outflows), *lim_in_imb*(*lim_out_imb*) represents the imbalance of aggressive limit order inflows(outflows) and *mkt_imb*(*lim_imb*) represents the imbalance of market(aggressive limit) orders. All imbalances are computed according to eq. (3). All variables comprise around 2.7 million observations across intervals $l \in \{0, 1, \dots, 60\}$. This number can roughly be derived as the product of 69 stocks, 750 trading days and 61 lags.

trading phase due to immediate execution. However, even the removal of limit orders may have an effect on the price as long as it is an aggressive order, i.e. a buy(sell) limit order above(below) the current uncrossing price *ival_price*.

In the next step, imbalances are computed for both market and limit orders in aggregation. For this purpose, the variable $mkt\_buys\_net_{s,t,l}$ is defined as the difference between $mkt\_buys\_in_{s,t,l}$ and $mkt\_buys\_out_{s,t,l}$, thus representing the net-inflow of market buy orders. The variables $mkt\_sells\_net_{s,t,l}$ as well as $lim\_buys\_net_{s,t,l}$ and $lim\_sells\_net_{s,t,l}$ are defined after the exact same logic. Using these net flows, the following aggregated imbalances are computed:

$$mkt\_imb_{s,t,l} = IMB\left(mkt\_buys\_net_{s,t,l},\ mkt\_sells\_net_{s,t,l}\right)$$
$$lim\_imb_{s,t,l} = IMB\left(lim\_buys\_net_{s,t,l},\ lim\_sells\_net_{s,t,l}\right)$$

Following these definitions of imbalance metrics, the data extraction process ultimately leads to the summary statistics on an interval level presented in table 1. One important observation from these summary statistics is that the imbalance metrics frequently take 0 as outcome. According to the definition in eq. (3), this indicates that there are many intervals without any order flow in either direction. The summary statistics specifically indicate that this lack of order flow is the case in at least 50% of all observations, except for *lim_imb*. For the two measures of market- and limit outflow imbalances, this is even more pronounced, e.g. with at least 80% of intervals without any flows relevant to *mkt_out_imb* and at least 65% for *lim_out_imb*. This implies that investors do not frequently cancel

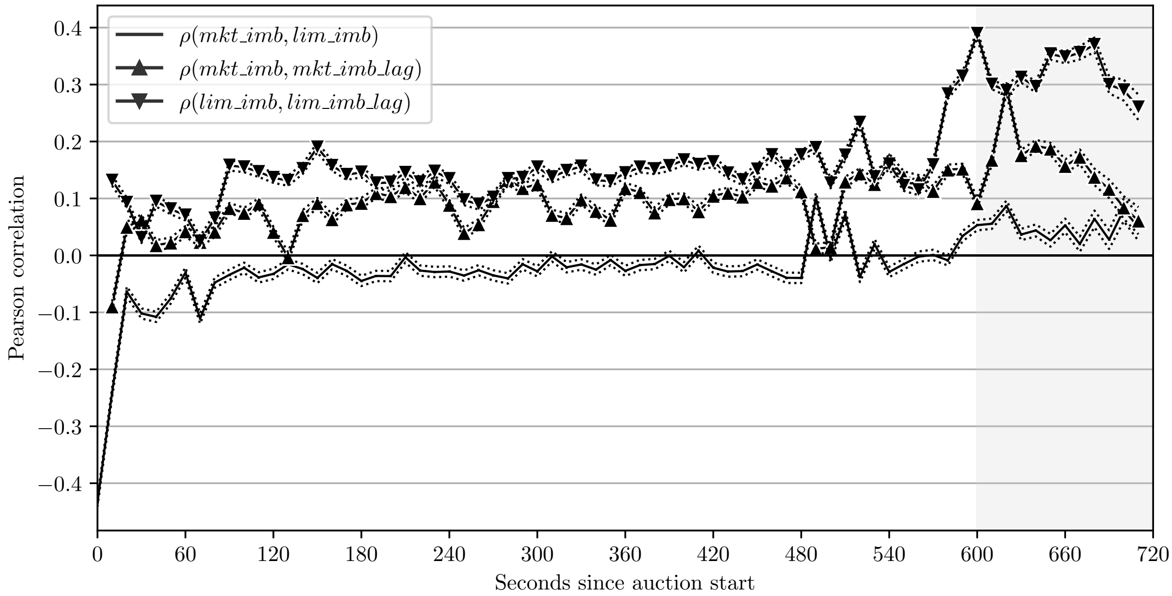


Figure 1: Correlation and autocorrelation of order imbalances. This figure shows correlations and autocorrelations of the variables *mkt_imb* and *lim_imb*. The solid line shows the correlation between both variables. The line with upward-(downward-)facing triangles represents the autocorrelation of market(limit) order imbalances over one lag. All correlations are estimated for each interval l individually. The dotted lines represent confidence intervals at 5% significance.

their orders but rather insert them into the book. Finally, looking at the return presented measures in the last two rows of table 1, the distribution is evidently symmetric for both measures. The statistics for *ival_rets* show that for at least 50% of observed intervals the prices do not change compared to the previous interval and in 95% of the cases the absolute price change is below 82 basis points. Overall, the summary statistics show a rather symmetric distribution without any apparent skew in any of the metrics shown in table 1. All these summary statistics are also computed based on four size-quantiles separately and are presented in table 7 in the appendix. The results are overall comparable, however there are two differences across quantiles. First, large stocks have larger absolute imbalances in both directions, which is driven by larger volumes. Second, there are much fewer intervals without any activity for large stocks.

In order to shed more light onto the interaction between some of the imbalances introduced, the correlations of and autocorrelations between market- and aggressive limit order imbalances are estimated in the next step. All measures are calculated based the Pearson correlation coefficient[12]. The correlations are calculated for each interval l individually

[12] The Pearson correlation between two variables x and y is defined as:

$$\rho(x,y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where each sum iterates through all observations i and both variables have the same number of observations.

in order to allow for varying estimates throughout the auction. The autocorrelations are calculated over one lag. The results of these estimations are presented in fig. 1. To begin with, the correlation between market- and aggressive limit order imbalances are heavily negative at interval $l = 0$, which represents the time span between when the continuous trading phase ends and when before the closing auction fully begins. Quickly afterwards, the correlation approximates 0. Despite statistical significance at a 5% level, the correlation coefficient does barely exceed a magnitude of 0.05 again after 90 seconds. In contrast to this, autocorrelations over one lag are positive for both types of order imbalances throughout the auction. Particularly *lim_imb* reaches autocorrelation of at least 0.2 after the 540 second mark. In summary, market- and aggressive limit order imbalances are significantly autocorrelated, whereas the correlation between both measures is much less prevalent.

4 Price Discovery Results

4.1 Predictability of Returns

Before considering any measures of order flow in the analysis, it is important to understand how prices evolve throughout the closing auction. In order to yield first insights into the price discovery process, the concept of *unbiasedness regressions* are introduced. This methodology has been first introduced by Biais et al. (1999), who analyzed the price discovery during the opening auction on the Paris stocks exchange by disentangling learning and noise. In their statistical approach, the authors analyze the relationship of two returns r_1 and r_2. The former measures the return between t_1 and t_2 and the latter from t_1 to t_3, where $t_1 < t_2 < t_3$. Importantly, the measurements of both returns start at the same time t_1. For the further stages of this analysis, let r_1 be the *inner return* and r_2 the *outer return*.

While Biais et al. (1999) compare apply their framework onto returns starting at the closing auction until various point into the opening auction, the approach taken here differs considerably due to the focus on closing auction price discovery. More specifically, the inner return is represented by the cumulative return as defined in eq. (2) until interval l. In contrast to this, the outer return is the return between the last observed LOB midquote and the closing price. This variable is determined for each day t, stock s and is referred to as $close\_return_{s,d}$. In addition to this, it is desirable to add another dimension to the analysis by taking stock size into account, as variations in price discovery patterns across stock sizes are plausible. For instance large stocks with more investor participation may react to significant price jumps more rapidly as opposed to small stocks, due to higher interest of ETFs who rebalance their portfolios in a correlated manner. In order to account for such effects, let $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_5\}$ be a set of dummy variables, where $Q_1(Q_5)$ represents the bucket containing stock-days with the lowest(highest) trading volumes during the continuous trading phase on the same day prior to the closing auction. The variables in $\mathcal{Q}$ are one if an observation falls into a given quantile and zero otherwise. Importantly, the quantiles are assigned within each stock, such that biases caused by unobserved heterogeneity within stocks are mitigated. These definitions result in the unbiasedness regression equations:

$$close\_return_{s,t} = \sum_{i=1}^{4} \beta_i \left(Q_i \times cumul\_rets_{s,t,l}\right) + \varepsilon_{s,t,l} \qquad \forall \quad l \in \{0, 1, \ldots, 72\} \tag{4}$$

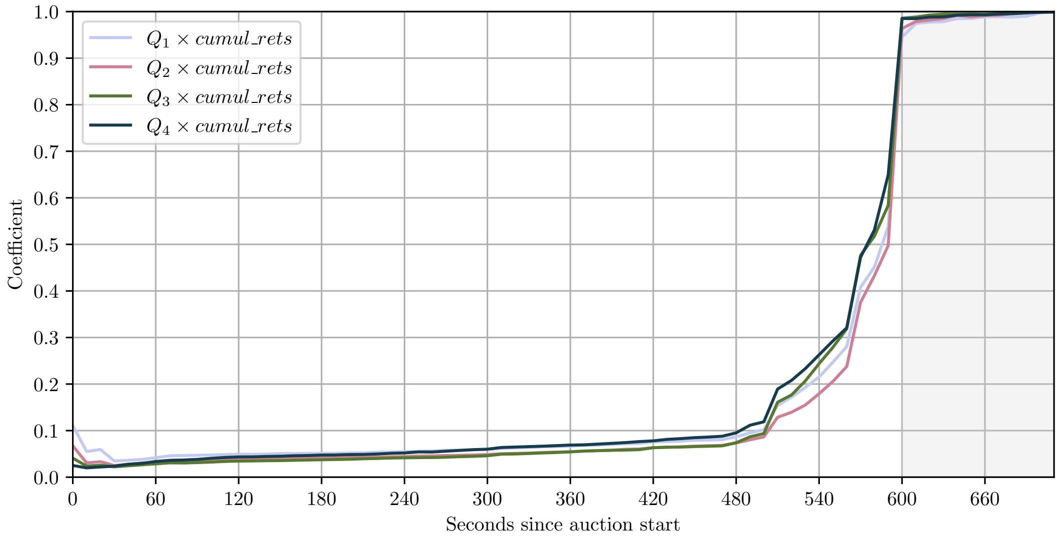
Importantly, this equation is estimated for each interval l resulting in 73 independent regressions. No intercept is included in the equation in order to effects price patterns that occur persistently across auctions. One example of such a pattern could be index funds introducing large amounts of market orders into the auction book as soon as the auction starts. This behavior would result in a consistently positive return towards the start of the auction, which would be captured by the intercept term. However, the main focus of this analysis is the isolation of patterns that are only dependent on the inner return[13].

The results from the unbiasedness regressions are presented in fig. 2. Panel A represents the estimates for all β coefficients throughout the auction. To begin with, there is no statistically significant difference between the coefficients interacted with all size quantiles. The confidence bands, which were derived from a covariance matrix double-clustered by individual stocks and trading days[14], are not shown due to reasons of clarity. For the sake of completeness, the coefficient estimates with confidence intervals by quantile are depicted, in fig. 10 in the appendix. Consequently, any hypothesis of equality between coefficients cannot be rejected at any interval, indicating that there are no effects due to stock size observed. Across all size quantiles, the coefficients are strictly below 0.1 for the first 480 seconds of the closing auction, however still significantly positive. This finding suggests that 90% of the price dislocation since the last midquote before the auction (i.e. *cumul_rets*) is reverted by the end of the auction. This indicates that closing auctions are successful at absorbing large liquidity shock in the beginning of the auction without allowing for large price dislocations. After the 480-second mark, the coefficients convert to 1.0 quickly, indicating that *cumul_rets* becomes an increasingly accurate predictor of

[13] The analysis has additionally been run with an intercept term for robustness purposes. The intercept does not deviate from zero with any statistical significance across all size quantiles. It also has no effect on either estimates of coefficients or R-Squared.

[14] The methodology is based on the presentation in Wooldridge (2012, p. 393).

Panel A: Coefficient estimates from unbiasedness regressions



Panel B: R-Squared estimates from unbiasedness regressions

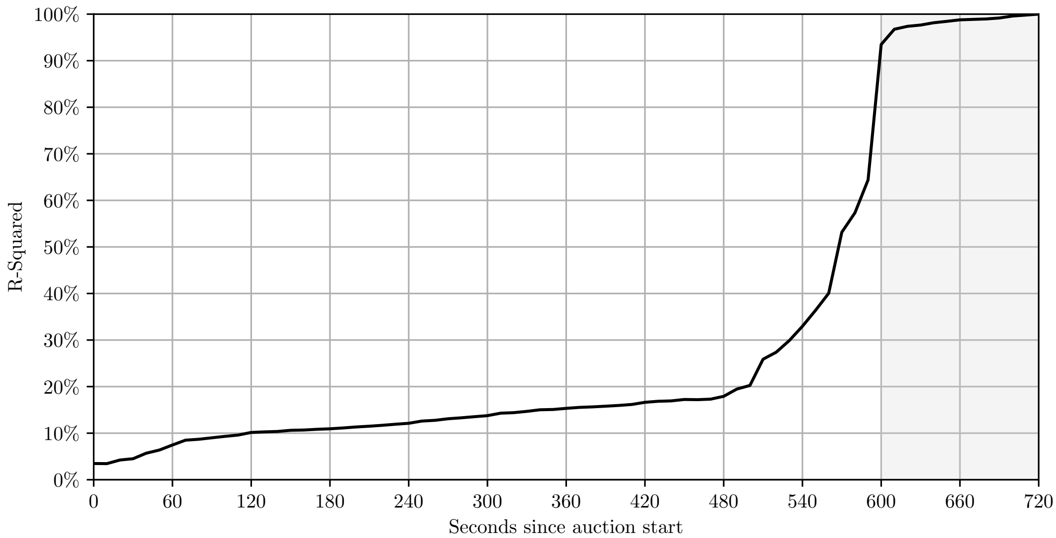


Figure 2: Results from unbiasedness regressions. This table presents the results from the unbiasedness regressions defined in eq. (4), which regress the return between the last midquote before the closing auction ($close\_return$) onto the return between the same starting point and the hypothetical uncrossing price at interval l within the auction ($cumul\_rets$). Panel A shows the coefficients estimated across all four size quantiles in $\mathcal{Q} = \{Q_1, Q_2, Q_3, Q_4\}$, where $Q_1(Q_4)$ represents stocks with small(large) traded volume during the continuous trading phase on the same day prior to the closing auction. Panel B shows the R-Squared of all models throughout the closing auction. The horizontal axis in both charts represents the seconds since auction start. The shaded area on the right side of each plot represents the period during which the auction is ended randomly.

the ultimate closing return. After the 600-second mark, all coefficients are virtually 1, indicating that there is no price discovery happening during the period of random closing. This finding is in line with Cordi et al. (2018) who find that randomized closing auctions are able to prevent price manipulation. Comerton-Forde and Putniņš (2011) and Hillion and Suominen (2004) have shown that some market participants engage in predatory behavior to manipulate the closing price in the last seconds of the trading day.

The R-Squared measures in Panel B indicate a similar pattern as the coefficient estimates. The low scores before the 480-second mark indicate that there is significant noise in prices during the first 8 minutes of the auction. The steeply rising R-Squared thereafter indicates that the cumulative return converges quickly to the final closing price during the last two minutes of the auction. Ultimately, these two plots lead to two main conclusions. First, cumulative returns overshoot the final closing return significantly even at $l = 0$ which is before the auction technically starts. Second, these large price movements are reverted almost fully until the point when the random closing window begins 600 seconds after the start of the auction.

4.2 Drivers of Order Flow

So far, this analysis has discussed to what degree prices are driven over the course of the auction. However, order flow has not been taken into consideration but only the previous price path. This section explores the interaction between the development of the hypothetical closing price as well as granular order flow. This enables statements about whether order flow appears randomly or in reaction to the current state of the LOB.

For this purpose, the first step of this section comprises the estimation of a panel model, estimating the effects of past returns on various order imbalances. The goal of this approach is the prediction of order imbalances based on the realized price path in the auction beforehand. This allows for statements about whether investors adjust their order submission strategies based on this information and actively steer price discovery into a certain direction. Therefore, certain variables defined previously such as $ival\_rets$ and $cumul\_rets$ are extended in order to separate potentially asymmetric effects of positive and negative realizations. Both of these variables are extended with the suffix $pos(neg)$ where the value is retained if positive(negative) and set to zero otherwise. In addition to this, the variable $initial\_ret_{s,t}$ is defined as the return between the last observed midquote before the auction start and the hypothetical uncrossing price at the auction start (i.e. at $l = 0$). This variable contains the information carried by imbalances of all MOC order which are submitted during the continuous trading phase, but are only activated at the beginning of the closing auction. Due to its nature, the variable is computed for each stock s and day t. These newly derived variables lead to the following panel regression equation:

$$imbal_{s,t,l} = \alpha_l + \gamma_t + \beta^{\intercal} \begin{bmatrix} initial\_ret_{s,t} \\ ival\_rets\_lag\_pos_{s,t,l} \\ ival\_rets\_lag\_neg_{s,t,l} \\ cumul\_rets\_lag\_pos_{s,t,l} \\ cumul\_rets\_lag\_neg_{s,t,l} \\ imbal_{s,t,l-1} \end{bmatrix} + \varepsilon_{s,t,l} \tag{5}$$

where $\beta \in \mathbb{R}^6$ represents a vector containing all the estimated coefficients. Additionally, $\alpha_l(\gamma_t)$ represents fixed effects for each interval(trading day). This approach allows for the distinction of potentially asymmetric effects of returns on order imbalances. Moreover, $imbal \in \{mkt\_imb, lim\_imb\}$ is a placeholder variable representing any type of imbalance measure defined previously[15]. For inference purposes, the covariance matrix has been clustered by stocks.

Table 2 presents the results of this panel regression for both market- and limit order imbalances[16]. To begin with, the imbalance of market orders presented in the first three models on the left is barely predictable. In all three models, the R-Squared is virtually zero and none of the independent variables is significant at the 5% level. These results suggest that market orders enter the order book randomly without dependence on the price path throughout the auction. In contrast to this, the three models on aggressive limit order imbalances on the right-hand side of the table show a very different picture. In the first model, all parameter estimates are highly significant and negative. This indicates a certain tendency to trade against realized returns, be it initial or lagged positive as well as negative. However, with an R-Squared below 1% the predictive power of this model remains low. In the second model only cumulative return metrics are considered, once again with highly significantly negative coefficients. In contrast to the previous model, the R-Squared is much higher with an explanatory power of more than 5%. This indicates that aggressive limit order flow has a corrective effect for both positive and negative cumulative returns. The final model containing all the variables results in a similar R-Squared, indicating that the predictive power of the model mostly driven by cumulative returns. This observation is also reinforced by the fact that both lagged interval returns become insignificant in the full model. Finally, a series of Wald tests for the equality of

[15] See table 1 for detailed overview over imbalance metrics.

[16] Table 8 in the appendix shows the same results but with pooled order flow for robustness purposes. More specifically, the order flows as part of the dependent imbalance variables has been calculated over 30 second time horizons forward for each interval l. This methodology mitigates the noise in the order flow and increases the goodness of fit accordingly. All results are roughly the same.

Dep. Variable	*mkt_imb*	*mkt_imb*	*mkt_imb*	*lim_imb*	*lim_imb*	*lim_imb*
initial_ret	-0.0665	-0.0733	-0.1502***	-0.5401***	-0.2191***	-0.1614***
	(-0.8035)	(-0.8446)	(-7.4257)	(-9.9872)	(-4.7146)	(-5.9339)
ival_rets_lag_pos	-0.0560	-0.0714	-0.2705***	-0.4576***	-0.0770	-0.2931***
	(-0.9676)	(-1.0860)	(-4.2623)	(-2.6559)	(-0.6944)	(-4.0867)
ival_rets_lag_neg	-0.0393	-0.0417	-0.2419***	-0.6081***	-0.1131	-0.3180***
	(-0.7593)	(-0.7066)	(-3.9551)	(-3.4805)	(-0.9182)	(-3.6677)
cumul_rets_lag_pos		0.0312	0.0310		-0.8637***	-0.7790***
		(0.9299)	(1.0152)		(-13.021)	(-12.175)
cumul_rets_lag_neg		0.0048	0.0084		-0.8847***	-0.7956***
		(0.1641)	(0.3284)		(-13.552)	(-12.624)
mkt_imb_lag			0.0657***			
			(9.6252)			
lim_imb_lag						0.1120***
						(7.4731)
N	2692536	2846378	2692536	2692536	2846378	2692536
R-Squared	0.0002	0.0002	0.0064	0.0067	0.0463	0.0591
Fixed Effects	Interval	Interval	Interval	Interval	Interval	Interval
Fixed Effects	Day	Day	Day	Day	Day	Day

Table 2: Results of panel regression of order imbalances. This table shows the results of the panel model where order imbalances of market- as well as aggressive limit orders are regressed onto several independent variables. *initial_ret* represents the return between the last pre-close midquote and the first LOB of the auction, *ival_ret_lag* represents the lagged interval return and *cumul_rets_lag* represents the lagged cumulative return from the last pre-close midquote until the current interval. The suffix *pos*(*neg*) represents the original value if it is positive(negative) and zero otherwise. The suffix *lag* the lagged value by one interval. Lagged variables are calculated such that there is no overlap with the order flow variables. The estimation is conducted with day- and interval-fixed effects, whereas t-statistics are presented in parentheses. The covariance matrix is clustered by interval and trading day to account for shocks within identical auctions across stocks. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively.

coefficients[17] for both *cumul_rets_lag_pos* and *cumul_rets_lag_neg* cannot reject the null-hypotheses. Therefore, the effects can be understood as symmetric for both positive and negative cumulative returns.

So far, it could be shown that aggressive limit order imbalances are reactive towards the past price path within the auction as opposed to market order imbalances that appear to be random. While this panel approach can give some reasonably rough implications between the two variables of interest, there is one important assumption inherently embedded in the model. The model treats all the intervals the same and estimates coefficients the same coefficients for the whole course of the auction. In practice, it has been shown that conditions may not be constant throughout the auction (Besson & Fernandez, 2021; Raillon, 2020). In order to account for this, it becomes necessary to allow for coefficients

[17]Wald tests were following the methodology outlined in Baltagi (2011, Section 7.9) and testing for the null-hypothesis $H_0 : \beta_1 - \beta_2 = 0$.

that are adjusting over time. In addition to this, it has been shown that order flow data without aggregation contains a lot of noise and may seem highly unpredictable[18]. Particularly the variance of order imbalances may be driven significantly by traders who need to get their orders filled due to idiosyncratic requirements such as mandatory portfolio rebalancing. Therefore, the direction of order imbalances and returns is more important than the extent thereof.

For these two reasons, the concept of *multinomial Logit regressions* is introduced into the analysis. This methodology has been laid out in Greene (2012, Section 18.2.2) and Baltagi (2011, Section 13.10.2). The essence of multinomial Logit regressions lies in the distinction of multiple discrete outcomes. Therefore, a set of discrete possible outcomes for the direction of $imbal_{s,t,l}$ as $\mathcal{J} = \{-1, 0, 1\}$ is defined. These outcomes cover positive, non-existent and negative imbalance respectively. Under these conditions, the probability of each outcome is defined as

$$P(y_i = j) = \frac{\exp(x_i'\theta_j)}{\sum_{k \in \mathcal{J}} \exp(x_i'\theta_k)} \qquad \forall \qquad j \in \mathcal{J} \tag{6}$$

where θ_j represents a parameter vector for each possible outcome j. $y_i(x_i)$ represents endogenous(exogenous) variables of observation i. Moreover, $\sum_{j \in \mathcal{J}} P(y_i = j) = 1$ must hold as a consequence of this definition.

Due to the non-linear nature of this regression problem, maximum likelihood estimation (MLE) is applied to maximize for the optimal set of parameters $\theta_j \ \forall \ j \in \mathcal{J}$. The log-likelihood function is derived by means of a multinomial distribution as

$$\mathcal{L}(\theta \mid y, x) = \sum_{i=1}^{N} \sum_{j \in \mathcal{J}} d_{ij} \cdot \ln P(y_i = j) \tag{7}$$

where d_{ij} is a dummy variable which is 1 if observation i takes outcome j and zero otherwise. As a consequence of the multinomial distribution only one d_{ij} is one per observation, as the outcomes are mutually exclusive.

In order to make the results more interpretable, the probabilities of two outcomes as defined in eq. (6) can be combined into the following form, following Greene (2012, Section 18.2.2) and Baltagi (2011, Section 13.10.2):

$$\ln\left(\frac{P(y_i = j)}{P(y_i = k)}\right) = x_i'(\theta_j - \theta_k) \qquad \forall \qquad j, k \in \mathcal{J} \text{ and } j \neq k. \tag{8}$$

This form is sometimes referred to as *log-odds* between two outcomes j and k. This

[18] Table 1 showed that even at a 10-second interval horizon many of the intervals are deprived of any order flow.

notation allows for one outcome, say k, to be specified as the base outcome by forcing $\theta_k = 0$, which leads the term $\exp(x_i'\theta_k)$ in eq. (6) to be 1. This leads to the application of this methodology to the log-odd form in eq. (8) but expressed in matrix form

$$\begin{bmatrix} \ln\left(\frac{P(imbal_{s,t,l}>0)}{P(imbal_{s,t,l}=0)}\right) \\ \ln\left(\frac{P(imbal_{s,t,l}<0)}{P(imbal_{s,t,l}=0)}\right) \end{bmatrix} = \begin{bmatrix} \theta_+^\intercal \\ \theta_-^\intercal \end{bmatrix} \begin{bmatrix} initial\_ret \\ ival\_rets\_lag_{s,t,l} \\ cumul\_rets\_lag\_pos_{s,t,l} \\ cumul\_rets\_lag\_neg_{s,t,l} \end{bmatrix} \tag{9}$$

where $\theta_+, \theta_- \in \mathbb{R}^4$ represent vectors containing all the coefficients. As laid out in eq. (5), *imbal* represents a placeholder variable that can be replaced with all types of imbalances defined in table 1. Due to the information in that table, it became apparent that in more than 50% of observations there is a lack of imbalance due to absence of order flow. This outcome of $imbal = 0$ is taken to be the *base outcome*, which will be used to calculate the log-odds as defined in eq. (8). The focus of this analysis will exclusively lie on the effect of lagged cumulative returns whereas lagged interval returns and initial returns represent control variables. Moreover, in the interest of preventing issues of multicollinearity among the three independent return variables the data had to be slightly adjusted at the start of the auction. More specifically, for the intervals $l \leq 1$, the variables *initial_ret* and *ival_rets_lag* are disregarded. With this adjustment, we can calculate the effect of cumulative returns during all auction intervals.

The coefficient estimates for *mkt_imb* as a dependent variable[19] are presented in fig. 3. In all four subplots, there is a spike right in the beginning of the auction that decays over the first 60 seconds. In the top row which depicts cases when the cumulative return is positive this spike is positive, indicating that when there is a large return the probability of either a positive or negative order imbalance is more likely versus the base-case of no imbalance. The same statement can also be made about the bottom row, where the same spike is depicted downward, however, the the implication is identical. Between the 60- and 480-second marks the estimates are rather stable. In this phase, positive(negative) cumulative returns decrease(increase) the probability of negative(positive) market order imbalances with statistical significance. This implies a tendency to trade into the direction of the cumulative return as a quasi-momentum strategy. In addition to this, all four subplots depict a spike at the 490-second mark which increases the probability for market order imbalance against the direction of the cumulative return. In other words this represents order flow that consistently enters the auction at the same time to counteract the cumulative return and therefore exerting a reverting force on the price dislocation.

[19]See fig. 11 for the same analysis with order flow pooled over 30 seconds to mitigate issues of noise.

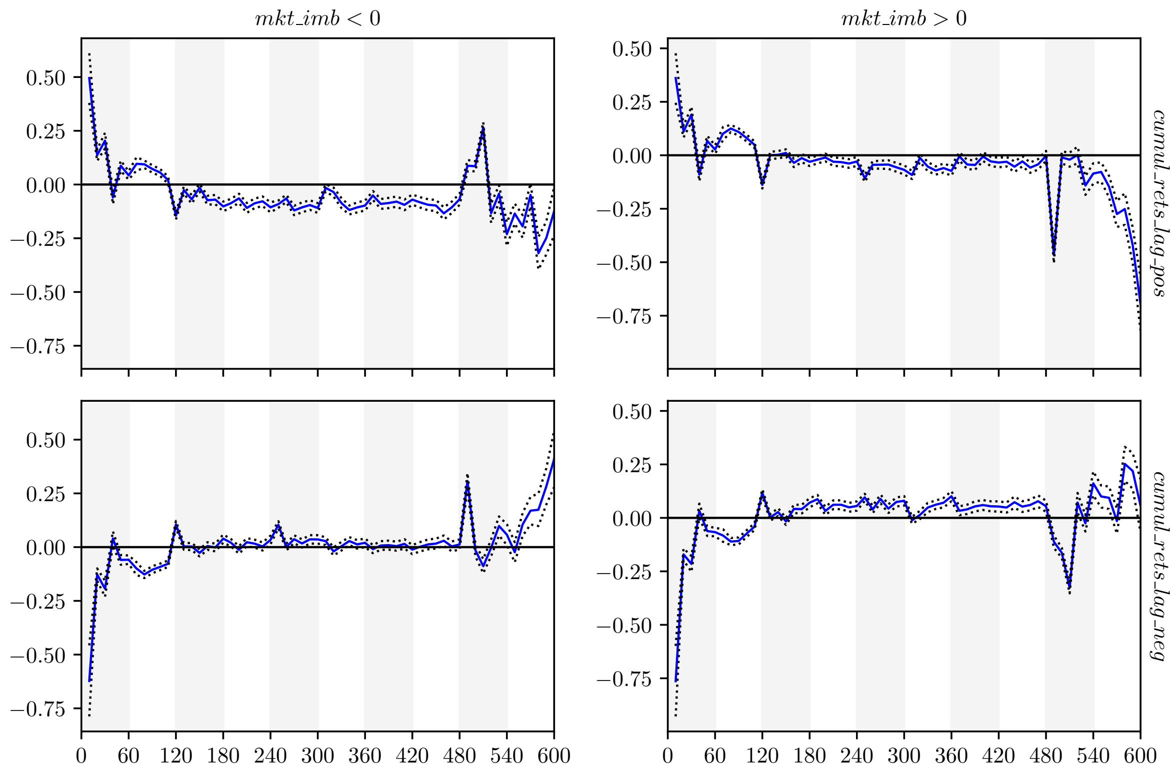


Figure 3: Prediction of market order imbalances by interval. This figure shows the predictive coefficients of the multinomial Logit approach as stated in eq. (9). The dependent variable mkt_imb represents the imbalance of market orders. The left(right) column of subplots depicts marginal effects for the case when mkt_imb is negative(positive). The solid lines represent the θ coefficients of the independent variables $cumul\_rets\_lag\_pos$ in the top row and $cumul\_rets\_lag\_pos$ in the bottom row. One model is estimated for every 10-second interval throughout the auction. The dotted lines around the coefficient estimates represent the 95% confidence interval for the parameter estimates. The horizontal axis represents the number of seconds since the start of the auction. The vertical axis representing the marginal effect of the independent variable is shared across all subplots within each column.

