Essays on trading and informational efficiency

Jiayuan Chen
ORCID: 0009-0004-8502-3988

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

Department of Finance
Faculty of Business and Economics
University of Melbourne
Australia

October 2024

Abstract

This thesis delves into the interplay between informed and uninformed trading activities, market structure, and their effects on informational efficiency. Financial markets function as a hub of collective wisdom, dynamically adjusting asset prices to reflect available information. Yet, the real-world mechanics of information assimilation and its implications for informational efficiency are intricate, involving diverse market participants and dynamic market structures. By focusing on informed trading activities such as short-selling and regulatory changes affecting European equity markets, this thesis helps shed light on how information is incorporated into prices and how market structure influences informational efficiency.

The first essay examines the impact of Exchange-Traded Fund (ETF) shorting activity on future returns and the informational efficiency of ETF's underlying securities. I document that ETF shorting activity negatively predicts future ETF returns, especially during periods of macroeconomic news releases, and future returns of ETF's underlying securities. This return predictability supports that ETF shorting activity is informed about market-wide information. I further corroborate that ETF shorting activity helps improve the informational efficiency of ETFs' underlying securities by helping incorporate market-wide information into stock prices rather than firm-specific information. Collectively, my results substantiate that informed traders short ETFs to speculate about their negative market-wide information, thereby improving the informational efficiency of ETFs' underlying securities.

The second essay distinguishes between liquidity-demanding and liquidity-supplying shorts and investigates their informative roles in the financial market. I document that liquidity-demanding shorts have predictive power over future stock returns and earnings surprises, whereas liquidity-supplying shorts have no such power. I find that liquidity-demanding shorts are associated with increased information acquisition and fewer hedge fund holdings, while liquidity-supplying shorts do not exhibit such an association. I show that liquidity-demanding shorts are associated with enhanced informational efficiency around negative earnings announcements, while liquidity-supplying shorts are associated with enhanced informational efficiency around positive earnings announcements. These findings collectively support that liquidity-demanding shorts are informed, while liquidity-supplying shorts are uninformed.

The third essay studies the recent European regulatory reforms that have elevated the role of Systematic Internalizers (SI) in European equity markets. We analyze the impact of SI sub-tick trading and SI public quote disclosure on market quality. We demonstrate that banning SI sub-tick trading increases competition for liquidity provision, as evidenced by narrower spreads and increased depth. We also find that this ban reduces the price-improvement benefits traders receive and undermines informational efficiency when the spread is constrained by the tick size. Mandating SI to provide more public quotes improves the lit market liquidity and informational efficiency but escalates the adverse selection costs borne by SI.

Declaration

I declare that

i this thesis comprises only my original work towards the PhD except where indicated in the Preface;

ii due acknowledgement has been made in the text to all other material used; and

iii this thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Jiayuan Chen

October 23, 2024

Preface

Chapter 2 and 3 comprise only my own original unpublished work.

Chapter 4 is an original unpublished work sponsored by the Plato Partnership and in collaboration with Carole Comerton-Forde. Carole Comerton-Forde (carole.comerton-forde@unimelb.edu.au) is at the Department of Finance, the University of Melbourne.

All co-authorships are in accordance with the Graduate Research Training Policy of the University of Melbourne.

I acknowledge fundings from the Australian Government Research Training Program and the Faculty of Business and Economics Graduate Research Abroad Travelling Scholarship.

Acknowledgements

I extend my deepest gratitude to the distinguished members of my supervisory committee: Carole Comerton-Forde, Bryan Lim, Garry Twite, and Qi Zeng, for their sage guidance, unwavering support, and invaluable advice throughout my doctoral odyssey. Their profound expertise and patience have been pivotal in sculpting my research and achieving my scholarly ambitions. This dissertation stands as a testament to their monumental influence and encouragement.

Furthermore, I owe a debt of gratitude to all faculty members of the Department of Finance at the University of Melbourne and the School of Banking and Finance at UNSW Business School. Their readiness to assist, coupled with their generous intellectual contributions and constructive critiques, have significantly enriched my academic endeavor. A heartfelt acknowledgement to Zhong Zhuo, James Brugler, Minsoo Kim, Peter Pham, and Breno Schmidt, whose exceptional academic mentorship and steadfast encouragement have been instrumental in my development.

I am equally thankful for the camaraderie and insights provided by all my fellow PhD students, especially my classmates Lei Chen, Sheng Huang, Ji Hyun Tak, Chunxiao Lu, Haoyi Luo, Fanqi Meng, Roham Rezaei, Yang Shi, Jiamu Sun, Chenfei Sun, Shuying Wu, Bochen Wu and Yuyang Zhang. This journey has been made significantly lighter and more fulfilling through your friendship and support.

My sincere appreciation also extends to the professional staff in the Department of Finance, whose behind-the-scenes efforts have created a conducive and productive research environment. Additionally, my profound thanks to the Department of Finance for their financial backing and comprehensive academic resources, which have been indispensable to my research pursuits.

I must express my deepest affection and gratitude to my parents, whose unconditional love and support have been the bedrock of my resilience and determination. And to my partner and best friend, Wen Yang, your presence in my life has been the greatest blessing. Embarking on and completing this PhD journey would have been unimaginable without you by my side.

Contents

List of Tables

Chapter 1

Introduction

In the era of big data, the increasing volume and complexity of information pose new challenges for the financial markets. This expansion complicates the process of information incorporation in the financial markets, where asset prices continuously adjust through trading to reflect information from diverse market participants. Central to this dynamic is the efficient market hypothesis, which posits that asset prices incorporate all available information. However, the practical process of how information is incorporated into prices and how information impacts the functioning of markets is complex and multifaceted. It involves various market participants, trading activities, and regulatory reforms, each adding to the market's intricacy and the overall price discovery process. Given the pivotal role that information incorporation plays in shaping market dynamics and informational efficiency,[1] a comprehensive exploration of this area is both timely and highly relevant. By exploring these three distinct but interconnected topics, this dissertation aims to shed light on the ongoing debate regarding how trading affects informational efficiency within financial markets. Furthermore, the findings will provide regulatory bodies, market participants, and academics with a deeper understanding of how trading and regulatory changes impact informational efficiency and the price discovery process, facilitating informed decision-making and potential market improvements.

[1]In this dissertation, I use the term "informational efficiency". However, it's worth noting that the literature also uses terms such as "pricing efficiency" and "information efficiency" to describe this concept.

In the first and second essays, I concentrate on a form of informed trading widely documented in literature: short selling. The first essay explores the increasing shorting activity in U.S. ETF markets. I explore why investors short ETFs and examine how this ETF shorting activity impacts on future returns and informational efficiency. By adopting the standard portfolio sorting and Fama-MacBeth regression methods, I find that the ETF shorting activity negatively predicts future returns of ETFs and, notably, this negative return predictability emerges during periods of macroeconomic news releases but not during non-macroeconomic news periods. These distinct results support that investors short ETFs to exploit their negative market-wide information, which occurs most often during macroeconomic news releases period. To reinforce this argument, I investigate the informational efficiency of the underlying securities of ETFs around the earnings announcements. Following Glosten, Nallareddy, and Zou (2021), I decompose future earnings information into two categories: the market-wide earnings information and firm-specific earnings information. The results demonstrate that ETF shorting activity helps incorporate more market-wide future earnings information into current period returns and improves the informational efficiency of the underlying securities. By exploiting the exogenous variation in ETF shorting activity introduced by the U.S. Securities and Exchange Commission (SEC) Fund of Fund arrangement, I confirm that the association between ETF shorting activity and informational efficiency is causal. Therefore, I conclude that investors short ETFs to exploit their negative market-wide information, which is subsequently transmitted in ETF's underlying securities and improves their informational efficiency.

The second essay explores the often-overlooked heterogeneity in informativeness among shorting activity in U.S. equity markets, challenging the conventional view that perceives short-selling as a uniform practice. Following Comerton-Forde, Jones, and Putniņš (2016b),I distinguish between liquidity-demanding shorts (LD) — defined as seller-initiated short volume — and liquidity-supplying shorts (LS) — defined as buyer-initiated short volume.[2] The analysis begins with an assessment of the informativeness of LD and LS through their return predictability. The findings reveal a stark contrast: LD are negatively predictive of future returns, highlighting their informed

[2]Whether a short is initiated by the seller or the buyer is determined using the Lee and Ready (1991) algorithm.

nature regarding future stock returns, while LS exhibit no such predictability, suggesting they are not informed. To further substantiate this argument, I explore the earnings predictability of LD and LS. Results confirm that LD negatively predict future earnings surprises and post-announcement Cumulative Abnormal Returns (CAR), whereas LS lack this predictive capability. Additionally, I find that hedge funds tend to avoid holding the stock with high LD.[3] The investigation extends to information acquisition patterns, using SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) data as a proxy for the information acquisition, I find that LD are positively associated with the information acquisition, reinforcing the informed nature of LD.[4] Collectively, these findings support that liquidity-demanding shorts are informed, while liquidity-supplying shorts are uninformed.

The third essay, co-authored, studies a popular trading platform: Single-Dealer Platforms (SDP), focusing on their proliferation as Systematic Internalizers (SI) in European equity markets following the introduction of a major regulatory reform - Markets in Financial Instruments Directive II (MiFID II). Our study concentrates on the two key features of SI: (i) SI sub-tick trading and (ii) the SI pre-trade transparency and analyzes their respective impacts on market quality. We observe that banning SI sub-tick trading improves liquidity, as indicated by reduced European Best Bid and Offer (EBBO) spreads in lit markets and narrower quoted spreads of SI public quotes. Conceptually, banning SI sub-tick trading mechanically prevents SI from providing sub-tick price improvement and lowering the overall price improvement. We find consistent evidence that banning SI sub-tick trading reduces the price-improvement benefits traders receive on SI. Moreover, we find that banning SI sub-tick undermines informational efficiency, supporting that banning SI sub-tick trading prevents investors from capturing within-tick valuations. Lastly, our study reveals that enforcing higher SI pre-trade transparency diverts informed traders from lit markets to SI.[5] We find that mandating SI to provide more public quotes results in lit markets less toxic while simultaneously makes SI more toxic, as evidenced by smaller (larger) spread, price impact and relative contribution to the price discovery process in lit markets (SI).

[3] I find no association between LS and hedge fund holdings

[4] I find no association between LS and the information acquisition

[5] Zhu (2014) demonstrate that informed traders primarily reside in lit markets.

Chapter 2

ETF shorting activity, return predictability and informational efficiency of underlying securities

2.1 Introduction

Shorting activity in U.S. Exchange Traded Fund (ETF) markets is high. Over the last two decades, the average short interest in ETFs represents 14% of ETF market capitalization, significantly exceeding the average short interest at the stock level (approximately 4% to 5% of market capitalization).[1] This level of shorting activity highlights the importance of understanding its effects on financial markets. Despite the growing significance of ETF shorting activity, there is a notable lack of literature that explores its effects. This chapter endeavors to bridge this gap by studying the effect of ETF shorting activity on future returns and the informational efficiency of ETFs' underlying securities.

Unlike purchasing ETFs, shorting ETFs incurs significant costs due to the high borrowing costs. The median rebate rate of ETFs is 300 basis points, which is eight times higher than the median rebate rate of ETFs' underlying securities (Bhojraj and Zhao

[1] I obtain short interest from Compustat and calculate the aggregate ETF short interest by aggregating all ETFs' short interest for each reporting period.

(2021)). As a result, the high borrowing costs discourage uninformed traders from participating in ETF shorting activity, leading to a high proportion of informed traders in ETF short-selling markets.[2] Considering the substantial costs associated with ETF shorting activity, one would expect that such shorting activity primarily comes from informed traders, making it informed.[3]

I begin by examining whether ETF shorting activity is informed in predicting future ETF returns. Shorting ETFs is often undertaken to bet against future market performance, and I label this type of shorting activity as *bearish-market shorting*. Existing research substantiates that traders take short positions to express their bearish view, as evidenced by the negative association between stock-level shorting activity and future stock returns.[4] In the same spirit, one would expect that ETF shorting activity negatively predicts future ETF returns when ETF shorting activity is mainly driven by bearish-market shorting. However, recent research by Evans, Moussawi, Pagano, and Sedunov (2022) finds that ETF shorting activity driven by liquidity suppliers lacks predictive power for future returns of ETF Net Asset Value (NAV). Additionally, Huang et al. (2021) argue that traders with positive information short industry ETFs to hedge against industry risks and find that industry ETF shorting activity positively predicts future ETF returns. Therefore, the direction — positive or negative — of return predictability depends on the strength of the predictive power of different ETF shorting activities.

I study the return predictability of the U.S. equity ETF shorting activity by employing standard portfolio sorting. I measure ETF shorting activity using the change in ETF short interest ratio. I find that a value-weighted long-short strategy based on ETF

[2]Diamond and Verrecchia (1987) show that, when the shorting cost is high, uninformed short sellers abstain from short selling. Boehmer, Huszár, Wang, Zhang, and Zhang (2022) find that shorts have stronger predictive power when shorting cost is higher.

[3]I refer ETF shorting activity as informed if it contains information about the ETF's underlying securities. Huang, O'Hara, and Zhong (2021) and Li and Zhu (2022) corroborate that ETF shorting activity is informed in predicting future ETF returns. Hasbrouck (2003) study the price discovery between ETF and ETF future contract and find that both of them contribute to the price discovery. Ernst (2021) show that ETFs contribute to the price discovery of ETF underlying securities. Bernile, Hu, and Tang (2016) report abnormal increases in ETF order imbalance prior to surprises in macroeconomic news (i.e., Federal Open Market Committee interest rate decisions, non-farm payrolls, consumer price index, gross domestic product), which also predict future ETF returns.

[4]See Boehmer, Jones, and Zhang (2008), Diether, Lee, and Werner (2009), Boehmer, Huszar, and Jordan (2010), and Rapach, Ringgenberg, and Zhou (2016)

shorting activity generates a negative excess return and a negative Fama and French (1993) and Carhart (1997) four-factor alpha.[5] To control for well-known return predictors, I estimate the fixed effects Ordinary Least Square (OLS) regressions of ETF future returns on ETF shorting activity and return predictors. I find that the negative return predictability of ETF shorting activity remains robust. The negative return predictability aligns with previous literature that short sellers possess superior and negative information about future returns. These results also suggest that ETF shorting activity, on average, carries negative market-wide information and traders mainly short ETFs to express their bearish view on the market (i.e., bearish-market shorting).

To reinforce the idea that bearish-market shorting mainly drives the negative return predictability of ETF shorting activity, I restrict my sample to market ETFs and find that the negative return predictability remains robust. I further analyze the return predictability of market ETF shorting activity on both macroeconomic news periods and non-macroeconomic news periods.[6] My results show that ETF shorting activity negatively predicts future ETF returns on macroeconomic news periods but not on non-macroeconomic news periods. This distinction in return predictability supports that ETF shorting activity mainly reflects the negative market-wide information, thereby contributing to the negative return predictability of ETF shorting activity.

The negative information in ETF shorting activity will eventually be transmitted to ETFs' underlying securities. This information transmission between ETFs and their underlying securities is assured by ETF arbitrage.[7] In some sense, shorting ETFs is equivalent to shorting on the ETFs' underlying securities. To capture this equivalent stock-level shorting activity of ETF underlying securities, I construct a measure called "ETF-based shorting activity" ($\textit{ETF-based } SA_{i,t}$) that translates the shorting activity in ETFs to its equivalent shorting activity on ETFs' underlying securities. In other words, $\textit{ETF-based } SA_{i,t}$ is the manifestation of ETF shorting activity at the stock level. $\textit{ETF-based } SA_{i,t}$

[5] The results are robust to the equally weighted long-short portfolio.

[6] I focus on ten major macroeconomic news events. For each bi-weekly short interest reporting period, I compute the number of macroeconomic news released within the bi-weekly period. I identify a bi-weekly period as a macroeconomic (non-macroeconomic) news period if the number of macroeconomic news released within the bi-weekly period is greater (less) than the 75 (25) percentile of the sample.

[7] ETF trading plays an important role in the price discovery of underlying securities (See, e.g., Hasbrouck (2003), Yu (2005), Fang and Sanger (2011), Ivanov, Jones, and Zaima (2013), Bhojraj, Mohanram, and Zhang (2020), Glosten et al. (2021) and Li and Zhu (2022))

captures the aggregate equivalent shorting activity from all ETFs that contain the stock i. With the information transmission between ETFs and their underlying securities, I expect that *ETF-based* $SA_{i,t}$ is also informed and, therefore, predicts future stock returns. I examine the return predictability of *ETF-based SA* at the stock level using the standard portfolio sorting and panel regressions. I find that *ETF-based SA* negatively predicts future stock returns in both approaches, showing that this *ETF-based SA* reflects negative information about future stock returns.

As *ETF-based SA* is informed, I hypothesize that an increase in *ETF-based SA* results in stock prices becoming more informationally efficient and enhances the informational efficiency of ETFs' underlying securities. I test this hypothesis by examining the association between informational efficiency and *ETF-based SA*. I focus on five informational efficiency proxies, including Future Earnings Response Coefficient (FERC), Weller (2018)'s jump ratio, autocorrelation, variance ratio and post-earnings announcement drift (PEAD). My results show that *ETF-based SA* helps incorporate future earnings information into current stock returns (i.e., a larger FERC) and is associated with a smaller jump ratio, autocorrelation and variance ratio and an attenuated PEAD. These results indicate that *ETF-based SA* increases the informational efficiency of ETFs' underlying securities.

Upon considering the improvement of informational efficiency, the question arises as to what information ETF shorting activity carries that makes the prices of ETFs' underlying securities more informationally efficient. To explore this question, I distinguish between two types of information: market-wide information versus firm-specific information.[8] I decompose the quarterly earnings into earnings related to market-wide information and earnings related to firm-specific information following Glosten et al. (2021) and Israeli, Lee, and Sridharan (2017). I find that the earnings related to market-wide information explain the improvement effect of *ETF-based SA* on informational efficiency but not the earnings related to firm-specific information, showing that *ETF-based SA* helps incorporate market-wide earnings information into stock returns but not firm-specific earnings information. These results support that ETF shorting

[8]Market-wide information refers to data or news that impacts the entire financial market or a broad range of assets, rather than specific individual securities. Some examples of market-wide information are Macroeconomic indicators, Central Bank policies, technological advancements and regulatory changes.

activity helps improve informational efficiency by enabling traders to price in market-wide information.

To draw causal inferences on the association between ETF shorting activity and informational efficiency, I exploit the exogenous variation in ETF shorting activity driven by the implementation of the U.S. Securities and Exchange Commission (SEC) Fund of Funds (FOF) arrangement. The SEC FOF arrangement eases the short-sale constraints on ETFs by increasing the supply of lendable ETF shares, and I expect that the SEC FOF arrangement will boost ETF shorting activity. I first confirm that the ETF shorting activity increases by 0.39% and ETF short flow increases by 6.28% after the SEC FOF arrangement. I subsequently explore this increased ETF shorting activity associated with the SEC FOF arrangement by examining its impact on informational efficiency. I find consistent results that ETF shorting activity enhances the informational efficiency of ETFs' underlying securities by facilitating the incorporation of market-wide information.

Although my results are consistent with bearish-market shorting improving informational efficiency, other alternative explanations are possible. First, hedging-motivated shorting could impact informational efficiency, as shown by Huang et al. (2021). However, I show that the improvement of informational efficiency is only present in stocks with low hedging demand but not in stocks with high hedging demand,[9] suggesting that the hedging component of ETF shorting activity is unlikely to explain the improvement of informational efficiency. Second, the improvement might result from the shorting activity that arises from the market makers' liquidity provision, where Evans et al. (2022) refer to as operational shorting. I compute the ETF shorting activity net of ETF operational shorting activity and re-estimate the results.[10] I find my results remain unchanged after filtering out ETF operational shorting. Third, the enhanced informational efficiency might originate from ETF activity rather than ETF shorting activity (Glosten et al. (2021)). I construct orthogonal ETF activity and show that ETF shorting activity continues to enhance informational efficiency after controlling

[9]I use CAPM beta to proxy for hedging demand. A high beta stock strongly co-moves with the market and has a high hedging demand.

[10]The ETF operational shorting activity measure is estimated using Evans et al. (2022) approach using the intraday ETF order imbalance and ETF share outstanding.

for orthogonal ETF activity. Finally, the improvement in informational efficiency may be driven by short sellers who trade on firm-specific information and short ETFs to circumvent short-sale constraints (Li and Zhu (2022)). I study the variation of short-sale constraints in the SEC Regulation SHO pilot program and find no differences between the effect of ETF shorting activity on stocks with and without short-sale constraints, indicating the improvement in informational efficiency is not driven by shorting that exploits firm-specific information.