During the final minute of the auction before the random closing phase, positive(negative) cumulative returns until this point decrease the relative probability of observing a positive(negative) market order imbalance. In other words, the probability of observing order imbalances that amplify cumulative returns reduces in both cases.

Analogously to this, fig. 4 depicts the results with lim_imb as the dependent variable in eq. (9)[20]. In contrast to the previous results, aggressive limit order imbalances are much more dependent on cumulative returns. The most striking patterns are observed in the top right and bottom left subplots. More specifically, after large positive(negative) cumulative returns aggressive limit order imbalances are much less(more) likely to be positive. In other words, limit orders are highly unlikely to reinforce cumulative return which becomes increasingly reinforced in the last minute before the random closing period. In contrast to this, during the first 480 seconds of the auction, the probability of aggressive order flow is close to zero. Only in the last two minutes of the auction, investors are actively submitting aggressive limit orders to counteract the cumulative returns and therefore absorbing the

[20]See fig. 12 for the same analysis with order flow pooled over 30 seconds to mitigate issues of noise.

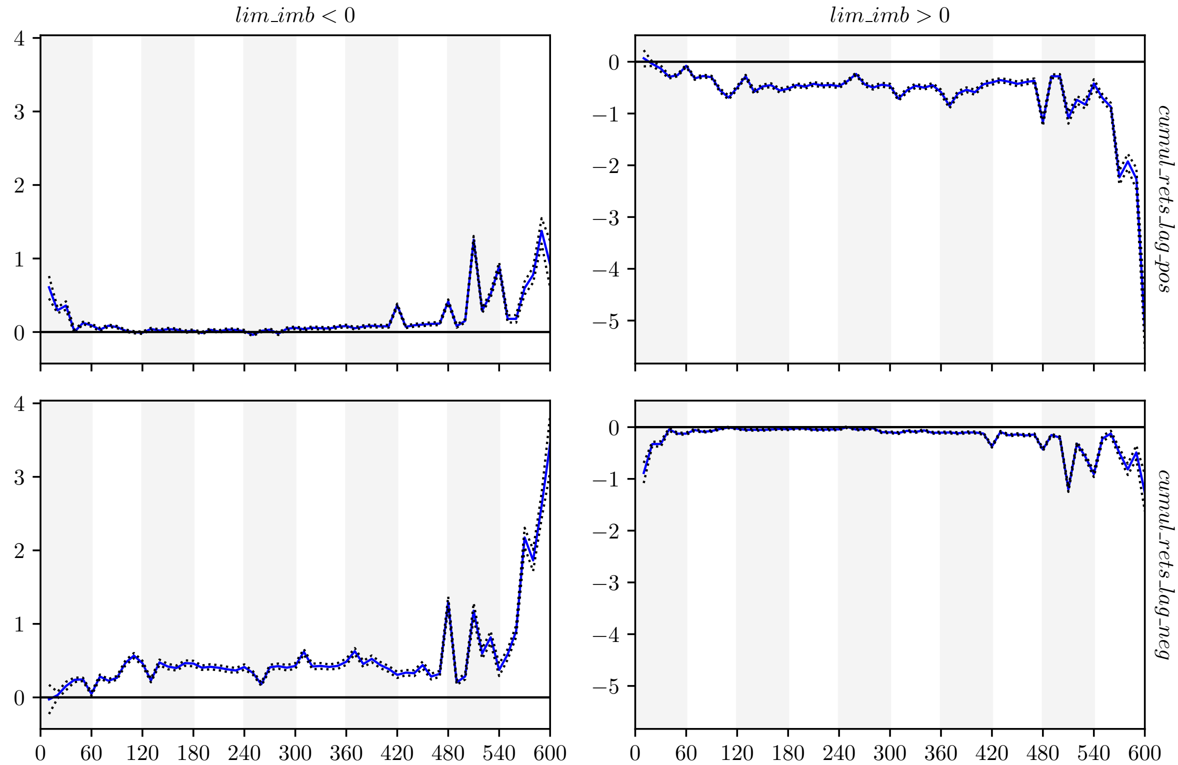


Figure 4: Prediction of aggressive limit order imbalances by interval. This figure shows the predictive coefficients of the multinomial Logit approach as stated in eq. (9). The dependent variable mkt_imb represents the imbalance of market orders. The left(right) column of subplots depicts marginal effects for the case when lim_imb is negative(positive). The solid lines represent the θ coefficients of the independent variables $cumul\_rets\_lag\_pos$ in the top row and $cumul\_rets\_lag\_pos$ in the bottom row. One model is estimated for every 10-second interval throughout the auction. The dotted lines around the coefficient estimates represent the 95% confidence interval for the parameter estimates. The horizontal axis represents the number of seconds since the start of the auction. The vertical axis representing the marginal effect of the independent variable is shared across all subplots within each column.

previous liquidity shock.

So far, the analysis has only considered the coefficients of cumulative returns, but not their predictive power as part of the model. Due to Logit models being estimated by maximizing the log-likelihood instead of minimizing squared errors, an alternative measurement for goodness of fit is required. The pseudo R^2 in question is derived from McFadden (1973, p. 121) and is based on the comparison of log-likelihood outcomes of two models estimated on the identical data:

$$\text{McFadden-}R^2 = 1 - \frac{\mathcal{L}_R}{\mathcal{L}_U}. \tag{10}$$

Thereby, $\mathcal{L}_U$ represents the log-likelihood of the unrestricted model estimated with all the parameters and under the existing model specifications. In contrast to this, $\mathcal{L}_R$ is a restricted version of this model with only one intercept. Consequently, the larger the log-likelihood improvement of the unrestricted model versus the restricted one entails a reduction in the second term becomes therefore a higher the goodness of fit. Importantly,

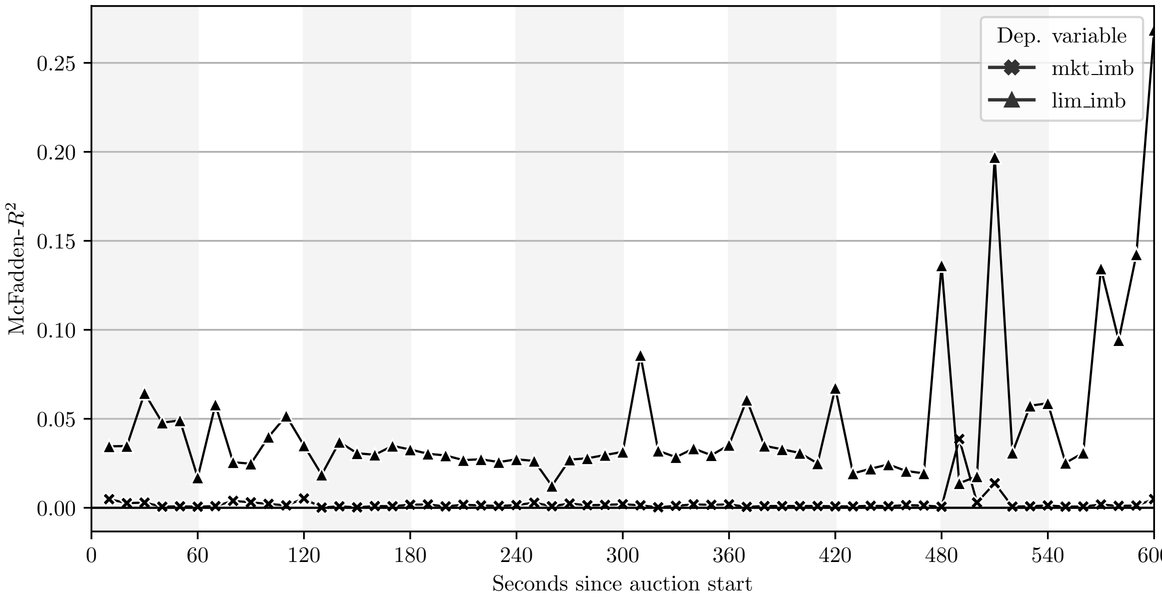


Figure 5: Comparison of McFadden R-Squared for imbalances. This figure shows the estimated McFadden R-Squared based on the multinomial Logit regression presented in eq. (9). The line with the crosses(triangles) represents the results of the model with the market (aggressive limit) order imbalances as dependent variable. The definition of the McFadden R-Squared is presented in eq. (10). The horizontal axis represents the number of seconds since the start of the auction.

this goodness of fit measure cannot be interpreted the same way as a traditional R^2. More specifically, the values for the McFadden-R^2 are generally lower to a considerable extent in comparison to the traditional R^2. According to McFadden (1979, p. 307), values of 0.2 to 0.4 for this measure already represent a very good fit.

Figure 5 presents the comparative results of the McFadden-R^2 between the models regressing market- and aggressive limit order imbalances. Similarly to the results in table 2, market order imbalances are not predicted more accurately using cumulative returns as opposed to the restricted intercept-only model. The only exception to this can be found in the two small spikes at 490 and 510 seconds after auction start. The estimates from the models of aggressive limit order imbalances are more dynamic throughout the auction. The relative improvement of the unrestricted model is consistently higher, indicating that cumulative returns in fact have predictive power over these imbalances. Particularly important is the observation that the McFadden-R^2 increases steeply during the last two minutes of the auction[21].

Until this point, order imbalances of market- as well as aggressive limit order flows have been addressed in detail. However, it is important to understand that both of these measures are computed based on the net inflows of orders during a given interval. One factor that is neglected in the academic literature is the distinction between inflows and

[21]This observation becomes even clearer in fig. 13 in the appendix, where the order flows as a basis for the imbalance measure are pooled over 30 seconds in order to smoothen out the irregular nature of order flow.

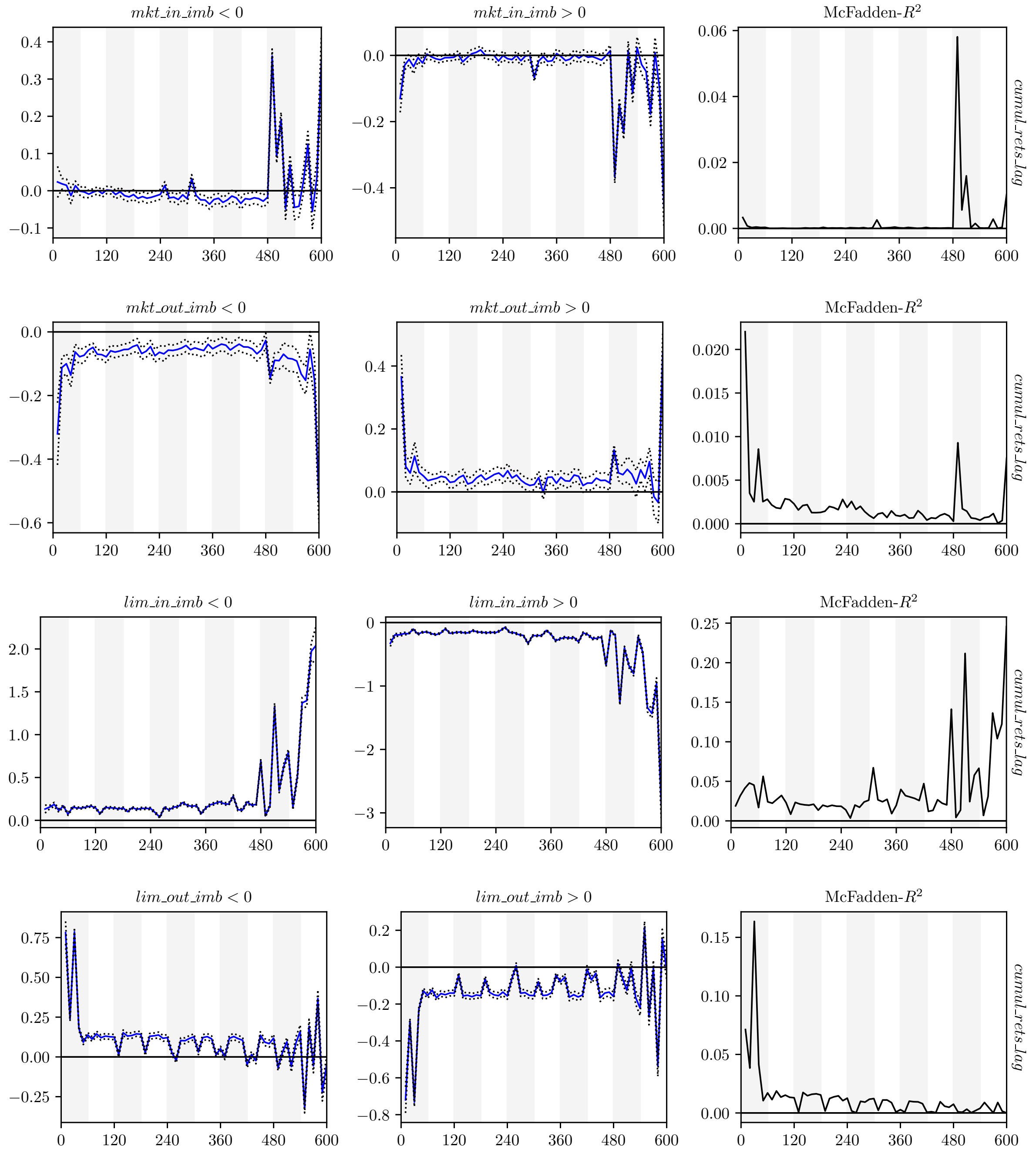


Figure 6: Prediction of inflow- and outflow imbalances. This figure shows the results of the multinomial Logit regression approach defined in eq. (9). The first row(second) row shows the results for the models with mkt_in_imb(mkt_out_imb) as dependent variable representing imbalances of market order inflows(outflows). The first third(fourth) row shows the results for the models with lim_in_imb(lim_out_imb) as dependent variable representing imbalances of aggressive limit order inflows(outflows). The left(center) column shows coefficient estimates in cases where the imbalance is negative(positive). The third column depicts the estimate for the McFadden R-Squared. The horizontal axis is shared among all subplots in this figure and represents the number of seconds since auction start.

outflows of orders. Particularly during the closing auction outflows of market orders are as important as inflows for the determination of the optimal uncrossing.

Consequently, fig. 6 depicts a new set of multinomial Logit models where the dependent variables are set to *mkt_in_imb*, *mkt_out_imb*, *lim_in_imb* and *lim_out_imb* respectively[22]. The estimations were otherwise conducted as previously. In the top row of the figure, the effect on market order inflows is presented. As in fig. 3, the predictive power of the model is very low with the exception of a few spikes between 490–510 seconds after auction start. These spikes represent an inflow of market orders to compensate for cumulative returns up this point. These spikes are barely visible in the second row, which depicts the imbalance of market order outflows. In these models one can observe that market order outflows are on average reinforcing the previous cumulative returns. For instance, a positive cumulative return entails an increasing probability that more buy orders are withdrawn as opposed to sell orders in terms of market orders. The opposite statement can be made when the outflow of market sells exceeds market buys. In both cases, these observations have a counteracting effect towards cumulative returns. However, the McFadden-R^2 barely exceeds 0.1% throughout the auction indicating a severe lack of predictive power.

In the lower half of fig. 6, the analogue set of results for aggressive limit flows are presented. The third row shows how inflows of aggressive limit orders counteract cumulative returns throughout the auction. However, the coefficients drastically increase in magnitude towards the end of the auction, amplifying the effect. Similarly, the goodness of fit increases in the last two minutes of the auction. This implies that investors submitting aggressive limit orders tend to trade against cumulative returns but only get serious about correcting mispricings towards the end. Outflows of aggressive limit orders represented by *lim_out_imb*, on the other hand show a different picture. During the first 60 seconds of the auction the withdrawals of aggressive limit orders have a momentum-reinforcing effect with considerable goodness of fit. In other words, positive cumulative returns at the beginning of the auction render negative *lim_out_imb* relatively more likely, which exerts even more upward pressure on the uncrossing price. Nonetheless, this reinforcement effect of limit order outflows is only prevalent during the first minute. Thereafter, the McFadden-R^2 consistently remains below 1%.

In summary, these findings show that the insertion and cancellation of market orders is driven by different factors than for aggressive limit orders, which are relevant for price discovery. Overall, the findings suggest that market orders arrive more randomly in the order book. While there are some consistent patterns, the goodness of fit measures are

[22] See fig. 14 in the appendix for the same analysis based on order flow pooled forward by 30 seconds in order to mitigate issues of noisy order flow data.

so low that these effects are barely relevant. Overall, market imbalances are more likely to occur after large price dislocations in the first 60 seconds of the auction. Afterwards, there persistent inflow of market orders which has a reverting effect on cumulative returns. In aggregation, imbalances of aggressive limit orders always have a reverting effect on cumulative returns. This effect is only reinforced during the last 2 minutes of the auction by incoming orders. Importantly, however, these results show that order flow is not independent, particularly aggressive limit order flow. Therefore, the assumption of IID distributed order flow in auction models such as Smith et al. (2003) and Derksen et al. (2020) cannot be supported after acknowledgment of these results. Overall, however, this result supports the notion in the model of Parlour (1998) that order flow is highly endogenous and that investors pay close attention to market conditions.

4.3 Drivers of Price Discovery

The focus of this analysis has so far been exclusively on the predictability of order flow imbalances based on the pre-existing price path. In this section, this logic is reversed. More specifically, the focus is shifted towards providing answers on the effect of order flow on price discovery.

For this purpose, a future returns variable is introduced which represents the logarithmic return from interval l until the end of the auction and is defined as

$$future\_ret_{s,t,l} = close\_return_{s,t} - cumul\_ret_{s,t,l} \tag{11}$$

where $close\_return_{s,t}$ is the logarithmic return between the last observed pre-close midquote and closing price and $cumul\_ret_{s,t,l}$ is the cumulative logarithmic return from the same starting point until interval l as defined in eq. (2).

In order to quantify the lagwise effects of order flow onto the future return, the same multinomial Logit approach reapplied with slight adjustments. In contrast to the previous section, the dependent variable is now defined as the future return. The reason for this is the motivation to assess the predictive power of order imbalances on the remaining price path from each interval l until the end of the auction. It is important to keep previous findings from this analysis in mind when specifying the model going forward. On the one hand, section 4 showed that cumulative returns are almost reverted for all $l < 48$, which coincides with the 8-minute mark in the auction. On the other hand, section 4.2 demonstrated that aggressive limit order imbalances are highly predictable based on cumulative returns, which is not applicable for market order imbalances. Consequently, controlling for cumulative returns in the regression alongside the order imbalance metrics enables the isolation of true contributions by variable. Importantly, the regression includes *cumul_rets_lag* as a control variable in order to avoid any timely overlap with other independent variables. As in the previous estimation, the base outcome is defined to be $future\_ret = 0$ such that the current interval price is already equal to the closing price and there is no price discovery left. This is not true for the remaining two possible outcomes. With this boundary condition, we arrive at the following two equations for log-odds:

$$\begin{bmatrix} \ln\left(\frac{P(future\_ret_{s,t,l}>0)}{P(future\_ret_{s,t,l}=0)}\right) \\ \ln\left(\frac{P(future\_ret_{s,t,l}<0)}{P(future\_ret_{s,t,l}=0)}\right) \end{bmatrix} = \begin{bmatrix} \theta_+^\intercal \\ \theta_-^\intercal \end{bmatrix} \begin{bmatrix} mkt\_imb_{s,t,l} \\ lim\_imb_{s,t,l} \\ cumul\_rets\_lag_{s,t,l} \end{bmatrix} \tag{12}$$

where $\theta_+, \theta_- \in \mathbb{R}^3$ both represent one vector each containing one unique set of coefficients for each discrete case respectively. The model has been estimated for each interval l individually, in order to avoid autocorrelation in observations. It is also important to note, that there is no timely overlap between dependent and independent variables in order to prevent issues of endogeneity, as returns and order flows are related closely. In addition to this, the models have been estimated for all four variations of order flow pooling in order to smoothen the results.

Figure 7 presents all estimates for all for θ coefficients from the model in eq. (12) graphically[23]. The first row in the figure shows the coefficients for market order imbalances. In the previous analysis, market order inflows as well as -outflows were shown to arise rather randomly without significant dependence on cumulative returns. In this analysis, market orders go against future returns consistently in the first half of the auction, despite controlling for lagged cumulative returns. This shows that market order imbalances are not predicted by lagged cumulative return and therefore contribute to the prediction of future returns. The fact that these imbalances are in the opposite direction compared to future returns indicates that the price dislocation caused by such market order imbalances are subsequently corrected until the end of the auction. Therefore, informed investors could use market orders in the beginning of the auction in order to hide their information since any dislocation will ultimately be reverted until the end of the auction.

The second row of the figure presents the coefficient estimates of aggressive limit order imbalances. In section 4.2 it has been demonstrated that these imbalances are highly dependent on cumulative returns in a counteracting manner. Particularly during the last two minutes of the auction, this effect becomes stronger. When cumulative returns are included in the regression alongside aggressive limit order imbalances, however, the im-

[23] More precise results are presented in fig. 15 in the appendix. These results show the results based on order imbalances calculated by pooling order flow over 30 seconds respectively in order to reduce noise as a consequence of the choice of small interval size.

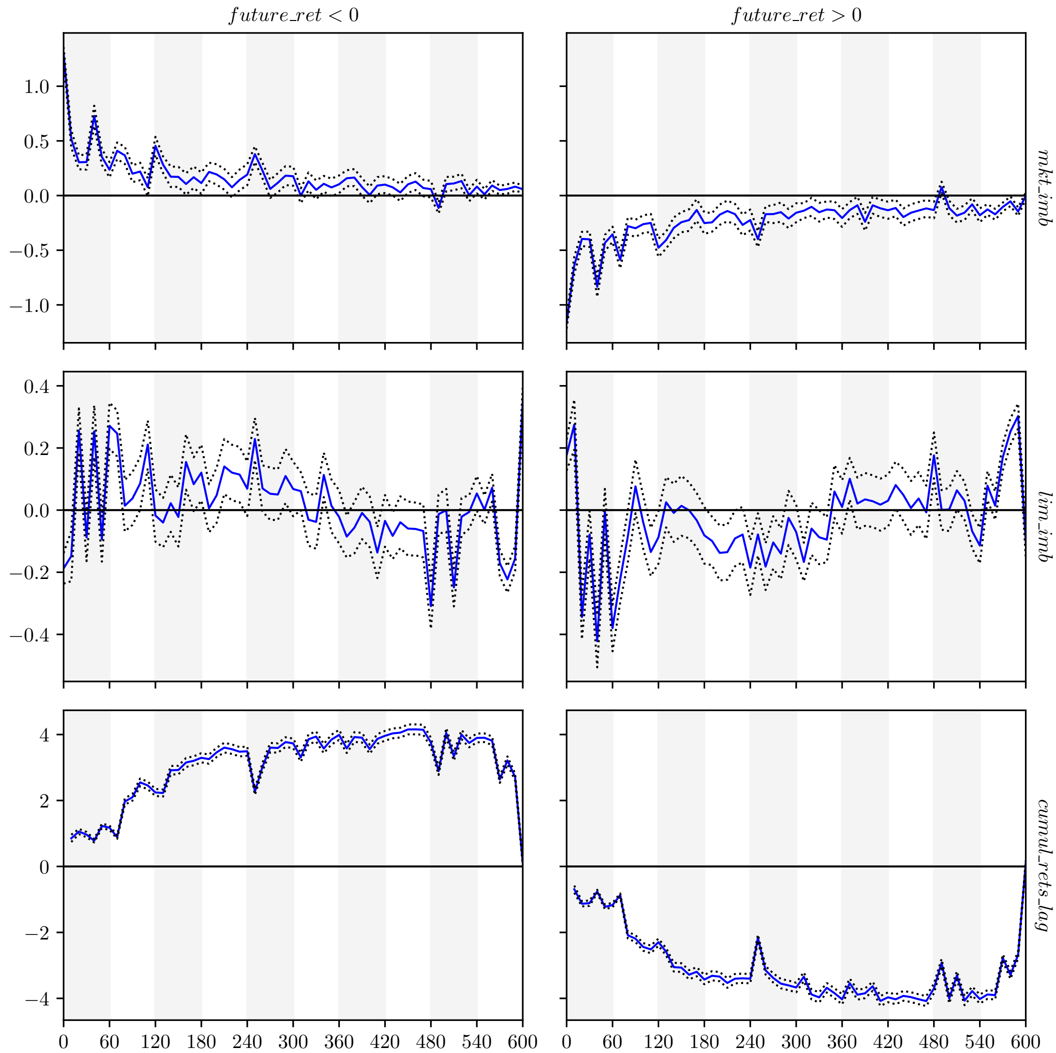


Figure 7: Prediction of future return using order imbalances by interval. This figure shows the predictive coefficients of the multinomial Logit approach as stated in eq. (12). The dependent variable $future\_ret$ represents the return between interval l and the end of the auction. The left(right) column of subplots depicts marginal effects for the case when $future\_ret$ is negative(positive). The independent variables shown are mkt_imb in the top row, lim_imb in the middle row and $cumul\_rets\_lag$ in the bottom row. One model is estimated for every 10-second interval throughout the auction. The dotted lines around the coefficient estimates represent the 95% confidence interval for the parameter estimates. The horizontal axis represents the number of seconds since the start of the auction. The vertical axis representing the marginal effect of the independent variable is shared across all subplots within each row.

balance loses its predictive power. There is consistent and statistically significant pattern visible. In addition to this, the coefficients are much lower compared to market order imbalances in the same model, despite the comparable distribution of the two variables[24]. This indicates that their contribution to the model has already been accounted for through the cumulative returns variable. Consequently, these imbalances do not bring any new information into the auction but are merely manifested in reaction to the previous price path. Finally, the last row of fig. 7 presents the coefficient estimates with respect to lagged cumulative returns. It is clearly visible that throughout the auction, future returns are

[24]See table 1 for more information.

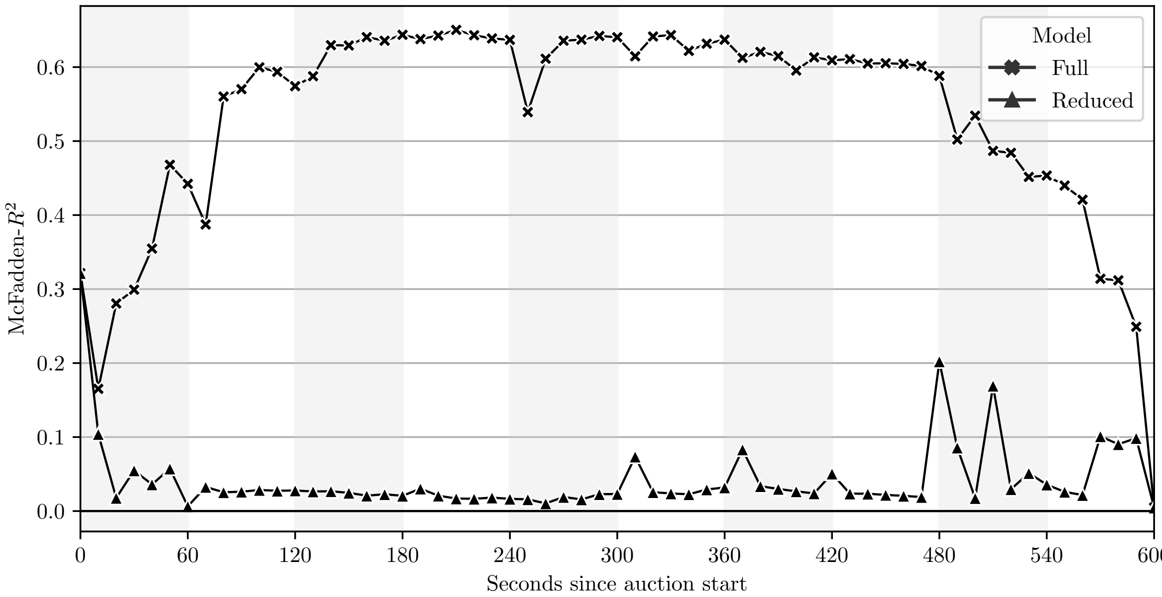


Figure 8: Comparison of McFadden R-Squared for future returns. This figure shows the estimated McFadden R-Squared based on two multinomial Logit regression models, both based on eq. (12). First, the full model includes all the variables. Second, the reduced variable does not include lagged cumulative returns and only relies on imbalance measures. The definition of the McFadden R-Squared is presented in eq. (10). The horizontal axis represents the number of seconds since the start of the auction.

going into the opposite direction of cumulative returns with high significance. This finding is in line with section 4.1, which showed that cumulative returns are mostly reverted towards the end of the auction.

The results predicting future returns so far have only included estimates of coefficients paired with their respective confidence bands. To complete the analysis, fig. 8 presents the estimates for the McFadden-R^2 throughout the auction for two types of models: Full and reduced[25]. The former is the exact model as defined in eq. (12). The latter consists of the same model but does not include *cumul_rets_lag* as an independent variable. The results show that there is a significant difference in predictive power between the two models, starting after 20 seconds into the auction. This finding implies that the lagged cumulative returns are strong predictors for future returns, much more so than order imbalances alone for the majority of intervals during the auction. However, right in the beginning of the auction at $l = 0$, when market order imbalances can crucially predict future returns through reversion.

[25]For robustness purposes, fig. 16 shows the same model with order imbalances pooled over 30 seconds backwards in order to mitigate noise in order flow data.

5 Overnight Price Efficiency

So far, this analysis was mostly focused on the order flow dynamics across multiple intervals throughout the closing auction. It has been shown that price dislocations caused by market order imbalances in the beginning of the auction are subsequently reversed towards the end of the auction. In other words, the auction is successful at absorbing large liquidity shocks in that investors are reverting these shocks using aggressive limit orders. This correction occurs in the last two minutes of the auction. Overall, this pattern of initial price dislocation and subsequent correction is very prevalent in all closing auctions. Nonetheless, these initial liquidity shocks may be of varying extent and may or may not include new information. If large cumulative returns caused substantial demand and supply imbalances are always reverted towards the end of the auction, some information may be lost in the process. Therefore, it is plausible that the overcompensation of price dislocations during the closing auction may have an impact overnight price efficiency subsequent to the auction.

Price efficiency is a broad term and can be used in various contexts. The most common definition used in academic literature originated in Fama (1970, p. 391). According to his definition, prices follow the weak version of market efficiency if they are martingale (i.e., follow a random walk). This implies that the price of any given security is the best prediction of every future price of the same security. For this purpose, the methodology of *unbiasedness regressions* by Biais et al. (1999) which has been introduced in section 4.1 is re-applied in a different context here. Previously, inner(outer) returns were assigned to cumulative(closing) returns. Given the definition of weak market efficiency provided by Fama (1970), the inner return must be the best prediction of the outer return, given both measurements start at the same point in time. Consequently, at the point in time when the inner measurement period of the inner return is terminated, the current price is the best prediction until the termination of the outer return as well. Is this condition violated, are prices not martingale by not following a random walk, presenting a violation of the weak efficient market hypothesis.

In the analysis presented in this section, the unbiasedness regressions are applied in order to determine under which conditions closing returns offer a contribution to price discovery. Therefore, the regressors are chosen to be inner returns, i.e. closing returns. In order to facilitate robust results, a set of various outer is chosen with all starting from the last pre-close midquote, representing the outer return. Those are the return from pre-close midquote to the opening price (*ret_to_open*), the return from pre-close midquote to pre-close midquote the next day (*ret_to_preclose*), the return from pre-close midquote to the closing price the next day (*ret_to_close*) and the return from pre-close midquote to the opening price two days later ($ret\_to\_open\_2$).

Besides the approach of unbiasedness regressions, there are also other approaches to measure price efficiency. For instance, variance ratio tests introduced by Lo and MacKinlay (1988) compare stock variances over various time horizons. In case of efficient prices, the variances must be additive due to serial independence of returns. Another measure of price efficiency is cross-autocorrelation between asset returns and lagged market returns, quantifying the time required to incorporate information into assets. This measure is used by Saffi and Sigurdsson (2011) who show that constraints on short-selling harm price efficiency and Bris et al. (2007) who find that negative information is incorporated into prices more quickly in countries where short-selling is allowed.

In order to quantify potential liquidity shocks ind the auction adequately, a new variable capturing the price reversion throughout an auction is defined. This variable is extracted by analyzing the price path of a stock throughout the auction and seeking out the peak or trough of the price curve and subsequently calculating the reversion to the final price. In algorithmic terms, we first locate the farthest dislocation of any hypothetical uncrossing price throughout the auction:

$$cumul\_max_{s,t} = cumul\_rets_{s,t,l^*} \qquad \text{where } l^* = \arg\max_l \left(\left| cumul\_rets_{s,t,l} \right| \right). \tag{13}$$

In the next step, the absolute dislocation from the maximum cumulative return is defined as the reversion variable:

$$reversion_{s,t} = close\_return_{s,t} - cumul\_max_{s,t} \tag{14}$$

due to the properties of logarithmic returns. Following this definition, a monotonously increasing or decreasing price path results in a *reversion* parameter 0. In contrast, it will yield highly positive results if there is an initial upward- or downward spike which is subsequently compensated towards the end of the auction. Since this section operates on aggregated data as opposed to before, table 3 represents updated summary statistics. There are two important changes with respect to the new data. First, we are now primarily looking at variables derived from the price path of assets. Second, the number of observations has decreased by a factor of approximately 61 due to the aggregation process.

Keeping in mind that the aim of this section is the analysis of the interaction between price discovery path and price efficiency, we need to define a framework allowing statistical tests. One common way to achieve this is by grouping the data into discrete quantiles, that are subsequently testable against one another. For this purpose, let $\mathcal{K} = \{K_1, K_2, \ldots, K_5\}$ be a set of dummy variables, where $K_1(K_5)$ represents the bucket containing stock-days

	N	μ	σ	5%	20%	40%	50%	60%	80%	95%
log_close_volume	40109	10.47	2.27	6.66	8.64	10.09	10.51	10.94	12.46	14.19
$initial\_ret$	35740	0.09	0.64	-0.57	-0.16	0.04	0.09	0.14	0.30	0.84
$close\_return$	40056	0.05	0.26	-0.32	-0.11	0.00	0.05	0.09	0.20	0.41
ret_to_open	38285	0.07	1.68	-1.41	-0.44	-0.06	0.09	0.23	0.60	1.61
$ret\_to\_preclose$	39929	-0.02	2.90	-3.12	-1.07	-0.24	0.06	0.34	1.13	2.92
ret_to_close	39929	0.02	2.91	-3.12	-1.02	-0.19	0.11	0.40	1.19	2.97
$cumul\_max$	35442	0.29	2.46	-3.67	-2.08	-0.52	0.58	1.06	2.61	4.39
$cumul\_max\_secs$	40111	93.37	117.66	0	10	40	40	60	150	360
$reversion$	40106	-0.24	2.35	-4.23	-2.41	-0.90	-0.44	0.44	1.98	3.59

Table 3: Descriptive statistics of aggregated closing auction data. This table presents the descriptive statistics over the data aggregated for each stock-day, including distributional information. log_close_volume represents the logarithmic volume traded at close, $initial\_ret$ represents the return between the last midquote of the continuous phase and the first LOB of the closing auction, $close\_return$ represents the return from the last midquote of the continuous phase to the closing price. ret_to_open, $ret\_to\_preclose$ and ret_to_close represent the return between the last midquote of the continuous phase and the following day's open, preclose midquote and closing price respectively. The $cumul\_max$ and $reversion$ parameters are defined in eqs. (13) and (14). max_cumul_secs represents the number of seconds after the beginning of the closing auction, at which $cumul\_max$ is reached.

with the lowest(highest) values. The variables in $\mathcal{K}$ are one if an observation falls into a given quantile and zero otherwise. Importantly, the quantiles are assigned within each stock, such that biases caused by unobserved heterogeneity within stocks are mitigated. For this analysis, the variables $reversion$ and $cumul\_max$ are used as a basis for the dissemination into quantiles. Consequently, stocks are distributed evenly across quantiles. Upon interaction with the unbiasedness regressions by Biais et al. (1999) we get the following regression equation:

$$outer\_return_{s,t} = \sum_{j=1}^{5} \beta_j (K_j \times close\_return_{s,t}) + \varepsilon_{s,t}. \tag{15}$$

where $outer\_return \in \{ret\_to\_open, ret\_to\_preclose, ret\_to\_close, ret\_to\_open\_2\}$ is a placeholder variable representing all previously introduced outer returns. In order to determine efficiency of prices, we can assess the individual coefficients. Prices are efficient if $close\_return$ is an unbiased predictor for $outer_return$ such that β equals one, constituting the null-hypothesis $H_0 : \beta = 1$. Consequently, the rejection of the null implies price inefficiency. In addition to this, two more Wald tests are conducted following Baltagi (2011, Section 7.9) in order to to assess the overall efficiency closing prices across quantiles. The first test is conducted to assess whether all the coefficients are jointly equal to one, which is referred to as the test for *joint efficiency*. Thereby, the test statistic follows a χ^2 distribution with five degrees of freedom. The second test assesses whether the sum of all coefficients equals the number of quantiles, which implies efficiency on average

Quantiles based on reversion variable				
Dep. Variable	*ret_to_open*	*ret_to_preclose*	*ret_to_close*	*ret_to_open_2*
N	38285	39929	39929	38214
R-squared	0.0142	0.0043	0.0061	0.0061
Quantiles	*reversion*	*reversion*	*reversion*	*reversion*
$K_1 \times close\_return$	0.7767*	0.8441	1.0252	1.0059
	(0.0573)	(0.3433)	(0.8793)	(0.9771)
$K_2 \times close\_return$	0.9109	0.6594**	0.8154	0.9621
	(0.2590)	(0.0343)	(0.2622)	(0.8297)
$K_3 \times close\_return$	0.8876	0.8205	0.9668	1.2960**
	(0.1209)	(0.2198)	(0.8305)	(0.0329)
$K_4 \times close\_return$	0.5691***	0.5824***	0.6637**	0.6052*
	(0.0000)	(0.0062)	(0.0308)	(0.0679)
$K_5 \times close\_return$	0.7277*	0.5895**	0.6845*	0.5237**
	(0.0561)	(0.0255)	(0.0889)	(0.0223)
Joint Efficiency	34.9437***	19.3526***	8.8856	13.1533**
	(0.0000)	(0.0017)	(0.1137)	(0.0220)
Total Efficiency	124.8849***	32.4652***	63.4125***	58.2112***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 4: Results of unbiasedness regressions quantiled by reversions. This table shows the results of the unbiasedness regressions defined in eq. (15). The dependent variables are represented by returns from preclose to open, preclose to preclose the next day and preclose to close the next day. The independent variable is the 10-minute closing return. Quantile-dummy variables $\{K_1, K_2, \ldots, K_5\}$ are assigned within each stock based on the $reversion$ variable defined in eq. (14), where $K_1(K_5)$ is one for observations with low(high) reversions and 0 otherwise. The covariance matrix is adjusted to be robust to heteroskedasticity. In contrast to other tables, the null-hypothesis for point estimates of each parameter is $H_0 : \beta = 1$. The lower part of the table contains two Wald tests. The first one test for joint efficiency, which tests for all coefficients being equal to one. The second one tests for the sum of coefficients to be equal to five. The quoted values for these two tests represent the test statistic, which follows a χ^2 distribution under the null-hypothesis. All values in parentheses are p-values. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively for all tests.

across quantiles. This is referred to as test for *total efficiency*. Finally, the covariance matrix for all regressions is corrected using the White (1980) procedure and thus robust to heteroskedastic residuals.

The results of the unbiasedness regressions interacted with quantiles based on the $reversion$ variable are presented in table 4. The upper part of the table shows the point estimates for coefficients of each quantiled parameter and tests them against the efficiency hypothesis $H_0 : \beta_j = 1$. From the point estimates we can see that the predictive power is very low with R-squared estimates of 1.4% at most[26]. All but one coefficient are estimated to be below one. The most significant point estimates are consistently clustered around the interactions with K_4 and K_5 across all four return horizons. In table 5, where the

[26]Hence, trading on any of the findings from these regressions would unlikely be a profitable strategy owing to the large variance of errors.

Quantiles based on maximum cumulative returns				
Dep. Variable	*ret_to_open*	*ret_to_preclose*	*ret_to_close*	*ret_to_open_2*
N	38285	39929	39929	38214
R-squared	0.0114	0.0037	0.0053	0.0047
Quantiles	*cumul_max*	*cumul_max*	*cumul_max*	*cumul_max*
$K_1 \times close\_return$	0.7140**	0.4722***	0.5386***	0.4601***
	(0.0159)	(0.0020)	(0.0078)	(0.0037)
$K_2 \times close\_return$	0.4109***	0.8159	0.9057	0.6936
	(0.0000)	(0.3753)	(0.6508)	(0.2535)
$K_3 \times close\_return$	0.7117***	0.9694	1.1976	1.2584
	(0.0002)	(0.8648)	(0.2759)	(0.1874)
$K_4 \times close\_return$	0.8972	0.9250	1.1130	1.0282
	(0.2139)	(0.5925)	(0.4283)	(0.8652)
$K_5 \times close\_return$	0.8712	0.7028*	0.8592	1.0375
	(0.2171)	(0.0839)	(0.4195)	(0.8488)
Joint Efficiency	39.5913***	13.5973**	9.7529*	11.5172**
	(0.0000)	(0.0184)	(0.0825)	(0.0420)
Total Efficiency	84.6176***	43.8890***	79.3454***	53.4488***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 5: Results of unbiasedness regressions quantiled by maximum cumulative returns. This table shows the results of the unbiasedness regressions defined in eq. (15). The dependent variables are represented by returns from preclose to open, preclose to preclose the next day and preclose to close the next day. The independent variable is the 10-minute closing return. Quantile-dummy variables $\{K_1, K_2, \ldots, K_5\}$ are assigned within each stock based on the $cumul\_max$ variable defined in eq. (13), where $K_1(K_5)$ is one for observations with low(high) maximum cumulative returns and 0 otherwise. The covariance matrix is adjusted to be robust to heteroskedasticity. In contrast to other tables, the null-hypothesis for point estimates of each parameter is $H_0 : \beta = 1$. The lower part of the table contains two Wald tests. The first one test for joint efficiency, which tests for all coefficients being equal to one. The second one tests for the sum of coefficients to be equal to five. The quoted values for these two tests represent the test statistic, which follows a χ^2 distribution under the null-hypothesis. All values in parentheses are p-values. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively for all tests.

quantiles are based on the variable $cumul\_max$, we can observe a different pattern. While the overwhelming majority of coefficients are likewise below zero, the significant deviations occur in K_1 across all return horizons.

These observations imply that certain price paths result in more efficient closing prices than others. More specifically, we see that reversions with highly positive values and $cumul\_max$ with low values individually underperform. This indicates that closing auctions with an initial negative spike, followed by a positive correction result in inefficient closing prices. More specifically, in both of both quantile dimensions the coefficients in these cases are significantly lower than one, which implies a correction with respect to each of the four analyzed return horizons.

In terms of price efficiency across coefficients, both tables 4 and 5 show similar results. The joint tests for efficiency, which tests for all coefficients being jointly equal to one, is rejected for all models except for the *ret_to_close* horizon. The tests for total efficiency, testing the sum of coefficients being equal to five, is highly rejected across all horizons. These findings indicate that closing prices are rather inefficient and are followed by price reversions over all horizons. One explanation as to why *ret_to_close* is not rejected in the test for joint efficiency could be explained by a positive correlation of closing returns caused by liquidity shocks by large financial institutions that are forced to rebalance their portfolios in sync. As shown in Raillon (2020), the market design of closing auction is very conducive for this type of order flow.

Importantly, this is not the first time in the literature that the inefficiency of closing prices has been pointed out. Bogousslavsky and Muravyev (2020) also point out that closing returns on both NYSE and NASDAQ revert overnight and therefore are detrimental to price discovery. They also find that closing prices lead to violated put-call parity due to liquidity-caused price dislocations. In addition to this, Hu and Murphy (2020) are concerned about the efficiency of a dealer-based system and show that call auctions comparatively perform better in terms of price efficiency.

6 Conclusion

The recent shift in trading volume to closing auctions has lead to an increasing reliance on closing prices, for instance in terms of benchmarking fund performances. For this reason, the closing price is the most important transaction price throughout the entire day, as it matches the larges amount of trading volume. Nonetheless, academic literature has not yet properly addressed the price discovery process of this price throughout the auction. Particularly large passive investors who follow a mandate and rebalance their portfolio raise questions about how much new information is actually revealed during these auctions and how much is just uninformed demand for liquidity. Consequently, this research is particularly relevant for regulators who have been looking into the introduction of additional rules concerning the fragmentation of such auctions or for regulators who are contemplating about their pricing strategies in order to incentivize certain types of order flow.

This paper presents an analysis which is based on granular order-level data of 69 Swiss equities, directly obtained from SIX Securities & Exchanges (SIX) over the course of three full calendar years 2018–2020. For this analysis, each closing auction is split into 10-second intervals in order to allow for time-varying effects throughout the auction. Over such a comprehensive data sample various conclusions are drawn. First, most of the price dislocations from the last observed price in the continuous auction are reversed until the end of the auction. This indicates that such auctions are working well in absorbing liquidity shocks and counteracting potential price impacts. Only during the last two minutes of the auction, the hypothetical uncrossing price approximates the ultimate clearing price of the closing auction.

The second main finding is concerned with the behavior of order flow during the auction and the distinction between imbalances of market- and aggressive limit orders. Specifically, investors submitting market orders are found to be non-reactive to the past price path, whereas aggressive limit orders are submitted aimed to reverse such price dislocations. This propensity monotonously increases over the course of the auction. Importantly, however, aggressive limit orders do not contribute new information to the closing auction but are merely reactive in nature. This is not true for market order imbalances which negatively predict the remaining return in the auction. This finding suggests that overall, imbalances in market orders lead to dislocations in price that are subsequently compensated by aggressive limit orders.

The third finding concerns overnight price efficiency as this pattern of price dislocation and subsequent reversion may accidentally hide informative order flow. The analysis finds that closing returns are on average inefficient in the context of overnight returns and tend to revert to a certain extent. This effect is particularly pronounced following auctions where prices spike downward and are subsequently reverting upward. This indicates that auctions are absorbing informative sell order flow.

The results in this analysis clearly suggest that price discovery is an important aspect of the closing auction. Particularly due of the importance of these auctions for the absorption of large liquidity shocks. This function becomes clear when looking at the granular order flow, when initial price shocks are compensated over the course of the auction potentially at the expense of informative order flow. Based on these results, further research should be conducted with respect to the design of incentives for both liquidity providers and -takers in such auctions. This may be important research in order to optimize the balance between the absorption of liquidity shocks and the incorporation of new information into closing prices.

References

Aitken, M., Comerton-Forde, C., & Frino, A. (2005). Closing call auctions and liquidity. *Accounting & Finance*, *45*(4), 501–518.

Anand, A., Chakravarty, S., & Martell, T. (2005). Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders. *Journal of Financial Markets*, *8*(3), 288–308.

Baltagi, B. H. (2011). *Econometrics* (5th ed.). Springer.

Barclay, M. J., & Hendershott, T. (2003). Price discovery and trading after hours. *The Review of Financial Studies*, *16*(4), 1041–1073.

Barclay, M. J., & Hendershott, T. (2008). A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance*, *15*(5), 839–849.

Barclay, M. J., Hendershott, T., & Jones, C. M. (2008). Order consolidation, price efficiency, and extreme liquidity shocks. *Journal of Financial and Quantitative Analysis*, 93–121.

Bellia, M., Pelizzon, L., Subrahmanyam, M. G., & Yuferova, D. (2017). *Coming Early to the Party*, Safe Working Papers.

Besson, P., & Fernandez, R. (2021). *Better trading at the close thanks to market impact models* (Report). Euronext Quantitative Research.

Biais, B., Hillion, P., & Spatt, C. (1999). Price discovery and learning during the preopening period in the paris bourse. *Journal of Political Economy*, *107*(6), 1218–1248.

Bloomfield, R., O'hara, M., & Saar, G. (2005). The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, *75*(1), 165–199.

Bogousslavsky, V., & Muravyev, D. (2020). *Should we use closing prices? institutional price pressure at the close* [Working Paper].

Bris, A., Goetzmann, W. N., & Zhu, N. (2007). Efficiency and the bear: Short sales and markets around the world. *The Journal of Finance*, *62*(3), 1029–1079.

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, *69*(5), 2045–2084.

Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial economics*, *65*(1), 111–130.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, *87*(2), 249–268.

Chordia, T., & Subrahmanyam, A. (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, *72*(3), 485–518.

Comerton-Forde, C., Lau, S. T., & McInish, T. (2007). Opening and closing behavior following the introduction of call auctions in singapore. *Pacific-Basin Finance Journal*, *15*(1), 18–35.

Comerton-Forde, C., & Putniņš, T. J. (2011). Measuring closing price manipulation. *Journal of Financial Intermediation*, *20*(2), 135–158.

Cordi, N., Félez-Viñas, E., Foley, S., & Putninš, T. (2018). *Closing time: The effects of closing mechanism design on market quality* (tech. rep.). Working paper.

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

Derksen, M., Kleijn, B., & De Vilder, R. (2020). Clearing price distributions in call auctions. *Quantitative Finance*, *20*(9), 1475–1493.

Ellul, A., Shin, H. S., & Tonks, I. (2005). Opening and closing the market: Evidence from the london stock exchange. *Journal of Financial and Quantitative Analysis*, *40*(4), 779–801.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *The review of financial studies*, *18*(4), 1171–1217.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, *14*(1), 71–100.

Goettler, R. L., Parlour, C. A., & Rajan, U. (2005). Equilibrium in a dynamic limit order market. *The Journal of Finance*, *60*(5), 2149–2192.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson.

Griffiths, M. D., Smith, B. F., Turnbull, D. A. S., & White, R. W. (2000). The costs and determinants of order aggressiveness. *Journal of Financial Economics*, *56*(1), 65–88.

Hagströmer, B., & Nordén, L. (2014). Closing call auctions at the index futures market. *Journal of Futures Markets*, *34*(4), 299–319.

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

Hillion, P., & Suominen, M. (2004). The manipulation of closing prices. *Journal of Financial Markets*, *7*(4), 351–375.

Holden, C. W., & Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance*, *69*(4), 1747–1785.

Hu, E., & Murphy, D. (2020). *Vestigial tails? floor brokers at the close in modern electronic markets* [Working Paper].

Ibikunle, G. (2015). Opening and closing price efficiency: Do financial markets need the call auction? *Journal of International Financial Markets, Institutions and Money*, *34*, 208–227.

Inci, A. C., & Ozenbas, D. (2017). Intraday volatility and the implementation of a closing call auction at borsa istanbul. *Emerging Markets Review*, *33*, 79–89.

Kalay, A., & Wohl, A. (2009). Detecting liquidity traders. *Journal of Financial and Quantitative Analysis*, *44*(1), 29–54.

Kandel, E., Rindi, B., & Bosetti, L. (2012). The effect of a closing call auction on market quality and trading strategies. *Journal of Financial Intermediation*, *21*(1), 23–49.

Kaniel, R., & Liu, H. (2006). So what orders do informed traders use? *The Journal of Business*, *79*(4), 1867–1913.

Kuo, W., & Li, Y.-C. (2011). Trading mechanisms and market quality: Call markets versus continuous auction markets. *International review of Finance*, *11*(4), 417–444.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The review of financial studies*, *1*(1), 41–66.

Madhavan, A. (1992). Trading mechanisms in securities markets. *The Journal of Finance*, *47*(2), 607–641.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press New York.

McFadden, D. (1979). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments. In D. Hensher & P. Stopher (Eds.), *Bahvioural travel modelling* (pp. 278–318). Groom Helm.

McInish, T. H., & Wood, R. A. (1992). An analysis of intraday patterns in bid/ask spreads for nyse stocks. *the Journal of Finance*, *47*(2), 753–764.

Pagano, M., Peng, L., & Schwartz, R. A. (2013). A call auction's impact on price formation and order routing: Evidence from the nasdaq stock market. *Journal of Financial Markets*, *16*(2), 331–361.

Pagano, M., & Schwartz, R. A. (2003). A closing call's impact on market quality at euronext paris. *Journal of Financial Economics*, *68*(3), 439–484.

Parlour, C. A. (1998). Price dynamics in limit order markets. *The Review of Financial Studies*, *11*(4), 789–816.

Pascual, R., & Veredas, D. (2009). What pieces of limit order book information matter in explaining order choice by patient and impatient traders? *Quantitative Finance*, *9*(5), 527–545.

Raillon, F. (2020). The growing importance of the closing auction in share trading volumes. *Journal of Securities Operations & Custody*, *12*(2), 135–152.

Rindi, B. (2008). Informed traders as liquidity providers: Anonymity, liquidity and price formation. *Review of Finance*, *12*(3), 497–532.

Roşu, I. (2009). A dynamic model of the limit order book. *The Review of Financial Studies*, *22*(11), 4601–4641.

Saffi, P. A., & Sigurdsson, K. (2011). Price efficiency and short selling. *The Review of Financial Studies*, *24*(3), 821–852.

Smith, E., Farmer, J. D., Gillemot, L., & Krishnamurthy, S. (2003). Statistical theory of the continuous double auction. *Quantitative finance*, *3*, 481–514.

Theissen, E., & Westheide, C. (2020). Call of duty: Designated market maker participation in call auctions. *Journal of Financial Markets*, *49*, 1–15.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–838.

Wooldridge, J. M. (2012). *Econometric analysis of cross section and panel data* (7th ed.). MIT press.

Appendix

6.1 Figures

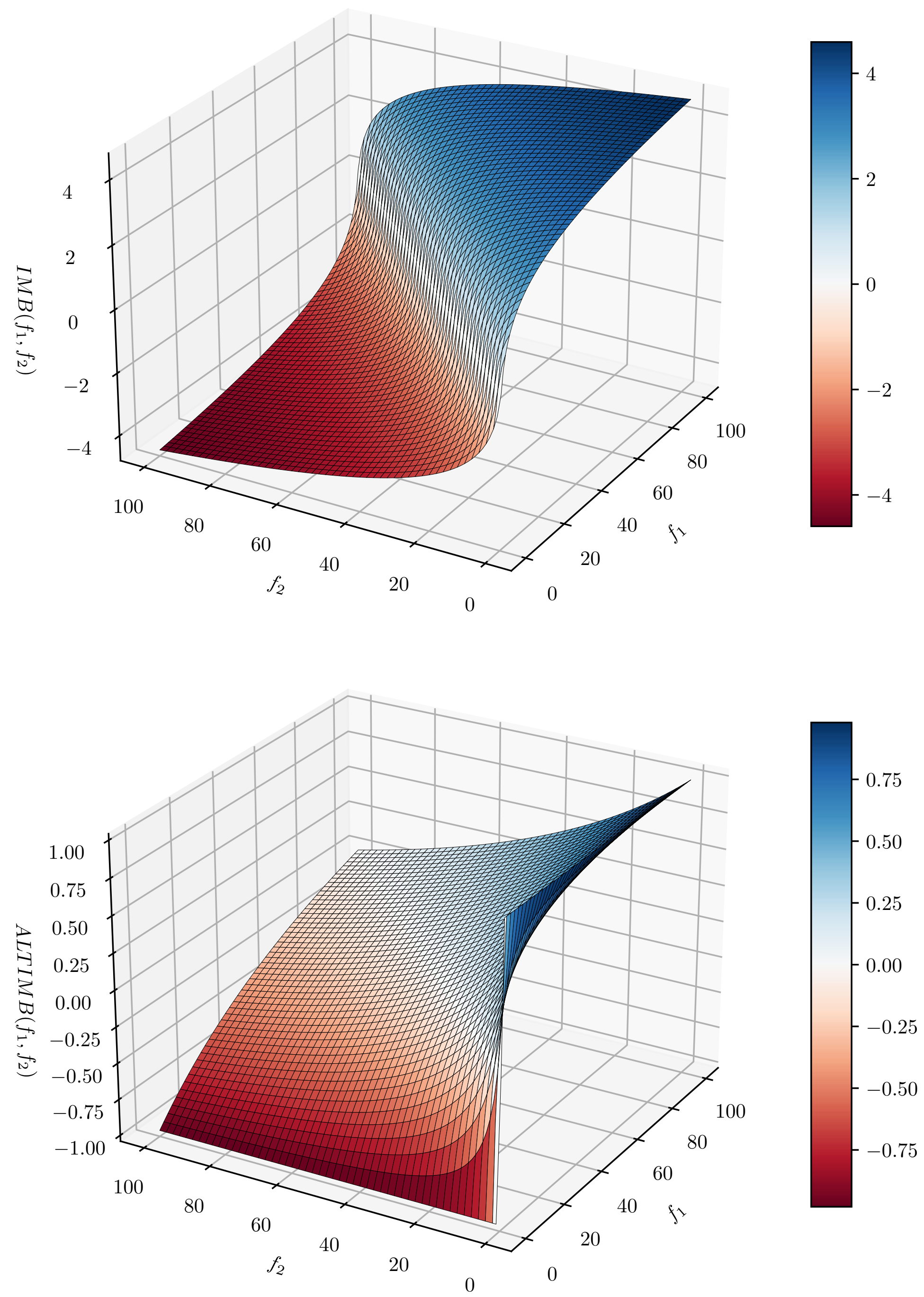


Figure 9: Visualization of imbalance measures. These two plots visualize the two imbalance measures defined in section 3. The top plot presents the imbalance measure used in this analysis, as defined in eq. (3). The bottom plot presents the traditional imbalance measure frequently used in the academic literature, which is presented defined in footnote 10. For both plots, f_1 and f_2 represent non-negative order flows. The vertical axis represents the resulting imbalance. For visualization purposes, the plane is color-mapped to the vertical axis to underline the outcome.

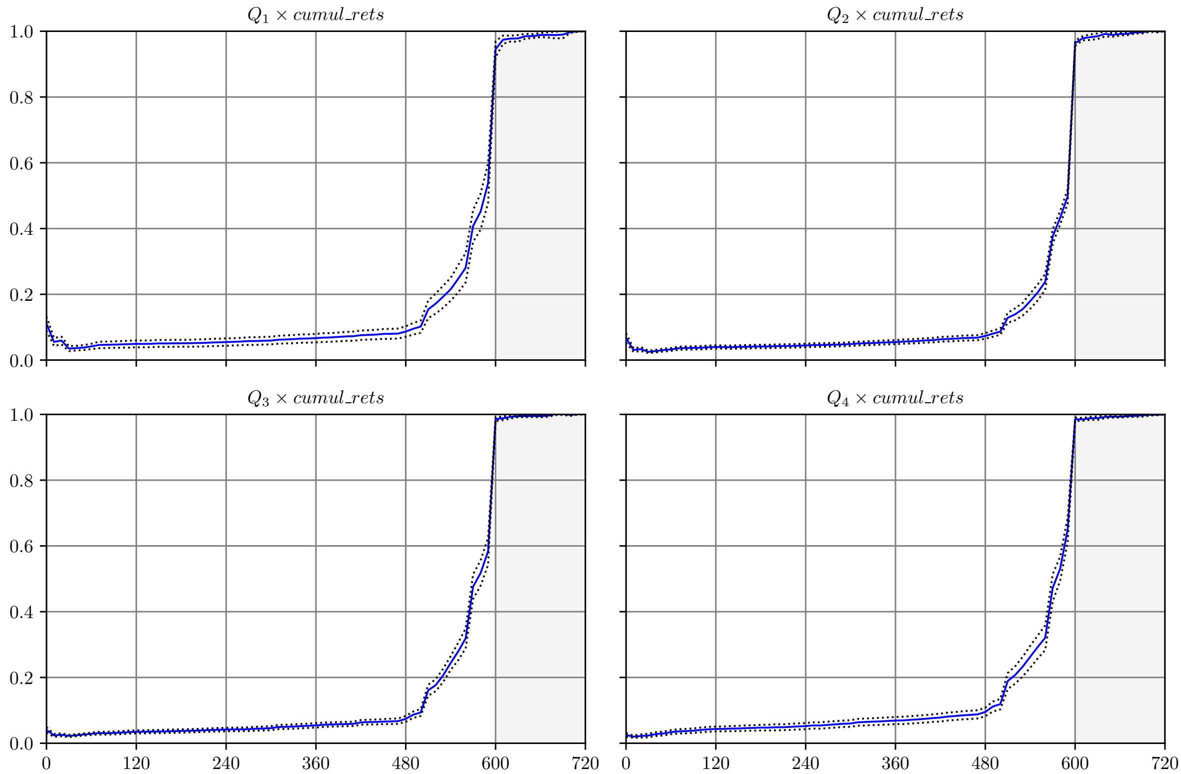


Figure 10: Results from unbiasedness regressions. This table presents the results from the unbiasedness regressions defined in eq. (15), which regress the return between the last midquote before the closing auction ($close\_return$) onto the return between the same starting point and the hypothetical uncrossing price at interval l within the auction ($cumul\_rets$). Each panel depicts the coefficient estimates interacted with a dummy variable for each size quantile in $\mathcal{Q} = \{Q_1, Q_2, Q_3, Q_4\}$, where $Q_1(Q_4)$ represents stocks with small(large) traded volume during the continuous trading phase on the same day prior to the closing auction. The shaded area on the right side of each plot represents the period during which the auction is ended randomly.