This study contributes to the literature on short-selling and information content. Empirical evidence on short-selling shows that short sellers are informed and their shorting activity predicts future returns (Chen, Zhu, and Chang (2019a) and Diether et al. (2009)). A typical interpretation of this predictability is that shorting activity is a proxy for the amount of negative firm-specific information excluded from the market price (Figlewski (1981)).[11] Given short sellers are informed, Boehmer and Wu (2013) and Saffi and Sigurdsson (2011) show that short-selling contributes to better firm-specific informational efficiency. My results extend this literature by showing that, in addition to firm-specific information (i.e., Li and Zhu (2022)), short sellers are also informed about market-wide information documented by the negative return predictability of ETF shorting activity on macroeconomic news periods and their ability to facilitate the incorporation of market-wide information. My results also add to the literature on whether ETF trading contributes to the price discovery in ETF markets and the price discovery of the underlying securities by showing that ETF shorting activity is informed in predicting future returns of ETFs and their underlying securities.

This study adds to the understanding of how shorting activity via ETFs contributes to the informational efficiency of the underlying securities. Huang et al. (2021) show that industry ETFs help improve informational efficiency by facilitating industry risk hedging for informed traders. Glosten et al. (2021) document a positive association between ETF activity and the short-run systematic informational efficiency for stocks with weak information environments. Israeli et al. (2017) show that ETF activity un-

[11]Numerous studies have found that firms with high levels of short-selling activity tend to experience lower future returns, subsequent financial misconduct, and credit rating downgrades, as well as downgrades by financial analysts.(See Chen et al. (2019a),Boehmer et al. (2008) and Diether et al. (2009), Karpoff and Lou (2010), Henry and Koski (2010), Christophe, Ferri, and Hsieh (2010) and Hou, Meng, Zhang, and Chan (2019))

dermines long-run pricing efficiency. Ben-David, Franzoni, and Moussawi (2018) find that ETFs increase the intraday quote spread and volatility of their underlying securities. Da and Shive (2018) document that ETF arbitrage activity is positively associated with return co-movement of their underlying securities. I complement their findings by providing additional evidence on how ETF enhances informational efficiency by allowing traders to incorporate negative market-wide information.

This study also contributes to the long-standing market timing literature. Prior research, including Busse (1999), Chen and Liang (2007) and Cao, Chen, Liang, and Lo (2013), document hedge funds' ability to time the market across various dimensions. Rapach et al. (2016) find that the aggregate short interest ratio negatively predicts future stock market returns, which is driven by short sellers' ability to anticipate future aggregate cash flows. This research complements this literature by demonstrating that ETF short sellers also possess a market-timing ability. This ability is evidenced by ETF short sellers' capacity to predict future ETF returns, particularly on macroeconomic news release periods.

2.2 Data and Measures

The ETF sample contains all U.S. domestic equity ETFs. I obtain the ETF sample from the Center for Research in Security Prices (CRSP) mutual fund database.[12] I extract the ETF holding data from the Thomson Reuters Mutual Fund holdings database (S12) and supplement it with the CRSP Mutual Fund holdings database. The sample period spans from January, 2002 to December, 2021. For each ETF and stock (i.e., ETF constituents), I obtain bi-weekly short interest data from the Compustat supplemental short interest file, and I obtain market capitalization, share price and industry classifi-

[12] I begin with mutual funds from the Center for Research in Security Prices (CRSP) mutual fund database with the ETF indicator ($etf\_flag = F$). Next, I intersect the CRSP mutual fund sample with securities from the CRSP daily stock file with a share code of 73. To isolate U.S. equity ETFs, I use the Lipper classification code along with parsing of fund names. Following Zhu (2020) and Li and Zhu (2022), I identify a fund as a U.S. equity ETF when the fund's Lipper classification code falls into one of the following categories: EIEI, G, LCCE, LCGE, LCVE, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE, SCVE, SP, SPSP. I also manually filter out ETFs by excluding non-U.S. and non-equity funds whose name contains non-U.S. region character. Following Li and Zhu (2022). I further refine the ETF sample by excluding ETFs containing fewer than 25 common stocks.

cation from the CRSP dataset.[13] For each stock, I obtain quarterly financial statement data from the Compustat dataset. In line with Shumway (1997), I replace the missing delisting return with -30%. To mitigate market microstructure noise, I exclude the observations with a bi-weekly closing price of less than \$2. Finally, I winsorize all variables at the 1% and 99% levels to alleviate the effects of outliers. The final ETF sample contains 598 equity ETFs and 256,064 ETF-bi-week observations. The final stock sample consists of 1,371,251 stock-bi-week observations and 270,984 stock-quarter observations.

The primary variable in my analyses is the stock-level bi-weekly ETF-based shorting activity (*ETF-based SA*). To construct this measure, I first calculate each ETF's bi-weekly change in shorted values by multiplying the change in ETF short interest ratio with ETF price. Subsequently, for each stock, I calculate the bi-weekly change in the shorted value via an ETF by multiplying the weight of the stock in an ETF with the change in ETF shorted value. Specifically, for a stock i in a bi-week period t, the bi-weekly change in stock-level shorted value via an ETF j (Δshorted value$_{i,t,j}$) is calculated as follows:

$$\Delta\text{shorted value}_{i,t,j} = (\text{short interest}_{j,t} - \text{short interest}_{j,t-1}) \times P_{j,t} \times \text{weight}_{i,t,j} \quad (2.1)$$

where $P_{j,t}$ is the ETF j's share price, weight$_{i,j,t}$ is the weight of stock i in ETF j.[14] Next, for each stock, I calculate the bi-weekly change in total shorted value via ETFs by aggregating Δshorted value$_{i,t,j}$ across all ETFs that contain stock i. To make this measure comparable across stocks, I scale the bi-weekly change in total shorted value by market capitalization:

$$\textit{ETF-based SA}_{i,t} = \frac{\sum_j \Delta\text{shorted value}_{i,t,j}}{MktCap_{i,t}} \quad (2.2)$$

where $MktCap_{i,t}$ is the market capitalization for stock i. By construction, *ETF-based SA*$_{i,t}$ is the manifestation of ETF shorting activity at the stock level. *ETF-based SA*$_{i,t}$ cap-

[13]The stock sample encompasses all common stocks listed on NYSE, NYSE MKT, and NASDAQ (CRSP share codes 10 and 11; exchange codes 1, 2, and 3.)

[14]I calculate the ETF weight as the proportion of value held by ETFs relative to ETF market capitalization, and the value held by ETFs is based on the number of shares held by ETF at the last month's end when ETF holding data is from CRSP Mutual Fund dataset and is based on the number of shares held by ETF at the last quarter end when ETF holding data is from Thomson Reuters Mutual Fund holdings database

tures the ETF shorting activity aggregated at stock i and reflects the collective changes in stock-level shorting demand originating from ETF shorting activity.[15] A positive *ETF-based* $SA_{i,t}$ means that the stock i is increasingly shorted via ETFs, and vice versa.

I use the change in ETF short interest rather than the level of ETF short interest to construct the measure because ETF short interest exhibits persistence and non-stationarity, which may lead to spurious regressions. I test whether ETF short interest is stationary using the Augmented Dicky-Fuller (ADF) test. The ADF test results are reported in Table 2.A2. I find that the aggregate time-series ETF short interest is not stationary. I also run the ADF test for the ten most shorted ETFs individually, which account for more than 90% of the total ETF short interest. I find that eight out of the ten most shorted ETFs have non-stationary short interest.[16] Consequently, following previous literature,[17] I take the first difference of ETF short interest to ensure that my results do not suffer from non-stationarity issues.

Panel A of Table 2.1 presents ETF characteristics. The mean (median) ETF market capitalization is \$ 1.89b (\$ 0.12b). The mean (median) ETF bi-weekly consolidated dollar volume is \$ 9.87m (\$ 0.13m). The mean (median) ETF short interest is 4.51% (0.7%) of market capitalization. The mean (median) change of ETF short interest is -0.03% (0%) of market capitalization. This means that the distribution of ETF short interest is highly left skewed, and most ETFs have low shorting activity. The top 10 most shorted ETFs account for around 94% of the total ETF short interest but only 42.4% of the total ETF market capitalization. The mean (median) ETF monthly return is 0.56% (0.8%). Panel B of Table 2.1 reports stock characteristics. The mean (median) market capitalization is \$ 5.15b (\$ 0.5b). The mean (median) quarterly consolidated volume is \$ 3.1b (\$ 0.23b). The mean (median) ETF-based shorting activity is 0.4% (0.1%). The mean (median) quarterly stock return is 1% (2%).

[Insert Table 2.1 near here]

[15]Mathematically, Li and Zhu (2022)'s ETF-based short ratio measure ($ETF\_SR_{i,t}$) is the integral of my *ETF-based* $SA_{i,t}$ measure (i.e., $ETF\_SR_{i,t} = \int_{j=0}^{t}$ *ETF-based* $SA_{i,j}$).

[16]I report the results of Augmented Dickey-Fuller test for ETF short interest in Table 2.A2 in the Appendix

[17]See Jiao, Massa, and Zhang (2015), Chen, Da, and Huang (2019b) and Huang et al. (2021)

2.3 ETF return predictability of ETF shorting activity

Traders short ETFs for various purposes. One possible reason to short ETFs is to bet against the market.[18] Existing literature demonstrates that traders short stocks to express bearish views and finds that the stock-level shorting activity negatively predicts future stock returns.[19] By the same token, I expect ETF shorting activity to negatively predict future ETF returns. However, Huang et al. (2021) document that informed traders use industry ETFs to hedge against their exposure to industry risk when having positive information about ETFs' underlying securities. They further observe that the industry ETF shorting activity positively predicts future ETF returns. Another important driver of ETF shorting activity is operational shorting. Evans et al. (2022) show that operational shorting is driven by liquidity provision and does not have predictive power in future returns of ETF NAV.

Therefore, whether ETF short activity positively or negatively predicts future ETF returns depends on the strength of the predictive power of different ETF shorting activities. I expect ETF shorting activity to (i) negatively predict future ETF returns when the return predictability of bear-market shorting dominates the return predictability of other shorting activity. (ii) positively predict future ETF returns when hedging-motivated shorting primarily drives return predictability or (iii) uncorrelated with future ETF returns when ETF shorting activity is mainly driven by operational shorting.

To empirically test this question, I employ the standard portfolio sorting approach. Since ETF short interest is not stationary, I use the bi-weekly change in ETF short interest ratio (ΔETF SI) to measure ETF shorting activity.[20] Subsequently, I use ETF shorting activity to sort ETFs into quintile portfolios each bi-week period. Following Boehmer et al. (2008), these portfolios are subsequently held for the next 20 trading days after skipping one trading day (i.e. from day 2 through day 21) and re-balance bi-weekly. Additionally, I construct a long-short ETF portfolio by taking a long position in ETFs

[18] A salient example was recently observed during Russia's recent invasion of Ukraine. Within a week, traders short three key Russian ETFs reaped approximately $120 million. The largest, the VanEck Russia ETF (RSX), saw a decline of almost 45%, along with iShares MSCI Russia ETF and Direxion Daily Russia ETF have dropped by 45% and 41%, respectively. (Wall Street Journal, 2022)

[19] Boehmer et al. (2008), Diether et al. (2009), Boehmer et al. (2010), and Rapach et al. (2016)

[20] ETF short interest ratio (ΔETF SI) is defined as the percentage ETF short interest relative to the ETF share outstanding.

with the highest ETF shorting activity and a short position in ETFs with the lowest ETF shorting activity. Finally, I calculate the equal-weighted and value-weighted excess returns and Fama and French (1993) and Carhart (1997) alphas for each ETF portfolio.

[Insert Table 2.2 near here]

Panels A and B of Table 2.2 report the equal-weighted and value-weighted excess returns and alphas of five ETF portfolios and the long-short portfolio, as well as t-statistics based on the Newey-West standard errors with four lags. In Panel A, I observe that the equal-weighted long-short ETF portfolio generates a negative and significant excess return of 30 basis points over the next 20 trading days with a t-stat of -5.66. After controlling for Fama and French (1993) and Carhart (1997) four factors, the long-short portfolio yields a negative alpha of 30 basis points with a t-stat of -6.43. The long-short strategy with value weights results in a negative excess return of 13 basis points (t=-1.72) and a negative Fama and French (1993) and Carhart (1997) four-factor alpha of 15 bps every two weeks (t = -2.27). The negative return predictability indicates that ETF shorting activity is informed about future ETF returns.

Next, I validate my results on the return predictability of ETF shorting activity using a regression approach. The regression analyses enable me to examine the return predictability of ETF shorting activity while controlling for other well-known factors. These factors include the size (Log(MktCap)), short-term reversal (REV), idiosyncratic volatility (IdiosynVol), put option activity (Put) and ETF flow (ETF flow).[21] Specifically, I estimate the following panel regression specification:

$$Ret_{i,t} = \beta_0 + \beta_1 \Delta \text{ETF SI}_{i,t} + \gamma X_{i,t} + \delta + \epsilon_{i,t}, \tag{2.3}$$

where i represents an ETF i, t represents a bi-weekly reporting date t. $Ret_{i,t+20}$ is the 20-day holding period return of an ETF after skipping one day, ΔETF SI$_{i,t}$ is the bi-weekly change in ETF short interest ratio, $X_{i,t}$ is a vector of factors described above, δ are ETF and bi-week fixed effects, The standard errors are clustered by ETF and bi-week.

[21] The put option activity is measured as the percentage change in open interest of put option. ETF flow is measured as the percentage change of ETF share outstanding. I do not include the book-to-market ratio (BTM) because there is no reliable measure of the book value for ETFs.

[Insert Table 2.3 near here]

The regression results are in Table 2.3. In Column (1), the coefficient estimator of ΔETF SI is negative and significant, consistent with ETF shorting activity negatively predicting future ETF returns. From Columns (2) to (4), I control for size, past return, idiosyncratic volatility and put option activity, and the coefficient estimators of ΔETF SI remain negative and statistically significant. On average, one standard deviation increase in ΔETF SI corresponds to a decrease in future ETF returns by 24 basis points. My results demonstrate that those well-known predictors cannot explain the negative predictive power of ΔETF SI. These results also suggest that the negative return predictability, stemming from bearish-market shorting, surpasses the positive return predictability from hedging-motivated shorting. Overall, I find that ETF shorting activity negatively predicts future ETF returns, signifying that ETF shorting activity is informed and, on average, reflects negative information.[22]

The negative return predictability of ETF shorting activity might come from traders switching between different ETFs to exploit their negative information about pricing factors (i.e., size and book-to-market), which is not necessarily market-wide information. To validate the argument that bearish-market shorting primarily drives the negative return predictability of ETF shorting activity, I restrict my ETF sample to the market ETFs.[23] The sample period spans from January, 2002 to December, 2021. The market ETF sample contains 153 market equity ETFs and 63,933 ETF-bi-week observations. This restricted sample allows me to filter out the ETF shorting activity from traders who switch between ETFs to exploit negative information about non-market pricing factors. I re-estimate the fixed effects regression specification in Eq. (2.3) using the market ETF sample to examine the return predictability of market ETF shorting activity.

[Insert Table 2.4 near here]

Columns (1)-(4) of Table 2.4 report the return predictability for market ETFs. The

[22]The results are robust when using the Fama and MacBeth (1973) cross-sectional regressions, as tabulated in Tables 2.A17

[23]I identify an ETF as a market ETF when the fund's Lipper classification code falls into one of the following categories: LCCE, LCGE, LCVE, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE, SCVE and SPSP.

coefficient estimators of ΔETF SI remain negative and significant, indicating that the ETF shorting activity negatively predicts future market ETF returns. This suggests that market ETF shorting activity is informed and, on average, contains negative market-wide information, supporting that the bearish-market shorting primarily drives the negative return predictability of ETF shorting activity.

Another possibility is that the negative return predictability in market ETF shorting activity may stem from synthetic shorting rather than bearish-market shorting. Synthetic shorting refers to the ETF shorting activity that exploits firm-specific information of a target stock by shorting ETFs and simultaneously longing all underlying stocks except for the target stock. Li and Zhu (2022) argue that short sellers use synthetic shorting to circumvent the stock-level short-sale constraints. To further distinguish the effects of these two types of ETF shorting activity, I examine the return predictability of market ETF shorting activity on macroeconomic news release periods. If the negative return predictability originates from bearish-market shorting, I anticipate that this return predictability should be concentrated on macroeconomic news periods rather than non-macroeconomic news periods. Conversely, if it arises from synthetic shorting, the negative return predictability should be indifferent between macroeconomic news periods and non-macroeconomic news periods.

I focus on the ten major macroeconomic news events examined extensively in prior literature: the Federal Open Market Committee (FOMC) meeting, the Non-Farm Payroll (NFP) release, Gross Domestic Product (GDP) release, the Consumer Price Index (CPI) release, Institute for Supply Management's Manufacturing Index (ISM) release, Consumer Sentiment Index (CSI) release, Industrial Production (IP) release, Personal Income (PI) release, Housing Starts (HST) release, and Producer Price Index (PPI) release.[24] For each bi-weekly observation, I compute the number of macroeconomic news released within the bi-weekly period. I identify a bi-weekly period as a macroeconomic (non-macroeconomic) news period if the number of macroeconomic news released within the bi-weekly period is greater (less) than the 75 (25) percentile of the sample. I estimate the OLS regression to examine the predictability of market ETF shorting activity on future market ETF returns on macroeconomic news periods.

[24]I manually collect the release dates of these macroeconomic news events from Bloomberg.

Columns (5)-(6) of Table 2.4 report the results with different controls on macroeconomic news periods.[25] I find that the coefficient estimators of ΔETF SI are negative and significant, indicating that the market ETF shorting activity negatively predicts future market ETF returns on macroeconomic news periods. In Columns (7) to (8), I repeat the above analysis for the non-macroeconomic news periods.[26] However, I fail to find any statistically significant return predictability of ΔETF SI on future market ETF returns on the non-macroeconomic news periods. The different return predictability between macroeconomic news periods and non-macroeconomic news periods supports that the bearish-market shorting drives the negative return predictability of market ETF shorting activity.

Additionally, I use the daily short volume data, which allows me to precisely identify macroeconomic/non-macroeconomic news dates, to examine the return predictability of market ETF shorting activity on macroeconomic news dates and non-macroeconomic news dates. I obtain daily short volume for all market ETFs from the Financial Industry Regulatory Authority (FINRA) daily short sale transaction file. The sample period spans from January 1, 2010 to December 31, 2021.[27] I am cautious about the difference between short volume and short interest. Comerton-Forde et al. (2016b) and Evans et al. (2022) show that there is a proportion of short volumes originates from the market makers' liquidity provision activity and is not necessarily associated with informed trading. To filter for the informed trading in the short volume, I follow Evans et al. (2022) to compute the shorting activity driven by liquidity suppliers and subtract this shorting activity from the total short volume. Subsequently, I divide the remaining

[25]The total numbers of observations in macroeconomic and non-macroeconomic news periods might not align precisely with the totals presented in Columns (3) and (4). This discrepancy arises from the methodology used to categorize these periods, which is based on the 25th and 75th percentiles. As a result, the combined number of observations from both macroeconomic and non-macroeconomic news periods is expected to approximate half of the total observations listed in Columns (3) and (4).

[26]The notable high R-square values observed in Table 2.4 and Table 2.5 can be attributed to the specific focus on market ETFs within both analyses. Given that most market ETFs tend to move in tandem due to their tracking of very similar indices. Consequently, there is small cross-sectional variation and the predominant source of variation in these cases arises from time-series fluctuations (i.e., change in market conditions) Adding time fixed effects in the regression models effectively captures these time-series variations, increasing the models' ability to explain variations over time. In Table 2.A23 of the Appendix, I re-estimate the regressions without time fixed effects, and I find a sharp decline in R-square values. Nevertheless, I find that ETF shorting activity continues to exhibit its negative return predictability.

[27]FINRA short volume data is available from August 1, 2010 onwards. To avoid the potential data corruption issue in the early release stage, I start my sample on January 1, 2010. Results are present in Table 2.A15 of the Appendix.

short volume by the total trading volume to obtain the adjusted ETF short volume ratio (ETF SVR). I estimate the panel regression of future ETF returns on ETF SVR with the following regression specification on macroeconomic news dates and non-macroeconomic news dates, respectively.[28]

$$Ret_{i,t+1} = \beta_0 + \beta_1 \text{ETF SVR}_{i,t} + \gamma X_{i,t} + \delta + \epsilon_{i,t}, \tag{2.4}$$

where i represents an ETF i, t represents a trading date t. $Ret_{i,t+1}$ is the next-day ETF return after skipping one day following Boehmer et al. (2008), ETF $\text{SVR}_{i,t}$ is the adjusted ETF short volume ratio, $X_{i,t}$ are the same factors as in Eq. (2.3), δ are ETF and date fixed effects, The standard errors are clustered by ETF and date.