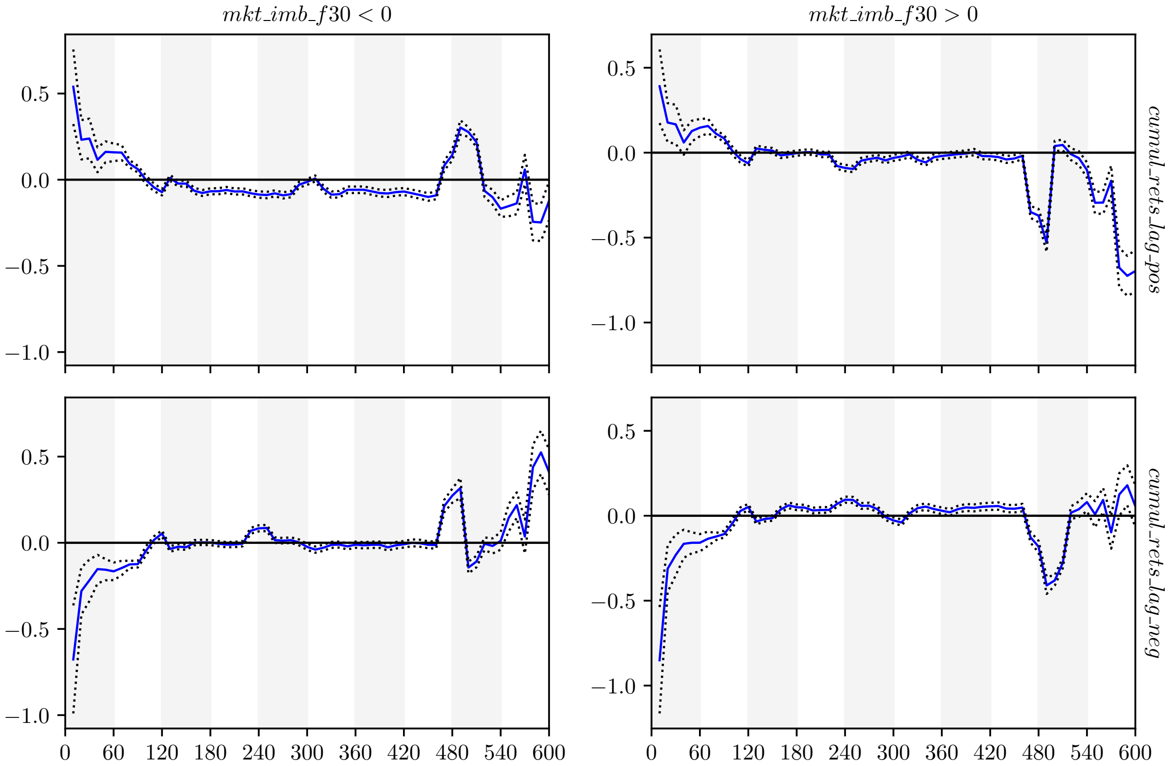


Figure 11: Prediction of pooled market order imbalances by interval. This figure shows the predictive coefficients of the multinomial Logit approach as stated in eq. (12). The dependent variable mkt_imb represents the imbalance of market orders which has been pooled forward over 30 seconds. The left(right) column of subplots depicts marginal effects for the case when mkt_imb is negative(positive). The solid lines represent the θ coefficients of the independent variables $cumul\_rets\_lag\_pos$ in the top row and $cumul\_rets\_lag\_pos$ in the bottom row. One model is estimated for every 10-second interval throughout the auction. The dotted lines around the coefficient estimates represent the 95% confidence interval for the parameter estimates. The horizontal axis represents the number of seconds since the start of the auction. The vertical axis representing the marginal effect of the independent variable is shared across all subplots within each column.

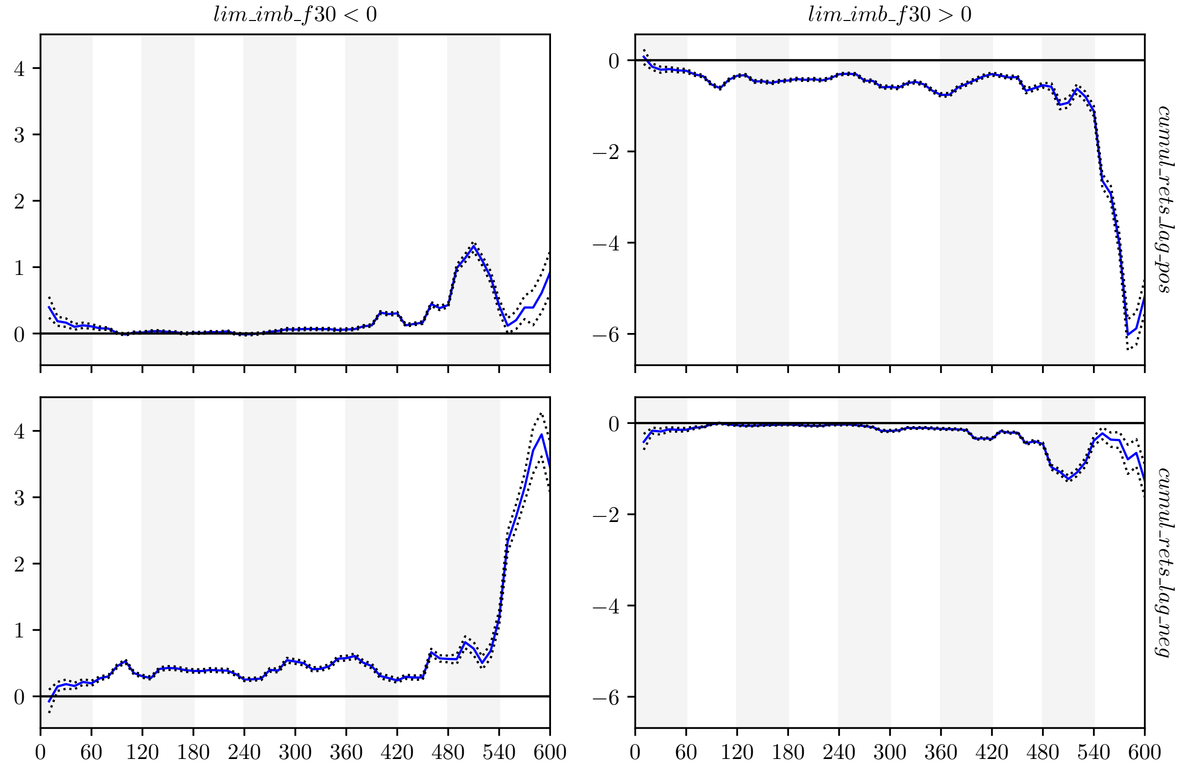


Figure 12: Prediction of pooled aggressive limit order imbalances by interval. This figure shows the predictive coefficients of the multinomial Logit approach as stated in eq. (12). The dependent variable lim_imb represents the imbalance of aggressive limit orders which has been pooled forward over 30 seconds. The left(right) column of subplots depicts marginal effects for the case when mkt_imb is negative(positive). The solid lines represent the θ coefficients of the independent variables $cumul\_rets\_lag\_pos$ in the top row and $cumul\_rets\_lag\_pos$ in the bottom row. One model is estimated for every 10-second interval throughout the auction. The dotted lines around the coefficient estimates represent the 95% confidence interval for the parameter estimates. The horizontal axis represents the number of seconds since the start of the auction. The vertical axis representing the marginal effect of the independent variable is shared across all subplots within each column.

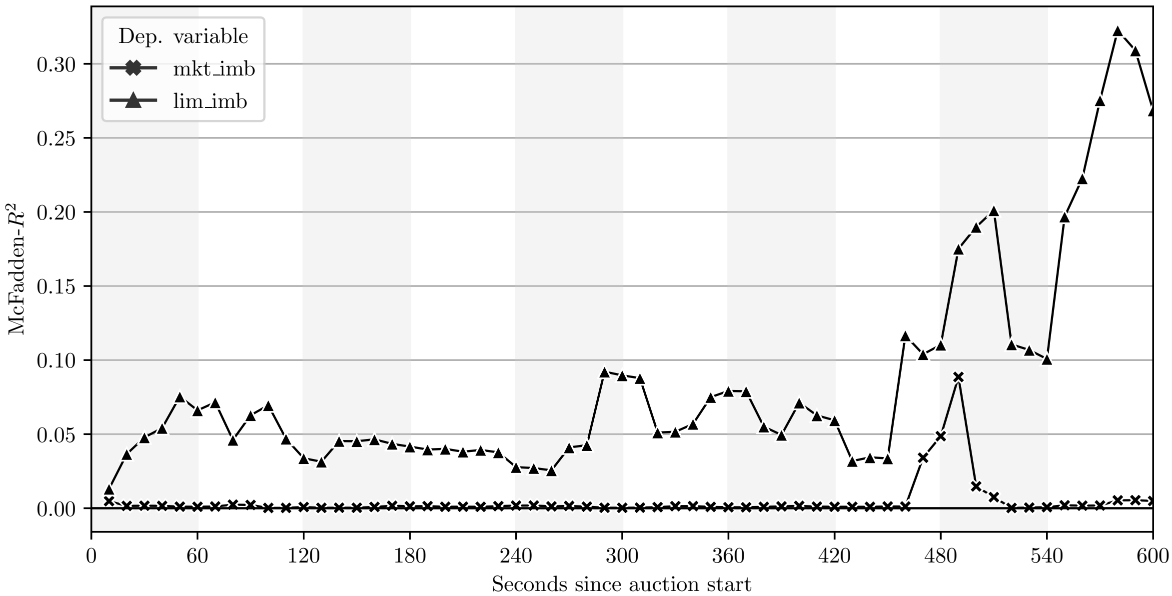


Figure 13: Comparison of McFadden R-Squared for pooled imbalances. This figure shows the estimated McFadden R-Squared based on the multinomial Logit regression presented in eq. (9). The line with the crosses(triangles) represents the results of the model with the market (aggressive limit) order imbalances pooled over 30 seconds forward as dependent variable. The definition of the McFadden R-Squared is presented in eq. (10). The horizontal axis represents the number of seconds since the start of the auction.

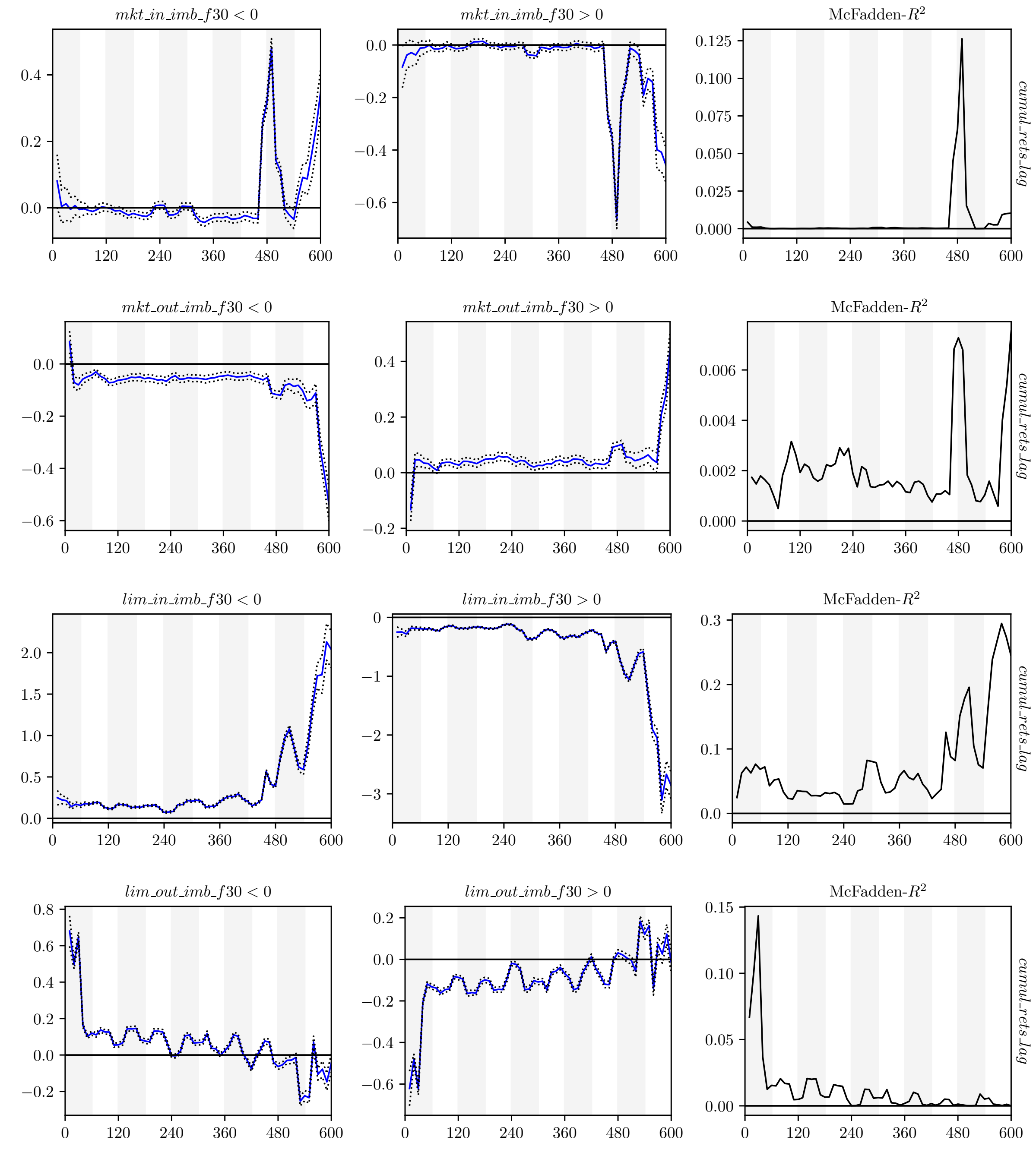


Figure 14: Prediction of pooled inflow- and outflow imbalances. This figure shows the results of the multinomial Logit regression approach defined in eq. (9). The first row(second) row shows the results for the models with $mkt\_in\_imb\_f30$($mkt\_out\_imb\_f30$) as dependent variable representing imbalances of market order inflows(outflows) pooled over 30 seconds forward. The first third(fourth) row shows the results for the models with $lim\_in\_imb\_f30$($lim\_out\_imb\_f30$) as dependent variable representing imbalances of aggressive limit order inflows(outflows) pooled over 30 seconds forward. The left(center) column shows coefficient estimates in cases where the imbalance is negative(positive). The third column depicts the estimate for the McFadden R-Squared. The horizontal axis is shared among all subplots in this figure and represents the number of seconds since auction start.

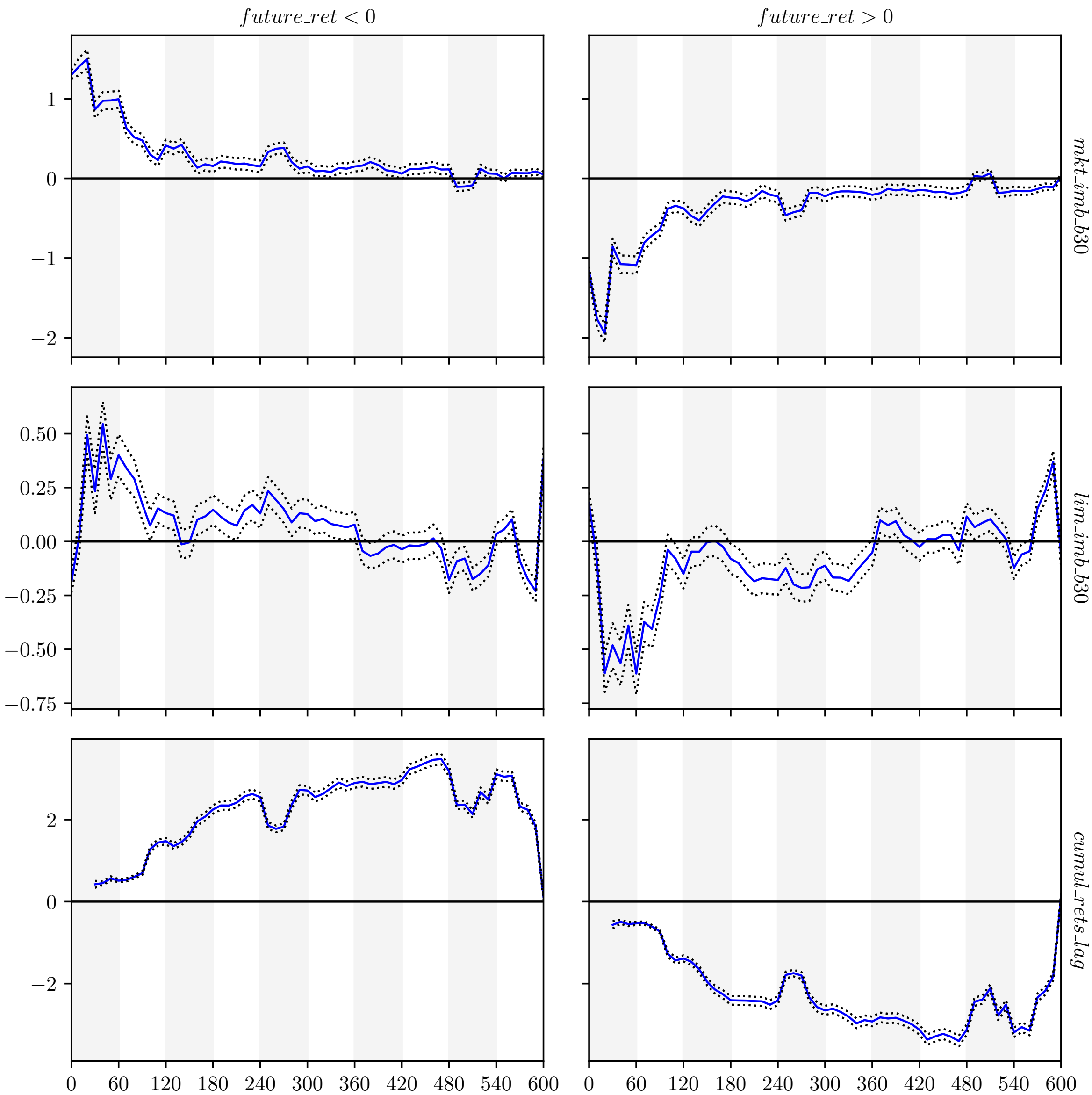


Figure 15: Prediction of future return using pooled order imbalances by interval. This figure shows the predictive coefficients of the multinomial Logit approach as stated in eq. (12). The dependent variable $future\_ret$ represents the return between interval l and the end of the auction. The left(right) column of subplots depicts marginal effects for the case when $future\_ret$ is negative(positive). The independent variables shown are mkt_imb in the top row, lim_imb in the middle row and $cumul\_rets\_lag$ in the bottom row. One model is estimated for every 10-second interval throughout the auction. All metrics of order imbalances are pooled 30 seconds backwards. The dotted lines around the coefficient estimates represent the 95% confidence interval for the parameter estimates. The horizontal axis represents the number of seconds since the start of the auction. The vertical axis representing the marginal effect of the independent variable is shared across all subplots within each row.

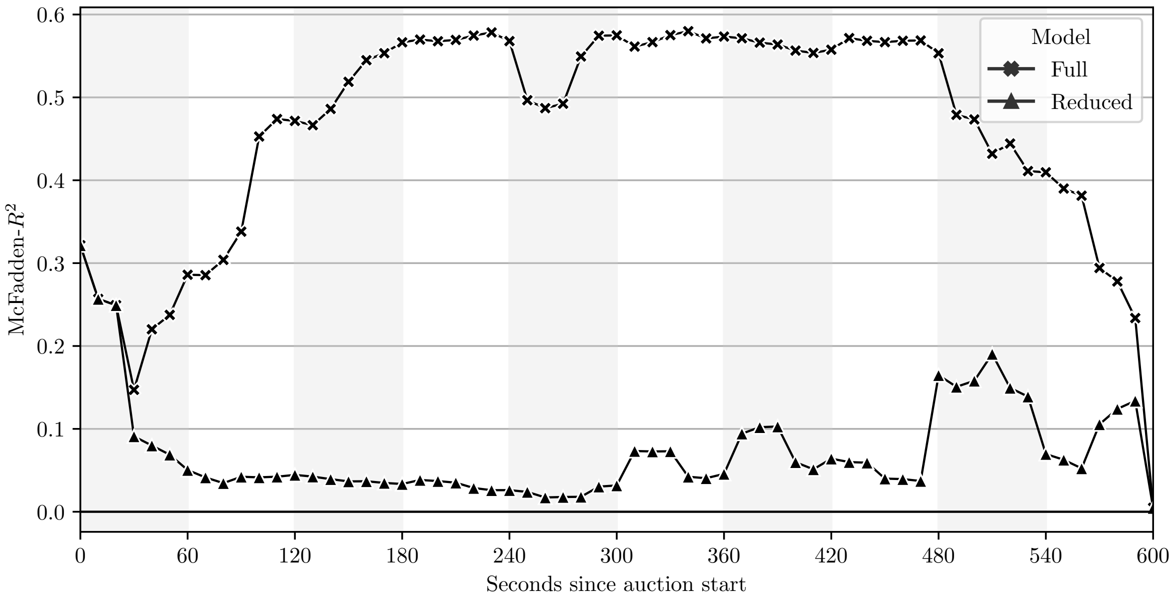


Figure 16: Comparison of McFadden R-Squared for future returns (pooled). This figure shows the estimated McFadden R-Squared based on two multinomial Logit regression models, both based on eq. (12). First, the full model includes all the variables. Second, the reduced variable does not include lagged cumulative returns and only relies on imbalance measures. The definition of the McFadden R-Squared is presented in eq. (10). The horizontal axis represents the number of seconds since the start of the auction.

6.2 Tables

		Closing volume		Continuous volume		Closing return		Size quantile		
	N	μ	σ	μ	σ	μ	σ	Q_1	Q_2	Q_3
ABBN	749	37.94	20.25	99.21	45.86	-0.00	0.22	0	0	749
ADEN	749	15.55	10.99	34.94	16.96	0.02	0.26	0	365	384
ALC	432	25.47	27.35	61.89	57.13	0.01	0.27	0	33	399
ALLN	749	1.39	1.07	2.65	1.83	0.04	0.21	742	7	0
AMS	749	7.98	5.52	43.37	30.08	0.01	0.44	0	289	460
ARYN	749	1.24	1.55	6.88	11.33	0.00	0.59	558	178	13
BAER	749	14.13	12.25	29.66	14.57	0.01	0.35	0	492	257
BALN	749	7.24	3.94	14.63	6.73	-0.03	0.39	30	705	14
BARN	749	6.71	6.35	12.20	10.41	0.01	0.24	163	576	10
BCVN	749	1.69	5.66	3.60	3.19	0.06	0.32	707	42	0
BEAN	749	1.18	3.38	2.81	2.13	0.02	0.39	733	16	0
BION	749	1.42	6.08	4.92	3.01	0.02	0.34	668	81	0
BUCN	749	1.91	1.75	5.85	3.44	0.02	0.25	591	158	0
CFR	749	37.49	17.69	91.60	48.09	0.01	0.30	0	0	749
CLN	749	7.80	4.75	23.79	18.95	0.01	0.31	11	621	117
CMBN	749	2.15	2.02	5.77	3.32	0.02	0.23	624	124	1
CSGN	749	32.65	18.96	102.26	45.09	0.01	0.25	0	0	749
DKSH	749	1.92	2.67	4.68	3.24	0.03	0.27	664	83	2
2DOKA	749	2.13	1.52	6.42	5.09	0.02	0.26	575	171	3
DUFN	749	7.15	5.55	23.26	15.99	0.02	0.37	5	618	126
EMSN	749	5.61	6.82	10.91	5.75	0.03	0.24	187	557	5
FHZN	749	3.81	2.11	9.79	6.07	0.01	0.24	282	462	5
FI-N	749	4.47	3.75	12.09	6.23	0.02	0.28	185	553	11
FORN	749	1.05	0.92	4.05	3.76	0.02	0.31	677	71	1
GALE	749	2.24	2.35	5.67	3.40	0.03	0.26	623	126	0

Continued on next page

		Closing volume		Continuous volume		Closing return		Size quantile		
	N	μ	σ	μ	σ	μ	σ	Q_1	Q_2	Q_3
GAM	749	1.12	1.20	4.77	5.24	-0.03	0.40	630	114	5
GEBN	749	19.34	9.96	42.64	24.55	0.01	0.24	0	177	572
GIVN	749	22.50	12.28	53.24	28.44	0.03	0.23	0	29	720
HELN	749	3.05	1.56	8.11	4.76	0.01	0.24	423	323	3
IDIA	749	1.66	1.94	7.92	6.34	0.02	0.35	473	270	6
KNIN	749	11.11	7.15	24.30	13.63	0.03	0.29	0	629	120
LAND	749	1.63	1.14	6.00	4.46	0.03	0.35	593	153	3
LHN	749	31.70	15.66	75.65	37.74	0.02	0.26	0	3	746
LISN	749	4.43	3.01	7.48	5.26	0.09	0.36	500	249	0
LISP	749	5.74	3.81	12.06	7.00	0.03	0.36	139	603	7
LOGN	749	11.22	24.75	37.54	27.07	0.01	0.23	0	393	356
LONN	749	28.13	14.77	78.15	42.55	-0.00	0.26	0	1	748
NESN	749	142.97	112.34	313.69	155.28	-0.01	0.27	0	0	749
NOVN	749	116.13	74.01	268.10	124.89	-0.01	0.30	0	0	749
OERL	749	3.32	2.20	9.38	6.54	-0.01	0.31	316	426	7
PARG	723	2.30	4.42	3.84	3.58	-0.06	3.11	676	47	0
PGHN	749	15.41	22.38	38.51	21.69	0.00	0.37	0	281	468
PSPN	749	4.76	3.46	8.67	6.04	0.03	0.22	367	381	1
PWTN	513	1.28	1.54	6.39	12.88	0.06	1.26	384	117	12
ROG	749	125.49	78.18	286.97	172.26	0.00	0.26	0	0	749
ROSE	749	1.01	2.90	6.11	9.13	0.07	0.46	566	164	19
SCHN	749	2.05	4.41	5.68	3.49	-1.40	9.94	620	129	0
SCHP	749	9.38	4.80	25.95	13.66	0.03	0.23	0	614	135
SCMN	749	21.98	12.15	49.10	30.54	0.01	0.22	0	87	662
SGSN	749	16.10	11.21	35.49	28.81	0.01	0.27	0	392	357
SIGN	563	3.71	7.29	8.30	11.85	0.10	0.75	339	213	11

Continued on next page

		Closing volume		Continuous volume		Closing return		Size quantile		
	N	μ	σ	μ	σ	μ	σ	Q_1	Q_2	Q_3
SIKA	639	21.03	14.53	58.47	31.77	0.02	0.27	0	36	603
SLHN	749	15.88	9.76	52.96	22.77	-0.00	0.24	0	24	725
SOON	749	13.75	7.98	35.44	20.18	-0.02	0.36	0	378	371
SPSN	749	7.14	5.11	10.93	7.71	0.02	0.27	219	526	4
SRAIL	429	2.80	4.80	9.44	19.64	0.05	0.42	256	164	9
SRCG	749	4.76	6.81	13.50	21.60	0.01	0.27	246	469	34
SREN	749	34.70	26.82	86.73	49.13	0.01	0.28	0	0	749
STMN	749	10.24	5.60	24.68	15.15	0.00	0.29	1	624	124
SUN	749	1.73	1.24	5.55	5.06	0.00	0.48	637	109	3
SWON	295	1.92	4.96	5.46	8.72	0.07	0.72	245	47	3
TECN	749	2.61	3.30	4.62	3.96	0.05	0.31	631	118	0
TEMN	749	11.38	22.47	30.85	23.63	0.01	0.38	0	525	224
UBSG	748	47.45	26.14	136.04	60.07	-0.00	0.26	0	0	748
UHR	749	20.22	9.05	51.01	28.38	0.01	0.29	0	142	607
UHRN	749	1.64	1.62	5.90	3.70	0.03	0.29	564	185	0
VACN	749	4.49	3.36	14.49	8.31	0.01	0.30	59	667	23
VIFN	749	7.26	4.70	24.09	14.74	0.05	0.29	0	634	115
ZURN	748	44.73	28.84	113.88	66.18	0.01	0.25	0	0	748

Table 6: List of all securities in the analysis. This table contains all the 69 equities analyzed. All of the securities were passed through the filters introduced in section 3. The table contains aggregated statistics for each security on a daily basis, including volumes and returns. The last three columns count the number of times a security falls into a size quantile across stocks within days where $Q_1(Q_3)$ represents stocks with the lowest(highest) continuous trading volume.

Panel A: Size quantile 1											
	μ	σ	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
mkt_in_imb	0.05	4.96	-11.29	-10.29	-8.12	0.00	0.00	0.00	8.36	10.37	11.33
mkt_out_imb	0.00	2.12	-5.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.75
lim_in_imb	-0.23	5.66	-11.73	-11.00	-9.75	0.00	0.00	0.00	9.25	10.72	11.50
lim_out_imb	-0.15	3.34	-10.70	-8.36	0.00	0.00	0.00	0.00	0.00	0.00	9.87
mkt_imb	0.06	5.32	-11.35	-10.48	-8.76	0.00	0.00	0.00	8.92	10.55	11.39
lim_imb	-0.31	5.85	-11.86	-11.14	-9.93	0.00	0.00	0.00	9.35	10.76	11.55
ival_rets	0.00	0.50	-0.63	-0.26	-0.06	0.00	0.00	0.00	0.06	0.27	0.63
cumul_rets	0.07	1.83	-4.07	-2.84	-1.34	-0.42	0.07	0.54	1.57	2.87	4.24

Panel B: Size quantile 2											
	μ	σ	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
mkt_in_imb	0.07	5.95	-12.26	-11.42	-9.87	0.00	0.00	0.0	10.00	11.48	12.29
mkt_out_imb	0.01	2.44	-7.64	0.00	0.00	0.00	0.00	0.0	0.00	0.00	7.83
lim_in_imb	-0.23	6.81	-12.60	-11.96	-10.99	0.00	0.00	0.0	10.67	11.76	12.44
lim_out_imb	-0.12	3.90	-11.56	-9.64	0.00	0.00	0.00	0.0	0.00	8.77	11.14
mkt_imb	0.08	6.33	-12.32	-11.55	-10.23	0.00	0.00	0.0	10.33	11.60	12.35
lim_imb	-0.29	6.95	-12.62	-12.01	-11.06	0.00	0.00	0.0	10.69	11.77	12.44
ival_rets	0.00	0.55	-0.80	-0.37	-0.10	0.00	0.00	0.0	0.10	0.37	0.81
cumul_rets	0.06	1.93	-4.51	-3.03	-1.80	-0.47	0.07	0.6	1.91	2.98	4.39

Table 7: Descriptive statistics of granular closing auction data. This table presents the descriptive statistics over all 10-second intervals over the closing auctions, including distributional information. *ival_rets* represents the return between two interval prices as defined in eq. (1), *mkt_in_imb*(*mkt_out_imb*) represents the imbalance of market order inflows(outflows), *lim_in_imb*(*lim_out_imb*) represents the imbalance of aggressive limit order inflows(outflows) and *mkt_imb*(*lim_imb*) represents the imbalance of market(aggressive limit) orders. All imbalances are computed according to eq. (3). All variables comprise around 3 million observations across intervals $l \in \{0, 1, \ldots, 60\}$.

Panel C: Size quantile 3											
	μ	σ	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
mkt_in_imb	0.05	7.54	-13.26	-12.61	-11.62	0.00	0.00	0.00	11.64	12.63	13.29
mkt_out_imb	-0.03	3.76	-11.60	-9.05	0.00	0.00	0.00	0.00	0.00	8.35	11.49
lim_in_imb	-0.18	8.26	-13.52	-12.92	-12.04	-8.52	0.00	6.83	11.89	12.79	13.39
lim_out_imb	-0.10	4.97	-12.52	-11.34	-7.30	0.00	0.00	0.00	0.00	11.07	12.31
mkt_imb	0.01	8.15	-13.31	-12.71	-11.87	-7.30	0.00	7.33	11.86	12.73	13.34
lim_imb	-0.21	8.36	-13.49	-12.88	-12.01	-8.93	0.00	7.92	11.85	12.74	13.35
ival_rets	0.00	0.61	-0.95	-0.45	-0.16	0.00	0.00	0.00	0.16	0.46	0.97
cumul_rets	0.02	1.96	-4.62	-3.04	-1.96	-0.49	0.05	0.59	1.96	2.93	4.26

Panel D: Size quantile 4											
	μ	σ	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
mkt_in_imb	-0.01	9.62	-14.69	-14.11	-13.24	-10.54	0.00	10.46	13.23	14.11	14.69
mkt_out_imb	-0.09	5.88	-13.63	-12.75	-10.59	0.00	0.00	0.00	9.42	12.66	13.56
lim_in_imb	-0.08	10.28	-14.85	-14.25	-13.42	-11.27	0.00	11.15	13.34	14.18	14.80
lim_out_imb	-0.06	6.98	-13.67	-12.82	-11.49	0.00	0.00	0.00	11.36	12.74	13.63
mkt_imb	-0.11	10.66	-14.81	-14.28	-13.53	-11.81	0.00	11.69	13.50	14.27	14.80
lim_imb	-0.06	10.39	-14.80	-14.19	-13.35	-11.26	0.00	11.17	13.29	14.12	14.75
ival_rets	0.00	0.54	-0.83	-0.42	-0.18	0.00	0.00	0.01	0.18	0.43	0.84
cumul_rets	-0.00	1.48	-3.06	-2.26	-1.37	-0.45	0.01	0.46	1.39	2.17	2.94

Table 7: (Continued)

Dep. Variable	*mkt_imb*	*mkt_imb*	*mkt_imb*	*lim_imb*	*lim_imb*	*lim_imb*
initial_ret	-0.1422**	-0.1293*	-0.1826***	-0.8226***	-0.2394***	-0.1779***
	(-2.1602)	(-1.9336)	(-7.0568)	(-16.531)	(-5.1996)	(-5.7441)
ival_rets_lag_pos	-0.1224*	-0.1118	-0.3216***	-0.2753	0.4177***	0.1360
	(-1.9389)	(-1.5126)	(-4.4161)	(-1.1984)	(3.0048)	(1.4211)
ival_rets_lag_neg	-0.0360	-0.0109	-0.2230***	-0.4603**	0.4371***	0.1714*
	(-0.5898)	(-0.1506)	(-3.0426)	(-2.0908)	(3.1976)	(1.8068)
cumul_rets_lag_pos		-0.0265	-0.0307		-1.5720***	-1.4609***
		(-0.4418)	(-0.5238)		(-17.587)	(-17.884)
cumul_rets_lag_neg		-0.0445	-0.0455		-1.6037***	-1.4846***
		(-0.8440)	(-0.8881)		(-17.687)	(-17.892)
mkt_imb_lag			0.0701***			
			(8.9412)			
lim_imb_lag						0.1450***
						(7.3243)
N	2692536	2846378	2692536	2692536	2846378	2692536
R-squared	0.0003	0.0004	0.0050	0.0079	0.0888	0.1038
Fixed Effects	Day	Day	Day	Day	Day	Day
Fixed Effects	Interval	Interval	Interval	Interval	Interval	Interval

Table 8: Results of panel regression of pooled order imbalances. This table shows the results of the panel model where order imbalances of market- as well as aggressive limit orders (pooled over 30 seconds forward from interval l) are regressed onto several independent variables. $initial\_ret$ represents the return between the last pre-close midquote and the first LOB of the auction, $ival\_ret\_lag$ represents the lagged interval return and $cumul\_rets\_lag$ represents the lagged cumulative return from the last pre-close midquote until the current interval. The suffix $pos(neg)$ represents the original value if it is positive(negative) and zero otherwise. The suffix lag the lagged value by one interval. Lagged variables are calculated such that there is no overlap with the order flow variables. The estimation is conducted with day- and interval-fixed effects, whereas t-statistics are presented in parentheses. The covariance matrix is clustered by interval and trading day to account for shocks within identical auctions across stocks. *, ** and *** denote significance at the 1%, 5% and 10% confidence level respectively.

Chapter 3

Sensitivity and Composition of Closing Order Books

Sensitivity and Composition of Closing Order Books*

Karl Frauendorfer† Louis Müller‡

June, 2022

Working Paper

Abstract

In current equity markets where closing auctions capture an increasing share of overall volume, price accuracy is paramount. Competition for order flow may be harmful for the price discovery process throughout the auction. In this paper, we analyse the final order books in closing auctions for a sample of Swiss equities and show that closing prices can be very sensitive towards removals of small percentages of volume. This is true for both limit- and market orders individually. In addition to this, we show that for large stocks consistently around 50% of executed orders are market orders whereas the order book composition of small stocks shows higher variance. With respect to overnight returns, we show that closing order books have no predictive power over the direction of overnight returns, however, a large share of market sell orders entails elevated volatility overnight. Finally, we look at price discovery contributions auctions and find that closing auctions are much less beneficial compared to overnight returns, indicating that only little new information is emerging during closing auctions.

Keywords: Closing auctions, price discovery, price efficiency, order imbalances

JEL Codes: G12, G14

*This work has been motivated by an ongoing research cooperation between SIX Swiss Exchange (SIX) and the Institute for Operations Research and Computational Finance (ior/cf-HSG) at the University of St.Gallen. We conducted research on several aspects of market microstructure. Most prominently discussed were topics of closing auctions and exchange performance in context of the European trading landscape. We would hereby like to thank SIX for their valuable inputs and for supplying detailed and granular order-level data of the Swiss stock exchange.

†University of St. Gallen, Switzerland

‡University of St. Gallen, Switzerland: louis.mueller@unisg.ch

1 Introduction

Throughout the European trading landscape, significant increases in market shares of closing auction have been observed over the last decade. On Euronext Paris for instance, CAC40 stocks have been trading more than 40% of their volume during closing auctions in the year 2019 (Raillon, 2020). The reasons for this are multi-faceted. First, stricter regulatory requirements around best execution have put brokers under increasing scrutiny with respect to the fulfillment of their client's orders. In this context, relying on the universally accepted closing prices mitigates the risk of misconduct. Second, investors increasingly prefer to deploy capital through low-cost passive strategies, that have been shown to perform on par with actively managed funds after fees (Easley et al., 2021; Sushko & Turner, 2018). Third, the rise of high-frequency traders increased adverse selection during the continuous trading phase, where sheer speed constitutes an advantage (Baldauf & Mollner, 2020; Biais et al., 2015; Budish et al., 2015). Finally, execution algorithms are learning that there are better execution opportunities during closing, since there is less adverse-selection (according to (3)). This in turn emphasizes end-of-day trading even stronger, eventually triggering a positive feedback-loop with volume clustering in those auctions (Pagano, 1989).

Such increasing reliance on closing auctions in comparison to intraday trading also raises certain questions. First and foremost, when more volume shifts into these auctions price discovery becomes more important, particularly given that the closing price marks the *reference price* to many market participants (Kandel et al., 2012). Simultaneously, it becomes more attractive for other venues who seek to take market share from the main exchange. Despite such behavior representing healthy competition between trading venues, there may be adverse consequences associated with it. For instance, the universal *reference price* for all other market closings is determined solely on the respective primary exchange. However, accurate price discovery crucially depends on the accumulation of all the participating volume, which in turn reveals the aggregation of all information to the entire market (Madhavan, 1992).

The question of market fragmentation and price discovery has been a very important one in academic literature in recent years. Despite several contradicting findings, the overall consensus views market fragmentation as positive for market quality if investors have simultaneous access to all venues and are interested in trading liquid large-cap stocks that usually have high volumes. Some relevant papers on this topic include Aitken et al. (2017), Degryse et al. (2015), Gomber et al. (2017), Haslag and Ringgenberg (2016), O'Hara and Ye (2011), and Pagano et al. (2013). Despite the overall favorable view with respect to fragmentation, all of these papers look at continuous trading in isolation, which is an order-driven market such that market participants can trade continuously. Closing auctions on the other hand are designed as call auctions following Madhavan (1992). During call auctions, orders are collected into one single order book for a predefined period of time and subsequently crossed, in order to obtain one single consensus price[1]. The fact that the universally accepted closing price is solely determined on one single venue[2] raises the question of whether fragmentation may have detrimental effects on the process of price discovery.

In order to understand the robustness of closing prices it is necessary to conduct a detailed investigation into the composition of relevant order books at the end of each auction. So far, the academic literature has not considered the order book of closing auctions but instead focused primarily on the ultimate auction price. One of the measures that cannot be captured when limiting the analysis solely onto prices are order imbalances. In the academic literature, however, it has been shown that order books and by an extension to this order imbalances during the continuous trading phase can contain information about the future price movement (Chordia et al., 2002, 2008; Chordia & Subrahmanyam, 2004; Su et al., 2012). Even though this has not yet been investigated formally in the context of closing auctions, it is conceivable that there are similar effects at play.

As a consequence of this, the first contribution of this paper lies in the granular analysis of the final order books as a part of closing auctions. Those order books manifest the basis for the determination of both closing price and volume traded. Of interest for this analysis are first and foremost the orders that are in fact executed after the uncrossing of the order book. Limit orders that have not been executed[3] have no impact on the closing price and are therefore irrelevant for the ultimate uncrossing. Nonetheless, there may be *opportunistic investors* placing significantly higher(lower) limit sell(buy) order in order to take advantage of sudden upward(downward) spikes in price. The second contribution lies in the analysis of the linkage between closing order books and returns both within the closing auction as well as overnight.

More specifically, this paper aims to address three main research questions. The first question relates to how sensitive closing prices are with respect to the outflow or absence of volume from the order book. For this purpose, multiple simulations are conducted where a certain percentage of orders is removed from the top of the order book, both symmetrically and asymmetrically. These outflows could represent an outflow of volume from the main exchange due to increasing fragmentation. Large adjustments in price

[1] This process is nothing else than crossing aggregated demand and supply to reach the optimal price and quantity.

[2] The decisive venue is usually the listing exchange of a given security.

[3] Limit buy(sell) orders with limit price below(above) the ultimate auction price expire without execution.

caused by volume outflows would indicate elevated sensitivity of prices, consequently implying that order flow should be concentrated on one venue. The results show that particularly in smaller (i.e. lower volume) stocks there are large amounts of volume beyond the closing price, indicating higher share of *opportunistic investors* relative to the total order book. The larger the stocks, the closer the volume is clustered around the final closing price. Another key result follows from the distribution of market orders at close. Moreover, symmetrically removing all market orders in larger(smaller) stocks leaders to smaller(larger) deviations in closing price, indicating that these orders to be more(less) balanced. However, one-sided removal of market orders in large(small) stocks leads to to larger(smaller) dislocations, indicating that the relative importance of market orders versus limit orders is greater in large stocks, as investors tend to use limit orders more frequently for small stocks.

The second research question under examination is about whether there is any predictive power in closing order books with respect to overnight returns. It has already been shown in the academic literature that closing auctions in isolation have a tendency to mean revert overnight in price (Bogousslavsky & Muravyev, 2020). However, it has not been examined whether these overnight reversions are significantly driven by the shape of the order book in addition to the closing return itself. Such an effect has already been documented empirically for market order imbalances during the continuous trading phase (Chordia et al., 2008; Chordia & Subrahmanyam, 2004). Similarly, it is conceivable that the same holds for the closing auction and the subsequent overnight return. However, the results of this study do not support such claims. More specifically, neither the imbalance nor the share of buy or sell market orders can meaningfully contribute to the prediction of overnight returns. A similar outcome is found when predicting overnight volatility instead of overnight returns. However, it has been found that closing auctions with a large share of market sell orders are followed by higher overnight volatility.

The third and final research question is about the quantification of price discovery attributable to closing auctions as well as the subsequent overnight return. The previously mentioned overnight reversion of closing returns by overnight returns indicates that both returns should have opposing effects on price discovery. For this reason, this study quantifies the Weighted Price Discovery Contribution (WPDC) following Barclay and Hendershott (2003), Barclay and Warner (1993), and Wang and Yang (2015) for both of those trading phases. Surprisingly, the results show that there is no opposing nature of the price discovery contribution of both measures. Instead, the results show that closing returns mainly consist of noise when put into context with the wider context of the trading day. In contrast to this, overnight returns have a significantly positive impact on price discovery. These results already account for the fact that overnight returns tend to be larger due to the longer time-span compared to the closing auction as well as the fact that earnings are released outside regular trading hours.

The remainder of this paper is structured as follows. In section 2, more information on the sample as well as institutional background is provided. Section 3 shows the first sensitivity results with respect to the removal of top-of-book volume from closing order books. Section 4 presents the results of the regression analysis aiming for the prediction of overnight volatility and returns based on the closing book. Section 5 shows the comparative analysis of WPDCs for both closing- and subsequent overnight returns. Finally, section 6 provides conclusive remarks and an outlook of potential further research in the realm of closing auctions.

2 Data

The analysis in this study relies on order data provided by SIX Securities & Exchanges (SIX) for their lit equity markets[4]. The data contains all orders that have been submitted or withdrawn across 294 stocks on the lit exchange. The data provides full visibility on normal as well as hidden orders[5]. Since the focus of this study is related to closing auctions, it is first necessary to extract the relevant closing order books from the order data. For this purpose, the design of an algorithm was required to interact with all orders in chronological order to reconstruct the closing order books at the end of the trading day, i.e. the closing auction. This approach is due to the fact that all incoming orders are interacting with the current state of the order book, which in turn depends on all previously submitted orders. This recursive procedure allows for the extraction of the full closing order books which is subsequently used to derive the closing price and -volume by crossing both sides of the book. This extensive data set allows for the deterministic reconstruction of order books at each point during the trading day and for each stock.

The closing auction on SIX is designed like a normal call auction, following Madhavan (1992). The auction starts at 17:20 following the continuous trading phase, such that both trading phases are strictly non-overlapping. Investors are allowed to submit two types of orders. The first type are market orders that are executed at the ultimate closing price. Limit orders on the other hand are submitted together with a limit price. Limit buy(sell) orders are only executed if the ultimate closing price is equal or below(above)

[4]Orders and trades that happen on the dark pool offering *SwissAtMid* that is also provided by SIX are not included in the data. Such dark pools is mostly targeted at large institutional investors who wish to execute their trades with minimal market impact. Therefore, this venue is usually chosen for the execution of block trades (Bloomfield et al., 2015; Buti et al., 2017; Comerton-Forde & Putniņš, 2015; Menkveld et al., 2017).

[5]One example of hidden orders are iceberg orders, where only a fraction of the full order is made visible to other market participants.

the limit price. The auction lasts for at least 10 minutes. Thereafter, the closing occurs randomly within a two-minute interval, in order to prevent manipulation of the closing price[6]. At the close, all market- and limit orders are gathered into aggregated demand- and supply. The closing price is determined where (a) crossed volume is maximized and (b) order imbalance is minimized.

The full sample period has been chosen to be between January 1, 2018 and June 30, 2021, consisting of 873 unique trading days. In addition to this, there were several filters allowed to the data. First, only stocks were considered that had at least 250 successful closing auctions, which represents one full year worth of trading. This filter has two main objectives. On the one hand, it enables more balanced panels for the later analysis by only taking into consideration stocks that have been public and on the exchange for a minimum of one year during the sample. On the other hand, it ignores stocks that are not frequently able to achieve a clearing and result in a closing price. Due to the nature of how closing auctions are crossed, it happens frequently that the there exists no decisive clearings as there are no crossing bid and ask orders. This is particularly prevalent in small stocks and has already been discussed in Ellul et al. (2005) and Ibikunle (2015) among others. Second, observations with absolute overnight returns of more than 10% were also disregarded from the sample. It can be assumed that for those observations there were some idiosyncratic events that led to such extreme price movements. Typical examples for this are extremely positive or negative earnings releases outside trading hours. An alternative explanation are stock splits, that are not explicitly marked in the data, but can be inferred through price reduction of more than 50% overnight Overall, this filter only disregards 15 observations over the entire sample and therefore has barely an impact on sample size. The third and final filter only considers the 100 largest stocks in the data that fulfill all of the above criteria. The size is determined by average closing volume per stock throughout the entire sample period. In combination, these three filters result in a cleaned dataset of around 84,400 valid observations. This indicates that it is only short around 2,900 observations from being a perfectly balanced panel. All the selected stocks as well as some stock-level descriptive statistics can be found in table 4 in the appendix.

In order to enhance the granularity of the analysis with consideration of stock size, a set of size-related quantiles are introduced. More specifically, all stocks are assigned into one of four size-quantiles in $\mathcal{Q} = \{Q_1, Q_2, Q_3, Q_4\}$ based on closing volume, i.e. traded volume at close. In this context, $Q_1(Q_4)$ represents the smallest(largest) stocks in the samples and the quantiles are reassigned on a daily basis. Therefore, the same stock can

[6] Some relevant studies with respect to random endings of auctions are Comerton-Forde and Putniņš (2011), Cordi et al. (2018), and Hillion and Suominen (2004).

fall into multiple different size quantiles throughout the sample[7]. This methodology has the advantage that within a trading day, all four quantiles are balanced. If size quantiles were only assigned based on overall volume, Q_4 would most likely be underrepresented due to small stocks being less likely to achieve a valid clearing. Moreover, on trading days where the panel is fully balanced, there will be exactly 25 stocks per quantile.

3 Order Book Sensitivities

In order to better understand the behavior of order books under the outflow of volume, this section focuses on the simulation of outflows from the top of the order book. In this context, a given percentage of volume is removed starting from the top of the order book[8]. In any case, the percentage must be calculated with respect to a given base value, which varies based on the removal algorithm applied. Consequently, all algorithms remove volume in a deterministic manner from the order book, always starting from the top. For the purpose of this work, there are three main algorithms applied in order to determine this base. The first two of them focus on the full- and partial order books including both limit- and market orders, whereas the third is only considering the latter.

There are two important points to be made before getting into the analysis. First, the simulated outflow of volume does assume a static order book. In reality, this may only be the case to a limited extent as other market participants have the ability to react to price dislocations caused by such an outflow. As Parlour (1998) models order books, there is a very high degree of endogeneity with respect to investors' decisions. More specifically, order submission strategies are dynamically adjusted in relation to the state of the order book and expectations for other investors' private information. Similarly, Pascual and Veredas (2009) develop a model in which all investors are fully aware of all the public information in relation to the state of the order book. Second, the outflows do not occur randomly but from the top of the book. Therefore, this analysis presents the worst-case scenario with respect to price dislocation for a given removal of volume. In contrast, if volume would be removed further down in the order book, for instance high(low) asks(bids), the price may not be affected at all since the concerned orders may not be executed anyway. Focusing on top-of-book volume allows for a certain degree of reproducibility as well as focus on orders that are most relevant to execution. Importantly, however, the exact percentage price dislocations are highly sensitive with respect to the

[7]Details to the assignment of quantiles per stock can be found in table 4 in the appendix.

[8]The top of the order book translates into attractiveness from the viewpoint of the opposite side. For instance, for buy orders the top of the book would be all market orders, followed by the limit orders with the highest limit price and so on.

chosen mode of removal.

The first algorithm presented here is called *liquidity-based algorithm*. Under this logic, the amount of volume to be removed is based on the average of the full order book on each side and thus including all market- and limit orders. The averaging of both sides of the book allows for a common base for the removal of both sides of the book. That means that an outflow of the same percentage from the same base always implies the same amount of volume in currency-terms. This algorithm is relevant as it captures volume beyond the clearing price on both sides. Particularly when there are large amounts of orders in the book that are far away from the clearing price and therefore unlikely to be executed. Such orders are not beneficial in terms of price discovery and consequently, a removal thereof has no impact on closing prices. Under this algorithm, it is quite possible that the resulting order book does not produce a viable clearing price anymore, as bid and ask may not have any overlapping orders remaining. Indeed, this starts to become an issue after removing around 40% of the volume for many of the stocks in the sample. In this analysis, the observations are disregarded as soon as clearing is made impossible. In practice, several studies have found issues of call auctions of smaller stocks with less investor interest, since order books may frequently not clear due to the composition of the order book. Examples of this include Ellul et al. (2005) and Ibikunle (2015) among others.

The results of this first algorithm can be seen in fig. 1 for simulated percentage removals of 5%–35%. The box-plots represent the distributions of price dislocations after removing a given percentage of volume from the top of the book. The box-plots are designed in such a way that the shaded area represents the inter-quartile range[9] which is separated by the median represented by a solid line. The whisks on each side cover all observations within the 5% and 95% percentiles of the distribution. The analysis is further segmented into the four size quantiles introduced in the previous section. The most obvious observation in all these distributions is that the smallest quartile of stocks is much more sensitive than all other quartiles. In fact, the plots manifest a pattern of decreasing sensitivity with increasing size. For instance, the simulated bid(ask) removal of 35% of orders leads to a dislocation of -636(+496) bps for small stocks versus only-123(+119) bps for high-volume stocks. Additionally, this example also underlines that large stocks are much more symmetric in terms of price dislocation, irrespective of whether bid or ask volume is removed. In contrast to this, small stocks with little volume are significantly more sensitive with respect to outflows bid volume throughout all simulated removals. This observation indicates that the order book for these stocks is not symmetric as it is for

[9]The inter-quartile range refers to the 50% of the observations between the 25% and 75% percentile of a given distribution.

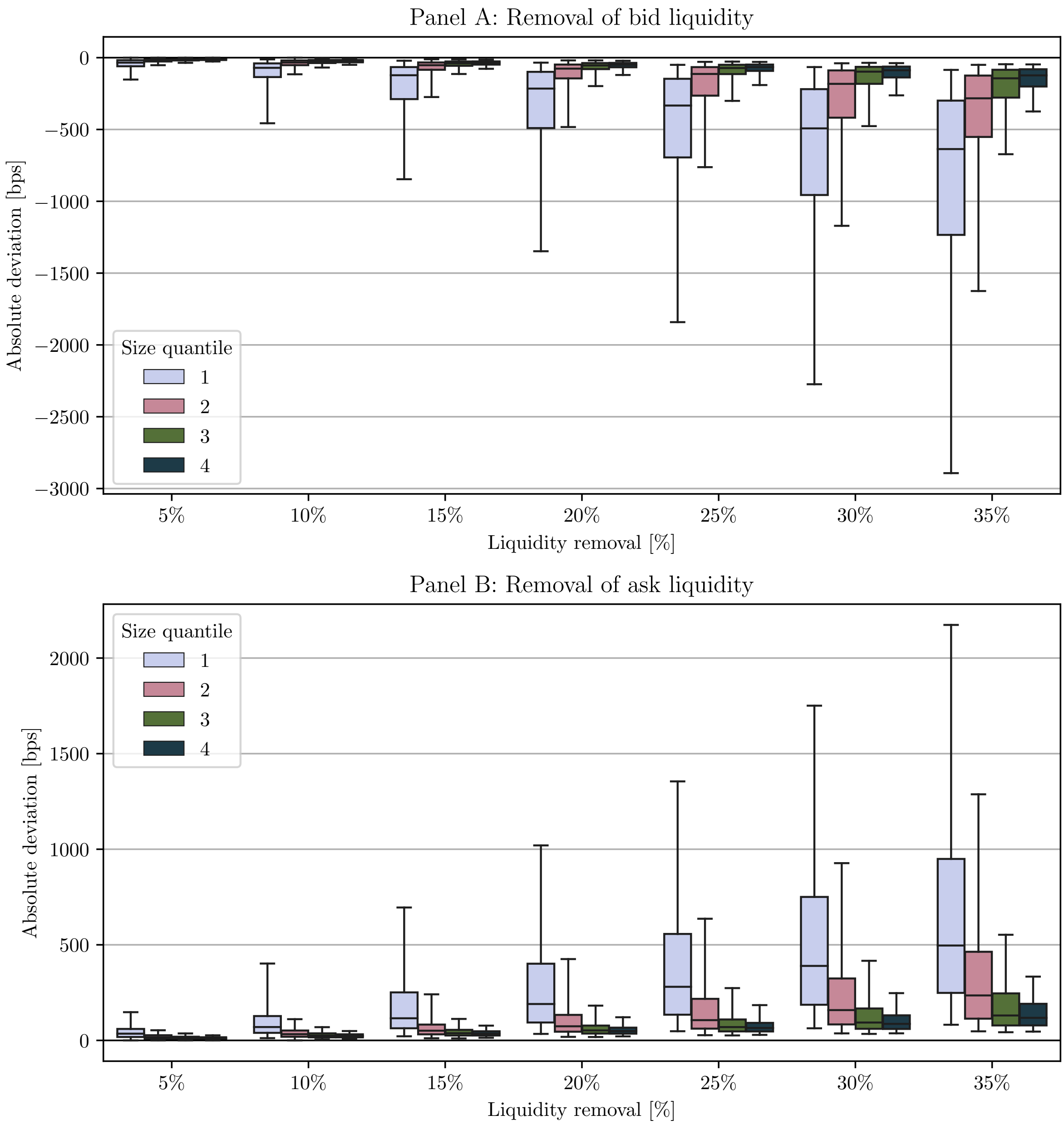


Figure 1: Distribution of price dislocations using the *liquidity-based algorithm*. The horizontal axis represents the percentage of volume removed and the vertical axis represents the price deviation in basis points. The data is presented in the form of box-plots, where the shaded area represents the inter-quartile range which is divided by a dark line representing the distribution median. The whisks represent the 95% prediction interval of the distribution. Panel A(B) shows the impact of percentage volume removal of bid(ask) limit orders. The results are presented for each of the size quartiles, which are computed based on continuous trading volume and reassigned on a daily basis.

large stocks. More specifically, the increased sensitivity towards bid removals indicates that there is generally an overhang of sell market orders beyond (i.e.ãbove) the clearing price. Such orders may be used by investors with a long position which they may want to close once the stock rallies sufficiently.

The *liquidity-based algorithm* takes into consideration outflows with respect to the full order book. However, not all orders are in fact beneficial to the discovery of the ultimate clearing price. More specifically, the removal of buy(sell) orders below(above) the clearing price has no impact on the price and is therefore less relevant. In particular, this algorithm includes orders that try to take advantage of large but short price movements, for instance to buy a stock during a downward spike. In order to account only for the orders that are in fact relevant for the clearing, the *execution-based algorithm* is introduced. In contrast to its previously introduced counterpart, this algorithm removes only volume that is in fact executed at the end of the auction. Assuming that the closing price does not deviate substantially from the pre-close midquote on most trading days, this algorithm mainly captures orders submitted during the auction, consisting of both market- and limit orders. Importantly, the execution volume is only a subset of all the market- and limit orders at the closing and highly dependent on the structure of the order book. Moreover, the executed volume must be equivalent on both sides of the book. Therefore, simultaneous removal of the same percentage from both bid- and ask top-of-book would not have any effect on the closing price. Moreover, there the auction will not lose its ability to clear unless 100% of volume is removed from both sides using this algorithm.

The results from the removal using this algorithm are shown in fig. 2. In contrast to the liquidity-based algorithms, the results here are much more comparable across all size quartiles. Nonetheless, large stocks are most affected by removal of top-of-book volume. One explanation for this may lie in a higher proportion of market orders as opposed to limit orders. Market orders are used by investors who want to execute at any given price and therefore rely on the auction's efficient determination of a clearing price. With less liquid stocks, investors may be more cautious and therefore prefer using limit orders due to fear of extreme adverse price movements. In addition to this, the extent of the price dislocation under the execution-based algorithm is significantly smaller given the same percentage removal. This is due to executed orders being a subset of order in the book. More specifically, removing 25% of execution volume entails a price dislocation of only 19.5(24.5)bps for small(large) stocks.