Column (1) of Table 2.5 presents the panel regression results of second-next-day returns (i.e., next-day returns after skipping on trading day) on ETF SVR_t on macroeconomic news dates. I find that the coefficient estimator of ETF SVR_t is negative and significant, indicating that the ETF SVR_t negatively predicts future ETF returns on macro news dates. However, this result might be influenced by the temporary price pressure caused by the large trading during important information events as suggested by Comerton-Forde, Do, Gray, and Manton (2016a). To alleviate this concern, I focus on the return predictability of ETF SVR over the previous five trading days (ETF $\text{SVR}_{t-5,t}$) on next-20-day returns, after skipping one trading day, on macro news dates. With the results shown in Column (2), I find the consistent negative return predictability of ETF $\text{SVR}_{t-5,t}$ on next-20-day returns on macro news dates. Columns (3) to (4) replicate the above analysis for the non-macroeconomic news dates. I find that the coefficient estimators of ETF SVR and ETF $\text{SVR}_{t-5,t}$ are statistically insignificant, indicating that the ETF short volume ratio has no predictive power on future ETF returns on non-macro news dates.[29] Overall, I find that the negative return predictability of ETF shorting activity appears only on macroeconomic news dates but not on non-macroeconomic news dates. This difference reinforces that the return predictability of market ETF shorting activity is concentrated on macroeconomic news dates. The concentration of return predictability on macroeconomic news periods supports that informed traders

[28]Non-macroeconomic news dates refer to days that are neither macroeconomic news dates nor 1 day after macroeconomic news dates.
[29]These results are robust in using Fama-Macbeth regressions, as shown in Table 2.A17.

short ETFs to exploit their negative market-wide information, which is mostly realized during macroeconomic news periods.

[Insert Table 2.5 near here]

2.4 Stock return predictability of ETF-based shorting activity

Shorting ETFs is equivalent to shorting the underlying securities of ETFs. The negative information in ETF shorting activity will eventually be reflected in ETFs' underlying securities through ETF arbitrage (Glosten et al. (2021)). By construction, the stock-level shorting activity via ETFs (*ETF-based SA*) captures negative information or bearish views from ETF short sellers. I expect a stock to yield a lower return when *ETF-based SA* is high compared to the return when *ETF-based SA* is low.

To empirically test this hypothesis, I once again employ the standard portfolio sorting approach. I sort stocks into quintile portfolios based on ETF-based shorting activity, *ETF-based SA*,[30] and hold these stock portfolios over the next 20 trading days and re-balance bi-weekly. I construct a long-short portfolio by taking a long position in stocks with the highest *ETF-based SA* and a short position in stocks with the lowest *ETF-based SA*. I calculate the equal-weighted and value-weighted stock portfolio's excess returns and Fama and French (1993) and Carhart (1997) alphas.

Panels C and D of Table 2.2 display the equal-weighted and value-weighted excess returns and alphas of five stock portfolios and the long-short portfolio, along with t-statistics based on the Newey-West standard errors with four lags. In Panel C, I find that the equal-weighted long-short portfolio generates a negative and significant excess return of 34 basis points with a t-stat of -2.30. After controlling for Fama and French (1993) and Carhart (1997) four factors, the long-short portfolio earns a negative alpha of 31 basis points with a t-stat of -2.06. Panel D reveals that the value-weighted long-short portfolio earns a negative excess return of 41 basis points (t=-2.46) and a negative

[30] The results are robust if I sort stocks into decile portfolios based on *ETF-based SA*, and I report decile portfolio results in Table 2.A16 of the Appendix.

Fama and French (1993) and Carhart (1997) four-factor alpha of 39 bps per bi-week period (t=-2.25). The negative return predictability of *ETF-based SA* indicates that ETF-based shorting activity is informed about future stock returns.

I strengthen the stock-level return predictability of *ETF-based SA* using the regression analyses. I control for the well-known predictors of stock returns, including the size ($Log(MktCap)$), book-to-market ratio (BTM), short-term reversal (REV), idiosyncratic volatility ($IdiosynVol$) and ETF flow (ETF-based Flow).[31] Specifically, I run the regression specification in Eq. (2.3) with the replacement of the Δ ETF SI to *ETF-based SA*.

Columns (5)-(8) of Table 2.3 report the regression results. In Column (5), the coefficient estimator of *ETF-based SA* is negative and significant, indicating that *ETF-based SA* negatively predicts future stock returns. Li and Zhu (2022) find that the stock-level short interest via ETFs has negative predictive power in future stock returns. Therefore, in Column (6), I control for Li and Zhu (2022)'s measure: the stock-level short interest via ETFs (ETF_SR), and the coefficient estimator of *ETF-based SA* remains negative and statistically significant,[32] indicating that the return predictability of *ETF-based SA* is distinct from the return predictability of ETF_SR documented by Li and Zhu (2022). In Columns (7) and (8), I control for well-known predictors of cross-sectional stock returns. The coefficient estimators of *ETF-based SA* remain negative and significant, indicating that the negative predictive power cannot be explained by those well-known predictors. On average, one standard deviation increase in *ETF-based SA* corresponds to a decrease in future stock returns by 12 basis points. My findings demonstrate that *ETF-based SA* negatively predicts future stock returns, suggesting that ETF shorting activity is informed and, on average, contains negative information about ETF underlying securities.[33]

---

[31] Lou (2012) find that mutual fund flow positively predicts future stock returns. To account for the effect of ETF flow, I construct the ETF-based flow that captures the stock-level price pressure derived from ETF flows, following Li and Zhu (2022), and I control for the ETF-based flow in Column (8).

[32] I construct the ETF-based short ratio (ETF_SR) following Li and Zhu (2022). Intuitively, my measure is the bi-weekly change in ETF-based short ratio (i.e., $ETF\_SR_{i,t} = \int_{j=0}^{t} \textit{ETF-based } SA_{i,j}$ and $\textit{ETF-based } SA_{i,t} = \frac{\partial ETF\_SR_{i,t}}{\partial t}$).

[33] The results are robust when using the Fama-Macbeth regressions, as tabulated in Table 2.A17.

2.5 ETF-based shorting activity and informational efficiency

Given that *ETF-based SA* is informed, the prices of ETFs' underlying securities become more informationally efficient as more *ETF-based SA* is incorporated. I hypothesize that *ETF-based SA* improves the informational efficiency of ETFs' underlying securities. I examine this hypothesis by using two approaches. In the first approach, I examine the effect of *ETF-based SA* on informational efficiency by estimating the Future Earnings Response Coefficient (FERC). Following Israeli et al. (2017), Glosten et al. (2021) and Brogaard and Pan (2022), I estimate the regression specification in Eq. (2.5).

$$\begin{aligned} \text{Return}_{i,t} = \beta_0 + \beta_1\text{Earn}_{i,t} + \beta_2\text{Earn}_{i,t+1} + \beta_3\textit{ETF-based SA}_{i,t} \\ + \beta_4\text{Earn}_{i,t+1} \times \textit{ETF-based SA}_{i,t} + \gamma Controls_{i,t} + \delta + \epsilon_{i,t} \end{aligned} \tag{2.5}$$

Where i represents a firm i and t represents a quarter t. $\text{Return}_{i,t}$ is the buy-and-hold return from quarter $t-1$ to quarter t. $\text{Earn}_{i,t}$ is quarterly seasonally adjusted earnings at quarter t following Glosten et al. (2021). *ETF-based* $SA_{i,t}$ is the ETF-based shorting activity from quarter $t-1$ to quarter t. I use standard control variables in my empirical analysis, including market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership.[34] δ are firm fixed effects and year-quarter fixed effects. The standard errors are clustered by firm and year-quarter. The FERC estimate (β_2) measures the extent to which the current stock return incorporates future earnings information. I focus on the coefficient (β_4) of the interaction between future earnings ($\text{Earn}_{i,t+1}$) and *ETF-based SA*. A positive coefficient shows that *ETF-based SA* is associated with higher FERC, meaning that *ETF-based SA* helps incorporate more future earnings information

[34] Boehmer and Kelley (2009) show that institutional ownership affects informational efficiency. Meanwhile, institutional ownership is a superset of ETF ownership and is correlated with ETF-based shorting activity. To isolate the effect of ETF shorting activity from the effect of institutional ownership, I orthogonalize the change in institutional ownership with respect to the ETF-based shorting activity. Following Glosten et al. (2021), for each stock, I estimate a time-series regression of the change in institutional ownership on ETF shorting activity. The residual from this regression captures the change in institutional ownership that cannot be explained by ETF shorting activity.

into current stock returns. Therefore, a positive coefficient estimate of β_4 indicates increased informational efficiency. For simplicity, I label the coefficient (β_4) as Incremental Future Earnings Response Coefficient (IFERC).

In the second approach, I use four measures to capture informational efficiency: (i)Weller (2018)'s jump ratio, (ii) post-earnings-announcement drift (PEAD), (iii) variance ratio and (iv) auto-correlation. The jump ratio, as proposed by Weller (2018), refers to the proportion of price adjustments during the earnings announcements relative to the total price adjustments around the earnings announcements as defined in Eq. (2.6).

$$Jump_{i,t}^{21,2} = \frac{CAR_{i,t}^{T-1,T+2}}{CAR_{i,t}^{T-21,T+2}} \tag{2.6}$$

Where i refers to firm i, t refers to quarter t, T refers to the earnings announcement date, and $CAR_{i,t}^{T-j,T+k}$ is the cumulative Fama-French three-factors adjusted return for firm i in quarter t from j trading days before the earnings announcements to k trading days after the earnings announcements.[35]

This jump ratio measure captures the proportion of earnings surprises (or unknown information) in the earnings announcements. When short sellers trade on their negative information, this information becomes partially revealed to market participants, leading to price adjustments toward the post-earnings-announcement fundamental value. Eventually, when the earnings are announced, there are fewer surprises in earnings announcements because some information has already been incorporated into the prices, leading to smaller price adjustments and a smaller jump ratio. Therefore, a smaller jump ratio indicates more information incorporated in the prices before earnings announcements, indicating better informational efficiency.

The post-earnings-announcement drift (PEAD) has been widely used in the literature (Chordia and Shivakumar (2006)). The PEAD is defined as the cumulative abnormal return from the first trading day after the earnings announcements to one

[35]Note that the jump ratio is unbound from below. Following Weller (2018), I drop the jump ratio when the $CAR_{i,q}^{T-21,T+2}$ is smaller than the standard deviation of daily returns over the preceding month of earnings announcements.

month after the earnings announcements, as shown in Eq. (2.7).[36]

$$PEAD_{i,t} = |CAR_{i,t}^{T+1,T+22}| \tag{2.7}$$

The efficient market hypothesis predicts that earnings announcements lead to instantaneous price adjustments. A small drift after the earnings announcements suggests that the earnings information is almost fully reflected in prices right after the earnings announcement. Therefore, a smaller PEAD means a more efficient stock price and better informational efficiency.

The variance ratio is defined as the absolute difference between one and the ratio of the variance of 2-day returns over two times the variance of 1-day returns as shown in Eq. (2.8). To maintain consistency with my jump ratio and PEAD measures, I compute the variance ratio based on returns from 21 trading days before to 2 trading days after quarterly earnings announcements.

$$\text{Variance Ratio} = |1 - \frac{\sigma^2_{\text{2-day}}}{2\sigma^2_{\text{1-day}}}| \tag{2.8}$$

If prices are efficient, prices follow a random walk with stable variance, and the variance ratio should be closer to zero. Therefore, a smaller variance ratio indicates a more efficient price. The last informational efficiency measure is the absolute value of autocorrelation of daily returns. An efficient stock price (or return) should be unpredictable by its past-day return. Therefore, when the stock price is more efficient, the autocorrelation of daily returns is closer to zero, and a smaller absolute value of autocorrelation indicates a more efficient stock price. Similarly, I compute the auto-correlation of daily returns from 21 trading days before to 2 trading days after quarterly earnings announcements to keep consistency across my informational efficiency measures.

Table 2.1 describes the summary statistics for these four informational efficiency measures. The mean (median) jump ratio is 42.7% (40.7%), suggesting that, on average, the earnings surprise accounts for 42.7% of price adjustments around earnings announcements. The mean (median) PEAD is 7.0% (3.8%), showing that, on average, the prices drift upward about one month after the earnings announcements. The mean

[36]Following Weller (2018), I assume that there are 22 trading days in a month.

(median) variance ratio is 0.33 (0.32). The mean (median) absolute value of autocorrelation is 0.19 (0.17) of market capitalization.

Using four informational efficiency measures to examine the effect of *ETF-based SA* requires a different regression setup. Specifically, I regress informational efficiency measures on *ETF-based SA* using the following regression specification shown in Eq. (2.9).

$$\text{InfoEff}_{i,t} = \alpha_0 + \alpha_1 \textit{ETF-based SA}_{i,t} + \gamma Controls_{i,t} + \delta + \epsilon_{i,t} \tag{2.9}$$

Where i represents a firm i and t represents a quarter t. $\text{InfoEff}_{i,q}$ is the informational efficiency measure. *ETF-based SA*$_{i,t}$ is the ETF-based shorting activity from quarter $t-1$ to quarter t. $Controls_{i,t}$ includes market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership. δ are firm fixed effects and year-quarter fixed effects. The standard errors are clustered by firm and year-quarter. I focus on the coefficient estimators of α_1, and a negative α_1 represents a positive association between *ETF-based SA* and informational efficiency.

[Insert Table 2.6 near here]

Table 2.6 reports the regression results. Columns (1) and (2) present the FERC results without and with controls. The coefficients for the interaction term $Earn_{i,t+1} \times$ *ETF-based SA* (IFERC) are positive and significant, indicating that *ETF-based SA* strengthens the returns-earnings association. This positive association suggests that *ETF-based SA* helps reflect more future earnings information on current stock returns, showing enhanced informational efficiency. Columns (3) to (6) present the results using informational efficiency measures. I find that the coefficients of *ETF-based SA* are negative and significant when informational efficiency is measured by PEAD, variance ratio and auto-correlation. The negative association indicates that *ETF-based SA* helps incorporate information into the prices of ETFs' underlying securities in a timely manner and makes the prices of ETFs' underlying securities move efficiently.[37] These results

[37]The association between *ETF-based SA* and informational efficiency is predicated on the conjecture that ETF short sellers increase their short positions when earnings fall short of expectations (i.e., bad earnings news). In Table 2.A12 of the Appendix, I split the sample into two subsamples based on the sign of earnings surprise (positive earnings news v.s. negative earnings news). I find little

support my hypothesis that ETF shorting activity improves the informational efficiency of ETFs' underlying securities and reinforce my previous finding that *ETF-based SA* is informed.

2.5.1 ETF-based shorting Activity and Incorporation of Components of Earnings Information

Given that *ETF-based SA* improves the informational efficiency of underlying securities, it is essential to understand the channel through which *ETF-based SA* improves the informational efficiency. The improvement of informational efficiency can be driven by the better incorporation of either market-wide or firm-specific information. Previously, I show that ETF shorting activity contains negative market-wide information, implying that the improvement of informational efficiency stems from the better incorporation of market-wide information. Moreover, ETFs provide traders with a convenient vehicle to trade a basket of stocks, allowing traders to easily incorporate their negative market-wide information by shorting ETFs. Glosten et al. (2021) support that informed ETF trades are driven by market-wide information rather than firm-specific information. In addition, shorting ETFs incurs high borrowing costs, which are significantly higher than the borrowing costs of ETFs' underlying securities (Bhojraj and Zhao (2021)). Therefore, it is challenging for ETF short sellers to trade profitably on ETFs with firm-specific information. Nonetheless, there is a possibility that the improvement of informational efficiency is driven by the better incorporation of firm-specific information.[38] Li and Zhu (2022) argue that traders with firm-specific information of a target stock short ETFs to circumvent short-sale constraints and form a synthetic short position in the target stock when the target stock is difficult to short. Therefore, it is an empirical question as to whether ETF shorting activity improves the informational efficiency of underlying securities by incorporating market-wide or firm-specific information. To answer this question, I focus on the FERC approach. I first decompose the quarterly earnings into systematic earnings ($Earn\_Sys$), which captures earnings related to

evidence that ETF shorting activity helps improve the informational efficiency of underlying securities in instances of positive earnings surprises. In other words, the improvement effect of *ETF-based SA* on informational efficiency mainly comes from the occasion of unfavorable earnings news.

[38] Eglīte, Patel, and Putniņš (2023) show that some traders with firm-specific information use ETF trades to conceal their trading intentions.

market-wide information, and firm-specific earnings ($Earn\_Firm$) following Glosten et al. (2021) and Israeli et al. (2017). Specifically, I estimate the time-series regression for each firm using:

$$\text{Earn}_{i,t} = \beta_0 + \beta_1 \text{Earn mkt}_t + \beta_2 \text{Earn Industry}_{t,y} + \epsilon_{i,t} \tag{2.10}$$

Where Earn mkt_t is the value-weighted market earnings and Earn $\text{Industry}_{t,y}$ is the value-weighted industry earnings based on the two-digit SIC code. $Earn\_Sys$ is the fitted value of Eq. (2.10), and $Earn\_Firm$ is the residual of Eq. (2.10). I then modify my FERC regression specification in Eq. (2.5) by replacing the seasonally adjusted earnings ($Earn$) with its two earnings components ($Earn\_Sys$ and $Earn\_Firm$) as shown in Eq.) (2.11).

$$\begin{aligned} Return_{i,t} = {} & \theta_0 + \theta_1 Earn\_Sys_{i,t+1} \times \textit{ETF-based SA}_{i,t} + \theta_2 Earn\_Firm_{i,t+1} \times \textit{ETF-based SA}_{i,t} \\ & + \theta_3 \textit{ETF-based SA}_{i,t} + \theta_4 Earn\_Sys_{i,t+1} + \theta_5 Earn\_Firm_{i,t+1} \gamma Controls_{i,t} + \delta + \epsilon_{i,t} \end{aligned} \tag{2.11}$$

By examining the IFERC associated with $Earn\_Sys$ and $Earn\_Firm$, I can determine whether ETF shorting activity improves informational efficiency by incorporating market-wide earnings information or firm-specific earnings information. When *ETF-based* $SA_{i,t}$ facilitates the incorporation of market-wide information, I expect the coefficient (θ_1) of the interaction term $Earn\_Sys_{i,t+1} \times$ *ETF-based* $SA_{i,t}$ to be positive. Conversely, I expect the coefficient (θ_2) of the interaction term $Earn\_Firm_{i,t+1} \times$*ETF-based* $SA_{i,t}$ to be positive when *ETF-based* $SA_{i,t}$ facilitates the incorporation of firm-specific information.

[Insert Table 2.7 near here]

Table 2.7 reports the results of IFERC associated with $Earn\_Sys$ and $Earn\_Firm$. I find that the IFERC associated with $Earn\_Sys$ (θ_1) is positive and significant, indicating that *ETF-based SA* helps incorporate market-wide earnings information into prices. Nonetheless, the IFERC associated with $Earn\_Firm$ (θ_2) is statistically insignificant, indicating that *ETF-based SA* does not help incorporate the firm-specific earnings information into prices. The different IFERCs indicate that the improvement in informational efficiency is attributable to the better incorporation of market-wide information rather than firm-specific information. This difference also aligns with my earlier findings that traders short ETF to exploit their negative market-wide information.

Information efficiency and information acquisition are typically interrelated, as informed traders collect relevant data to assess firm value before engaging in informed trading. An increase in information acquisition generally leads to a corresponding rise in informed trading activity, which in turn enhances information efficiency. To further support the argument that the observed improvement in information efficiency is not driven by the incorporation of firm-specific information into stock prices, I investigate the association between ETF shorting activity and firm-specific information acquisition, using EDGAR search volume as a proxy. If ETF shorting activity does not facilitate the incorporation of firm-specific information, I expect no significant association between ETF shorting activity and firm-specific information acquisition.

The results, presented in the Appendix,[39] reveal no statistically significant association between ETF shorting activity and either prior or subsequent firm-specific information acquisition. This lack of association indicates that ETF shorting activity does not facilitate the incorporation of firm-specific information into prices. These findings are consistent with earlier results that the improvement in information efficiency is driven by the more effective incorporation of market-wide information rather than firm-specific information, reinforcing the conclusion that traders predominantly engage in ETF shorting to exploit negative market-wide signals.

2.6 The causal relation between ETF-based shorting activity and informational efficiency

Despite my efforts to control for firm characteristics that influence informational efficiency, some unobservable factors that affect informational efficiency might exist. To

[39] Tables 2.A22 in the Appendix presents the results of ETF shorting activity and information acquisition.

address this omitted variable problem, I use the implementation of the SEC Funds of Funds (FOF) arrangement to identify the exogenous variation in *ETF-based SA*.

I exploit the variation of ETF shorting activity driven by the SEC FOF arrangement that eases the short-sale constraints in ETF markets. Bhojraj and Zhao (2021) demonstrate that one of the primary reasons why shorting ETFs is so costly is due to the supply constraints of ETF shares. This limited supply of ETF lendable shares is attributable to the SEC Section 12(d)(1) rule, which prohibits a fund or an institution from investing in other investment funds for: (i) more than 3% of an investment fund's voting shares, or (ii) more than 5% of a single investment fund assets or (iii) more than 10% of its assets on all investment funds. As a consequence, the SEC section 12(d)(1) results in a limited number of institutional ETF holdings, restricts institutions' ability to lend ETF shares to short sellers and constrains ETF shorting activity. On October 7, 2020, SEC adopted a new FOF arrangement and allowed funds to invest in other funds in excess of the above limits, subject to certain conditions.[40] Therefore, this SEC FOF arrangement facilitates funds and institutions to invest in ETFs more easily, increasing institutional holdings in ETFs. To examine the first-order impact of the SEC FOF arrangement on institutional holdings in ETFs, I estimate the time-series regression of institutional holdings in ETFs using the following specification.

$$\text{IO}_{i,t} = \alpha_0 + \alpha_1 Post_t + Controls_{i,t} + \sigma + \epsilon_{i,t} \tag{2.12}$$

Where i represents a ETF i and t represents a quarter t. $IO_{i,t}$ is the institutional ownership in the ETF i. $Post_t$ is the post-SEC FOF indicator that equals one if trading day t is after the effective date of the SEC FOF arrangement and 0, otherwise.[41] $Controls_{i,t}$ includes the inverse of the close price and the logarithm of dollar volume. δ are ETF fixed effects. The standard errors are clustered by ETF and quarter.

[Insert Table 2.8 near here]

[40] There are two conditions. First, the acquiring fund should not control the acquired fund. Second, the acquiring fund is required to mirror vote shares in the acquired fund.

[41] The effective date of the SEC FOF arrangement is 19 January, 2021. My sample period covers one and a half years before the effective date of the SEC FOF arrangement and one and a half years after the effective date of the SEC FOF arrangement so that I have sufficient quarterly observations.

Column (1) of Table 2.8 reports the results. The coefficient of the post-SEC FOF indicator is positive and significant. This positive coefficient indicates a 2.72% increase in institutional holdings in ETFs after the SEC FOF arrangement, indicating a larger ETF lending supply and relaxation of ETF short-sale constraints. I subsequently investigate the consequence of the increased ETF lending supply on ETF shorting activity. I expect more ETF shorting activity when the ETF lending supply is higher and when ETFs are cheaper to borrow. I use the change in ETF short interest (ΔETF SI) to measure ETF shorting activity, and I also use ETF short volume (ETF SV) as an alternative measure of ETF shorting activity. I run the time-series regressions with the regression specification in Eq. (2.12) and replace the dependent variable with ETF SV and ΔETF SI, respectively.[42]

Columns (2) and (3) report the results. I find that, after the SEC FOF arrangement, there are a 0.39% increase in ETF shorting activity and a 6.28% increase in ETF short volume, supporting that the SEC FOF arrangement alleviates ETF short-sale constraints and increases ETF shorting activity. This increased ETF shorting activity will mechanically translate into increases in stock-level *ETF-based SA*. Yet, this time-series increase in *ETF-based SA* can potentially associate with unobservable time-varying characteristics, some of which can be associated with the error terms and bias the results.

I address this concern by exploring the cross-sectional difference in the effect of the SEC FOF arrangement on ETF shorting activity. As this increased ETF shorting activity translates into stock-level *ETF-based SA*, mechanically, the increase in *ETF-based SA* will be larger among high ETF ownership stocks than low ETF ownership stocks. I exploit this cross-sectional difference by comparing the impact of the SEC FOF arrangement on stocks with high ETF ownership to stocks with low ETF ownership, which helps control for unobservable time-varying characteristics. However, the variation in ETF ownership is not random and is highly correlated with market capitalization, meaning that this cross-sectional difference captures the differences in both ETF ownership and market capitalization. To address this concern, following the literature,[43] I focus on

[42]ETF short volume (ETF SV) is measured as a percentage of ETF short volume relative to the total ETF volume. The change in ETF short interest is measured as a percentage of the change in ETF short interest relative to the ETF share outstanding.

[43]See Crane, Michenaud, and Weston (2016) and Appel, Gormley, and Keim (2016)

around 200 stocks at the bottom of the Russell 1000 constituents and 200 stocks at the top of the Russell 2000 constituents, which are similar in market capitalization but differ in ETF coverage. I expect that the increase in *ETF-based SA* will be larger among top Russell 2000 stocks than bottom Russell 1000 stocks. In Column (4) of Table 2.8, I find that the coefficient of the interaction term between the post-SEC FOF indicator and the Russell 2000 indicator is positive and significant. This confirms that top Russell 2000 stocks have larger increases in *ETF-based SA* than bottom Russell 1000 stocks after the SEC FOF arrangement.

Subsequently, I compare the change in informational efficiency after the SEC FOF arrangement for bottom Russell 1000 stocks with the change in informational efficiency for top Russell 2000 stocks after the SEC FOF arrangement. By doing so, I can control for time-varying unobservables and draw causal inferences about the association between ETF shorting activity and informational efficiency. I formalize this comparison by estimating the IFERCs together with a difference-in-difference (diff-in-diff) setting in Eq. (2.13) as shown below. My sample period covers the first quarter of 2020 (one year before the SEC FOF implementation) to the first quarter of 2022 (one year following the SEC FOF implementation).

$$\begin{aligned} Return_{i,t} = {} & \beta_0 + \beta_1 Earn_{i,t+1} \times Rusl2k_i \times Post_t + \beta_2 Earn_{i,t+1} + \beta_3 Rusl2k_i \times Post_t \\ & + \gamma Controls_{i,t} + \delta + \epsilon_{i,t} \end{aligned} \tag{2.13}$$

Where i refers to firm i and t refers to quarter t. Return$_{i,t}$ is the buy-and-hold return from quarter $t-1$ to quarter t. Earn$_{i,t}$ is quarterly seasonally adjusted earnings at quarter t following Glosten et al. (2021). $Rusl2k_i$ is an indicator that equals one if firm i is in the top of Russell 2000 constituents, and zero if firm i is in the bottom of Russell 1000 constituents. $Post_t$ is the post-SEC FOF indicator that equals one if quarter t is after the implementation date of FOF arrangement and 0, otherwise. $Controls_{i,t}$ includes market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership. δ are firm fixed effects and year-quarter fixed effects. The standard errors are clustered by firm and year-quarter.

[Insert Table 2.9 near here]

Table 2.9 depicts the results. Column (1) reports a positive and significant coefficient of the interaction term $Earn_{i,t+1} \times Rusl2k_i \times Post_t$, indicating a larger IFERC for top Russell 2000 firms than for bottom Russell 1000 firms after the SEC FOF arrangement implementation. By decomposing quarterly earnings into systematic and firm-specific components and replacing quarterly earnings with its two components, I find that the coefficient estimate on the interaction term $Earn\_Sys_{i,t+1} \times Rusl2k_i \times Post_t$ is positive and significant, while the coefficient estimate on the interaction term $Earn\_Firm_{i,t+1} \times Rusl2k_i \times Post_t$ is statistically insignificant. This distinction supports that ETF shorting activity facilitates the incorporation of market-wide information into stock prices but not firm-specific information.

Meanwhile, I examine the effect of ETF shorting activity on four informational efficiency measures by adopting a diff-in-diff analysis as shown in Eq. (2.14).

$$\text{InfoEff}_{i,t} = \alpha_0 + \alpha_1 Rusl2k_i \times Post_t + Controls_{i,t} + \sigma + \epsilon_{i,t} \tag{2.14}$$

Where i refers to firm i and t refers to quarter t. $\text{InfoEff}_{i,t}$ is the informational efficiency measure. $Rusl2k_i$ is an indicator that equals one if firm i is in the top of Russell 2000 constituents, and zero if firm i is in the bottom of Russell 1000 constituents. $Post_t$ is the post-SEC FOF indicator that equals one if quarter t is after the implementation date of FOF arrangement and 0, otherwise. $Controls_{i,t}$ includes market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership. δ are firm fixed effects and year-quarter fixed effects. The standard errors are clustered by firm and year-quarter.

Columns (3) to (6) of Table 2.9 report the results using informational efficiency proxies. I find that the coefficients on the interaction term $(Rusl2k_i \times Post_t)$ are negative, indicating prices for top Russell 2000 stocks become more informationally efficient than prices for bottom Russell 1000 stocks after the SEC FOF arrangement implementation. Overall, these results are consistent with the increase in ETF shorting activity improving the informational efficiency of ETF underlying securities.

Notice that, as shown in Table 2.9, the coefficient estimators in Columns (4) and (5) are statistically insignificant. I provide two potential explanations for the insignificant results. First, my sample size is relatively small, making it challenging to filter out the noise in the regression analysis. Focusing on informational efficiency around quarterly earnings announcements with the short period surrounding the SEC FOF implementation reduces the sample size. Second, the SEC FOF sample coincides with the volatile COVID-19 eruption period around March 2020. To alleviate these concerns, I shift my focus to the daily informational efficiency measures generated by using intraday trade and quote data. Using daily informational efficiency measures allows for increased sample size and greater variation in regression analyses, and the large sample size also enables me to focus on the relatively short sample period to mitigate disruptions caused by the COVID-19 eruption period. I obtain the intraday trade and quote data from the Refinitiv Datascope Select dataset and compute two daily informational efficiency measures: the 15-second National Best Bid and Offer (NBBO) midpoint return autocorrelation and the variance ratio based on 1-minute and 15-second variances of NBBO midpoint returns.

The sample of daily informational efficiency regression covers the sample period from six months before the effective date of the SEC FOF arrangement, 19 July, 2020, to six months after the effective date of the SEC FOF arrangement, 19 July, 2021. I examine the effect of ETF shorting activity on two daily informational efficiency measures using a diff-in-diff analysis with the following regression specification:

$$\text{InfoEff}_{i,t} = \alpha_0 + \alpha_1 Rusl2k_i \times Post_t + Controls_{i,t} + \sigma + \epsilon_{i,t} \tag{2.15}$$

Where i refers to firm i and t refers to trading day t. $\text{InfoEff}_{i,t}$ is the daily informational efficiency measure. $Rusl2k_i$ is an indicator that equals one if firm i is in the top of Russell 2000 constituents, and zero if firm i is in the bottom of Russell 1000 constituents. $Post_t$ is the post-SEC FOF indicator that equals one if business date t is after the implementation date of FOF arrangement and 0, otherwise. $Controls_{i,t}$ includes market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership. δ are firm fixed effects and date fixed effects. The standard errors are clustered by firm and date.

[Insert Table 2.10 near here]

Table 2.10 reports the results using daily informational efficiency measures. The coefficients of the interaction term are negative and significant, indicating that prices for top Russell 2000 stocks become more efficient than those for bottom Russell 1000 stocks after the SEC FOF arrangement's implementation. These results are in line with my previous observations that ETF shorting activity improves informational efficiency for ETFs' underlying securities.

Additionally, I examine the "parallel trend" assumption in the diff-in-diff analysis in both quarterly and daily informational efficiency measures. I find that the effect of the SEC FOF arrangement on informational efficiency does not display significant differences between the top Russell 2000 stocks and the bottom Russell 1000 stocks in each quarter preceding the implementation of the SEC FOF arrangement.[44] This observation suggests that the improvement effect on informational efficiency in the diff-in-diff analysis is attributable to the implementation of the SEC FOF arrangement rather than the time trend. In conclusion, I confirm that the association between ETF shorting activity and informational efficiency is causal.

2.7 Robustness check

This section presents additional tests conducted to validate and strengthen the robustness of my findings. Although the association between informational efficiency and ETF shorting activity aligns with the hypothesis that traders short ETFs to speculate about their negative information, other alternatives are possible. In this section, I explore the alternative channels, through which ETF shorting activity could potentially impact the informational efficiency of ETFs' underlying securities.

2.7.1 Hedging-motivated Shorting

The first alternative channel, hedging-motivated shorting, is documented by Huang et al. (2021). They propose that ETFs enable informed traders to easily hedge their

[44]The "parallel trend" results are reported in Tables 2.A13 and 2.A14 of the Appendix.

exposure to industry risks by shorting industry ETFs. This allows informed traders to exploit the information without suffering from industry risks, thereby improving informational efficiency. To differentiate the effects of bearish-market shorting and hedging-motivated shorting, I compare the effect of ETF shorting activity on informational efficiency on stocks with high hedging demand to the effect on stocks with low hedging demand. If the improvement in informational efficiency is mainly driven by the hedging-motivated shorting, I expect that the effect of ETF shorting activity concentrates on stocks with high hedging demand rather than stocks with low hedging demand. I proxy the hedging demand using the CAPM beta as it captures the degree of co-movement with market returns. Intuitively, stocks that strongly co-move with the market are exposed to a high level of market risks and, therefore, have high hedging demand.[45] Specifically, for each stock and each quarter, I use daily returns to estimate the CAPM beta. Subsequently, I assign a stock to the high-(low-) beta group if its CAPM beta is above (below) the quarterly median of CAPM beta. I re-estimate the panel regression for this sample partition.

[Insert Table 2.11 near here]

Table 2.11 reports the regression results for the sample partition with different hedging demands. Columns (1) to (5) show the results for low-beta stocks, and I find that *ETF-based SA* increases the informational efficiency as evidenced by a stronger return-earnings association, a closer-to-one variance ratio, lower autocorrelation and a smaller PAED. Moreover, in Column (2), I find consistent evidence that the improvement in informational efficiency is driven by the better incorporation of market-wide information rather than firm-specific information. In contrast, Columns (6) to (10) do not exhibit such improvement in informational efficiency for high-beta stocks. Overall, I document the effect of ETF shorting activity in improving informational efficiency on low-beta stocks but not on high-beta stocks. These results indicate that the effect of ETF shorting activity on informational efficiency is not driven by the hedge function of ETF shorting activity.

[45]Precisely, the hedging demand in this section refers to the hedging demand from the downside risk. The stock that moves in the inverse direction of the market return also has high hedging demand. However, this hedging demand is not related to ETF shorting activity.

2.7.2 Operational shorting

Second, the improvement in informational efficiency could be driven by shorting activity that comes from market makers' liquidity provision activity. Evans et al. (2022) further show that ETF operational shorting accounts for around 20% of ETF short interest, and find that ETF operational shorting is associated with improved informational efficiency in ETFs' underlying securities. Nevertheless, operational shorting is unlikely to drive my previous results. This is because I use the change in ETF short interest rather than the level of ETF short interest to construct *ETF-based SA*, and taking the first difference in ETF short interest helps filter out a large proportion of operational shorting since operational shorting is persistent (Evans et al. (2022)).

To further alleviate this concern, I use the ETF shorting activity net of operational shorting to construct the adjusted ETF shorting activity: $\widehat{\textit{ETF-based SA}}$. I first form the ETF operational shorting activity measure following Evans et al. (2022).[46] Next, I subtract the ETF operational shorting activity measure from the total ETF shorting activity. I then use this remaining ETF shorting activity to form the $\widehat{\textit{ETF-based SA}}$ using Eq. (2.1) and Eq. (2.2), which does not contain ETF operational shorting activity. I re-estimate my previous regressions using this $\widehat{\textit{ETF-based SA}}$.

[Insert Table 2.12 near here]

Table 2.12 reports the regression results for $\widehat{\textit{ETF-based SA}}$. Columns (1) to (3) report the IFRECs. I find that $\widehat{\textit{ETF-based SA}}$ helps incorporate more market-wide earnings information into stock returns. Columns (4) to (7) report the results using four informational efficiency measures. I find that $\widehat{\textit{ETF-based SA}}$ is associated with a closer-to-one variance ratio, smaller autocorrelation and an attenuated PEAD. All these results are consistent with my previous findings, supporting that the effect of ETF shorting activity on the informational efficiency of underlying securities remains significant after filtering out ETF operational shorting activity.

[46] The calculation of ETF operational shorting activity is based on the daily buy/sell imbalance and the change in ETF share outstanding. I obtain the intraday ETF trade and quote data from the Refinitiv Datascope Select and compute the daily buy/sell imbalance for each ETF, where the trade sign is assigned based on Lee and Ready (1991) algorithm.

2.7.3 The association with ETF activity

Another alternative channel to consider is that the effect of improving informational efficiency might come from ETF activity (ETFA) rather than ETF shorting activity. Glosten et al. (2021) find that the ETF activity (measured by the change in ETF ownership) increases the informational efficiency of ETFs' underlying securities. By construction, the variation of *ETF-based SA* comes from two sources: (i) ETF constituents' weights and (ii) the change in ETF short interest, the former of which is associated with ETF activity.[47] It is crucial to distinguish whether the improvement in informational efficiency is driven by ETF activity or ETF shorting activity.

I examine this by controlling for the effect of ETF activity. I first construct the orthogonal ETF activity with respect to ETF shorting activity to isolate the effect of ETF shorting activity. Specifically, for each quarter, I estimate the following regression:

$$\text{ETFA}_{i,t} = \alpha_0 + \alpha_1 \textit{ETF-based SA}_{i,t} + \epsilon_{i,t} \tag{2.16}$$

Where i refers to firm i and t refers to fiscal quarter t. $\text{ETFA}_{i,t}$ is the ETF activity measured by the quarterly change in ETF ownership. The orthogonal ETF activity is obtained from residual $\epsilon_{i,t}$, which captures the variation in ETFA that cannot be explained by *ETF-based SA*. By controlling for the orthogonal ETF activity, I can separate the effect of ETF shorting activity on informational efficiency that is not confounded by ETF activity. I re-estimate the results controlling for orthogonal ETF activity.[48]

[Insert Table 2.13 near here]

Table 2.13 reports the regression results after controlling for ETF activity. Columns (1) to (2) report the IFRECs of ETF-based shorting activity. I find consistent results that *ETF-based SA* increases the return-earnings association, showing enhanced informational efficiency. Columns (3) to (6) report the results using four informational

[47]The variation of ETF activity comes from the variation in ETF market capitalization and ETF constituents' weight, while the variation of ETF shorting activity comes from the variation in ETF short interest and ETF constituents' weights.

[48]Specifically, I re-estimate the results with the modification that replaces the orthogonal institutional ownership with the orthogonal ETF activity because they are highly correlated, and including both in the regression leads to the multi-colinearity problem.

efficiency measures. I find that *ETF-based SA* is associated with a variance ratio closer to one, smaller autocorrelation and an attenuated PEAD. These results show that ETF shorting activity improves the informational efficiency of underlying securities even after controlling for ETF activity, indicating that the effect of ETF shorting activity in improving informational efficiency is distinct from the effect of ETF activity.

2.7.4 Synthetic shorting and short-sale constraints

Li and Zhu (2022) demonstrate that some traders exploit negative firm-specific information by forming a synthetic short position when the target stock is hard to short. A synthetic short position of a target stock is formed by shorting an ETF containing the target stock and going long on all ETF constituents except for the target stock. They show that synthetic short interest via ETFs can negatively predict future returns of short-constrained stocks, indicating synthetic shorting contains negative firm-specific information. As a result, synthetic shorting also allows informed traders to incorporate their firm-specific information into prices and improve informational efficiency. However, my previous results show that the improvement of informational efficiency is attributable to the better incorporation of market-wide earnings information rather than firm-specific information, suggesting that synthetic shorting is not the main channel through which ETF shorting activity improves informational efficiency around earnings announcements.

To reinforce this argument, I examine whether synthetic shorting enhances informational efficiency. If synthetic shorting does enhance informational efficiency, I expect that the effect of ETF shorting activity is more pronounced on short-constrained stocks than the effect on short-unconstrained stocks because synthetic shorting is most active among stocks that are difficult to short (i.e., short-constrained stocks). Otherwise, I expect that there are no differences in the effect of ETF shorting activity between short-constrained stocks and short-unconstrained stocks.