The discrepancy of outcomes between the liquidity-based and execution-based algorithm shows that particularly for low-volume stocks, a comparatively large portion of volume is located beyond the clearing price which is only captured by the liquidity-based algorithm. However, neither of these two algorithms takes into consideration the composition of the order book in terms of different order types. This is particularly important since academic research has shown that investors use different order types based on their objectives and information. For instance, Brown and Zhang (1997) show in a theoretical model that market orders can reveal significant information to the wider market. In addition to this Foucault et al. (2005), Goettler et al. (2005), and Roşu (2009) find that impatient investors use market orders since they value immediacy higher as opposed to getting the optimal price. This argument should generally hold for price-inelastic in-

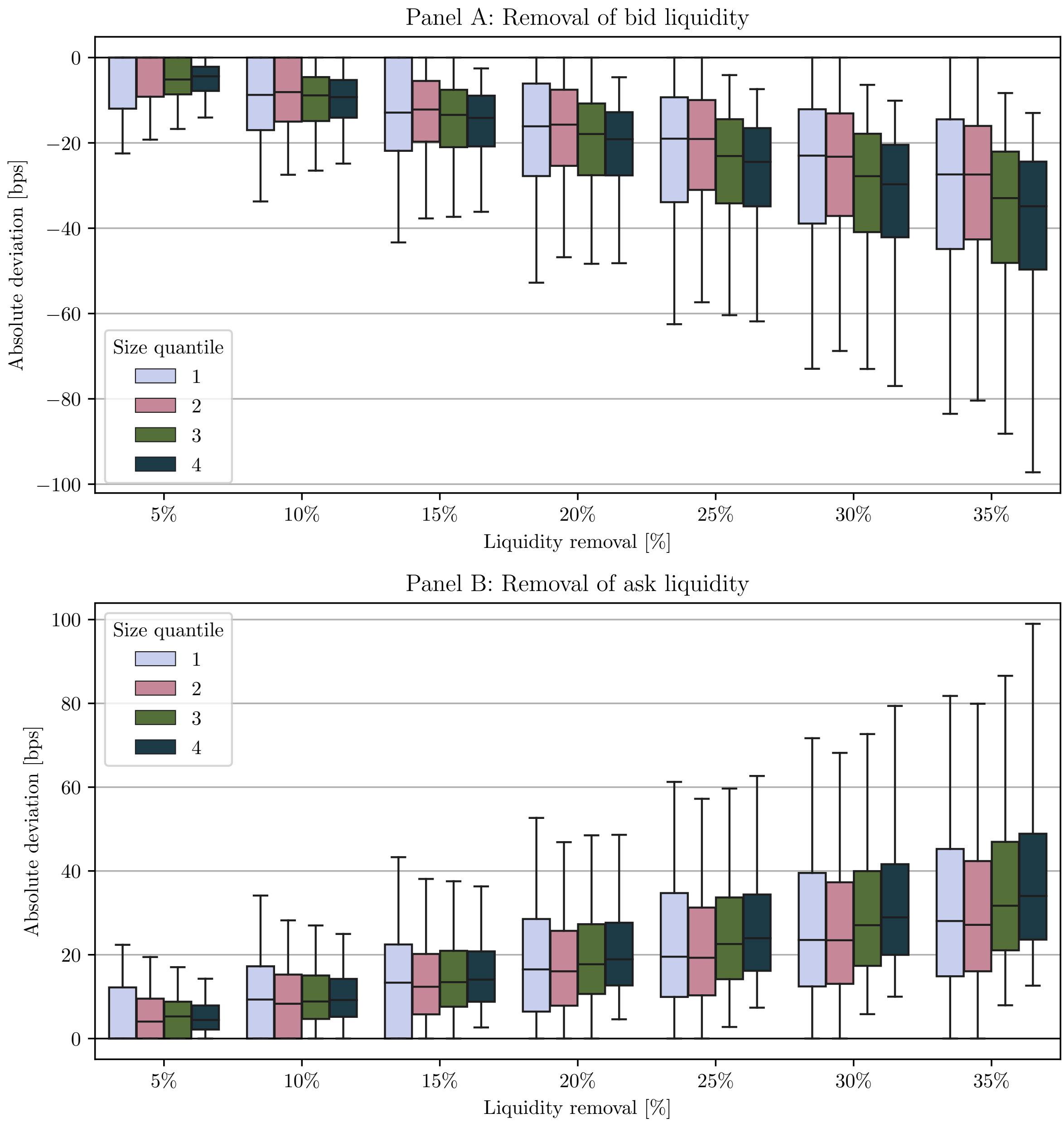


Figure 2: Distribution of price dislocations using the *execution-based algorithm*. The horizontal axis represents the percentage of volume removed and the vertical axis represents the price deviation in basis points. The data is presented in the form of boxplots, where the shaded area represents the inter-quartile range which is divided by a dark line representing the distribution median. The whisks represent the 95% prediction interval of the distribution. Panel A(B) shows the impact of percentage volume removal of bid(ask) limit orders. The results are presented for each of the size quartiles, which are computed based on continuous trading volume and reassigned on a daily basis.

vestors, such as index funds who are rebalancing based on a given benchmark or investors who want to unload their inventory to avoid overnight price risk (Cartea & Jaimungal, 2015).

For this purpose, the final algorithm considered here focuses on the role of market orders in the closing auction, given by the motivation of investors using them. The *market-based algorithm* removes only market orders but leaves limit orders unaffected. The market orders can be submitted both during the closing auction as well as during

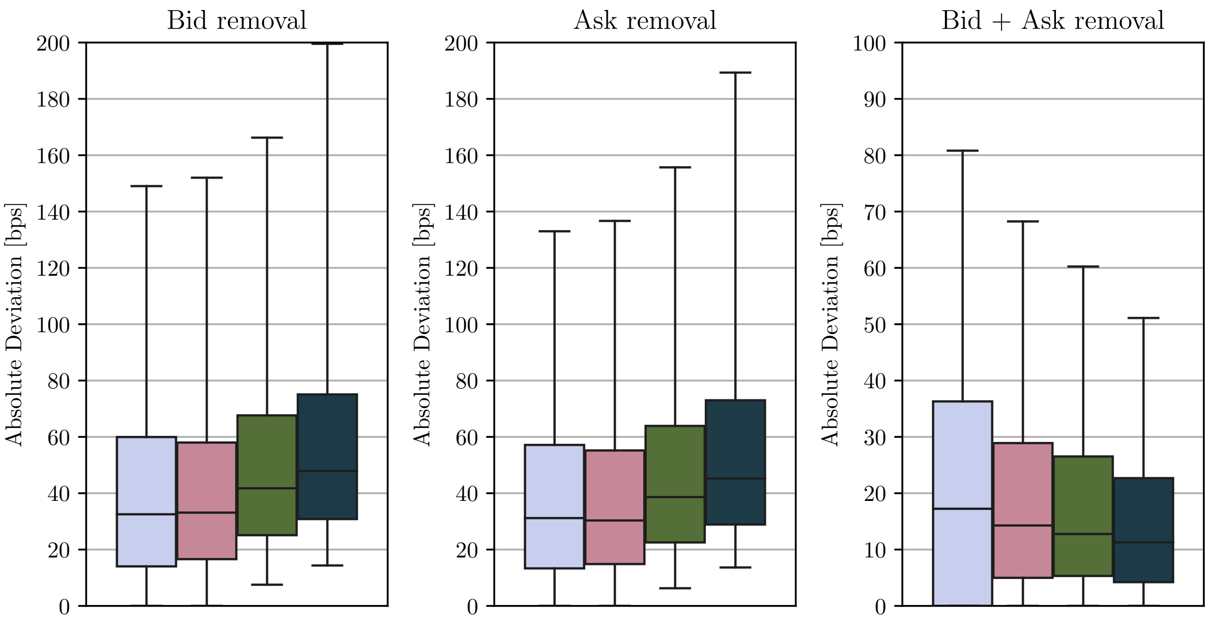


Figure 3: Distribution of absolute price dislocation using the *market-based algorithm*. The vertical axis represents the price deviation in basis points. The data is presented in the form of boxplots, where the shaded area represents the inter-quartile range which is divided by a dark line representing the distribution median. The whisks represent the 95% prediction interval of the distribution. The results are presented for each of the size quartiles, which are computed based on continuous trading volume and reassigned on a daily basis, where the fill-color of the boxes represents each quartile. The lightest(darkest) box represents the least(most) liquid stocks.

the continuous trading phase[10]. In contrast to the other two algorithms, the *market-based algorithm* is not necessarily remove the same number of shares on both sides, due to the possibility of order imbalances. Such imbalances occur when there is an overhang of either buy- or sell market orders. Consequently, the simultaneous removal of a given percentage from both sides of the book is likely to lead to a dislocation of the closing price. This has not been possible under the previous two algorithms.

The dislocations after the removal of all market orders from the book are depicted in fig. 3. To begin with, the one-sided removal of all market orders has fairly symmetric effect. In all cases, large stocks are affected the most with a median absolute deviation of around 45bps. The smallest stocks on the volume spectrum on other hand only deviate around 32bps when removing all bid or ask market orders. Moreover, for stocks in quartile 1(2), such a removal does not cause any deviation from the original clearing price in 15%(10%) of observations. This indicates that market orders tend to have a smaller influence than in larger stocks. When looking at the right panel, the pattern is reversed such that large stocks are affected the least when market orders are removed, with a median of around 11bps of absolute dislocation. The the 95% prediction intervals are also considerably smaller as opposed to the other size quantiles. Meanwhile, the smallest size quartile in

[10]Such orders are submitted to be executed at-close. Until the beginning of the closing auction, these orders remain invisible to other investors and are only activated once the auction begins. From this point onward, they are treated equal to market orders submitted after the start of the auction.

particular shows the largest variance.

In summary, these results lead to the following two observations. First, large stocks react less to one-sided removal of market orders than small stocks. Second, small stocks react more to a removal of two-sided removal of market orders than small large stocks. These two observations indicate that the composition of order books at execution varies across size quartiles. The following metric is defined as market ratio MR and computed for each stock s trading day d in order to to capture the share of market orders executed on any given side of the book:

$$MR_{d,s}^{(side)} = \frac{MKTVOL_{d,s}^{(side)}}{CLVOL_{d,s}} \qquad \forall \quad side \in \{buy, sell\} \tag{1}$$

The variable $MKTVOL^{(buy)}(MKTVOL^{(sell)})$ represents the volume initiated using buy(sell) market orders. $CLVOL$ stands for the total volume traded in the closing auction. As closing auctions cross both market- and limit orders at the optimal uncrossing $MKTVOL_{d,s}^{(side)} \leq CLVOL_{d,s}$ must hold. Consequently, the market ratio measure is bound by $MR \in [\,0, 1\,]$. Moreover, the measure is calculated for each side of the book individually due to their independence.

In order to visualize the joint distribution of $MR_{d,s}^{(buy)}$ and $MR_{d,s}^{(sell)}$, the concept of bi-variate kernel density estimation (KDE) is introduced. KDE is a method which is used to approximate non-parametric distributions from empirical data. As the name suggests, the distributions are flexible and are not subject to a set of underlying parameters. The methodology was first introduced by Rosenblatt (1956) and Parzen (1962). To explain the concept of KDEs where $m \in \mathbb{N}$ is the number of dimensions, a kernel $K_{\mathbf{H}}(x)$ is a function that takes a vector $x \in \mathbb{R}^m$ as input. The kernel function returns the density of a multinomial distribution with zero mean across all dimensions and covariance matrix $\mathbf{H} \in \mathbb{R}^{m \times m}$ as parameters[11], where $\mathbf{H}$ is a diagonal and positive-semidefinite matrix. The covariance matrix is estimated using Scott's rule $\sqrt{\mathbf{H}_{jj}} = n^{-1/(m+4)}\sigma_j$, where σ_j is the standard deviation of the jth variable and all off-diagonal elements are zero (Scott, 1979). In a next step, the average density can be calculated for any input value x based on the proximity of all observations in the sample:

$$\hat{f}_{\mathbf{H}}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(x - x_i) \tag{2}$$

[11]The formula of the multivariate kernel is given by:

$$K_{\mathbf{H}}(x) = (2\pi)^{-m/2} \det{(\mathbf{H})}^{-1/2} \exp\left(-\frac{x^{\intercal}\mathbf{H}^{-1}x}{2}\right)$$

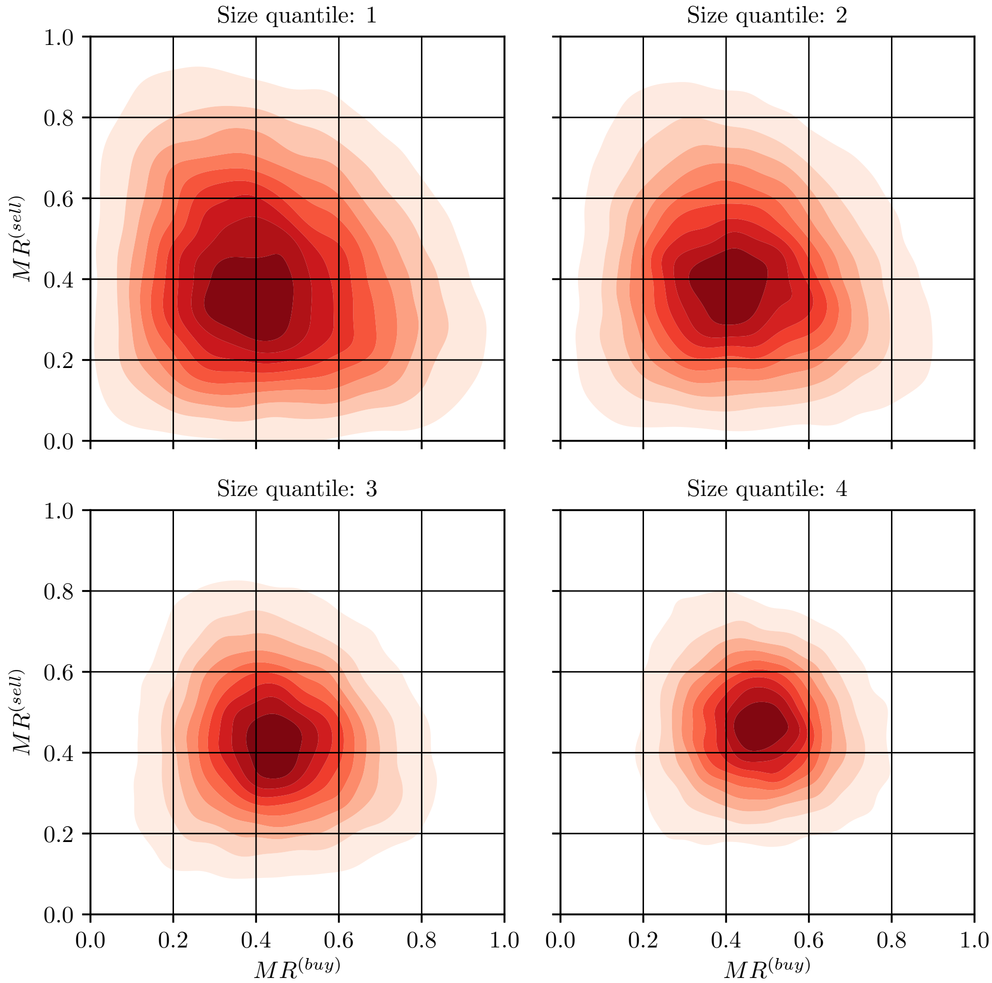


Figure 4: Non-parametric distributions of market order ratios. This figure shows the results of the multivariate KDEs with respect to market buy- and market sell ratios. Each panel represents presents the estimation for one size quantile respectively, where size quantile 1(4) represents the smallest(largest) stocks in the sample based on closing volume as defined in section 2. All estimations is based on a Gaussian kernel with covariance matrix approximated using Scott's rule. The grid has 200 steps in each dimension, resulting in 40,000 estimations per panel and 160,000 across the four panels.

In this equation $x_i \in \mathbb{R}^2$ is a two-dimensional vector containing $MR^{(buy)}$ and $MR^{(sell)}$ as components, such that $m = 2$. The estimation in eq. (2) is repeated for each point of the input space ranging from 0 to 1, given by the boundary condition of the market ratio variables. Along each of the two dimensions, the input space is separated into 200 equally-sized steps, resulting in a grid of 40,000 unique points. The resulting plot can essentially be interpreted as a two-dimensional histogram visualizing the empirical distribution.

The results of the KDE estimation are presented in fig. 4, for each size quartile individually. The plot very quickly shows that there are significant differences in order book composition between sizes. To begin with, small stocks in quantile 1 show the most broad distribution. From both the bid- and ask- side of the book there are days auctions when one side is almost entirely determined by market orders. However, this does not occur simultaneously. This can be partly explained that an auction cannot clear with only market orders on both sides, as the price cannot be determined without presence of at least one limit order. In contrast to this, it is also common that there are barely any market orders and the clearing is purely driven by limit orders from both sides. Overall, the mode of the distribution for the smallest stocks is at around 39% market order share in both dimensions. Larger stocks in size quantile 4 show a much more balanced picture. Specifically the distribution is much more contained in the center of the plot. The resulting distribution flattens out quickly for market order ratios outside of the 20%–80% range. This indicates that for these stocks, closing auctions are driven by both market orders and limit orders on both sides of the book. Moreover, the mode of the distribution is at around 47% in each dimension, indicating that investors are more comfortable using market orders in large stocks.

This observation can be explained by the risk of smaller stocks not clearing properly at the close due to lack of liquidity. Moreover, in these stocks a small amount of incoming volume has much greater effect on the ultimate closing price. Similar observations have also been made by Ellul et al. (2005) and Ibikunle (2015) who found that call auctions of small stocks can be less reliable for those reasons. Ellul et al. (2005) also shows that investors are less likely to participate in such small stock call auctions when they anticipate volume to be low.

4 Effects on Overnight Returns

So far, the focus of the analysis has mostly been on the descriptive properties of closing auction books. To contrast this, this section is going to shed light on the effects of order book composition on overnight returns after the clearing. For this reason, a new set of variables is being introduced. To begin with, let $P^{(OP)}_{d,s}$, $P^{(CL)}_{d,s}$ and $P^{(PR)}_{d,s}$ on day d and for stock s be the stock prices after the opening auction, the closing auction and the last price before the beginning of the closing auction respectively. Following this understanding, overnight returns are defined as:

$$ONROP^{(h)}_{d,s} = \ln\left(\frac{P^{(OP)}_{d+h,s}}{P^{(CL)}_{d,s}}\right) \qquad ONRCL^{(h)}_{d,s} = \ln\left(\frac{P^{(CL)}_{d+h,s}}{P^{(CL)}_{d,s}}\right) \tag{3}$$

which represents the logarithmic overnight return starting at the closing price. In both equations, $h \in \mathbb{N}^+$ stands for how many days in the future the endpoint of the interval lies. For instance, $ONRCL^{(1)}_{d,s}$ represents the overnight return starting at the closing price on day d and finishing on the closing price the next day.

In addition to this, returns that are preceding the closing are defined as well. More specifically, these are intraday returns ($IR_{d,s}$) between the opening and the pre-close midquote as well as closing returns ($CR_{d,s}$) between the pre-close midquote and the closing price. Both measures are expressed in the form of logarithmic returns are defined as follows.

$$IR_{d,s} = \ln\left(\frac{P_{d,s}^{(PR)}}{P_{d,s}^{(OP)}}\right) \qquad\qquad CR_{d,s} = \ln\left(\frac{P_{d,s}^{(CL)}}{P_{d,s}^{(PR)}}\right)$$

In addition to only considering returns, volumes are also an important driver of price changes according to the literature (Campbell et al., 1993; Chen et al., 2001; Chordia & Swaminathan, 2000; Frieder & Subrahmanyam, 2004; McMillan, 2007). For this reason, the closing volume will also be considered in more detail during the further analysis. However, in order to account for the heterogeneity between different stocks with respect to volume, the observations are transformed accordingly by demeaning the data. More specifically:

$$DEMVOL_{d,s} = \ln\left(CLVOL_{d,s}\right) - \overline{\ln\left(CLVOL\right)}_s$$

where $CLVOL_{d,s}$ represents tha closing volume of stock s on date d in currency terms and the second term on the right-hand side represents the arithmetic mean of daily closing volume of stock s. Moreover, the leftmost term of the equation stands for the expected logarithmic volume for stock s. This measure is achieved by computing the arithmetic average across all recorded trading days of a given stock. Consequently, this measure for demeaned volume is centered around zero with respect to each individual stock. This mitigates the requirement to account for heterogeneity in volumes across stocks in a later stage of the analysis.

In addition to the metrics of market ratio defined in eq. (1), the imbalance of market orders is another important factor when analyzing returns in conjunction with order flow. Market ratios only consider the amount of market orders in comparison to the full executed volume. However, it does not consider the relationship between market orders on the opposite sides of the book. For this purpose, it is crucial to introduce an additional metric measuring the imbalance of such orders. In academic literature there has been evidence that order imbalance can lead price movements. Examples of this include Chordia et al. (2005, 2008). What is not yet clear, however, is how order imbalances at the closing auction influences the outcome of overnight return. For this purposes, order imbalances at the end of the closing auction ($IMBAL$) are computed following Chordia et al. (2002), Chordia and Subrahmanyam (2004), and Holden and Jacobsen (2014) for each stock s on day d, where:

$$IMBAL_{d,s} = \frac{MKTVOL_{d,s}^{(buy)} - MKTVOL_{d,s}^{(sell)}}{MKTVOL_{d,s}^{(buy)} + MKTVOL_{d,s}^{(sell)}}. \tag{4}$$

The variable $MKTVOL$ represents the volume available for sale or purchase without limit price (i.e. the volume of all market orders entered into the closing auction). Based on this definition, the result is bound by $IMBAL \in [-1, 1]$. One advantage of this measure comes from it following comparable distribution across stocks, regardless of market capitalization. Moreover, there are only 120 observations in the data where $IMBAL$ takes either +1 or -1. This implies that for the remaining 99.86% of the data, there is no situation where there are only market orders from one side of the book.

Out of all the metrics defined so far, table 1 shows the descriptive statistics as well as distributional information[12]. The first observation with respect to the table comes from looking at the return metrics. It becomes apparent that overnight returns from closing to opening auction are on average +0.07%. By computing the standard error[13] the t-score amounts to be approximately 20, showing significant deviation of the mean from zero. However, the overnight return until close ($ONRCL$) has an average of zero again, indicating a reversal of sorts. Indeed, the intraday returns IR is at -0.07%, offsetting the average overnight return $ONROP^{(1)}$. Closing returns are also show a slightly positive average of 0.02% for each day and stock. The same observation can be made for market order imbalances, indicating that both of these metrics are related, which has also been documented by Besson and Fernandez (2021). Finally, the variable $DEMVOL$ is centered around zero, due to its definition. In terms of distributional shape, most variables are rather symmetrical with the exception of intraday returns and overnight returns ending on a closing price. All of these are skewed to the left-hand side, indicating that negative shocks materialize more significantly than positive ones. On the other hand, closing returns are more skewed to the positive side. Interestingly, all of these observations are observed to varying extent across all size quantiles as shown in table 5 in the appendix. However, smaller stocks have these tendencies more extreme. For instance, small(large) stocks have an average overnight returns to open of +0.11%(+0.04%) and average intraday returns of -0.13%(-0.01%).

[12] The same information for the individual size quantiles defined in section 2 is presented in table 5 in the appendix.

[13] The standard error is calculated as

$$SE = \frac{\sigma}{\sqrt{N}}$$

where σ is the standard deviation and N represents the number of observations.

	N	μ	σ	1%	5%	10%	25%	50%	75%	90%	95%	99%
$ONROP^{(1)}$	78257	0.07	0.97	-1.87	-1.04	-0.70	-0.27	0.07	0.44	0.87	1.21	1.88
$ONRCL^{(1)}$	84292	0.00	2.57	-5.88	-2.99	-1.99	-0.84	0.04	0.93	1.98	2.87	5.56
$ONROP^{(2)}$	78157	0.10	2.30	-5.50	-2.96	-2.00	-0.82	0.13	1.08	2.20	3.09	5.70
$ONRCL^{(2)}$	84292	0.00	2.57	-5.88	-2.99	-1.99	-0.84	0.04	0.93	1.98	2.87	5.56
IR	78346	-0.07	1.65	-4.78	-2.58	-1.80	-0.83	-0.03	0.74	1.63	2.34	4.26
CR	84392	0.02	0.38	-0.89	-0.45	-0.31	-0.14	0.01	0.17	0.34	0.49	0.96
$DEMVOL$	84391	-0.00	0.69	-1.81	-0.98	-0.71	-0.35	-0.01	0.35	0.75	1.07	1.86
$MR^{(buy)}$	84392	0.44	0.17	0.07	0.16	0.22	0.32	0.44	0.56	0.67	0.74	0.86
$MR^{(sell)}$	84392	0.43	0.17	0.06	0.15	0.20	0.31	0.43	0.55	0.66	0.73	0.85
$IMBAL$	84392	0.02	0.31	-0.74	-0.49	-0.37	-0.18	0.02	0.21	0.41	0.54	0.77

Table 1: Descriptive statistics of relevant metrics. This table presents the descriptive statistics over the metrics of interest. $ONROP^{(1)}(ONROP^{(2)})$ represents overnight returns starting at the closing and ending at the open one(two) days later. The same logic applies to $ONRCL$ with the exception that the interval ends at the closing auction of the respective day. IR and CR represent intraday- and closing returns respectively. $DEMVOL$ represents the logarithmic volume, demeaned within stocks. $MR^{(buy)}/MR^{(sell)}$ and $IMBAL$ represent market ratio and market order imbalances respectively.

After having defined the full set of variables of interest[14], the analysis proceeds with the relevant regression equation. Due to the two-dimensional nature of the data, a panel regression is the appropriate means for further procedure. It has been shown repeatedly in academic literature that simultaneous returns across stocks are correlated due to common reactions to newly emerging macro news. For instance, Baker and Wurgler (2006) show that market-wide investor sentiment affects stocks to different extent based on their features, however, all the shocks are correlated. In other work, Fang and Peress (2009) show that stocks with larger(smaller) media coverage perform worse(better) in the cross-section of stocks.

In addition to fixed effects, the analysis does also contain lagged observations of both overnight returns ending the next day at open and closing returns. For this purpose, the notation of the lag operator[15] is applied. More specifically, the overnight return ending in next day's opening as well as the closing return are lagged, resulting in the variables $LONROP^{(1)}_{d,s}$ and $LCR_{d,s}$ respectively. Following these restrictions, the following vector

[14]These variables of interest include $ONROP^{(h)}$, $ONRCL^{(h)}$, IR, CR, $DEMVOL$, $IMBAL$, $MR^{(buy)}$ and $MR^{(sell)}$.

[15]The lag operator is used to lag a given variable by an arbitrary number of periods. The notation of the capital L is taken from Baltagi (2011, p. 137) and Verbeek (2017, p. 262), which leads to the expression:

$$Lx_t = x_{t-1}.$$

In the academic literature, this is often also referred to as the backshift operator and denoted with a B.

of control variables is defined:

$$CTRL_{d,s} = \begin{bmatrix} LCR_{d,s} \\ LONROP^{(1)}_{d,s} \\ IR_{d,s} \\ CR_{d,s} \\ DEMVOL_{d,s} \end{bmatrix} \tag{5}$$

For the further course for the analysis, the same set control variables are reused in order to account for potential inefficiencies in market returns as well as for asymmetric influences of market return. This leads to the following two regression equations that constitute the basis for the following analysis:

$$ONR_{d,s} = \alpha_d + \beta_1 MR^{(buy)}_{d,s} + \beta_2 MR^{(sell)}_{d,s} + \gamma' CTRL_{d,s} + \varepsilon_{d,s} \tag{6}$$

$$ONR_{d,s} = \alpha_d + \beta_3 IMBAL_{d,s} + \gamma' CTRL_{d,s} + \varepsilon_{d,s} \tag{7}$$

where $ONR \in \{ONROP^{(1)}, ONRCL^{(1)}, ONROP^{(2)}, ONRCL^{(2)}\}$ is a placeholder variable for illustrative purposes and $\gamma \in \mathbb{R}^5$ represents a vector containing all the coefficients to the control variables. Moreover, day-fixed effects are denoted as α_d. Under these two equations, problems of endogeneity are mitigated by timely separation, such that all dependent variables are determined strictly after all independent variables have been measured. This setup thus represents a Granger (1969) type form of causality. For inference purposes, the covariance matrices of all panel regressions were computed following the methodology laid out in Driscoll and Kraay (1998). This procedure is an extension to Newey and West (1987) and is based on the usage of a Bartlett kernel function as recommended by Hoechle (2007). The resulting standard errors are robust to both cross-sectional correlation as well as autocorrelation of residuals.

The results of the regression equations (6) and (7) are presented in table 2. The first panel on top shows the results of constrained models which are limited to the inclusion of the control variables defined in eq. (5). The results show that across the four time horizons for overnight returns, there are certain consistencies. To begin with, the lagged overnight returns are all statistically significant at the 5% or even 1% level with consistently positive coefficients. This indicates positive autocorrelation between overnight returns. This observation is amplified when looking at the coefficient in the third model, which indicates that positive one percent of additional overnight return coincides with +0.07% return for $ONROP^{(2)}$. In contrast to this, intraday returns only affect the overnight return until the next day's opening with a significantly negative coefficient. This indicates that intra-

Panel A: Control variables only

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.3325	0.2797	0.2624	0.2119
Effects	Day	Day	Day	Day
LCR	-0.0324	0.0065	-0.0028	-0.0065
	(-0.9398)	(0.1749)	(-0.0611)	(-0.1243)
$LONROP^{(1)}$	0.0404**	0.0455***	0.0706***	0.0508**
	(2.4048)	(3.0963)	(4.9041)	(2.0100)
IR	-0.0133***	0.0193*	0.0051	0.0251
	(-3.0648)	(1.6486)	(0.4738)	(1.6110)
CR	-0.3281***	-0.2434***	-0.4228***	-0.2756***
	(-6.8016)	(-2.8644)	(-8.5424)	(-2.7208)
$DEMVOL$	-0.0295*	-0.0456**	-0.0668***	-0.1070***
	(-1.8723)	(-2.1153)	(-2.6936)	(-3.0748)

Panel B: Market buy- and sell ratios

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.3327	0.2797	0.2625	0.2119
Effects	Day	Day	Day	Day
Controls	Yes	Yes	Yes	Yes
$MR^{(buy)}$	-0.0697***	-0.0120	-0.0790	-0.0595
	(-2.9775)	(-0.1965)	(-1.3832)	(-0.7320)
$MR^{(sell)}$	0.0267	0.0399	0.0117	0.1496
	(0.6681)	(0.5588)	(0.1580)	(1.4338)

Panel C: Market order imbalance

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.3326	0.2797	0.2624	0.2119
Effects	Day	Day	Day	Day
Controls	Yes	Yes	Yes	Yes
$IMBAL$	-0.0408	-0.0005	-0.0130	-0.0564
	(-1.6065)	(-0.0134)	(-0.3061)	(-1.0180)

Table 2: Results of panel regressions on overnight returns. This table shows the regressions of overnight returns over multiple time horizons as defined in eq. (3) onto other variables that are listed as follows. LCR represents the previous' day closing return whereas $LONROP^{(1)}$ stands for the return between the previous day's close and the same day opening price. IR and CR represent intraday- and closing return on the same trading day respectively. $DEMVOL$ represents the logarithmic volume, demeaned within stocks. $MR^{(buy)}/MR^{(sell)}$ and $IMBAL$ represent the market buy/sell ratios and market order imbalances respectively. All panel models were estimated using day-fixed effects. Reported standard errors are derived using Driscoll-Kraay covariance matrices. *, ** and *** denote significance at the 1%, 5% and 10% level respectively.

day returns are partly reverted overnight. Similarly, closing returns are also significantly negatively correlated with overnight returns, albeit to much larger extent. The results indicate that on average between 27% and 42% of closing returns are reverted over the four overnight return horizons analyzed. In contrast to intraday returns, this negative effect persists over at least two full days. Consequently, closing returns must be considered to be inefficient which is in line with the results in Bogousslavsky and Muravyev (2020). Note that the mere inefficiency of prices does not necessarily lead to profitable trading strategies, as these opportunities may not be systematically exploitable for profit due large amounts of noise[16]. Finally, the demeaned volume also shows significantly negative coefficients over all four models. This implies that abnormally large closing volume is followed by negative returns. This can be explained by sell-offs in stocks, as during such sell-offs many investors want to liquidate their positions simultaneously.

The same regressions in Panel A were re-estimated without inclusion of any fixed effects and are presented in table 6 in the appendix. Even without the day-fixed effects, the coefficients are directionally identical with one difference for the variable IR. Specifically, when including day-fixed effects, intraday returns are insignificant over a time horizon longer than $ONROP^{(1)}$. When pooling the regression as in table 6, intraday-returns are reinforced instead of reverted over time-horizons of $ONRCL^{(1)}$ and longer. Another major difference is the explanatory power of the models. Whereas the day-fixed effects models yield R-Squared values between 21% and 33%, the pooled models only achieve 0.2% to 1.8%. Consequently, a large part of the variance in the dependent variable is explained by market-wide movements in prices.