To test this alternative explanation, I study the variation of short-sale constraints in the Regulation SHO (Reg SHO) pilot program implemented in 2005. The Reg SHO removes short-sale constraints for 1000 randomly selected stocks (treatment stocks) in Russell 3000 constituents and keeps the short-sale constraints the same for the remaining Russell 3000 stocks (control stocks). This Reg SHO pilot program allows me to examine how short-sale constraints alter the association between ETF shorting activity and informational efficiency. The sample of the Reg SHO pilot program study contains all Russell 3000 stocks constituted based on the market capitalization in May 2004. I focus on the sample period from one year before the Reg SHO pilot program implementation, May 2004, to one year after the Reg SHO implementation, May 2006. I compare the effect of the Reg SHO pilot program on treatment stocks with the effect on control stocks using a diff-in-diff setting. Specifically, I compare the change in the IFERC for treatment stocks with the change in the IFERC for control stocks after the Reg SHO implementation using the following regression specification Eq. (2.17).

$$\begin{aligned} Return_{i,t} = \beta_0 + \beta_1 Earn_{i,t+1} \times \textit{ETF-based SA} \times Treatment_i \times Post_t + \beta_2 Earn_{i,t} + \\ \beta_3 \textit{ETF-based SA} + \beta_4 Treatment_i \times Post_t + \gamma Controls_{i,t} + \delta + \epsilon_{i,t} \end{aligned} \tag{2.17}$$

Where i represents a firm i and t represents a quarter t. $Treatment_i$ is the treatment indicator that equals one if firm i is a treatment firm in the Reg SHO pilot program and zero otherwise. $Post_t$ is the post-Reg SHO indicator that equals one if quarter t is after the effective date of the Reg SHO pilot program and zero otherwise. $Controls_{i,t}$ includes market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership. δ are firm fixed effects and year-quarter fixed effects. The standard errors are clustered by firm and year-quarter. I focus on the coefficient estimator of the quadruple interaction term β_1. A negative β_1 indicates that short-sale constraints strengthen the improvement effect of ETF shorting activity on informational efficiency. Additionally, I examine how the Reg SHO alters the effect of ETF shorting activity on four informational efficiency measures using the following Eq. (2.18):

$$\begin{aligned} \text{InfoEff}_{i,t} = \alpha_0 + \alpha_1 Treatment_i \times Post_t \times \textit{ETF-based SA}_{i,t} + \alpha_1 \textit{ETF-based SA} \\ \alpha_2 Treatment_i \times Post_t + Controls_{i,t} + \sigma + \epsilon_{i,t} \end{aligned} \tag{2.18}$$

I focus on the coefficient estimator of the triple interaction term α_1. A positive α_1 indicates that short-sale constraints strengthen the improvement effect of ETF shorting activity on informational efficiency. $Controls_{i,t}$ includes market capitalization, book-to-market ratio, idiosyncratic return volatility, seasonally adjusted earnings, past returns and orthogonal institutional ownership. δ are firm fixed effects and date fixed effects.

[Insert Table 2.14 near here]

I estimate regressions following the above regression specifications, and the results are reported in 2.14. Column (1) reports the IFERC of ETF-based shorting activity ratio in the Reg SHO pilot program. I find that the coefficient of the quadruple interaction term ($Earn_{i,t+1}$ × *ETF-based SA*$_{i,t}$ × $Treatment_i$ × $Post_t$) is statistically insignificant, indicating that there are no significant differences in the change in IFERC between treatment and control stocks after the implementation of the Reg SHO pilot program. This result do not provide evidence that short-sale constraints facilitate the incorporation of earnings information into the prices of ETFs' underlying securities. In Columns (2) to (5), the triple interaction term (*ETF-based SA*$_{i,t}$ × $Treatment_i$ × $Post_t$) coefficient estimators are also statistically insignificant. The insignificant coefficient estimators again support that short-sale constraints do not alter the association between ETF shorting activity and informational efficiency. These results support that the effect of ETF shorting activity in improving informational efficiency is not driven by the stocks where synthetic shorting is the most active, suggesting that synthetic shorting does not drive the improvement in informational efficiency.

2.7.5 Market and Industry ETFs

The key variable in this study, ETF-based SA, aggregates the shorting activity across all ETFs, a proportion of which are non-market ETFs. Integrating non-market ETFs shorting activity could potentially introduce complications, as it is improbable that traders with market-wide information would exploit their information using non-market ETFs. Consequently, the inclusion of non-market ETFs may add noise and bias originating from non-market ETFs short sellers. However, the inclusion of non-market ETFs is unlikely to influence my results because ETF shorting activity predominantly concentrates on market ETFs.[49] To further mitigate this concern, I re-calculated the ETF-based

[49]The 10 most shorted ETFs in my sample are all market ETFs and account for more than 90% of the total ETF short interest.

SA using only market ETF shorting activity and re-estimate all the regression results, inclusive of the regressions featured in robustness checks. As detailed in the Appendix, I find that all my results remain qualitatively consistent.[50]

For completeness, I also investigate a key subset of non-market ETFs: industry ETFs, which are widely used as investment vehicles for hedging industry or sector-specific risks. Huang et al. (2021) provide evidence that the industry ETFs short interest increases concurrently with hedge fund holdings in constituent stocks ahead of positive earnings surprises, reflecting a long-stock and short-industry-ETF hedging strategy. To explore this further, I start by identifying industry ETFs in U.S. equity markets using Lipper classification codes from the CRSP mutual fund dataset. I re-calculated the ETF-level industry ETF shorting activity and stock-level industry ETF-based shorting activity (iETF-based SA) using only shorting activity from industry ETFs and re-estimated the main regression results.[51] At the ETF level, I find that industry ETF shorting activity positively predicts future ETF returns. Similarly, at the stock level, industry ETF shorting activity positively predicts future stock returns. This positive return predictability aligns with Huang et al. (2021), which argue that industry ETFs are often shorted to hedge against industry risks when underlying stocks experience favorable news. However, these findings do not necessarily contradict my earlier results. First, industry ETF shorting activity constitutes a relatively small portion of the total short interest. As noted previously, the ten most shorted ETFs in the U.S. equity markets are all market ETFs, comprising over 90% of total ETF short interest. Second, the magnitude of the positive return predictability is relatively modest compared to the stronger negative return predictability found in market ETFs.

From the perspective of information efficiency, I find that industry ETF-based shorting activity facilitates the incorporation of future earnings information into current stock returns and is associated with lower variance ratio and a reduction in PEAD, indicating enhanced information efficiency. The improvement is consistent with Huang et al. (2021) that industry ETFs, through their hedging capabilities, enable informed traders to more effectively and easily manage sector-specific risks, thereby contributing to improved information efficiency.

[50] Tables 2.A3 - 2.A11 in the Appendix present the results of market ETF shorting activity.
[51] Tables 2.A20 and 2.A21 in the Appendix present the results of industry ETF shorting activity.

To sum up, I examine several alternatives that may account for the improvement in information efficiency, and I find that all alternative explanations above are difficult to explain the improvement in information efficiency. Therefore, I conclude that my results collectively support that bearish-market shorting causes an improvement in the information efficiency of ETFs' underlying securities.

2.8 Conclusion

The rapid growth in ETF size and trading volume has led to an increasing trend in academic literature examining the effects of ETF activity on financial markets. However, most of the literature focuses on long-position ETF activity, leaving studies on ETF shorting activity relatively scarce. In this chapter, I explore the information content in ETF shorting activity and find that it negatively predicts future returns of ETFs and their underlying securities, showing that ETF shorting activity is informed. I further examine the association between ETF shorting activity and the informational efficiency of ETFs' underlying securities, finding that ETF shorting activity facilitates the incorporation of negative market-wide information but not firm-specific information into these securities. Investigating several alternative explanations of ETF shorting activity that might potentially account for the improvement in informational efficiency, such as shorting for hedging purposes, operational shorting, or synthetic shorting, I find my results remain robust to all these explanations. In conclusion, my findings support that ETF short sellers short ETFs to exploit their negative market-wide information, which is later transmitted in ETF's underlying securities through ETF arbitrage and improves their informational efficiency.

Table 2.1: Summary statistics

This table summarizes the number of observations, mean, median, standard deviation, and the 25th and 75th percentiles of variables used in my regression analysis. Panel A reports ETF characteristics at the ETF-bi-week level, and Panel B reports stock characteristics at the stock-quarter level. ETF SI is the percentage of ETF short interest scaled by the market capitalization. Δ ETF SI is the percentage change in ETF short interest scaled by the market capitalization. Price is the ETF close price. MktCap is the market capitalization in $1,000,000,000. DollarVolume is the bi-weekly consolidated dollar volume in the ETF sample in $1,000,000 and quarterly consolidated dollar volume in stock sample in $1,000,000,000. ETF-based SA is the ETF-based shorting activity within a quarter. Return is the quarterly return in between two consecutive quarterly earnings reporting dates. Jump is the Weller(2018) jump ratio. PEAD is the post-earnings announcement drift measured by the abnormal return from the first day to one month after the earnings announcements. |1-VR| is the absolute value of one minus the ratio of variance of 2-day returns divided by two times the variance of 1-day returns from 21 days before the earnings announcements to 2 days after the earnings announcements. |AC| is the absolute value of daily returns auto-correlation from 21 days before the earnings announcements to 2 days after the earnings announcements. Earn is the seasonally adjusted earnings innovation scaled by price. IO is the orthogonal institutional ownership with respect to ETF-based SA. IdiosynVol is the idiosyncratic volatility estimated based on the Fama-French (1993) three-factor model. BTM is the book-to-market ratio. Loss is a dummy variable that takes the value of one if income before extraordinary items is negative, and zero otherwise. ETF_SR is the percentage of ETF-based short ratio following Li and Zhu (2022). The sample period covers from January, 2002 to December, 2021.

Panel A: Summary Statistics (ETF sample: bi-weekly observations)

Variable	Count	Mean	Std	25%	Median	75%
ETF SI (%)	254,829	4.51	11.99	0.19	0.7	2.84
Δ ETF SI (%)	252,526	-0.03	4.32	-0.25	0.0	0.25
Monthly Return (%)	253,078	0.56	6.63	-2.19	0.8	3.74
Price ($)	254,829	51.17	38.41	25.29	38.2	63.97
MktCap ($ billions)	254,829	1.89	9.36	0.02	0.12	0.64
DollarVolume ($ millions)	254,829	9.87	100.38	0.02	0.13	0.9

Panel B: Summary Statistics (Stock sample: quarterly observations)

Variable	Count	Mean	Std	25%	Median	75%
ETF-based SA (%)	322,092	0.02	0.16	-0.0	0.0	0.02
SI (%)	286,446	4.98	24.65	0.9	2.56	5.95
MktCap ($ billions)	322,091	5.15	28.95	0.13	0.5	2.12
DollarVolume ($ billions)	300,089	3.1	190.62	0.03	0.23	1.34
Quarterly Return (%)	325,814	0.01	0.23	-0.09	0.02	0.12
Jump (%)	259,487	42.72	80.53	0.85	40.73	83.81
PEAD (%)	232,722	7.01	9.36	0.14	3.82	9.80
|1-VR|	322,066	0.33	0.12	0.24	0.32	0.4
|AC|	322,065	0.19	0.14	0.08	0.17	0.28
Earn	272,578	-0.0	0.05	-0.01	0.0	0.01
IO (%)	241,789	0.07	3.29	-1.34	-0.2	1.22
IdiosynVol	324,480	0.03	0.02	0.01	0.02	0.03
BTM	291,968	0.63	0.57	0.26	0.48	0.81
Loss	325,814	0.3	0.46	0.0	0.0	1.0
ETF_SR (%)	322,092	0.41	0.58	0.0	0.1	0.69

Table 2.2: ETF and stock Portfolio Returns Sorted on the ETF shorting activity

This table reports the monthly average excess returns, CAPM alpha, Fama and French (1993) 3-factor alpha, and Fama and French (1993) and Carhart (1997) 4-factor alpha (in percentage) for ETF and stock portfolios, as well as for the long-short portfolios (High-Low). Panels A and B report for five quintile ETF portfolios sorted by the ETF shorting activity (defined as the percentage change in ETF short interest: ΔETF SI). Panels C and D report for five quintile stock portfolios sorted by the ETF-based shorting activity (ETF-based SA). At the end of each month, all securities are sorted into quintile portfolios based on either ETF shorting activity at ETF-level or ETF-based shorting activity at the stock level, and a long-short portfolio is formed by buying the lowest quintile and shorting the highest quintile portfolio. Portfolios are rebalanced monthly and portfolio returns are computed over the next month. Panels A and C report results for equally weighted portfolios, and Panels B and D show results for value-weighted portfolios. t-statistics from errors with Newey-West correction of four lags are reported. The sample runs from January 2002 to December 2021.

Panel A: Performance of Equal-weighted ETF portfolio sorted by ΔETF SI

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	Carhart(97)	t-stat
Low	-0.24	-1.19	-0.54	-2.24	-0.55	-2.30	-0.53	-2.14
2	0.40	1.67	0.10	0.38	0.09	0.33	0.13	0.46
3	0.40	1.56	0.08	0.28	0.07	0.24	0.11	0.35
4	0.31	1.33	-0.00	-0.01	-0.01	-0.05	0.02	0.08
High	-0.49	-2.34	-0.79	-2.97	-0.80	-2.92	-0.79	-2.74
High-Low	-0.25	-2.08	-0.31	-2.60	-0.30	-2.64	-0.31	-2.70

Panel B: Performance of Value-weighted ETF portfolio sorted by ΔETF SI

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	Carhart(97)	t-stat
Low	0.45	1.86	0.16	0.56	0.14	0.51	0.17	0.61
2	0.49	2.04	0.20	0.74	0.19	0.67	0.22	0.74
3	0.57	2.15	0.28	0.87	0.26	0.79	0.30	0.90
4	0.60	2.41	0.30	1.05	0.28	1.00	0.32	1.10
High	0.27	1.09	-0.03	-0.09	-0.05	-0.16	-0.02	-0.06
High-Low	-0.18	-1.79	-0.16	-1.99	-0.17	-2.09	-0.18	-2.14

Panel C: Performance of Equal-weighted stock portfolio sorted by ETF-based SA

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	FFC(97)	t-stat
Low	1.31	3.56	0.42	2.31	0.35	3.25	0.37	3.35
2	1.08	2.93	0.20	0.99	0.12	1.07	0.17	1.46
3	0.96	2.64	0.07	0.45	0.00	0.06	0.07	0.96
4	1.11	2.91	0.20	1.18	0.12	1.28	0.19	2.15
High	0.99	2.71	0.11	0.57	0.04	0.44	0.07	0.79
High-Low	-0.32	-2.12	-0.31	-1.90	-0.31	-2.00	-0.30	-1.92

Panel D: Performance of Value-weighted stock portfolio sorted by ETF-based SA

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	FFC(97)	t-stat
Low	1.41	4.32	0.61	3.91	0.56	4.42	0.53	3.98
2	1.24	3.90	0.45	3.56	0.40	4.29	0.37	4.17
3	0.89	2.90	0.10	1.02	0.08	0.85	0.09	0.93
4	0.70	2.22	-0.10	-1.04	-0.13	-1.29	-0.14	-1.41
High	0.97	3.06	0.17	1.20	0.14	1.21	0.12	1.00
High-Low	-0.44	-2.57	-0.43	-2.51	-0.42	-2.40	-0.41	-2.33

Table 2.3: The return predictability of bi-weekly ETF shorting activity

This table presents coefficient estimates from panel regressions of ETF returns on the ETF shorting activity (ΔETF SI) and the stock-level ETF-based shorting activity (ETF-based SA) and controls. Columns (1) to (4) report the coefficient estimates of ETF-level regressions. Columns (5) to (8) report the coefficient estimates of stock-level regressions. For dependent variable(s), Ret_{t+20} is next-20-day returns (in percentage), after skipping one trading day. For independent variable(s), the variables of interest are the ETF shorting activity (ΔETF SI) in Columns (1) to (4) and the ETF-base shorting activity(ETF-based SA) in Columns (5) to (8). For control variables, ETF_SR is the ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. IdiosynVol the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF-based Flow is the aggregate flow via ETFs scaled by the market capitalization. ETF Flow is the change of ETF outstanding shares scaled by the total shares outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	ETF sample				Stock sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}
ΔETF SI	-0.0218**	-0.0226**	-0.0206**	-0.0237**				
	(-3.14)	(-3.23)	(-2.41)	(-2.66)				
ETF-based SA					-0.871***	-1.014***	-0.776***	-0.772***
					(-5.86)	(-6.72)	(-4.86)	(-4.84)
ETF_SR						0.211***	-0.367***	-0.365***
						(4.32)	(-6.81)	(-6.77)
BTM							-1.419***	-1.420***
							(-44.60)	(-44.63)
Log(MktCap)		-0.310***	-0.313***	-0.308***			-2.295***	-2.295***
		(-17.11)	(-17.24)	(-17.02)			(-59.96)	(-59.95)
IdiosynVol		-35.32***	-29.71***	-28.74***			0.283	1.180
		(-5.62)	(-4.80)	(-4.65)			(0.14)	(0.55)
REV		0.152	0.145	0.560			-2.636***	-2.638***
		(0.34)	(0.33)	(1.23)			(-17.25)	(-17.27)
Turnover			-0.00392***	-0.00391***				-0.00215
			(-9.78)	(-9.78)				(-1.17)
ETF Flow				-1.434***				
				(-5.90)				
Put				0.236				
				(1.12)				
ETF-based Flow								0.00000344***
								(7.59)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bi-week fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	138,374	134,799	134,799	134,715	926,341	926,341	789,894	789,893
Adjusted R2	0.320	0.324	0.325	0.325	0.230	0.230	0.245	0.245

Table 2.4: The return predictability of market ETF shorting activity on macroeconomic news period

This table presents coefficient estimates from panel regressions of market ETF returns on the market ETF shorting activity (ΔETF SI), and controls. Columns (1) to (4) report the coefficient estimates during the full sample. Columns (5) to (6) report the coefficient estimates on the macroeconomic new periods. Columns (7) to (8) report the coefficient estimates on the non-macroeconomic new periods. For dependent variable(s), Ret_{t+20} is next-20-day returns (in percentage), after skipping one trading day. For independent variable(s), the variables of interest are the ETF shorting activity(Δ ETF SI). For control variables, Log(MktCap) is the logarithm of market capitalization. IdiosynVol is the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF Flow is the change of ETF outstanding share scaled by the total share outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Market ETF Sample				Macro news periods		Non-Macro news periods	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}
Δ ETF SI	-0.00170***	-0.00175***	-0.00500**	-0.00577**	-0.0129**	-0.0129**	0.00509	0.00507
	(-2.75)	(-2.76)	(-2.06)	(-2.34)	(-2.26)	(-2.26)	(0.84)	(0.84)
Log(MktCap)		-0.143***	-0.143***	-0.319***	-0.159***	-0.268***	-0.103***	-0.288***
		(-9.17)	(-9.18)	(-4.83)	(-6.93)	(-2.77)	(-4.33)	(-2.97)
IdiosynVol		-9.041	-9.129	-41.10	-7.242	-263.8***	-7.009	64.37
		(-1.47)	(-1.48)	(-0.79)	(-0.74)	(-2.93)	(-0.67)	(1.56)
REV		0.367	0.323	-0.652	7.484***	6.500*	5.336***	4.954**
		(0.44)	(0.38)	(-0.50)	(3.71)	(1.71)	(3.88)	(2.20)
Turnover			-0.000219**	-0.000352	-0.000264**	-0.00204*	-0.000156	0.000485
			(-2.37)	(-0.55)	(-2.21)	(-1.74)	(-0.30)	(0.51)
ETF Flow				0.0230		-0.0510		0.947
				(0.20)		(-0.43)		(1.40)
Put				-0.00388		0.00502		-0.00617
				(-0.71)		(0.56)		(-0.74)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bi-week fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,426	32,757	30,240	30,240	8,412	8,412	5,843	5,843
Adjusted R2	0.843	0.844	0.850	0.851	0.811	0.811	0.804	0.804

Table 2.5: Return predictability of ETF short volume on Macro news dates

This table presents coefficient estimates from panel regressions of ETF returns on ETF short volume ratio and controls. For dependent variable(s), Ret_{t+1} is next-day stock returns, after skipping one trading day. Ret_{t+20} is next-20-day stock returns (in percentage), after skipping one trading day. For independent variable(s), the variables of interest are the adjusted short ratio (ETF SVR_t) and the average adjusted short ratio over the last five trading days ($\text{ETF SVR}_{t-5,t}$). For control variables, Log(MktCap) is the logarithm of market capitalization. MOM is the cumulative return measured from 12 months prior to two months ago. REV is the short-term reversal measure that equals to the lagged monthly return. IdiosynVol is the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. Turnover is the daily consolidated volume scaled by the market capitalization. t-statistics from robust standard errors clustered by stock and date are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Macro news dates		Non-Macro news dates	
	(1)	(2)	(3)	(4)
	Ret_{t+1}	Ret_{t+20}	Ret_{t+1}	Ret_{t+20}
ETF SVR_t	-0.0191***		0.00706	
	(-2.75)		(0.79)	
$\text{ETF SVR}_{t-5,t}$		-0.0202**		0.0102
		(-2.20)		(1.15)
Log(MktCap)	-0.0175***	-0.253***	0.0113**	-0.301***
	(-4.90)	(-16.23)	(2.50)	(-19.84)
MOM	0.00968	-1.224***	-0.0780*	-1.083***
	(0.29)	(-9.00)	(-1.92)	(-8.18)
REV	0.226**	-2.724***	-0.738***	-1.661***
	(2.08)	(-6.62)	(-6.23)	(-4.27)
IdiosynVol	-5.495***	-66.59***	-1.980	-30.77***
	(-2.66)	(-8.21)	(-0.90)	(-4.11)
Turnover	-0.0824**	-0.644***	0.00439	-0.342**
	(-2.09)	(-3.57)	(0.17)	(-2.45)
Stock fixed effects	Yes	Yes	Yes	Yes
Date fixed effects	Yes	Yes	Yes	Yes
Observations	145,677	145,392	239,888	239,396
Adjusted R2	0.788	0.796	0.780	0.812

Table 2.6: ETF-based shorting activity and informational efficiency

This table presents coefficient estimates from panel regressions of the ETF-based shorting activity (ETF-based SA) on informational efficiency. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variables of interest are the interaction term between the seasonally adjusted earnings innovation at t+1 ($Earn_{t+1}$) and ETF-based SA in Columns (1) and (2), which captures the IFERC, and ETF-based SA in Columns (3) to (6). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), lag ETF-based shorting activity, lag seasonally adjusted earnings innovation, scaled by price, ($Earn_t$), the interaction term between $Earn_t$ and ETF-based SA, quarterly stock returns at t+1 (Ret_{t+1}), the interaction term between $Earn_t$ and ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price, $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Ret_t	(3) Jump	(4) |1-VR|	(5) $\|AC\|$	(6) PEAD
$Earn_{t+1}$ × ETF-based SA	0.512***	0.356**				
	(3.34)	(2.14)				
$Earn_{t+1}$	0.421***	0.397***				
	(33.93)	(23.37)				
ETF-based SA	0.0186***	0.00592	-1.284	-0.00644***	-0.00535**	-0.656***
	(5.14)	(1.17)	(-0.83)	(-2.94)	(-2.29)	(-2.79)
IO		0.00683***	0.00702	-0.000300***	-0.000394***	0.274***
		(35.57)	(0.10)	(-3.48)	(-3.85)	(25.17)
Log(MktCap)		0.0662***	-0.676	-0.00458***	-0.00891***	1.317
		(44.56)	(-1.40)	(-7.31)	(-12.09)	(17.59)
IdiosynVol		-0.726***	-503.6***	0.0459	-0.229***	-14.56***
		(-7.28)	(-22.95)	(1.50)	(-6.38)	(-3.09)
BTM		-0.00457*	1.181*	0.000235	0.00302***	-3.227***
		(-1.85)	(1.68)	(0.25)	(2.72)	(-24.89)
Loss		-0.00470*	-1.659**	-0.000703	-0.00175	-0.775***
		(-1.69)	(-2.49)	(-0.78)	(-1.64)	(-7.63)
SI		-0.00914***	2.635***	-0.00247***	-0.00481***	-0.0270
		(-6.67)	(4.53)	(-3.27)	(-5.41)	(-0.35)
Ret_{t+1}		-0.0702***	-0.258	-0.00530***	-0.00424*	
		(-12.20)	(-0.17)	(-2.62)	(-1.77)	
$\|Earn_{t+1}\|$		0.337***	16.79***	-0.0234***	-0.0147**	-0.401***
		(20.82)	(4.12)	(-4.25)	(-2.23)	(9.96)
$\|Earn_{t+1}\|$ × ETF-based SA		0.690***	120.9*	0.161*	0.0788	25.52**
		(3.27)	(1.83)	(1.77)	(0.78)	(2.16)
$Earn_{t-1}$		0.0399**				
		(2.56)				
$Earn_{t-1}$ × ETF-based SA		0.119				
		(0.74)				
$Earn_t$		0.198***				
		(12.36)				
$Earn_t$ × ETF-based SA		0.223				
		(1.31)				
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	273,345	170,637	161,540	200,467	200,467	198,891
Adjusted R2	0.011	0.074	0.003	0.001	0.002	0.028

Table 2.7: ETF-based shorting activity and informational efficiency: market-wide v.s. firm-specific information

This table presents coefficient estimates from a panel regression of the ETF-based shorting activity (ETF-based SA) on informational efficiency based on two earnings components: systematic earnings and firm-specific earnings. The dependent variable (Ret_t) is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. For independent variable(s), the variables of interest are the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and ETF-based SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and ETF-based SA. Control variables include seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), the idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), the interaction term between $EarnSys_t$ and ETF-based SA, quarterly stock returns at t+1, the interaction term between $Earn\_Firm_t$ and ETF-based SA, quarterly stock returns at t+1, the interaction term between $Earn_t$ and ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) Ret_t
$Earn\_Sys_{t+1}$ × ETF-based SA	0.595**
	(2.12)
$Earn\_Firm_{t+1}$ × ETF-based SA	0.212
	(1.38)
$Earn\_Sys_{t+1}$	0.713***
	(21.41)
$Earn\_Firm_{t+1}$	0.339***
	(20.15)
ETF-based SA	-0.00908***
	(-2.90)
$Earn\_Sys_t$	0.0934***
	(2.79)
$Earn\_Firm_t$	0.256***
	(14.88)
IO	0.00665***
	(34.51)
Log(MktCap)	0.0626***
	(46.71)
IdiosynVol	-0.0977
	(-1.02)
BTM	0.0112***
	(4.70)
Loss	-0.00209
	(-1.17)
SI	-0.0154***
	(-11.02)
$Return_{t+1}$	-0.0792***
	(-16.90)
$Earn\_Sys_t$ × ETF-based SA	0.175
	(0.63)
$Earn\_Firm_t$ × ETF-based SA	0.322**
	(2.10)
$Return_{t-4,t-1}$	0.0121***
	(7.06)
Firm fixed effects	Yes
Quarter fixed effects	Yes
Observations	173,181
Adjusted R2	0.067

Table 2.8: The SEC FOF arrangement implementation and ETF shorting activity

This table presents coefficient estimates from panel regressions on the effect of the SEC FOF arrangement. For dependent variable(s), IO is the quarterly institutional ownership in percentage on ETFs, ETF SV is the percentage of ETF short volume relative to the total ETF volume, ΔETF SI is the percentage of change in ETF short interest relative to the ETF share outstanding and ETF-based SA is the ETF-based shorting activity. For independent variable(s), the variables of interest are a post-SEC FOF indicator variable (Post_t) that takes value of 1 if after the SEC FOF arrangement, and 0 otherwise for Columns (1) to (3), and the interaction term between Post_t and Rusl2k_i, the latter of which is a treatment indicator variable that takes value of 1 if stock i is in the top Russell 2000 stocks, and 0 if stock i is in the bottom Russell 1000 stocks for Column (4). Control variables include the inverse of the close price ($1/Price$), the logarithm of consolidated dollar volume (Log(Dollar Volume)), the realized volatility based on daily stock returns, ETF fixed effects, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and ETF in Columns (1) - (3) and Year-quarter and firm in Column (4) are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) IO	(2) ETF SV	(3) ΔETF SI	(4) ETF-based SA
$Post_t$	2.718***	6.284***	0.386***	
	(3.36)	(16.18)	(6.24)	
$\text{Rusl2k}_i \times \text{Post}_t$				0.00542***
				(2.93)
1/Price	-11.57***	181152.3	-9.842***	-0.0133***
	(-6.11)	(1.48)	(-6.54)	(-2.62)
Log(Dollar Volume)	0.122***	19643.9***	0.194***	0.00108
	(9.67)	(33.98)	(18.94)	(1.34)
RealVol				0.00140**
				(2.44)
ETF fixed effects	Yes	Yes	Yes	No
Firm fixed effects	No	No	No	Yes
Year-quarter fixed effects	No	No	No	Yes
Observations	5,028	327,470	26,038	21,192
Adjusted R2	0.119	0.060	0.045	0.023

Table 2.9: The SEC FOF arrangement implementation and informational efficiency (quarterly)

This table presents coefficient estimates from panel regressions of the effect of the SEC FOF arrangement on information efficiency (quarterly metrics). In the diff-in-diff analysis, I study the implementation of the SEC FOF arrangement. I assign stocks in the top of Russell 2000 constituents in the treatment group and stocks in the bottom of Russell 1000 constituents in the control group. I examine the effect of the SEC FOF arrangement in the (-1year,+1year) window around the effective date of the SEC FOF arrangement. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variables of interest are the triple interaction term among the treatment indicator (Rusl2k), the post-SEC FOF indicator (Post) and the future seasonally adjusted earnings ($Earn$) in Column (1), the triple interaction term among the treatment indicator (Rusl2k), the post-SEC FOF indicator (Post) and the seasonally adjusted systematic earnings innovation ($Earn\_Sys$) and the triple interaction term among the treatment indicator (Rusl2k), the post-SEC FOF indicator (Post) and the seasonally adjusted idiosyncratic earnings innovation ($Earn\_Firm$) in Column (2), and the interaction term among the treatment indicator (Rusl2k) and the post-SEC FOF indicator (Post) in Columns (3) to (6). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by date and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	The SEC FOF arrangement (Quarterly)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Ret_t	Ret_t	Jump	|1-VR|	$\|AC\|$	PEAD
$Earn_{t+1}$ × $Rusl2k_i$ × $Post_t$	1.769^{**}					
	(1.97)					
$Earn\_Sys_{t+1}$ × $Rusl2k_i$ × $Post_t$		5.185^{***}				
		(3.12)				
$Earn\_Firm_{t+1}$ × $Rusl2k_i$ × $Post_t$		1.402				
		(1.46)				
$Rusl2k_i$ × $Post_t$	-0.139^{***}	-0.136^{***}	-13.58^{*}	0.0108	0.00363	-1.263^{*}
	(-8.47)	(-7.22)	(-1.71)	(1.00)	(0.29)	(-1.90)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls×$Rusl2k_i$, Control×$Post_t$	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,677	1,618	1,423	1,865	1,865	1,391
Adjusted R2	0.240	0.248	0.014	0.021	0.027	0.016

Table 2.10: The SEC FOF arrangement implementation and informational efficiency (daily)

This table presents coefficient estimates from panel regressions of the effect of the SEC FOF arrangement on information efficiency (daily metrics). For dependent variable(s), |1-VR| is the absolute difference between one and the variance ratio based on 15-second and 1-minute EBBO midpoint return. $|AC|$ is the 15-second NBBO midpoint return auto-correlation. For independent variable(s), the variable of interest is the interaction term between Post_t and Rusl2k_i. Post_t is a post-event indicator variable that takes value of 1 if after the SEC FOF arrangement, and 0 otherwise, Rusl2k_i is an indicator variable that takes value of 1 if stock i is in the top Russell 2000 stocks, and 0 if stock i is in the bottom Russell 1000 stocks. the variable of interest is the interaction term between Rusl2k_i and Post_t. Control variables include the logarithm of dollar consolidated volume (Log(Dollar Volume)), the inverse of the close price (1/Price), 15-second NBBO midpoint realized volatility (RealVol), firm fixed effects and data fixed effects. t-statistics from robust standard errors clustered by date and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	The SEC FOF arrangement (Daily)	
	(1)	(2)
	|1-VR|	|AC|
$\text{Rusl2k}_i \times \text{Post}_t$	-0.00385***	-0.00292***
	(-2.97)	(-3.29)
RealVol	0.0293***	0.0225***
	(14.39)	(15.60)
Log(Dollar Volume)	0.00115	0.000530
	(1.40)	(0.95)
1/Price	-0.794***	-0.544***
	(-6.60)	(-6.38)
Firm fixed effects	Yes	Yes
Date fixed effects	Yes	Yes
Observations	82,960	82,960
Adjusted R2	0.029	0.036

Table 2.11: ETF-based shorting activity and informational efficiency by hedging demand

This table presents coefficient estimates from panel regressions of ETF-based shorting activity (ETF-based SA). Columns (1) to (5) present the results for stocks with low market beta, and Columns (6) to (10) present the results for stocks with high market beta. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variables of interest are the interaction term between the seasonally adjusted earnings innovation ($Earn$) at t+1 and ETF-based SA in Columns (1) and (6), which captures the FERC, the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and ETF-based SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and ETF-based SA in Columns (2) and (7), and ETF-based SA in Columns (3) to (5) and Columns (8) to (10). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), the interaction term between $Earn_t$ and ETF-based SA, quarterly stock returns at t+1, the interaction term between $Earn_t$ and ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	low market beta						high market beta					
	(1) Ret$_t$	(2) Ret$_t$	(3) Jump	(4) |1-VR|	(5) $\|AC\|$	(6) PEAD	(7) Ret$_t$	(8) Ret$_t$	(9) Jump	(10) |1-VR|	(11) $\|AC\|$	(12) PEAD
$Earn_{t+1}$ × ETF-based SA	0.463***						0.333					
	(2.59)						(1.48)					
$Earn\_Sys_{t+1}$ × ETF-based SA		0.656**						0.321				
		(2.27)						(0.84)				
$Earn\_Firm_{t+1}$ × ETF-based SA		0.183						0.105				
		(1.11)						(0.51)				
ETF-based SA	0.0176**	0.00804	-2.010	-0.0101***	-0.00830**	-0.739**	-0.0269***	-0.0250***	0.221	-0.00272	-0.00196	-0.0419
	(2.56)	(1.25)	(-0.93)	(-2.94)	(-2.50)	(-2.40)	(-3.77)	(-4.41)	(0.10)	(-0.91)	(-0.56)	(-0.09)
$Earn_{t+1}$	0.304***					0.369***						
	(17.96)					(16.90)						
$Earn\_Sys_{t+1}$		0.454***					0.744***					
		(13.64)					(16.35)					
$Earn\_Firm_{t+1}$		0.240***					0.340***					
		(13.33)					(15.90)					
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls ×$Earn$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	102,505	101,647	91,819	112,841	112,841	111,966	67,315	72,829	69,087	86,818	86,818	86,101
Adjusted R2	0.043	0.042	0.002	0.001	0.001	0.007	0.088	0.081	0.004	0.001	0.003	0.028

Table 2.12: The operational-shorting-adjusted ETF-based shorting activity and informational efficiency

This table presents coefficient estimates from panel regressions of the adjusted ETF-based shorting activity ($\widehat{\text{ETF-based SA}}$) by operational shorting. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variables of interest are the interaction term between the seasonally adjusted earnings innovation at t+1 and $\widehat{\text{ETF-based SA}}$ in Columns (1) and (2), which captures the FERC, the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and $\widehat{\text{ETF-based SA}}$ and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and $\widehat{\text{ETF-based SA}}$ in Column (3), and $\widehat{\text{ETF-based SA}}$ in Columns (4) to (7). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the adjusted ETF-based shorting activity ($\widehat{\text{ETF-based SA}}$), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag adjusted ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Ret_t	(3) Ret_t	(4) Jump	(5) |1-VR|	(6) |AC|	(7) PEAD
$Earn_{t+1} \times \widehat{\text{ETF-based SA}}$	0.353**	0.318*					
	(2.14)	(1.83)					
$Earn\_Sys_{t+1} \times \widehat{\text{ETF-based SA}}$			0.588**				
			(1.98)				
$Earn\_Firm_{t+1} \times \widehat{\text{ETF-based SA}}$			0.184				
			(1.14)				
$\widehat{\text{ETF-based SA}}$	0.0198***	-0.00114	-0.0147***	0.382	-0.00378**	-0.00334	-0.311**
	(6.26)	(-0.30)	(-4.53)	(0.23)	(-2.02)	(-1.62)	(-2.29)
$Earn_{t+1}$	0.382***	0.388***					
	(29.20)	(24.19)					
$Earn\_Sys_{t+1}$			0.694***				
			(19.20)				
$Earn\_Firm_{t+1}$			0.338***				
			(15.81)				
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls × Earn	No	Yes	Yes	No	No	No	No
Observations	199,169	145,495	173,116	161,076	199,470	199,470	199,763
Adjusted R2	0.011	0.074	0.087	0.003	0.001	0.002	0.027

Table 2.13: ETF-based shorting activity, ETF activity, and informational efficiency

This table presents coefficient estimates from panel regressions of the ETF-based shorting activity (ETF-based SA) by controlling the effect of ETF activity. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variables of interest are the interaction term between the seasonally adjusted earnings innovation at t+1 and ETF-based SA in Columns (1) and (2), which captures the FERC, the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and ETF-based SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and ETF-based SA in Column (3), and ETF-based SA in Columns (4) to (7). Control variables include the orthogonal ETF activity ($ETFA$), seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the ETF-based shorting activity (ETF-based SA), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Ret_t	(3) Jump	(4) |1-VR|	(5) |AC|	(6) PEAD
$Earn_{t+1}$ × ETF-based SA	0.386**					
	(2.24)					
$Earn\_Sys_{t+1}$ × ETF-based SA		0.616**				
		(1.99)				
$Earn\_Firm_{t+1}$ × ETF-based SA		0.247				
		(1.52)				
ETF-based SA	0.00878	-0.00507	-1.002	-0.00598**	-0.00504**	-0.531**
	(1.39)	(-0.87)	(-0.63)	(-2.27)	(-2.10)	(-2.00)
$Earn_{t+1}$	0.406***					
	(24.72)					
ETFA	0.00601***	0.00710***	0.543	-0.000760*	-0.000905*	0.213***
	(7.03)	(8.06)	(1.61)	(-1.69)	(-1.74)	(4.13)
$ETFA \times Earn_{t+1}$	0.0225					
	(1.17)					
$ETFA \times Earn\_Sys_{t+1}$		-0.00549				
		(-0.13)				
$ETFA \times Earn\_Firm_{t+1}$		-0.00729				
		(-0.34)				
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls × Earn	Yes	Yes	Yes	Yes	Yes	Yes
Observations	158,897	146,317	144,977	178,199	178,199	178,268
Adjusted R2	0.062	0.055	0.003	0.000	0.001	0.015

Table 2.14: The Regulation SHO pilot program, ETF shorting activity and informational efficiency (short-sale constraint)

This table presents coefficient estimates from panel regressions of the effect of the SEC Reg SHO pilot program on the association between ETF shorting activity and informational efficiency. SEC randomly select 1000 treatment stocks from Russell 3000 index constituents and 2000 control stocks from the remaining Russell 3000 index constituents. I examine the effect of the SEC Regulation SHO pilot program in the (-1year,+1year) window around the effective date of the SEC Regulation SHO pilot program. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), Post_t is a post-event indicator variable that takes value of 1 if after the implementation of Regulation SHO pilot program, and 0 otherwise, Treatment_i is the treatment indicator that takes value of 1 if stock i is in the treatment group of the SEC Regulation SHO pilot program, and 0 otherwise. the variables of interest are the quadruple interaction term among Post_t, Treatment_t, ETF-based SA and $Earn_{t+1}$ in Column (1), which captures the IFERC, and the triple interaction term among Post_t, Treatment_t and ETF-based SA in Columns (2) to (4). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by date and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Jump	(3) |1-VR|	(4) |AC|	(5) PEAD
$Earn_{t+1}$ × ETF-based SA × Treatment_i × Post_t	-1.457				
	(-0.72)				
ETF-based SA × Treatment_i × Post_t	0.0346	0.805	-0.0214	-0.0486	-0.805
	(1.27)	(0.03)	(-0.63)	(-1.18)	(-0.20)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Controls×Treatment_i, Control×Post_t	Yes	Yes	Yes	Yes	Yes
Observations	12,439	10,918	13,177	13,177	10,247
Adjusted R2	0.206	0.007	0.003	0.002	0.088

2.9 Appendix

2.9.1 Stationarity in ETF short interest

In Table 2.A2, I examine whether the ETF short interest is stationary. I focus on the ten most shorted ETFs, whose short interests account for more than 95% of total short interest in the ETF market. I find that eight out of ten ETFs have a unit root, suggesting that the ETF short interest is non-stationary. Additionally, I construct the time-series aggregate ETF short interest by summing up all the short interest in the ETF market. I again find that the aggregate ETF short interest is non-stationary.

2.9.2 Good v.s. Bad earnings news

Table 2.A12 reports the panel OLS regressions of informational efficiency on ETF-based shorting activity on the partition sample of stocks with good/bad earnings news. I define that the stocks with good earnings news are those with positive earnings surprises and the stocks with bad earnings news are those with negative earnings surprises. I find that the improvement effect on informational efficiency mainly appears when bad earnings news.

2.9.3 Return predictability of ETF-based SA

Literature has documented that there are several ETFs missing in CRSP mutual fund dataset and the Thomson Reuters mutual fund holdings (S12) dataset. To address this problem, I constrain my sample period from 2010 onwards, where ETF holdings are almost complete.[52] Table 2.A15 reports the fixed effects panel OLS regressions of future stock returns on the ETF-based shorting activity and controls in the post-2010 sample. I find consistent evidence that the ETF-based shorting activity negatively predicts future stock returns.