The lower two panels of table 2 include all the independent variables in addition to the control variables in Panel A as discussed in the previous paragraph. In the first panel, market order ratios show little significance in predicting overnight returns. The only exception to this is a significantly negative effect of market buy ratio on overnight returns ending at the next opening. Despite the lack of consistent significance across time-horizons, all coefficients are are having the same direction. More specifically, all coefficient estimates for market buy(sell) ratio negative(positive). This indicates that there tend to be reverting forces overnight, but not sufficiently strong to grant statistical significance. Similarly, the measure of market imbalance presents insignificantly negative coefficients.

Both of these results indicate that overnight returns go against large one-sided pressures of market orders during the closing auctions but not enough to cause statistical significance. One explanation lies in the inclusion of closing returns as a control variable.

[16]In a time when high-frequency traders are increasingly active, markets have also become more efficient (Brogaard et al., 2014; Budish et al., 2015; Carrion, 2013; Manahov et al., 2014). For this reason there may be less exploitation of such patterns.

In fact, when removing CR from the regression, the outcomes for the other independent variables change as presented in table 7. Specifically, $MR^{(buy)}(MR^{(buy)})$ becomes significantly positive(negative) and $IMBAL$ becomes significantly negative. This finding suggests that market orders are an important driver of closing returns, but offer limited explanatory power beyond in terms over overnight return predictability.

So far, it has been shown that the composition of closing order books does not have any explanatory power on overnight returns when controlling for closing returns themselves. However, the previous analysis was only considering directional predictions. Another important aspect of overnight returns is their volatility, particularly after periods of sustained buying- or selling pressures. For instance, Chan and Fong (2000) and Su et al. (2012) have found a positive relationship between order imbalances and return volatility during continuous trading hours. For closing auctions, this phenomenon has not been discovered yet. This is also partly due to the fundamentally different characteristics of closing auctions compared to the continuous phase.

In a next step, the previous analysis is extended in that it now focuses on absolute returns as the dependent variable instead of directional ones in oder to measure volatility. In the academic literature, it has been shown that volatility is highly autocorrelated (Adam et al., 2016; Aggarwal et al., 1999; French et al., 1987; Schwert, 1989). This phenomenon has multiple explanations, such as elevated volatility around earnings release dates or general periods of heightened market uncertainty. For this purpose, one new variable is introduced, the volatility of each securities during the previous ten trading days. More specifically, it measures the volatility of returns from the closing on the previous day until the closing of the day of the observation. Thus it is defined as:

$$VOLA_{d,s}^{(m)} = \sqrt{\frac{\sum_{j=0}^{m-1}\left(LONROP_{d-j,s}^{(1)} + IR_{d-j,s} + CL_{d-j,s}\right)^2}{m-1}} \tag{8}$$

where $m \in \mathbb{N}^+$ represents the number of trading days considered prior to d in the calculation of the volatility. During the further course of the analysis, the parameter m will be set to represent the previous ten trading days[17]. Due to the logarithmic nature of returns, the sum of the three variables in the numerator represents the return from closing price to closing price on each day. Before defining the regression equation, let $ABSCTRL$ be

[17]The analysis has also been conducted with other choices of m. The results have shown to be robust with respect to the choice of $m^* \in \{5, 10, 20, 50\}$ since the results do not deviate.

a vector containing the absolute values of all control variables considered in this step:

$$ABSCTRL_{d,s} = \begin{bmatrix} |LONROP^{(1)}| \\ |IR_{d,s}| \\ |CR_{d,s}| \\ DEMVOL_{d,s} \\ VOLA^{(10)}_{d,s} \end{bmatrix}$$

In a next step, fixed effects are added to the analysis. On the one hand, day-fixed effects account for the fact that some trading days experience more volatility caused by external factors. It has already been shown in table 2 that such effects are important to enhance the predictive power of the model. On the other hand, this analysis introduces stock-fixed effects that have not been considered before. The reason for this is that in this analysis the dependent variable is represented by absolute returns, introducing a certain degree of asymmetry. As seen in table 5 in the appendix, small stocks tend to experience higher volatility than larger ones, while both have average returns close to zero. This unobserved heterogeneity can be well accounted for by stock-fixed effects, denoted as μ_s. This logic leads to the following regression equation:

$$|ONR_{d,s}| = \alpha_d + \mu_s + \gamma' ABSCTRL_{d,s} + \beta_1 MR^{(buy)}_{d,s} + \beta_2 MR^{(sell)}_{d,s} + \beta_3 |IMBAL_{d,s}| + \varepsilon_{d,s} \quad (9)$$

As in the previous analysis, ONR is a placeholder variable representing overnight returns over four different time horizons and $\gamma \in \mathbb{R}^5$ is a vector containing all the coefficients with respect to absolute control variables. As previously, all covariance matrices have been computed following Driscoll and Kraay (1998) in order to obtain HAC standard errors for inference.

The results of the regression equation in (9) are laid out in table 3. Panel A presents the results of all models with only stock-fixed effects μ_s but without day-fixed effects. The results show that volatility is persistent through various time horizons, given by the positive coefficients for volatility-based control variables. Also the coefficient for demeaned volume is highly significant and positive with the exception of the models with overnight returns until the next morning's open. This indicates that over longer time-horizons larger volumes during closing auctions is connected to higher volatility in the coming days. The important part of this panel, however, lies in the three variables on the bottom that are related to the closing order book. Most coefficients are insignificant with the exception for $MR^{(sell)}$, which are significantly positive at the 1% level. Panel B shows similar results, albeit to a lesser extent, due to the introduction of day-fixed effects α_d.

Panel A: One-way fixed effects

Dep. Variable	$\lvert ONROP^{(1)}\rvert$	$\lvert ONRCL^{(1)}\rvert$	$\lvert ONROP^{(2)}\rvert$	$\lvert ONRCL^{(2)}\rvert$
N	73115	73115	69763	73026
R-squared	0.5184	0.5450	0.5038	0.3861
Effects	Stock	Stock	Stock	Stock
$\lvert LONROP^{(1)}\rvert$	0.1074***	0.1980***	0.2040***	0.3175***
	(5.8092)	(5.1250)	(4.1367)	(4.1195)
$\lvert IR\rvert$	0.0468***	0.1088***	0.1044***	0.1597***
	(8.0869)	(9.5110)	(7.1433)	(6.7907)
$\lvert CR\rvert$	0.3071***	0.3888***	0.4126***	0.4339***
	(9.3743)	(6.3996)	(7.2385)	(4.5636)
$DEMVOL$	-0.0006	0.0445***	0.0342**	0.0473*
	(-0.0873)	(2.9716)	(2.0766)	(1.7386)
$VOLA^{(10)}$	0.0174**	0.0571***	0.0590***	0.0824**
	(2.4576)	(2.8029)	(2.5797)	(2.4549)
$\lvert IMBAL\rvert$	-0.0060	0.0003	0.0312	0.0458
	(-0.4989)	(0.0111)	(0.7538)	(0.8082)
$MR^{(buy)}$	0.0032	0.0738**	0.0980*	0.1509**
	(0.2120)	(2.2594)	(1.8358)	(2.0011)
$MR^{(sell)}$	0.0883***	0.1409***	0.1791***	0.3200***
	(4.6273)	(3.5521)	(3.8716)	(4.0410)

Panel B: Two-way fixed effects

Dep. Variable	$\lvert ONROP^{(1)}\rvert$	$\lvert ONRCL^{(1)}\rvert$	$\lvert ONROP^{(2)}\rvert$	$\lvert ONRCL^{(2)}\rvert$
N	73115	73115	69763	73026
R-squared	0.5997	0.6143	0.5660	0.4491
Effects	Stock/Day	Stock/Day	Stock/Day	Stock/Day
$\lvert LONROP^{(1)}\rvert$	0.0380***	0.0453*	0.0576*	0.0809**
	(3.7771)	(1.8707)	(1.9296)	(2.0769)
$\lvert IR\rvert$	0.0317***	0.0737***	0.0677***	0.0971***
	(7.8303)	(9.3399)	(7.0463)	(7.2672)
$\lvert CR\rvert$	0.2462***	0.2640***	0.2783***	0.1977***
	(7.6698)	(6.3423)	(5.5113)	(3.1167)
$DEMVOL$	-0.0197***	0.0389***	0.0317**	0.0468**
	(-3.1266)	(3.0836)	(2.2647)	(2.0080)
$VOLA^{(10)}$	0.0073**	0.0318***	0.0340**	0.0406*
	(2.1295)	(2.5822)	(2.3000)	(1.9160)
$\lvert IMBAL\rvert$	0.0010	0.0096	0.0305	0.0601
	(0.0926)	(0.3410)	(0.7553)	(1.0576)
$MR^{(buy)}$	-0.0043	0.0534**	0.0584	0.0826
	(-0.3690)	(2.0092)	(1.1639)	(1.3790)
$MR^{(sell)}$	0.0639***	0.0865***	0.1198***	0.1905***
	(4.0539)	(2.7560)	(2.9451)	(3.5097)

Table 3: Results of panel regressions on overnight volatility. This table shows the regressions of absolute overnight returns over multiple time horizons as defined in eq. (3) onto other variables that are listed as follows. The regression equation is stated in eq. (9), based on the following variables: $|LONROP^{(1)}|$ stands for the absolute return between the previous day's close and the same day opening price. $|IR|$ and $|CR|$ represent absolute intraday- and closing return on the same trading day respectively. $DEMVOL$ represents the logarithmic volume, demeaned within stocks. $MR^{(buy)}/MR^{(sell)}$ and $|IMBAL|$ represent the market buy/sell ratios and absolute market order imbalances respectively. The top panel presents models estimate using only stock-fixed effects, whereas the bottom panel shows the same model with both stock- and day-fixed effects. Reported standard errors are derived using Driscoll-Kraay covariance matrices. *, ** and *** denote significance at the 1%, 5% and 10% level respectively and t-scores are in parentheses.

More specifically, $|LONROP^{(1)}|$ and $VOLA^{(10)}$ have reduced t-scores as opposed to Panel A. This decrease in t-score implies that much of the variance could be accounted for by simultaneous effects in the cross-section of stocks. Nonetheless, the positive coefficients for the market sell ratios persist, particularly over time horizons between 24 and 48 hours.

What is striking about this finding is that market sell ratios have no predictive power with respect to directional overnight returns as presented in table 2. However, these ratios are a predecessor of increased overnight volatility. This indicates that a high share of market sell orders entails the presence of uninformed investors, attempting to reduce their positions in a correlated manner. This increase in volume could explain the elevated volatility as also observed by Lee et al. (1994), Louhichi (2011), Odean (1998), and Xu et al. (2006). Importantly, the order flow is uninformed since it does not contain actual information about future outcomes of returns. Interestingly, this is not the same on the buy-side of the order book. This could be explained by the fact that investors tend to sell-off securities in a correlated manner, whereas buying is less correlated (Abreu & Brunnermeier, 2003; Griffin et al., 2011; Huang & Wang, 2009). The same reasoning explains why market corrections downwards happen more abrupt than upwards, implying negative skewness of returns (Alles & Kling, 1994; Harris et al., 2004). What is important to note is the lack of predictive power of market order imbalances with respect to overnight returns, both in terms of direction as well as extent. While Besson and Fernandez (2021) find that market order imbalances have positive correlation with and are thus drive closing returns, the same is not true for overnight returns.

5 Contribution to Price Discovery

The main focus of the analysis has so far been on the interaction between closing order books and overnight returns. In this final section of the paper, the analysis will examine the importance of the closing auction to the price contribution of the closing auction with respect to the remainder of the trading day. This is important since it is not yet clear whether the closing auction does benefit to price discovery or whether it is solely a platform for large orders to be traded. The previous results in table 2 suggest that closing returns tend to be reverted overnight, which would make closing returns inefficient. However, price discovery must always be put in perspective in that it is the comparison of two different returns.

In order to quantify price discovery in the context of this paper, the WPDC measure is applied. This metric has been frequently used in the academic literature to illustrate the contribution of specific returns to more broad price developments. Examples of this can be found in Barclay and Hendershott (2003), Barclay and Warner (1993), and Wang and Yang (2015). Generally speaking, this measure considers two trading periods, one outer return and one inner return, where the latter a strict sub-period of the former. Consequently, the WPDC measure aims to calculate contribution of the inner return to the outer return, underlining the previously mentioned relative nature of price discovery. For the purpose of this analysis, the outer return is defined as the logarithmic return from opening to the opening price the next day. Due to the logarithmic nature of returns, the outer return is the sum of intraday returns, closing returns and overnight returns until the open:

$$OUTR_{d,s} = IR_{d,s} + CR_{d,s} + ONROP^{(1)}_{d,s}$$

This outer return therefore covers price changes from opening price to opening price the next day within a 24-hour time frame[18]. In a next step, a function calculating the WPDC is defined. The function takes one input, which is representing the inner return INR:

$$WPDC_{d,s} = \underbrace{\frac{|OUTR_{d,s}|}{\sum_{\hat{s}\in\mathcal{S}}|OUTR_{d,\hat{s}}|}}_{weighting} \times \underbrace{\frac{INR_{d,s}}{OUTR_{d,s}}}_{contribution} \tag{10}$$

This definition can be separated into two terms. First, the weighting term puts absolute returns across stocks into relation to each other within the same day, where $\mathcal{S}$ represents the set of all stocks in the sample. Consequently, stocks with large(small) absolute open-to-open price changes are weighted more(less). By definition, all weights across stocks must sum up to 1. Second, the contribution term captures the price discovery contribution of the inner return r with respect to the outer return. In contrast to the weighting term, the contribution term is directional and results in a positive(negative) number if inner- and outer returns do(do not) align. For stability reasons, observations where $OUTR_{d,s} \in (-\varepsilon, \varepsilon)$ are disregarded, where ε is a sufficiently small positive number. The reason for this is to prevent the denominator in the second term from becoming to small and distorting moments of the resulting distribution. For the purpose of this analysis a value of $\varepsilon = 0.01\%$ is chosen, leading to the exclusion of 2.76% of all valid observations. Other values for ε have been tested for robustness purposes and have been found to yield similar results[19].

[18]Based on the definition in eq. (3), this can also be defined based on $P^{(OP)}_{d,s}$ representing the opening price of stock s on day d:

$$OUTR_{d,s} = \ln\left(\frac{P^{(OP)}_{d+1,s}}{P^{(OP)}_{d,s}}\right)$$

[19]Specifically, ε has been set to 0.05%, 0.1%, 0.5% resulting in the exclusion of 3.73%, 6.92%, 31.08% of valid observations.

In order to make the price discovery measures statistically meaningful, the weighted price discovery contributions are aggregated into t-scores for each stock $s \in \mathcal{S}$. For this purpose, the an aggregation function taking x as an input is defined, whereas x is indexed in both the s and d dimensions:

$$\Phi_s(x) = \frac{\overline{x}_s}{\sigma_s(x)/\sqrt{D}} \tag{11}$$

where $\overline{x}_s = \sum_{\hat{d}=1}^{D} x_{\hat{d},s}/D$ represents the average for stock s across all days d and D stands for the number of trading days in the sample. Furthermore, $\sigma_s(x)$ represents the standard deviation of variable x for stock $s \in \mathcal{S}$. The above equation effectively calculates the t-statistic for testing the null-hypothesis of no price contribution on average $H_0 : \overline{x}_s = 0$, since the denominator of the fraction equates to the standard error through the division by $\sqrt{D}$. Another advantage of this standardization lies in the ability to compare price discovery contribution of both closing- and overnight returns on an even playing field, despite the latter most likely being larger than the former.

This method of aggregation is subsequently applied to the weighted price discovery contribution and the order imbalance as defined in eq. (4), resulting in $\Phi_s(WPDC)$ and $\Phi_s(IMBAL)$ respectively. Figure 5 shows two scatter plots where the vertical axis presents the former and the horizontal axis represents the latter. In the top panel, the $WPDC$ is based on closing returns (CR) as inner returns whereas the lower panel relies on overnight returns ($ONROP^{(1)}$) as inner returns. For illustrative purposes, the axes are scaled identically for both plots. There are two main observations to be made. First, market order imbalances tend to be rather positive than negative in the cross-section of stocks. Exactly half of all stocks in the sample are significantly positive at the 5% level with respect to market order imbalances, whereas only 5 are negative. Nonetheless, the largest stocks as measured by closing return are more balanced in terms of market orders and are not significantly deviating from zero in the horizontal dimension. One notable exception to this is Roche with a t-score of 4.3.

Second and more importantly, there are substantial differences when comparing the t-scores of weighted price discovery contributions based on closing- and overnight returns, shown on the vertical axis of the plots. In the top panel, 49(5) of the 100 stocks have significantly positive(negative) results for WPDC, with no stock reaching a t-score of above 6. This indicates that for 49 of the stocks, the closing auction constitutes an important contribution to the overall price discovery, whereas for 46 of the stocks the closing return is pure noise with respect to the remainder of the trading day. For the 5 stocks with negative results, this means that the auction returns tend to go against the wider price discovery, i.e. constituting negative contribution. In contrast to this,

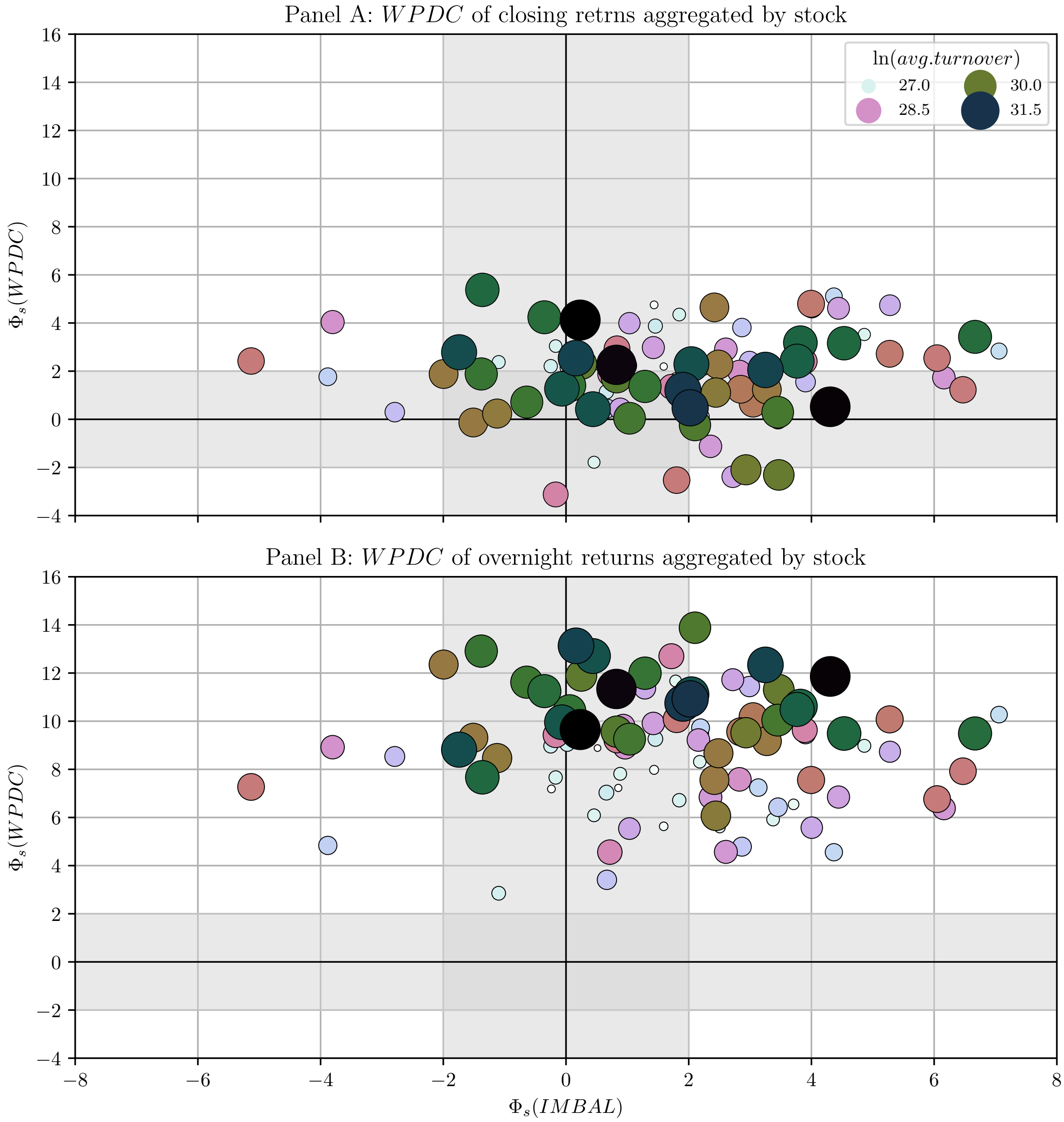


Figure 5: Contribution to price discovery aggregated by stock. This figure shows a bubble plot depicting the intersection between order imbalances and weighted price discovery contribution. The horizontal axis represents the order imbalance at the closing auction. The vertical axis represents the weighted price discovery contribution for the closing(overnight) return with respect to the entire trading day, as defined in eq. (10). Both axes have been aggregated for each stock and normalized following the definition in eq. (11). The gray shaded area mark results below the 5% confidence level of 1.96, rendering observations falling into said interval as insignificantly different from zero. The color as well as the size of the bubbles is indicative of the logarithm of the average closing turnover of each stock throughout the sample. Each of the 100 bubbles represents one of the stocks in the sample.

in all stocks without exception the overnight return is a positive contribution to price discovery. Particularly large stocks in terms of closing volume yield t-scores between 8 and 14, indicating extremely high statistical significance beyond the 0.1% level of significance. These results clearly show that overnight returns are very important for price discovery, much more so than closing returns, for all stocks in the sample.

In order to test for robustness of the results, the same analysis was also conducted with aggregation of the $WPDC$ metric on a daily basis instead of on a stock level. Similarly to the definition in eq. (11), this methodology standardizes the data for each day d.

$$\Phi_d(x) = \frac{\overline{x}_d}{\sigma_d(x)/\sqrt{n(\mathcal{S})}} \tag{12}$$

where $n(\mathcal{S})$ represents the number of stocks in the data set with valid observations on day d.

The results for this adjusted aggregation under $\Phi_d(x)$ are depicted in fig. 6. Each plot shows 872 observations, i.e. one for each trading day in the 3.5 years of the sample. As with the previous results, imbalances of market orders are rather well balanced across trading days, which is expressed by the even distribution around zero on the horizontal axis. The vertical axis, depicting aggregated $WPDC$ on a daily basis is much less balanced, however. For the $WPDC$ of closing returns shown in the top panel there is no significant difference from zero on 611 days, which amounts to approximately 70% of trading days. Of the remaining trading days, 194(67) are significantly positive(negative). Only 13 trading days result in a value larger than 6. In contrast to this, the bottom panel depicting $WPDC$ of overnight returns is more dispersed in the vertical dimension. In this panel, only 218 of the trading days in the sample are insignificant. Out of the remaining trading days 632(22) show a significantly positive(negative) contribution of overnight returns to price discovery.

Ultimately, both figs. 5 and 6 show similar results. When aggregated on both stock- and day-level, closing auctions contribute less to price discovery compared to overnight returns, which is manifested through the larger number of positive standardized $WPDC$ measurements of the latter compared to the former. In addition to this, there are no visual patterns indicating that either the direction or the extent of market order imbalances at close have any impact on price discovery, both during the close and overnight. This is particularly noteworthy with respect to closing returns. As Besson and Fernandez (2021) pointed out, closing returns are mainly driven by market order imbalances. However, the evidence presented here indicates that the generated returns are mainly constituting noise when compared to the trading day as a whole. More specifically, the price discovery happens only after hours, when closing returns are digested by the markets.

6 Conclusions

Regulatory changes, rising popularity of passive investment vehicles and the predatory nature of some high-frequency trading strategies have all led to increasing relative impor-

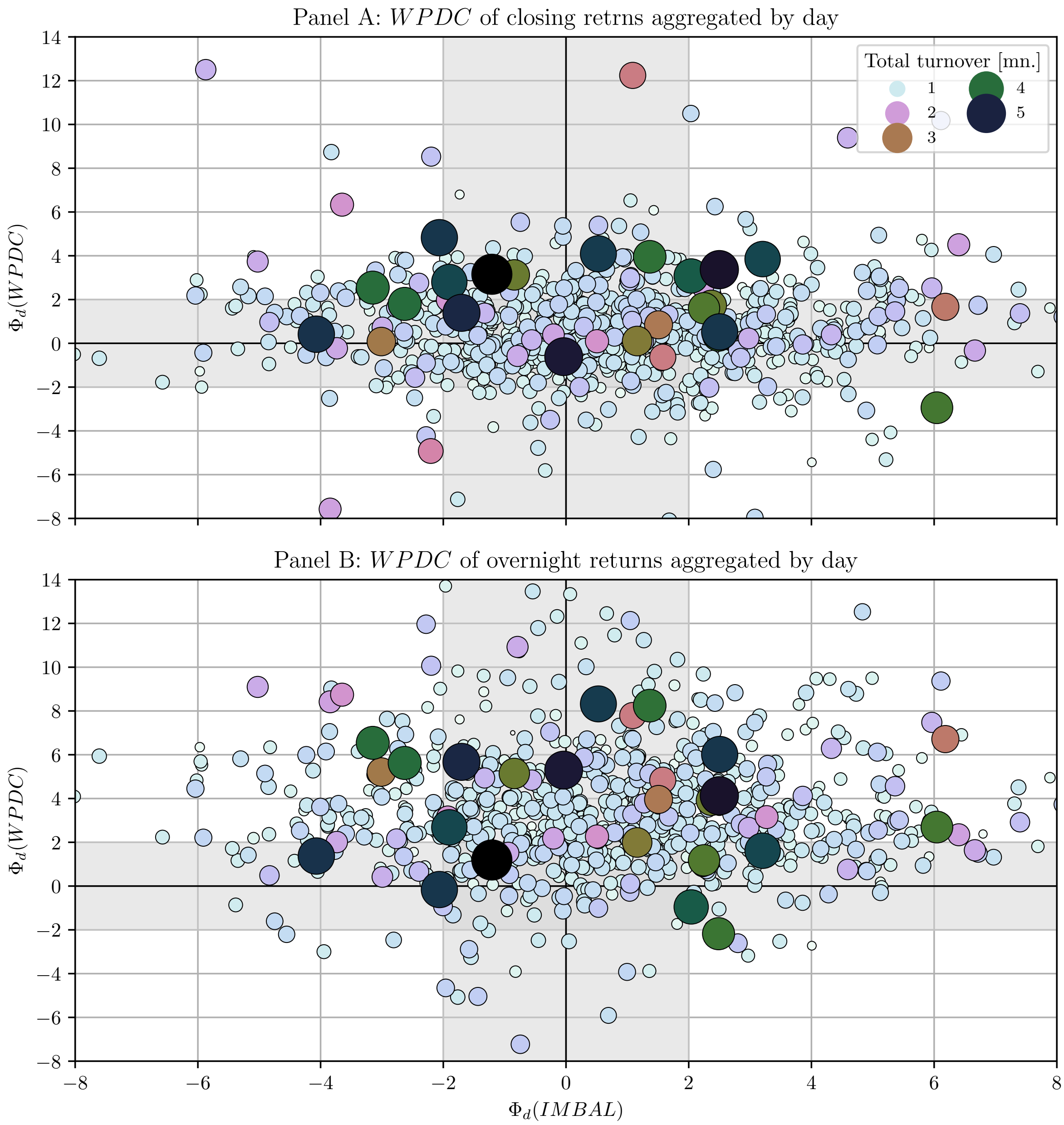


Figure 6: Contribution to price discovery aggregated by day. This figure shows a bubble plot depicting the intersection between order imbalances and weighted price discovery contribution. The horizontal axis represents the order imbalance at the closing auction. The vertical axis represents the weighted price discovery contribution for the closing(overnight) return with respect to the entire trading day, as defined in eq. (10). Both axes have been aggregated for each trading day and normalized following the definition in eq. (12). The gray shaded area mark results below the 5% confidence level of 1.96, rendering observations falling into said interval as insignificantly different from zero. The color as well as the size of the bubbles is indicative of the total turnover of a given day's closing auction across all stocks. Each of the 872 bubbles represents one trading day in the sample.

tance of closing auctions versus the remainder of the trading day. This has reached a point where the closing price is the single most important daily price as it matches the largest volume. Where academic literature has unambiguously shown that the introduction of closing auctions improves all metrics of market quality in the context of an event study, it has not yet examined the composition of order books at close. To address this gap, this paper analyses granular order-level data from SIX Securities & Exchanges (SIX) for 100 of the most liquid equities over a period of 3.5 years. The added value of this analysis lies in the recursive reconstruction of limit order books throughout the day, resulting in full visibility over the final order book in the closing auction, being the main focus of this study.

The first finding in this paper relates to the sensitivity of order books at close. More specifically, closing order books for low-volume stocks are composed quite differently from their high-volume counterparts. This observation is manifested through two main observations. On the one hand, small stocks have significant volume that is beyond the clearing price, i.e. bid(ask) limit orders below(above) said price. In contrast to this, large stocks have limit orders clustered much closer around the actual closing price, such that removal thereof barely affects prices compared to less liquid stocks. On the other hand, the statistical distribution of the market ratio metric also differs materially between large and small stocks. For the large stocks, the ratio of market order volume versus total closing volume is very stable around 0.5 across all auctions whereas for small stocks, this ratio varies anywhere from 0–1 and averages at around 0.39.

The second finding of this paper relates the effects of order book composition onto overnight returns for the same stock over various horizons. The results show that the composition closing order books has no predictive power with respect to the direction of overnight returns. However, this is not true when attempting to predict overnight volatility, approximated by means of absolute returns. Even when controlling for past volatility, trading days with high concentration of market sell orders entail elevated overnight volatility over various time horizons. The same observation cannot be made for the concentration of market buy orders, which are found to have no effect.

The third and final finding in this paper relates to the contribution of closing returns to price discovery with respect to a longer time horizon. The results suggest that the closing auction offers less contribution to price discovery as opposed to overnight returns. This can be explained by closing auctions being a very attractive trading facility to investors who are looking to rebalance their portfolios for regulatory reasons or to track an index, resulting in huge flows of volume that do not contain any information on the fair asset value. Overnight returns on the other hand may serve as a correction for these deviations from the efficient price caused by such shocks during the closing auction.

Overall, these findings further underline the importance of closing auctions with respect to the remainder of the trading day. The finding that the composition of closing order books does not contain any information with respect to the direction of overnight returns as well as closing returns tend to be reverted overnight both indicate that no new information is disseminated during closing auctions. Even if there are certain investors who trade on private information during these auctions, they are sinking in comparison to other investors who trade based on non-informative constraints. Therefore, the closing auction fulfills its purpose to serve as a liquid trading facility for large investors who trade for reasons other than private information. This was also the original intention of Madhavan (1992) when he initially proposed this format almost three decades ago, a time before any presence of high-frequency traders and before the boom in passive investing.

References

Abreu, D., & Brunnermeier, M. K. (2003). Bubbles and crashes. *Econometrica*, *71*(1), 173–204.

Adam, K., Marcet, A., & Nicolini, J. P. (2016). Stock market volatility and learning. *The Journal of Finance*, *71*(1), 33–82.