To address potential oversight of risk effects when using raw returns, I replicate the regression in Table 2.3 with the modification that replaces the dependent variable

[52] I am able to match the ETFs and their holdings, which account for more than 90% of ETF short interest

of raw returns with Fama-French three factors alphas. Table 2.A18 reports the return predictability of ETF shorting activity based on Fama-French three factors alphas. I find that ETF shorting activity negatively predict Fama-French three factors alphas, supporting that ETF shorting activity is informed

Furthermore, Table 2.A19 considers the long-term return predictability of ETF shorting activity. I replace the dependent variable in Table 2.3 with next-60-day returns. I find that ETF shorting activity negatively predicts next-60-day returns, corroborating that ETF shorting activity is informed in predicting return in an extended horizon.in

Table 2.A1: Variable definitions

Variable	**Definitions**
$Δ *ETF SI*	The change of ETF short interest as a percentage of market capitalization
ETF-based SA	The equivalent stock-level shorting activity of ETF underlying securities from ETF shorting markets
ETF_SR	ETF-based short ratio based on Li and Zhu (2022)
MktCap	Market value of common shares
BTM	Book value of common shares divided by the market value of equity
MOM	The cumulative returns from month t-12 to t-2
REV	The lagged monthly return
IdiosynVol	The standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock return
Put	The put option volume scaled by consolidated trading volume of underlying asset
ETF-based Flow	The equivalent stock-level flow of ETF underlying securities from ETF flow
Turnover	The daily consolidated volume scaled by the market value of common shares
ETF flow	The ETF flow measured as the change of ETF shares outstanding in dollar term
ETF SVR	The FINRA short volume of ETF as a percentage of consolidated trading volume in TRF and ADF
Earn	The standard unexpected earnings based on Livnat and Mendenhall (2006)
IO	institutional holdings scaled by the total shares outstanding
Loss	An indicator variable that takes the value of 1 if the earnings surprise is negative, and 0 otherwise (Loss)
SI	The stock-level short interest as a percentage of market capitalization
ETFA	The ETF activity measured as the percentage change of ETF ownership
Jump	The jump ratio based on Weller(2018)
AC	Daily 15-second NBBO midpoint return auto-correlation
VR	The ratio of the 60-second NBBO midpoint return variance over four times 15-second NBBO midpoint return variance
PEAD	the cumulative abnormal return from the first trading day after the earnings announcements to one month after the earnings announcements

Table 2.A2: The Augmented Dickey-Fuller test for ETF short interest

This table summarizes the Augmented Dickey-Fuller (ADF) test results for the shorted value of top ten most shorted ETFs and the aggregate ETF shorted from January 2002 to December 2021.I include constant and trend in the ADF test, and I select the lag by minimizing Akaike's 'An Information Criterion (AIC). P-value is computed under one-tail test against the null hypothesis that a time series does not have a unit root.I report the 1%/5%/10% critical value based on the number of observations.

ETF name	Test Statistic	pvalue	CV(1%)	CV(5%)	CV(10%)	#Lags	Obs.	
SPDR Trust	-2.96	0.14	-3.98	-3.42	-3.13	9	441	
iShares Russell 2000 Index Fund	-3.17	0.09	-3.98	-3.42	-3.13	5	441	
Invesco QQQ Trust	0.55	1.00	-3.98	-3.42	-3.13	15	433	
SPDR Dow Jones Industrial Average ETF	-4.98	0.00	-3.98	-3.42	-3.13	1	449	
MidCap SPDR Trust	-2.67	0.25	-3.98	-3.42	-3.13	3	447	
iShares S&P 500 Index Fund	-1.27	0.90	-3.98	-3.42	-3.13	9	437	
iShares Russell 2000 Growth Index Fund	-2.94	0.15	-3.98	-3.42	-3.13	1	443	
iShares Russell 1000 Index Fund	-2.85	0.18	-3.98	-3.42	-3.13	0	437	
Vanguard 500 Index Fund	-0.87	0.96	-3.99	-3.43	-3.14	6	272	
iShares Russell 2000 Value Index Fund	-2.30	0.43	-3.98	-3.42	-3.13	12	431	
Aggregate ETF short interest	-2.93	0.15	-3.98	-3.42	-3.13	7	443	-

Table 2.A3: Summary statistics (market ETFs)

This table summarizes the number of observations, mean, median, standard deviation, and the 25th and 75th percentiles of variables for market ETFs. Panel A reports market ETF characteristics at the ETF-month level, and Panel B reports stock (market ETF constituents) characteristics on the stock-quarter level. ETF SI is the percentage of ETF short interest scaled by market capitalization. Price is the monthly ETF close price. MktCap is the market capitalization in $ 1,000,000,000. DollarVolume is the monthly consolidated dollar volume in ETF sample in $ 1,000,000 and quarterly consolidated dollar volume in stock sample in $ 1,000,000,000. ETF-based SA is the ETF-based shorting activity. Return is the quarterly return in between two consecutive quarterly earnings reporting dates. Jump is the Weller(2018) jump ratio. PEAD is the post-earnings announcement drift measured by the abnormal return from the first day to one month after the earnings announcement. |1-VR| is the absolute value of one minus the ratio of the variance of 2-day returns divided by two times the variance of 1-day returns from 21 days before the report date of quarterly earnings to 2 days after the report date of quarterly earnings. |AC| is the absolute value of daily returns auto-correlation from 21 days before the report date of quarterly earnings to 2 days after the report date of quarterly earnings. Earn is the seasonally adjusted earnings innovation scaled by price. IO is the orthogonal institutional ownership with respect to ETF-based SA. IdiosynVol is the idiosyncratic volatility estimated based on the Fama-French (1993) three-factor model. BTM is the book-to-market ratio. Loss is a dummy variable that takes the value of one if income before extraordinary items is negative, and zero otherwise. ETF_SR is the percentage of ETF-based short ratio following Li and Zhu (2022). The sample period covers from January, 2002 to December, 2021.

Panel A: Summary Statistics (market ETF sample: bi-weekly observations)

Variable	Count	Mean	Std	25%	Median	75%
ETF SI (%)	58,876	3.44	10.55	0.12	0.44	1.58
Monthly Return (%)	58,049	0.88	6.17	-1.27	1.59	3.9
Price($)	58,876	56.96	45.28	28.61	43.41	69.29
MktCap ($ billions)	58,876	3.53	20.25	0.02	0.12	0.71
DollarVolume ($ millions)	58,876	245.17	2,382.47	0.07	0.48	4.01

Panel B: Summary Statistics (Stock sample: quarterly observations)

Variable	Count	Mean	Std	25%	Median	75%
ETF-based SA (%)	171,572	0.0	0.17	-0.06	0.0	0.07
SI (%)	217,891	5.26	25.18	1.37	3.06	6.42
Quarterly Return (%)	224,569	1.35	22.13	-8.44	2.32	13.62
MktCap ($ billions)	220,463	5.73	32.6	0.19	0.98	2.89
Dollar Volume ($ millions)	184,577	3.02	18.91	0.06	0.49	1.53
Jump (%)	179,025	47.6	82.87	8.65	39.8	82.96
|1-VR|	218,180	0.34	0.11	0.26	0.33	0.4
|AC|	218,180	0.19	0.14	0.08	0.17	0.29
PEAD (%)	216,879	-0.28	15.5	-5.98	-0.08	6.12
Earn	163,151	0.0	0.07	-0.01	0.0	0.01
IO (%)	154,447	0.1	2.96	-1.49	-0.11	1.35
IdiosynVol	224,509	0.03	0.02	0.01	0.02	0.03
BTM	199,985	0.63	0.68	0.23	0.42	0.78
Loss	224,569	0.35	0.48	0.0	0.0	1.0
ETF_SR (%)	174,934	0.62	0.45	0.22	0.61	0.92

Table 2.A4: Return predictability: the market ETF-based shorting activity

This table presents the results from the panel regressions of market ETF and stock returns on the stock-level market ETF-based shorting activity (market ETF-based SA) and controls. For dependent variable(s), Ret_{t+1} is monthly stock returns in percentage at t+1. For independent variable(s), , the variable of interest is the market ETF-base shorting activity(market ETF-based SA). For control variables, ETF_SR is the market ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. Log(Dvol) is the logarithm of daily dollar trading volume. MOM is the cumulative return from t-12 to t-2. BTM is the logarithm of the book-to-market ratio. IdiosynVol is the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. market ETF-based Flow is the aggregate flow via market ETFs scaled by market capitalization. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Ret_{t+1}	(2) Ret_{t+1}	(3) Ret_{t+1}	(4) Ret_{t+1}
market ETF-based SA	-0.709***	-0.915***	-1.095***	-1.118***
	(-3.50)	(-4.43)	(-5.31)	(-5.32)
ETF_SR		0.417***	1.258***	1.439***
		(5.05)	(14.43)	(16.38)
Log(MktCap)			2.194***	2.838***
			(34.67)	(42.74)
BTM			-2.077***	-2.082***
			(-32.64)	(-30.53)
MOM			-1.850***	-1.805***
			(-10.43)	(-9.78)
IdiosynVol				97.53***
				(25.76)
REV				-3.476***
				(-7.84)
Beta				0.914***
				(14.66)
market ETF-based Flow				-1412.1***
				(-4.47)
Stock fixed effects	Yes	Yes	Yes	Yes
Date fixed effects	Yes	Yes	Yes	Yes
Observations	694,866	694,866	634,558	634,558
Adjusted R2	0.212	0.212	0.246	0.259

Table 2.A5: Market ETF-based shorting activity and informational efficiency

This table presents coefficient estimates from panel regressions of the market ETF-based shorting activity (market ETF-based SA) on informational efficiency. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variable of interest is the interaction term between the seasonally adjusted earnings innovation at t+1 ($Earn_{t+1}$) and market ETF-based SA in Columns (1) and (2), which captures the IFERC, and market ETF-based SA in Columns (3) to (6). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), lag ETF-based shorting activity, lag seasonally adjusted earnings innovation, scaled by price, ($Earn_t$), the interaction term between $Earn_t$ and ETF-based SA, quarterly stock returns at t+1 (Ret_{t+1}), the interaction term between $Earn_t$ and ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price, $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Ret_t	Ret_t	Jump	|1-VR|	$\|AC\|$	PEAD
$Earn_{t+1}$ × market ETF-based SA	0.672^{***}	0.747^{***}				
	(3.31)	(3.17)				
$Earn_{t+1}$		0.182^{***}				
		(7.79)				
market ETF-based SA	-0.0222^{***}	-0.0229^{**}	-6.612^{**}	-0.0291^{***}	0.00886	-6.777^{***}
	(-3.25)	(-2.17)	(-2.07)	(-3.80)	(1.57)	(-9.82)
IO		0.00526^{***}	-0.280^{**}	-0.000276	0.000207	0.175^{***}
		(13.44)	(-2.20)	(-1.02)	(0.88)	(8.92)
Log(MktCap)		0.0856^{***}	-0.866	0.00158	-0.00364^{***}	-3.617^{***}
		(32.47)	(-0.99)	(1.17)	(-2.61)	(-23.29)
IdiosynVol		-0.00508	-578.5^{***}	0.189^{*}	-0.218^{**}	44.02^{***}
		(-0.02)	(-13.42)	(1.84)	(-2.29)	(3.33)
BTM		-0.00000199	-0.000449^{*}	0.00000136^{***}	0.00000224^{***}	0.0000527
		(-1.22)	(-1.70)	(3.06)	(4.11)	(0.58)
Loss		0.00926^{**}	-0.723	0.00742^{*}	-0.00389	-1.545^{***}
		(2.15)	(-0.65)	(1.69)	(-1.60)	(-6.16)
SI		0.0191^{***}	-2.412	-0.000841	0.00330	-1.576^{***}
		(4.91)	(-1.38)	(-0.35)	(1.24)	(-5.91)
Leverage		0.0432^{***}	2.621	-0.0105^{**}	0.00625	-3.679^{***}
		(5.31)	(0.82)	(-2.37)	(1.28)	(-6.67)
Amihud		-0.0110	35.28	0.0140	0.0655^{***}	-2.724
		(-0.96)	(1.37)	(0.58)	(2.91)	(-1.01)
$Ret_{t-4,t-1}$		0.0103^{***}	-0.580	0.000909	-0.000349	-0.414^{**}
		(3.13)	(-0.87)	(0.66)	(-0.22)	(-2.04)
$Earn_t$		0.0224	1.505	-0.00399	-0.0217^{**}	-5.460^{***}
		(0.99)	(0.24)	(-0.40)	(-2.26)	(-3.44)
$\|Earn_{t+1}\|$ × ETF-based SA		0.150	6.372	0.639^{***}	-0.0610	78.00^{***}
		(0.43)	(0.08)	(3.35)	(-0.46)	(3.31)
Ret_{t+1}		-0.0259^{**}	-6.479^{**}	-0.00154	0.00212	
		(-2.16)	(-2.47)	(-0.30)	(0.38)	
$Earn_t$ × ETF-based SA		0.131				
		(0.52)				
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	118,452	91,853	74,501	90,228	90,228	91,952
Adjusted R2	0.229	0.300	0.050	0.107	0.115	0.087

Table 2.A6: Market ETF-based shorting activity and informational efficiency: market-wide v.s. firm-specific information

This table presents the association between market ETF-based shorting activity (market ETF-based SA) and different earnings components. The dependent variable (Ret_t) is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. For independent variable(s), the variable of interest is the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and market ETF-based SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and ETF-based SA. Control variables include seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), market ETF-based shorting activity (market ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), the idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag market ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), the interaction term between $EarnSys_t$ and market ETF-based SA, quarterly stock returns at t+1, the interaction term between $Earn\_Firm_t$ and market ETF-based SA, quarterly stock returns at t+1, the interaction term between $Earn_t$ and market ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and market ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Ret_t
$Earn\_Sys_{t+1}$ × market ETF-based SA	2.060***
	(5.12)
$Earn\_Firm_{t+1}$ × market ETF-based SA	-0.00160
	(-0.01)
$Earn\_Firm_{t+1}$	0.358***
	(7.88)
$Earn\_Firm_{t+1}$	0.212***
	(8.56)
market ETF-based SA	-0.0232**
	(-2.16)
IO	0.00522***
	(13.50)
Log(MktCap)	0.0856***
	(32.35)
IdiosynVol	-0.000263
	(-0.00)
BTM	-0.00000191
	(-1.19)
Loss	0.0106**
	(2.27)
SI	0.0190***
	(4.91)
Return$_{t+1}$	-0.0266**
	(-2.22)
$\|Earn_{t+1}\|$ × market ETF-based SA	0.0577
	(0.17)
Amihud	-0.00857
	(-0.73)
Return$_{t-4,t-1}$	0.00963***
	(2.89)
Leverage	0.0430***
	(5.26)
Lag $Earn\_Sym$, $Earn\_Firm$ and interaction	Yes
Firm fixed effects	Yes
Quarter fixed effects	Yes
Observations	91,755
Adjusted R2	0.301

Table 2.A7: The SEC FOF arrangement implementation effect on market ETF shorting activity

This table presents coefficient estimates from panel regressions on the effect of the SEC FOF arrangement. For dependent variable(s), IO is the quarterly institutional ownership in percentage on market ETFs, ETF SV is the percentage of market ETF short volume relative to the total market ETF volume, ΔETF SI is the percentage of change in market ETF short interest relative to the market ETF share outstanding and market ETF-based SA is the market ETF-based shorting activity. For independent variable(s), the variables of interest are a post-SEC FOF indicator variable (Post_t) that takes value of 1 if after the SEC FOF arrangement, and 0 otherwise for Columns (1) to (3), and the interaction term between Post_t and Rusl2k_i, the latter of which is a treatment indicator variable that takes value of 1 if stock i is in the top Russell 2000 stocks, and 0 if stock i is in the bottom Russell 1000 stocks for Column (4). Control variables include the inverse of the close price ($1/Price$), the logarithm of consolidated dollar volume (Log(Dollar Volume)), the realized volatility based on daily stock returns, ETF fixed effects, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and ETF in Columns (1) - (3) and Year-quarter and firm in Column (4) are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	The SEC FOF arrangement and ETF short sale			
	(1)	(2)	(3)	(4)
	IO	ETF SV	ΔETF SI	Market ETF-based SA
$Post_t$	2.905**	3.971***	0.393***	
	(2.50)	(10.73)	(4.11)	
$\text{Rusl2k}_i \times \text{Post}_t$				0.00542***
				(2.93)
1/Price	-571.7***	-439.1***	-51.09***	-0.133***
	(-3.42)	(-11.39)	(-2.87)	(-2.62)
Log(Dollar Volume)	6.833***	-0.150*	0.127***	0.00108
	(5.21)	(-1.72)	(4.09)	(1.34)
RealVol				0.00140**
				(2.44)
ETF fixed effects	Yes	Yes	Yes	No
Firm fixed effects	No	No	No	Yes
Month fixed effects	No	No	No	Yes
Observations	3,232	208,600	14,635	21,013
Adjusted R2	0.692	0.254	0.144	0.023

Table 2.A8: Market ETF-based shorting activity and informational efficiency by hedging demand

This table presents coefficient estimates from panel regressions of market ETF-based shorting activity (market ETF-based SA). For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. 1_{Hedge} is an indicator variable that takes the value of 1 if the stock has high hedging demand, and 0 otherwise. For independent variable(s), the variable of interest is the triple interaction term between the seasonally adjusted earnings innovation at t+1, marketETF-based SA and 1_{Hedge} in Columns (1) and (2), and the interaction term between marketETF-based SA and 1_{Hedge} in Columns (3) to (6). the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and market ETF-based SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and market ETF-based SA in Columns (2) and (7), and market ETF-based SA in Columns (3) to (5) and Columns (8) to (10). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the market ETF-based shorting activity (market ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag market ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), the interaction term between $Earn_t$ and market ETF-based SA, quarterly stock returns at t+1, the interaction term between $Earn_t$ and market ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and market ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Ret_t	(3) Jump	(4) |1-VR|	(5) $\|AC\|$	(6) PEAD
$Earn_{t+1}$ × market ETF-based SA ×1_{Hedge}	-0.761					
	(-0.64)					
$Earn\_Sys_{t+1}$ × market ETF-based SA × 1_{Hedge}		0.837				
		(0.51)				
$Earn\_Firm_{t+1}$ × market ETF-based SA × 1_{Hedge}		-0.125				
		(-0.21)				
market ETF-based SA ×1_{Hedge}			6.795	0.00441	-0.0146	-0.0282
			(1.36)	(0.16)	(-1.18)	(-0.01)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	98,998	98,903	79,596	97,373	97,373	99,097
Adjusted R2	0.303	0.302	0.050	0.107	0.115	0.087

Table 2.A9: The adjusted market ETF-based shorting activity and informational efficiency (operational short)

This table presents coefficient estimates from panel regressions of the adjusted market ETF-based shorting activity ($\widehat{\text{market ETF-based}}$ SA). For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variable of interest is the interaction term between the seasonally adjusted earnings innovation at t+1 and $\widehat{\text{market ETF-based}}$ SA in Columns (1) and (2), which captures the FERC, the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and $\widehat{\text{market ETF-based}}$ SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and $\widehat{\text{market ETF-based}}$ SA in Column (3), and $\widehat{\text{market ETF-based}}$ SA in Columns (4) to (7). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the adjusted market ETF-based shorting activity ($\widehat{\text{market ETF-based}}$ SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag adjusted market ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ret_t	Ret_t	Ret_t	Jump	|1-VR|	|AC|	PEAD
$Earn_{t+1}$ × $\widehat{\text{market ETF-based}}$ SA	0.671***	0.729***					
	(3.26)	(3.05)					
$Earn\_Sys_{t+1}$ × $\widehat{\text{market ETF-based}}$ SA			2.026***				
			(5.01)				
$Earn\_Firm_{t+1}$ × $\widehat{\text{market ETF-based}}$ SA			-0.0231				
			(-0.10)				
$\widehat{\text{market ETF-based}}$ SA	-0.0157**	-0.0162	-0.0165	-6.853**	-0.0302***	0.00656	-6.731***
	(-2.29)	(-1.54)	(-1.55)	(-2.11)	(-4.06)	(1.15)	(-9.66)
$Earn_{t+1}$		0.181***					
		(7.77)					
$Earn\_Sys_{t+1}$			0.357***				
			(7.87)				
$Earn\_Firm_{t+1}$			0.212***				
			(8.56)				
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	118,452	91,853	91,758	74,501	90,228	90,228	91,952
Adjusted R2	0.229	0.299	0.301	0.050	0.107	0.115	0.087