Aggarwal, R., Inclan, C., & Leal, R. (1999). Volatility in emerging stock markets. *Journal of financial and Quantitative Analysis*, *34*(1), 33–55.

Aitken, M., Chen, H., & Foley, S. (2017). The impact of fragmentation, exchange fees and liquidity provision on market quality. *Journal of Empirical Finance*, *41*, 140–160.

Alles, L. A., & Kling, J. L. (1994). Regularities in the variation of skewness in asset returns. *Journal of financial Research*, *17*(3), 427–438.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The journal of Finance*, *61*(4), 1645–1680.

Baldauf, M., & Mollner, J. (2020). High-frequency trading and market performance. *The Journal of Finance*, *75*(3), 1495–1526.

Baltagi, B. H. (2011). *Econometrics* (5th ed.). Springer.

Barclay, M. J., & Hendershott, T. (2003). Price discovery and trading after hours. *The Review of Financial Studies*, *16*(4), 1041–1073.

Barclay, M. J., & Warner, J. B. (1993). Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics*, *34*(3), 281–305.

Besson, P., & Fernandez, R. (2021). *Better trading at the close thanks to market impact models* (Report). Euronext Quantitative Research.

Biais, B., Foucault, T., & Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial economics*, *116*(2), 292–313.

Bloomfield, R., O'Hara, M., & Saar, G. (2015). Hidden liquidity: Some new light on dark trading. *The Journal of Finance*, *70*(5), 2227–2274.

Bogousslavsky, V., & Muravyev, D. (2020). *Should we use closing prices? institutional price pressure at the close* [Working Paper].

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, *27*(8), 2267–2306.

Brown, D. P., & Zhang, Z. M. (1997). Market orders and market efficiency. *The Journal of Finance*, *52*(1), 277–308.

Budish, E., Cramton, P., & Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, *130*(4), 1547–1621.

Buti, S., Rindi, B., & Werner, I. M. (2017). Dark pool trading strategies, market quality and welfare. *Journal of Financial Economics*, *124*(2), 244–265.

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, *108*(4), 905–939.

Carrion, A. (2013). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, *16*(4), 680–711.

Cartea, A., & Jaimungal, S. (2015). Optimal execution with limit and market orders. *Quantitative Finance*, *15*(8), 1279–1291.

Chan, K., & Fong, W.-M. (2000). Trade size, order imbalance, and the volatility–volume relation. *Journal of Financial Economics*, *57*(2), 247–273.

Chen, J., Hong, H., & Stein, J. C. (2001). Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of financial Economics*, *61*(3), 345–381.

Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial economics*, *65*(1), 111–130.

Chordia, T., Roll, R., & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, *76*(2), 271–292.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, *87*(2), 249–268.

Chordia, T., & Subrahmanyam, A. (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, *72*(3), 485–518.

Chordia, T., & Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance*, *55*(2), 913–935.

Comerton-Forde, C., & Putniņš, T. J. (2011). Measuring closing price manipulation. *Journal of Financial Intermediation*, *20*(2), 135–158.

Comerton-Forde, C., & Putniņš, T. J. (2015). Dark trading and price discovery. *Journal of Financial Economics*, *118*(1), 70–92.

Cordi, N., Félez-Viñas, E., Foley, S., & Putninš, T. (2018). *Closing time: The effects of closing mechanism design on market quality* (tech. rep.). Working paper.

Degryse, H., de Jong, F., & van Kervel, V. (2015). The impact of dark trading and visible fragmentation on market quality. *Review of Finance*, *19*(4), 1587–1622.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, *80*(4), 549–560.

Easley, D., Michayluk, D., O'Hara, M., & Putniņš, T. (2021). The active world of passive investing. *Review of Finance*, *25*(5), 1433–1471.

Ellul, A., Shin, H. S., & Tonks, I. (2005). Opening and closing the market: Evidence from the london stock exchange. *Journal of Financial and Quantitative Analysis*, *40*(4), 779–801.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023–2052.

Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *The review of financial studies*, *18*(4), 1171–1217.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, *19*(1), 3–29.

Frieder, L., & Subrahmanyam, A. (2004). Nonsecular regularities in returns and volume. *Financial Analysts Journal*, *60*(4), 29–34.

Goettler, R. L., Parlour, C. A., & Rajan, U. (2005). Equilibrium in a dynamic limit order market. *The Journal of Finance*, *60*(5), 2149–2192.

Gomber, P., Sagade, S., Theissen, E., Weber, M. C., & Westheide, C. (2017). Competition between equity markets: A review of the consolidation versus fragmentation debate. *Journal of economic surveys*, *31*(3), 792–814.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Griffin, J. M., Harris, J. H., Shu, T., & Topaloglu, S. (2011). Who drove and burst the tech bubble? *The Journal of Finance*, *66*(4), 1251–1290.

Harris, R. D., Coskun Küçüközmen, C., & Yilmaz, F. (2004). Skewness in the conditional distribution of daily equity returns. *Applied Financial Economics*, *14*(3), 195–202.

Haslag, P., & Ringgenberg, M. (2016). *The causal impact of market fragmentation on liquidity*, Working Paper.

Hillion, P., & Suominen, M. (2004). The manipulation of closing prices. *Journal of Financial Markets*, *7*(4), 351–375.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal*, *7*(3), 281–312.

Holden, C. W., & Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance*, *69*(4), 1747–1785.

Huang, J., & Wang, J. (2009). Liquidity and market crashes. *The Review of Financial Studies*, *22*(7), 2607–2643.

Ibikunle, G. (2015). Opening and closing price efficiency: Do financial markets need the call auction? *Journal of International Financial Markets, Institutions and Money*, *34*, 208–227.

Kandel, E., Rindi, B., & Bosetti, L. (2012). The effect of a closing call auction on market quality and trading strategies. *Journal of Financial Intermediation*, *21*(1), 23–49.

Lee, C. M., Ready, M. J., & Seguin, P. J. (1994). Volume, volatility, and new york stock exchange trading halts. *The Journal of Finance*, *49*(1), 183–214.

Louhichi, W. (2011). What drives the volume–volatility relationship on euronext paris? *International Review of Financial Analysis*, *20*(4), 200–206.

Madhavan, A. (1992). Trading mechanisms in securities markets. *The Journal of Finance*, *47*(2), 607–641.

Manahov, V., Hudson, R., & Gebka, B. (2014). Does high frequency trading affect technical analysis and market efficiency? and if so, how? *Journal of International Financial Markets, Institutions and Money*, *28*, 131–157.

McMillan, D. G. (2007). Non-linear forecasting of stock returns: Does volume help? *International Journal of forecasting*, *23*(1), 115–126.

Menkveld, A. J., Yueshen, B. Z., & Zhu, H. (2017). Shades of darkness: A pecking order of trading venues. *Journal of Financial Economics*, *124*(3), 503–534.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation. *Econometrica*, *55*(3), 703–708.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The journal of finance*, *53*(6), 1887–1934.

O'Hara, M., & Ye, M. (2011). Is market fragmentation harming market quality? *Journal of Financial Economics*, *100*(3), 459–474.

Pagano, M. (1989). Trading volume and asset liquidity. *The Quarterly Journal of Economics*, *104*(2), 255–274.

Pagano, M., Peng, L., & Schwartz, R. A. (2013). A call auction's impact on price formation and order routing: Evidence from the nasdaq stock market. *Journal of Financial Markets*, *16*(2), 331–361.

Parlour, C. A. (1998). Price dynamics in limit order markets. *The Review of Financial Studies*, *11*(4), 789–816.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*, 1065–1076.

Pascual, R., & Veredas, D. (2009). What pieces of limit order book information matter in explaining order choice by patient and impatient traders? *Quantitative Finance*, *9*(5), 527–545.

Raillon, F. (2020). The growing importance of the closing auction in share trading volumes. *Journal of Securities Operations & Custody*, *12*(2), 135–152.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*, 832–837.

Roşu, I. (2009). A dynamic model of the limit order book. *The Review of Financial Studies*, *22*(11), 4601–4641.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The journal of finance*, *44*(5), 1115–1153.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, *66*(3), 605–610.

Su, Y.-C., Huang, H.-C., & Lin, S.-F. (2012). Dynamic relations between order imbalance, volatility and return of top gainers. *Applied Economics*, *44*(12), 1509–1519.

Sushko, V., & Turner, G. (2018). The implications of passive investing for securities markets. *BIS Quarterly Review, March*.

Verbeek, M. (2017). *A guide to modern econometrics* (5th ed.). John Wiley & Sons Inc.

Wang, J., & Yang, M. (2015). How well does the weighted price contribution measure price discovery? *Journal of Economic Dynamics and Control*, *55*, 113–129.

Xu, X. E., Chen, P., & Wu, C. (2006). Time and dynamic volume–volatility relation. *Journal of Banking & Finance*, *30*(5), 1535–1558.

Appendix

			Close volume		Continuous volume		Size quantile			
		N	μ	σ	μ	σ	Q_1	Q_2	Q_3	Q_4
1	NESN	873	142.38	107.30	303.80	149.42	0	0	0	873
2	ROG	873	126.77	76.88	278.94	163.37	0	0	0	873
3	NOVN	873	115.24	70.95	258.89	120.02	0	0	0	873
4	UBSG	872	46.39	25.86	131.77	58.34	0	0	0	872
5	ZURN	872	44.74	28.25	110.20	63.32	0	0	0	872
6	ABBN	873	38.80	24.10	98.76	44.29	0	0	0	873
7	CFR	873	37.82	19.37	88.19	46.42	0	0	0	873
8	SREN	873	33.59	25.98	83.35	47.43	0	0	1	872
9	CSGN	873	32.41	19.08	101.52	49.82	0	0	5	868
10	LHN	837	31.52	15.88	74.21	36.68	0	0	0	837
11	LONN	873	29.45	15.21	76.96	40.96	0	0	0	873
12	ALC	556	25.17	26.58	55.60	52.12	0	0	13	543
13	GIVN	873	23.40	12.63	52.00	27.02	0	0	3	870
14	SIKA	763	22.04	14.30	56.80	30.10	0	0	6	757
15	SCMN	873	21.83	11.93	47.80	29.40	0	0	5	868
16	GEBN	873	19.68	10.00	41.18	23.67	0	0	20	853
17	UHR	873	19.45	8.95	48.11	27.59	0	0	120	753
18	PGHN	873	16.69	25.70	38.08	20.71	0	0	127	746
19	SGSN	873	16.29	11.56	34.07	27.19	0	0	103	770
20	SLHN	873	15.96	10.99	51.59	21.90	0	0	109	764
21	ADEN	873	15.34	10.65	32.64	16.90	0	0	229	644
22	SOON	873	14.16	8.05	34.45	19.48	0	0	243	630
23	BAER	873	14.02	11.49	27.77	14.45	0	0	280	593
24	LOGN	873	12.95	25.24	40.69	29.02	0	0	482	391
25	KNIN	873	12.22	9.02	24.48	13.13	0	0	448	425

Continued on next page

			Close volume		Continuous volume		Size quantile			
		N	μ	σ	μ	σ	Q_1	Q_2	Q_3	Q_4
26	TEMN	873	11.35	20.88	29.76	22.66	0	0	561	312
27	STMN	873	10.80	5.91	23.94	14.60	0	0	550	323
28	SCHP	873	9.69	5.31	24.58	13.32	0	0	667	206
29	AMS	873	8.16	5.77	42.14	29.30	1	5	685	182
30	CLN	873	7.64	4.70	21.92	18.23	0	5	761	107
31	VIFN	873	7.27	5.08	22.87	14.34	0	0	821	52
32	BALN	873	7.26	4.70	13.79	6.71	0	2	819	52
33	DUFN	873	7.24	5.70	22.86	15.49	0	24	758	91
34	SPSN	873	7.11	5.13	10.18	7.43	0	2	829	42
35	BARN	873	7.09	7.34	11.87	11.02	0	4	814	55
36	EMSN	873	5.77	7.45	10.25	5.64	0	5	861	7
37	LISP	873	5.71	3.83	11.29	6.90	0	15	846	12
38	PSPN	873	4.79	3.41	8.15	5.82	0	48	823	2
39	VACN	873	4.74	3.29	14.00	7.99	0	38	824	11
40	FI-N	873	4.46	3.66	11.23	6.27	0	59	809	5
41	LISN	873	4.44	3.18	7.00	5.08	0	78	794	1
42	SRCG	813	4.40	6.65	12.46	21.03	94	44	647	28
43	SIGN	686	4.28	6.95	8.49	10.91	113	165	392	16
44	FHZN	873	3.84	2.13	9.21	5.87	0	131	739	3
45	OERL	873	3.18	2.10	8.57	6.41	5	290	574	4
46	TECN	873	3.11	4.28	4.88	3.99	9	496	360	8
47	HELN	873	3.09	1.64	8.04	4.56	1	250	622	0
48	SRAIL	553	2.80	4.28	8.62	17.51	16	324	209	4
49	GALE	873	2.36	2.25	5.67	3.33	3	560	310	0
50	PARG	720	2.31	4.43	3.83	3.57	94	318	305	3
51	CMBN	873	2.28	2.02	5.71	3.27	0	606	267	0

Continued on next page

			Close volume		Continuous volume		Size quantile			
		N	μ	σ	μ	σ	Q_1	Q_2	Q_3	Q_4
52	RO	873	2.23	22.96	10.15	8.92	376	404	90	3
53	SWON	419	2.13	4.27	5.66	8.04	89	261	66	3
54	SCHN	854	2.09	4.29	5.39	3.40	8	705	139	2
55	DOKA	873	2.04	1.46	5.84	4.95	24	575	274	0
56	BUCN	873	1.98	1.74	5.55	3.33	7	671	195	0
57	DKSH	873	1.92	2.51	4.42	3.13	14	684	174	1
58	BCVN	873	1.86	5.33	3.50	2.99	90	635	147	1
59	SUN	873	1.78	1.23	5.42	4.84	20	700	153	0
60	IDIA	873	1.76	1.86	7.55	6.01	62	676	132	3
61	LAND	873	1.69	1.18	5.72	4.33	56	671	146	0
62	UHRN	873	1.65	1.86	5.47	3.66	68	687	118	0
63	ROSE	873	1.64	3.29	8.49	12.00	520	159	189	5
64	ALLN	873	1.54	1.31	2.69	1.80	33	786	54	0
65	BION	873	1.38	5.64	4.98	2.96	208	617	45	3
66	BEAN	873	1.31	3.17	2.78	2.07	314	512	46	1
67	PWTN	515	1.27	1.53	6.37	12.87	116	330	67	2
68	ARYN	873	1.16	1.47	6.57	10.64	397	349	125	2
69	SFZN	873	1.10	1.12	4.42	2.88	247	607	18	1
70	FORN	873	1.04	0.90	3.76	3.57	225	626	22	0
71	GAM	873	0.99	1.16	4.21	5.05	417	349	106	1
72	SFSN	873	0.94	0.58	2.82	1.65	214	649	10	0
73	VONN	873	0.91	0.55	3.06	1.94	233	631	9	0
74	EMMN	873	0.89	0.63	2.97	2.27	305	553	15	0
75	BKW	873	0.87	1.05	2.76	1.99	398	468	7	0
76	MOBN	873	0.79	0.96	1.70	1.22	382	488	3	0
77	DAE	873	0.64	0.47	2.22	1.36	554	315	4	0

Continued on next page

			Close volume		Continuous volume		Size quantile			
		N	μ	σ	μ	σ	Q_1	Q_2	Q_3	Q_4
78	IFCN	873	0.64	0.77	2.37	1.40	556	314	3	0
79	KARN	873	0.63	1.19	1.73	1.12	630	238	5	0
80	UBXN	873	0.58	0.57	3.63	3.74	594	270	9	0
81	CEVA	335	0.56	1.95	3.55	11.23	259	65	10	2
82	INRN	873	0.54	0.85	2.31	1.77	687	183	3	0
83	ALSN	873	0.54	0.58	2.01	1.51	669	200	4	0
84	VATN	873	0.52	0.47	1.22	0.74	679	193	1	0
85	VALN	873	0.51	0.32	2.89	2.24	641	232	0	0
86	BSLN	873	0.51	0.37	2.94	2.48	642	226	5	0
87	KOMN	873	0.51	0.63	2.66	2.43	697	171	5	0
88	MBTN	873	0.50	1.35	5.10	6.27	717	146	9	1
89	BANB	873	0.46	0.76	2.11	2.23	694	177	2	0
90	HUBN	873	0.45	0.50	1.72	1.41	717	154	2	0
91	COTN	873	0.43	0.70	2.64	2.69	758	113	2	0
92	BOSN	873	0.42	0.29	2.29	1.81	765	108	0	0
93	SQN	873	0.37	0.39	3.02	2.51	774	99	0	0
94	CON	832	0.36	0.38	1.48	1.35	769	63	0	0
95	IMPN	873	0.34	0.28	2.31	2.24	769	103	1	0
96	LEON	873	0.34	0.34	2.14	2.93	756	110	7	0
97	AUTN	873	0.33	0.24	2.73	2.46	801	71	1	0
98	MOVE	559	0.33	0.99	1.16	6.26	521	30	7	1
99	SWTQ	873	0.32	0.25	1.49	0.90	837	36	0	0
100	BCHN	873	0.32	0.22	1.47	1.63	800	72	1	0

Table 4: List of all securities in the analysis. This table contains all the 69 equities analyzed. All of the securities were passed through the filters introduced in section 2. The table contains aggregated statistics for each security on a daily basis, including volumes and returns. The last three columns count the number of times a security falls into a size quantile across stocks within days where $Q_1(Q_4)$ represents stocks with the lowest(highest) continuous trading volume.

Panel A: Size quantile 1

	N	μ	σ	1%	5%	10%	25%	50%	75%	90%	95%	99%
$ONROP^{(1)}$	18630	0.11	0.85	-1.86	-1.14	-0.77	-0.30	0.07	0.52	1.02	1.40	2.11
$ONRCL^{(1)}$	21416	0.01	2.42	-6.58	-3.38	-2.33	-1.02	0.00	1.09	2.36	3.44	6.62
$ONROP^{(2)}$	18635	0.16	2.49	-6.18	-3.33	-2.33	-1.00	0.11	1.31	2.69	3.81	7.14
$ONRCL^{(2)}$	21416	0.01	2.42	-6.58	-3.38	-2.33	-1.02	0.00	1.09	2.36	3.44	6.62
IR	18689	-0.13	1.94	-5.47	-3.07	-2.20	-1.09	-0.09	0.84	1.94	2.78	4.96
CR	21445	0.02	0.46	-1.02	-0.51	-0.35	-0.15	0.00	0.19	0.39	0.55	1.11
$DEMVOL$	21444	-0.37	0.83	-3.20	-1.66	-1.27	-0.76	-0.31	0.12	0.51	0.76	1.27
$MR^{(buy)}$	21445	0.43	0.21	0.02	0.11	0.16	0.27	0.42	0.58	0.72	0.80	0.92
$MR^{(sell)}$	21445	0.41	0.21	0.03	0.10	0.15	0.25	0.40	0.56	0.70	0.78	0.90
$IMBAL$	21445	0.02	0.39	-0.88	-0.64	-0.50	-0.25	0.02	0.30	0.54	0.67	0.89

Panel B: Size quantile 2

	N	μ	σ	1%	5%	10%	25%	50%	75%	90%	95%	99%
$ONROP^{(1)}$	19530	0.09	0.70	-1.74	-1.02	-0.69	-0.26	0.08	0.44	0.88	1.20	1.77
$ONRCL^{(1)}$	20928	-0.01	3.29	-5.77	-2.97	-1.99	-0.84	0.00	0.93	1.97	2.82	5.47
$ONROP^{(2)}$	19479	0.10	2.35	-5.47	-2.96	-2.01	-0.81	0.12	1.08	2.20	3.09	5.47
$ONRCL^{(2)}$	20928	-0.01	3.29	-5.77	-2.97	-1.99	-0.84	0.00	0.93	1.97	2.82	5.47
IR	19574	-0.08	1.70	-4.96	-2.67	-1.87	-0.88	-0.05	0.75	1.68	2.41	4.52
CR	20951	0.03	0.38	-0.87	-0.45	-0.31	-0.13	0.03	0.18	0.35	0.50	0.96
$DEMVOL$	20951	0.16	0.63	-1.13	-0.77	-0.58	-0.26	0.11	0.53	0.97	1.27	1.95
$MR^{(buy)}$	20951	0.43	0.18	0.07	0.14	0.20	0.29	0.42	0.56	0.68	0.75	0.87
$MR^{(sell)}$	20951	0.41	0.19	0.05	0.12	0.17	0.27	0.40	0.54	0.67	0.74	0.86
$IMBAL$	20951	0.03	0.34	-0.74	-0.53	-0.41	-0.21	0.03	0.26	0.47	0.59	0.79

Table 5: Descriptive statistics of relevant metrics by size quantile. This table presents the descriptive statistics over the metrics of interest. Each panel represents a distinct set of statistics based on size quantile, where size quantile 1(4) represents small(large) stocks, as defined in section 2.$ONROP^{(1)}(ONROP^{(2)})$ represents overnight returns starting at the closing and ending at the open one(two) days later. The same logic applies to $ONRCL$ with the exception that the interval ends at the closing auction of the respective day. IR and CR represent intraday- and closing returns respectively. $DEMVOL$ represents the logarithmic volume, demeaned within stocks. $MR^{(buy)}/MR^{(sell)}$ and $IMBAL$ represent market ratio and market order imbalances respectively.

Panel C: Size quantile 3

	N	μ	σ	1%	5%	10%	25%	50%	75%	90%	95%	99%
$ONROP^{(1)}$	20226	0.06	1.37	-1.85	-0.99	-0.66	-0.26	0.07	0.41	0.81	1.13	1.80
$ONRCL^{(1)}$	21248	-0.01	2.58	-5.91	-2.93	-1.88	-0.78	0.06	0.88	1.86	2.68	5.32
$ONROP^{(2)}$	20190	0.07	2.55	-5.49	-2.87	-1.90	-0.77	0.12	1.00	2.06	2.90	5.42
$ONRCL^{(2)}$	21248	-0.01	2.58	-5.91	-2.93	-1.88	-0.78	0.06	0.88	1.86	2.68	5.32
IR	20222	-0.05	1.55	-4.61	-2.39	-1.66	-0.76	0.00	0.71	1.54	2.17	3.98
CR	21272	0.02	0.37	-0.85	-0.44	-0.30	-0.13	0.02	0.17	0.34	0.48	0.90
$DEMVOL$	21272	0.14	0.61	-0.93	-0.65	-0.50	-0.26	0.04	0.41	0.89	1.27	2.27
$MR^{(buy)}$	21272	0.45	0.15	0.12	0.21	0.25	0.34	0.44	0.55	0.64	0.71	0.81
$MR^{(sell)}$	21272	0.43	0.15	0.11	0.19	0.24	0.32	0.43	0.53	0.63	0.70	0.80
$IMBAL$	21272	0.02	0.26	-0.59	-0.41	-0.31	-0.16	0.02	0.19	0.35	0.45	0.63

Panel D: Size quantile 4

	N	μ	σ	1%	5%	10%	25%	50%	75%	90%	95%	99%
$ONROP^{(1)}$	19871	0.04	0.80	-2.16	-1.03	-0.69	-0.28	0.07	0.41	0.80	1.11	1.89
$ONRCL^{(1)}$	20700	0.02	1.75	-5.21	-2.63	-1.72	-0.72	0.08	0.84	1.73	2.41	4.51
$ONROP^{(2)}$	19853	0.08	1.75	-4.88	-2.62	-1.79	-0.74	0.14	0.99	1.91	2.64	4.49
$ONRCL^{(2)}$	20700	0.02	1.75	-5.21	-2.63	-1.72	-0.72	0.08	0.84	1.73	2.41	4.51
IR	19861	-0.00	1.37	-3.85	-2.10	-1.45	-0.67	0.03	0.70	1.43	2.01	3.47
CR	20724	0.01	0.31	-0.81	-0.41	-0.28	-0.13	0.01	0.14	0.29	0.40	0.83
$DEMVOL$	20724	0.08	0.48	-0.81	-0.55	-0.42	-0.21	0.03	0.30	0.60	0.83	1.68
$MR^{(buy)}$	20724	0.47	0.13	0.19	0.27	0.31	0.39	0.47	0.56	0.64	0.69	0.77
$MR^{(sell)}$	20724	0.47	0.13	0.18	0.26	0.30	0.38	0.46	0.55	0.63	0.68	0.77
$IMBAL$	20724	0.01	0.20	-0.47	-0.32	-0.25	-0.12	0.01	0.15	0.27	0.34	0.48

Table 5: (Continued)

Control variables only without fixed effects

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.0183	0.0042	0.0042	0.0022
Effects	None	None	None	None
$constant$	0.0770***	0.0263***	0.1053***	0.0188*
	(27.802)	(4.0303)	(14.327)	(1.6684)
LCR	0.0044	0.0922***	0.0846***	0.1597***
	(0.3569)	(3.1832)	(2.6392)	(3.7291)
$LONROP^{(1)}$	0.0168**	0.0227	0.0505***	0.0402
	(2.1357)	(1.3117)	(3.6042)	(1.6315)
IR	-0.0066**	0.0279***	0.0186**	0.0559***
	(-2.0104)	(3.9895)	(2.2496)	(4.8901)
CR	-0.3008***	-0.2892***	-0.3455***	-0.2456***
	(-15.927)	(-8.1582)	(-7.5294)	(-4.5792)
$DEMVOL$	-0.0044	-0.0364***	-0.0288**	-0.0823***
	(-0.8916)	(-2.9841)	(-2.0893)	(-4.1248)

Table 6: Regression results of overnight returns without fixed effects. This table shows the regressions of overnight returns over multiple time horizons as defined in eq. (3) (without the fixed effects) onto other variables that are listed as follows. LCR represents the previous' day closing return whereas $LONROP^{(1)}$ stands for the return between the previous day's close and the same day opening price. IR and CR represent intraday- and closing return on the same trading day respectively. $DEMVOL$ represents the logarithmic volume, demeaned within stocks. $MR^{(buy)}/MR^{(sell)}$ and $IMBAL$ represent the market buy/sell ratios and market order imbalances respectively. Reported standard errors are derived using Driscoll-Kraay covariance matrices. *, ** and *** denote significance at the 1%, 5% and 10% level respectively.

Panel A: Control variables only

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.3140	0.2765	0.2589	0.2108
Effects	Day	Day	Day	Day
LCR	-0.0702*	-0.0207	-0.0535	-0.0373
	(-1.8595)	(-0.5129)	(-1.0565)	(-0.6620)
$LONROP^{(1)}$	0.0429**	0.0477***	0.0738***	0.0533**
	(2.4906)	(3.2494)	(5.0409)	(2.1111)
IR	-0.0072*	0.0240**	0.0126	0.0303**
	(-1.6631)	(2.2135)	(1.1775)	(2.0180)
IR	-0.0072*	0.0240**	0.0126	0.0303**
	(-1.6631)	(2.2135)	(1.1775)	(2.0180)
$DEMVOL$	-0.0328**	-0.0484**	-0.0709***	-0.1102***
	(-2.1012)	(-2.2700)	(-2.8706)	(-3.1699)

Panel B: Market buy- and sell ratios

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.3168	0.2768	0.2594	0.2110
Effects	Day	Day	Day	Day
Controls	Yes	Yes	Yes	Yes
$MR^{(buy)}$	-0.1959***	-0.1096*	-0.2414***	-0.1632**
	(-9.6340)	(-1.7959)	(-4.1735)	(-2.1109)
$MR^{(sell)}$	0.1614***	0.1427**	0.1870**	0.2588**
	(4.3857)	(2.1360)	(2.5416)	(2.5615)

Panel C: Market order imbalance

Dep. Variable	$ONROP^{(1)}$	$ONRCL^{(1)}$	$ONROP^{(2)}$	$ONRCL^{(2)}$
N	73969	73969	70601	73880
R-squared	0.3163	0.2767	0.2591	0.2109
Effects	Day	Day	Day	Day
Controls	Yes	Yes	Yes	Yes
$IMBAL$	-0.1406***	-0.0788**	-0.1439***	-0.1406***
	(-6.6586)	(-2.1106)	(-3.3820)	(-2.7422)

Table 7: Reduced regression results of overnight returns. This table shows the regressions of overnight returns over multiple time horizons as defined in eq. (3) onto other variables that are listed as follows. LCR represents the previous' day closing return whereas $LONROP^{(1)}$ stands for the return between the previous day's close and the same day opening price. IR represents intraday returns on the same trading day. $DEMVOL$ represents the logarithmic volume, demeaned within stocks. $MR^{(buy)}/MR^{(sell)}$ and $IMBAL$ represent the market buy/sell ratios and market order imbalances respectively. All panel models were estimated using day-fixed effects. Reported standard errors are derived using Driscoll-Kraay covariance matrices. *, ** and *** denote significance at the 1%, 5% and 10% level respectively.

Curriculum Vitae

Louis Robin Müller

Born: 19.11.1993
Seestrasse 50, 8942 Oberrieden, Switzerland muellerlouis@outlook.com | +41 79 947 66 26

EXPERIENCE

SIGNA Group of Companies, *Zurich, Switzerland*
Corporate Finance Associate *Mar. 2022 – present*

- Structured finance and M&A transactions in prime real estate and luxury retail

Trium Corporate Finance AG, *Zug, Switzerland*
Senior M&A Advisor (part-time) *Jul. 2016 – Feb. 2022*

- Advised family offices based in Europe, Russia and the Middle East on small-cap M&A transactions with enterprise values of around USD 25m on both buy- and sell-side

Institute for Operations Research and Computational Finance, *St. Gallen, Switzerland*
Research Associate (part-time) *Aug. 2018 – Dec. 2021*

- Took the lead in a multi-year research cooperation between the University of St. Gallen and SIX Swiss Exchange (SIX) for the quantitative analysis of their competitiveness among European trading platforms

Swiss Armed Forces, *Switzerland*
Non-commissioned officer *Oct. 2012 – Apr. 2013*

- Instructor for chemical, biological, radioactive and nuclear threats, ranked as Corporal.

EDUCATION

University of St. Gallen — School of Finance, *St. Gallen, Switzerland*
Ph.D. in Finance *Aug. 2018 – Dec. 2021*

University of St. Gallen — School of Finance, *St. Gallen, Switzerland*
Master of Arts in Banking and Finance (MBF) — GPA: 5.71/6 *Sep. 2016 – Jul. 2018*

Hong Kong University of Science and Technology — Business School, *Hong Kong, China*
Exchange Semester *Jan. 2018 – Jul. 2018*

University of St. Gallen, *St. Gallen, Switzerland*
Bachelor of Arts in Business Administration — GPA: 5.59/6 *Sep. 2013 – May. 2016*

University of Southern California — Marshall School of Business, *Los Angeles, United States*
Exchange Semester *Aug. 2015 – Dec. 2015*

SKILLS & INTERESTS

CFA: Passed Level I
Languages: Swiss German/German (native), English (proficient), French (conversational)
IT/Programming: MS Office, VBA, Python, R, SQL
Interests: Personal investing, photography, diving, surfing, golf