Table 2.A10: Market ETF-based shorting activity and informational efficiency

This table presents coefficient estimates from panel regressions of the market ETF-based shorting activity (market ETF-based SA) by controlling the effect of ETF activity. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after earnings announcements to one month after earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before earnings announcements to two days after earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before earnings announcements to two days after earnings announcements. For independent variable(s), the variable of interest is the interaction term between the seasonally adjusted earnings innovation at t+1 and market ETF-based SA in Columns (1) and (2), which captures the FERC, the interaction term between seasonally adjusted systematic earnings innovation ($Earn\_Sys$) at t+1 and market ETF-based SA and the interaction term between seasonally adjusted idiosyncratic earnings innovation at t+1 ($Earn\_Firm$) and market ETF-based SA in Column (3), and market ETF-based SA in Columns (4) to (7). Control variables include the orthogonal ETF activity ($ETFA$), seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the market ETF-based shorting activity (market ETF-based SA), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag market ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Ret_t	(3) Ret_t	(4) Jump	(5) |1-VR|	(6) |AC|	(7) PEAD
$Earn_{t+1}$ × market ETF-based SA	0.672***	0.641***					
	(3.31)	(2.66)					
$Earn\_Sys_{t+1}$ × market ETF-based SA			1.880***				
			(4.63)				
$Earn\_Firm_{t+1}$ × market ETF-based SA			-0.0288				
			(-0.13)				
market ETF-based SA	-0.0222***	-0.0260**	-0.0261**	-6.126*	-0.0283***	0.00844	-6.857***
	(-3.25)	(-2.47)	(-2.45)	(-1.89)	(-3.74)	(1.47)	(-9.86)
$Earn_{t+1}$		0.170***					
		(7.21)					
$Earn\_Sys_{t+1}$			0.357***				
			(7.84)				
$Earn\_Firm_{t+1}$			0.200***				
			(8.05)				
ETFA		0.00674***	0.00673***	1.115*	-0.00232***	-0.00132	0.211**
		(5.19)	(5.19)	(1.93)	(-2.77)	(-1.46)	(2.26)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	118,452	90,813	90,722	73,684	89,191	89,191	90,851
Adjusted R2	0.229	0.304	0.305	0.050	0.109	0.116	0.090

Table 2.A11: The Regulation SHO pilot program and informational efficiency (short-sale constraint)

This table presents the diff-in-diff analysis on the impact of short-sale constraints on the effect of market ETF-based shorting activity. In the diff-in-diff analysis, I focus on the implementation of the SEC Regulation SHO pilot program. The SEC randomly selects 1000 treatment stocks from Russell 3000 index constituents and 2000 control stocks from the remaining Russell 3000 index constituents. I examine the effect of the SEC Regulation SHO pilot program in the (-1year,+1year) window around the effective date of the SEC Regulation SHO pilot program. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after earnings announcements to one month after earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before earnings announcements to two days after earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before earnings announcements to two days after earnings announcements. For independent variable(s), Post$_t$ is a post-event indicator variable that takes value of 1 if after the implementation of Regulation SHO pilot program, and 0 otherwise, Treatment$_i$ is the treatment indicator that takes value of 1 if stock i is in the treatment group of the SEC Regulation SHO pilot program, and 0 otherwise. the variable of interest is the quadruple interaction term among Post$_t$, Treatment$_t$, market ETF-based SA and $Earn_{t+1}$ in Column (1), which captures the IFERC, and the triple interaction term among Post$_t$, Treatment$_t$ and market ETF-based SA in Columns (2) to (4). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the market ETF-based shorting activity (market ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag market ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by date and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Ret$_t$	(2) Jump	(3) |1-VR|	(4) |AC|	(5) PEAD
$Earn_{t+1}$ × market ETF-based SA × Treatment$_i$ × Post$_t$	-2.954				
	(-1.13)				
market ETF-based SA × Treatment$_i$ × Post$_t$	0.0224	-15.06	0.0202	0.0659	-3.885
	(0.71)	(-0.47)	(0.60)	(1.36)	(-1.09)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	11,580	10,333	12,014	12,014	12,253
Adjusted R2	0.352	0.039	0.043	0.028	0.090

Table 2.A12: ETF-based shorting activity and informational efficiency by Earning news type

This table presents coefficient estimates from panel regressions of the ETF-based shorting activity (ETF-based SA) on informational efficiency. Columns (1) to (4) report the coefficient estimates when good news is announced, and columns (5) to (8) report the coefficient estimates when bad news is announced. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variable of interest is the interaction term between the seasonally adjusted earnings innovation at t+1 ($Earn_{t+1}$) and ETF-based SA in Columns (1) and (2), which captures the IFERC, and ETF-based SA in Columns (3) to (6). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), lag ETF-based shorting activity, lag seasonally adjusted earnings innovation, scaled by price, ($Earn_t$), the interaction term between $Earn_t$ and ETF-based SA, quarterly stock returns at t+1 (Ret_{t+1}), the interaction term between $Earn_t$ and ETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price, $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and ETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	Good News					Bad News				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Ret_t	Jump	|1-VR|	|AC|	PEAD	Ret_t	Jump	|1-VR|	|AC|	PEAD
$Earn_{t+1}$ × ETF-based SA	-26.61**					25.14**				
	(-2.26)					(2.04)				
ETF-based SA	-0.0414***	-4.357	-0.00365	0.0111	2.462***	-0.0293**	-9.031*	-0.0177***	0.00608	-6.374***
	(-3.00)	(-0.95)	(-0.53)	(1.53)	(2.83)	(-2.47)	(-1.84)	(-2.78)	(0.82)	(-8.76)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,113	40,614	48,111	45,822	49,156	40,313	31,752	39,736	38,337	40,360
Adjusted R2	0.299	0.054	0.127	0.086	0.119	0.354	0.072	0.143	0.183	0.119

Table 2.A13: The SEC FOF arrangement: parallel trend assumption (quarterly)

This table presents the dynamics of average treatment effects of the SEC FOF arrangement. I generate four event time indicators (Qtr($-j$)), where j takes values of (-2, -1, 1, 2) to indicate the j quarter after the SEC FOF arrangement implementation. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), Post$_t$ is a post-event indicator variable that takes value of 1 if after the implementation of Regulation SHO pilot program, and 0 otherwise, Treatment$_i$ is the treatment indicator that takes value of 1 if stock i is in the treatment group of the SEC Regulation SHO pilot program, and 0 otherwise. the variable of interest is the quadruple interaction term among Post$_t$, Treatment$_t$, ETF-based SA and $Earn_{t+1}$ in Column (1), which captures the IFERC, and the triple interaction term among Post$_t$, Treatment$_t$ and ETF-based SA in Columns (2) to (4). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), seasonally adjusted systematic earnings innovation at t+1, scaled by price ($Earn\_Sys_{t+1}$), seasonally adjusted idiosyncratic earnings innovation at t+1, scaled by price ($Earn\_Firm_{t+1}$), the ETF-based shorting activity (ETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), the lag ETF-based shorting activity, seasonally adjusted earnings innovation at t, scaled by price, ($Earn_t$), quarterly stock returns at t+1 ($Return_{t+1}$), the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price $|Earn_{t+1}|$ and firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by date and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Ret$_t$	(2) Jump	(3) |AC|	(4) |1-VR|	(5) PEAD
$Earn_{t+1}$ × Rusl2k$_i$ × Qtr(-2)	0.0465				
	(0.06)				
$Earn_{t+1}$ × Rusl2k$_i$ × Qtr(-1)	-0.510				
	(-0.58)				
$Earn_{t+1}$ × Rusl2k$_i$ × Qtr($+1$)	2.529*				
	(1.96)				
$Earn_{t+1}$ × Rusl2k$_i$ × Qtr($+2$)	-0.748				
	(-0.64)				
Rusl2k$_i$ × Qtr(-1)	-0.0373*	4.828	-0.00516	-0.0268	0.760
	(-1.68)	(0.42)	(-0.34)	(-1.52)	(0.86)
Rusl2k$_i$ × Qtr(-2)	-0.00436	10.35	0.00248	-0.0258	0.0554
	(-0.19)	(0.84)	(0.16)	(-1.46)	(0.07)
Rusl2k$_i$ × Qtr($+1$)	0.113***	-1.637	-0.0340**	-0.0492**	-1.223*
	(4.82)	(-0.13)	(-2.13)	(-2.49)	(-1.68)
Rusl2k$_i$ × Qtr($+2$)	-0.0264	-20.58*	0.00739	-0.00646	-1.076
	(-1.29)	(-1.69)	(0.47)	(-0.34)	(-1.55)
Controls	Yes	Yes	Yes	Yes	Yes
Controls×Treatment$_i$, Control×Post$_t$	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	2,069	1,649	2,171	2,171	1,634
Adjusted R2	0.312	0.030	0.025	-0.005	0.145

Table 2.A14: The SEC FOF arrangement: parallel trend assumption (daily)

This table presents the dynamics of average treatment effects of the SEC FOF arrangement. I generate four event time indicators (Qtr(j)), where j takes values of (-2, -1, 1, 2) to indicate the j quarter after the SEC FOF arrangement implementation. For dependent variable(s), $\|1-VR\|$ is the absolute difference between one and the variance ratio based on 15-second and 1-minute EBBO midpoint return. $|AC|$ is 15-second NBBO midpoint return auto-correlation. For independent variable(s), the variable of interest is the interaction term between Rusl2k$_i$ and Qtr(j), where Rusl2k$_i$ is an indicator variable that takes value of 1 if stock i is in the top Russell 2000 stocks, and 0 if stock i is in the bottom Russell 1000 stocks. Control variables include the logarithm of dollar consolidated volume (Log(Dollar Volume)), the inverse of close price (1/Price), 15-second NBBO midpoint realized volatility (RealVol), firm fixed effects and data fixed effects. t-statistics from robust standard errors clustered by date and firm are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) |1-VR|	(2) |AC|
Rusl2k$_i$ × Qtr(-2)	-0.00323	-0.00165
	(-1.43)	(-1.06)
Rusl2k$_i$ × Qtr(-1)	0.00215	0.00237
	(0.88)	(1.39)
Rusl2k$_i$ × Qtr($+1$)	-0.00900***	-0.00530***
	(-3.71)	(-3.19)
Rusl2k$_i$ × Qtr($+2$)	-0.0234***	-0.0167***
	(-10.41)	(-10.81)
RealVol	0.0386***	0.0285***
	(26.02)	(27.05)
Log(Dollar Volume)	-0.000881	-0.00100
	(-0.87)	(-1.42)
1/Price	-0.613***	-0.414***
	(-5.57)	(-5.30)
Firm fixed effects	Yes	Yes
Date fixed effects	Yes	Yes
Observations	125,293	125,293
Adjusted R2	0.799	0.806

Table 2.A15: The return predictability of bi-weekly ETF-based shorting activity after 2010

This table presents the results from the panel regressions of stock returns on the stock-level market ETF-based shorting activity (market ETF-based SA) and controls during the post-2010 period. For dependent variable(s), Ret_{t+20} is next-20-day stock returns in percentage. For independent variable(s), , the variable of interest is the market ETF-base shorting activity(market ETF-based SA). For control variables, ETF_SR is the market ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. IdiosynVol is the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF-based Flow is the aggregate flow via ETFs scaled by the market capitalization. ETF Flow is the change of ETF outstanding share scaled by the total share outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) RET_{t+20}	(2) RET_{t+20}	(3) RET_{t+20}	(4) RET_{t+20}
market ETF-based SA	-1.110***	-1.237***	-0.878***	-0.876***
	(-6.54)	(-7.18)	(-4.78)	(-4.77)
ETF_SR		0.191***	-0.459***	-0.459***
		(3.25)	(-6.84)	(-6.84)
Log(MktCap)			-2.694***	-2.705***
			(-59.48)	(-59.02)
BTM			-1.521***	-1.521***
			(-41.65)	(-41.67)
REV			-2.575***	-2.569***
			(-15.78)	(-15.75)
IdioVol				-0.987
				(-0.43)
ETF-based Flow				0.00000356***
				(7.77)
Stock fixed effects	Yes	Yes	Yes	Yes
Bi-week fixed effects	Yes	Yes	Yes	Yes
Observations	824,381	824,381	697,819	697,818
Adjusted R2	0.231	0.231	0.247	0.247

Table 2.A16: Stock Portfolio Returns Sorted on the ETF-based SA

This table reports the monthly average excess returns, CAPM alpha, Fama and French (1993) 3-factor alpha, and Fama and French (1993) and Carhart (1997) 4-factor alpha (in percentage) for stock portfolio, as well as for the long-short portfolio (High-Low). Panel A reports for ten decile equal-weighted stock portfolios sorted by the ETF-based SA. Panel B reports for ten decile value-weighted stock portfolios sorted by the ETF-based SA. At the end of each month, all securities are sorted into decile portfolios based on ETF-based shorting activity at the stock level, and a long-short portfolio is formed by buying the lowest decile and shorting the highest decile portfolio. Portfolios are rebalanced monthly and portfolio returns are computed over the next month. t-statistics from errors with Newey-West correction of four lags are reported. The sample runs from January 2002 to December 2021.

Panel A: Performance of Equal-weighted stock portfolio sorted by ETF-based SA

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	Carhart(97)	t-stat
Low	1.32	3.81	1.19	3.29	1.18	3.38	1.17	3.23
2	1.04	2.94	0.89	2.41	0.88	2.44	0.88	2.34
3	0.94	2.66	0.79	2.00	0.78	2.04	0.77	1.95
4	0.93	2.68	0.79	2.06	0.78	2.10	0.77	2.03
5	0.89	2.59	0.76	2.10	0.74	2.15	0.74	2.03
6	0.70	2.01	0.54	1.47	0.52	1.47	0.54	1.48
7	0.95	2.69	0.80	2.23	0.79	2.27	0.80	2.20
8	0.82	2.29	0.67	1.78	0.66	1.83	0.66	1.75
9	0.84	2.40	0.69	1.84	0.67	1.89	0.69	1.85
High	0.89	2.61	0.75	2.13	0.74	2.20	0.73	2.10
High-Low	-0.43	-2.83	-0.44	-2.75	-0.45	-2.77	-0.44	-2.72

Panel B: Performance of Value-weighted stock portfolio sorted by ETF-based SA

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	Carhart(97)	t-stat
Low	2.04	6.28	1.92	5.93	1.90	6.14	1.91	5.87
2	2.07	6.47	1.91	5.81	1.89	5.98	1.93	5.77
3	1.85	5.93	1.74	5.31	1.72	5.48	1.73	5.30
4	1.91	5.89	1.79	5.37	1.77	5.46	1.80	5.34
5	1.71	5.46	1.58	4.80	1.57	4.97	1.57	4.79
6	1.39	4.53	1.22	4.09	1.21	4.12	1.24	4.14
7	1.40	4.54	1.23	4.41	1.22	4.43	1.24	4.42
8	1.40	4.50	1.23	4.09	1.22	4.05	1.22	3.97
9	1.61	5.25	1.47	4.86	1.45	4.93	1.48	4.79
High	1.75	5.63	1.59	5.16	1.57	5.30	1.58	5.19
High-Low	-0.29	-1.58	-0.33	-1.98	-0.34	-2.02	-0.33	-1.97

Table 2.A17: Fama-MacBeth Regression: the bi-weekly ETF shorting activity

This table presents the results from the Fama-Macbeth cross-sectional regressions of ETF and stock returns on the bi-weekly ETF shorting activity (ΔETF SI), the stock-level ETF-based shorting activity (ETF-based SA) and controls. Columns (1) to (4) reports the coefficient estimates of Fama-Macbeth regressions for ETFs. Columns (5) to (8) reports the coefficient estimates of Fama-Macbeth regressions for stocks. For dependent variable(s), Ret_{t+20} is next-20-day stock returns in percentage, after skipping one trading day. For independent variable(s), the variables of interest are the bi-weekly ETF shorting activity (ΔETF SI) in Columns (1) to (4) and the bi-weekly ETF-base shorting activity(ETF-based SA) in Columns (5) to (8). For control variables, ETF_SR is the ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. Log(Dvol) is the logarithm of daily dollar trading volume. BTM is the logarithm of the book-to-market ratio. IdiosynVol the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF Flow is the change of ETF outstanding share scaled by the total share outstanding. ETF-based Flow is the aggregate flow via ETFs scaled by the market capitalization. t-statistics from errors with Newey-West correction of four lags are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	ETF sample				Stock sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}	RET_{t+20}
ΔETF SI	-4.856**	-4.542**	-3.849*	-3.431*				
	(-2.59)	(-2.46)	(-1.81)	(-1.80)				
ETF-based SA					-3.694***	-6.385***	-8.524***	-7.796***
					(-3.69)	(-3.66)	(-4.98)	(-5.39)
ETF_SR						0.747	1.219**	1.192**
						(1.11)	(2.42)	(2.21)
BTM							-0.382***	-0.435***
							(-3.11)	(-3.61)
Log(MktCap)		11.00***	6.692***	2.341*			0.345***	0.264***
		(4.57)	(3.53)	(1.66)			(5.60)	(5.45)
REV			543.8*	-83.43			-1.377**	-1.510***
			(1.68)	(-0.25)			(-2.23)	(-2.83)
IdiosynVol				-5865.6***				-41.95***
				(-3.72)				(-4.82)
Turnover				-0.399***				-0.0162*
				(-4.16)				(-1.88)
Put				-8.440				
				(-0.53)				
Flow				202.2				
				(1.00)				
ETF-based Flow								0.000175
								(0.76)
Constant	17.92	-108.3***	-93.10***	-0.00203	0.276	0.383	-5.317***	-3.377***
	(0.62)	(-3.19)	(-3.31)	(-0.00)	(0.66)	(0.98)	(-4.35)	(-3.66)
Obs.	134,600	134,600	134,600	134,516	926,375	926,375	789,971	789,970
R-Square	0.016	0.036	0.263	0.518	0.010	0.016	0.048	0.072

Table 2.A18: The return predictability of bi-weekly ETF shorting activity (Fama-French alphas)

This table presents coefficient estimates from panel regressions of Fama-French three factor alphas on the ETF shorting activity (ΔETF SI) and the stock-level ETF-based shorting activity (ETF-based SA) and controls. Columns (1) to (4) reports the coefficient estimates of ETF-level regressions. Columns (5) to (8) reports the coefficient estimates of stock-level regressions. For dependent variable(s), $Alpha_{t+20}$ is the next-20-day Fama-French three factor alpha in percentage, after skipping one trading day. For independent variable(s), the variables of interest are the ETF shorting activity($ETFSI$) in Columns (1) to (4) and the ETF-base shorting activity(ETF-based SA) in Columns (5) to (8). For control variables, ETF_SR is the ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. IdiosynVol the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF-based Flow is the aggregate flow via ETFs scaled by the market capitalization. ETF Flow is the change of ETF outstanding share scaled by the total share outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	ETF sample				Stock sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Alpha_{t+20}$	$Alpha_{t+20}$	$Alpha_{t+20}$	$Alpha_{t+20}$	$Alpha_{t+20}$	$Alpha_{t+20}$	$Alpha_{t+20}$	$Alpha_{t+20}$
ΔETF SI	-0.0158***	-0.0159***	-0.0151***	-0.0152***				
	(-3.51)	(-3.52)	(-3.31)	(-3.32)				
ETF-based SA					-2.420***	-1.478***	-1.550***	-1.601***
					(-14.24)	(-8.67)	(-8.76)	(-9.04)
ETF_SR						-1.627***	-1.068***	-1.082***
						(-33.86)	(-20.38)	(-20.64)
BTM							-1.851***	-1.872***
							(-61.35)	(-61.81)
Log(MktCap)		-0.209***	-0.212***	-0.211***			2.114***	2.049***
		(-17.53)	(-17.85)	(-17.78)			(58.71)	(56.58)
IdiosynVol		-52.74***	-50.67***	-50.45***			-9.354***	-12.21***
		(-9.33)	(-8.97)	(-8.93)			(-4.18)	(-5.18)
REV		-0.131	-0.127	-0.0857				-1.238***
		(-0.46)	(-0.45)	(-0.29)				(-9.43)
Turnover			-0.000524***	-0.000524***				-0.00249*
			(-2.86)	(-2.86)				(-1.66)
ETF Flow				-0.185				
				(-1.17)				
Put				-0.0627				
				(-0.47)				
ETF-based Flow								-0.00000109***
								(-2.77)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	179,048	174,458	174,458	174,345	935,949	935,949	792,590	789,778
Adjusted R2	0.065	0.073	0.073	0.073	0.025	0.027	0.059	0.059

Table 2.A19: The long-term return predictability of bi-weekly ETF shorting activity

This table presents coefficient estimates from panel regressions of long-term returns on the ETF shorting activity (ΔETF SI) and the stock-level ETF-based shorting activity (ETF-based SA) and controls. Columns (1) to (4) reports the coefficient estimates of ETF-level regressions. Columns (5) to (8) reports the coefficient estimates of stock-level regressions. For dependent variable(s), Ret_{t+60} is next-60-day returns in percentage, after skipping one trading day. For independent variable(s), the variables of interest are the ETF shorting activity($ETFSI$) in Columns (1) to (4) and the ETF-base shorting activity(ETF-based SA) in Columns (5) to (8). For control variables, ETF_SR is the ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. IdiosynVol is the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF-based Flow is the aggregate flow via ETFs scaled by the market capitalization. ETF Flow is the change of ETF outstanding share scaled by the total share outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	ETF sample				Stock sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RET_{t+60}	RET_{t+60}	RET_{t+60}	RET_{t+60}	RET_{t+60}	RET_{t+60}	RET_{t+60}	RET_{t+60}
ΔETF SI	-0.0329***	-0.0356***	-0.0334***	-0.0342***				
	(-3.01)	(-3.24)	(-3.02)	(-3.07)				
ETF-based SA					-0.824***	-1.001***	-0.567**	-0.564**
					(-3.88)	(-4.64)	(-2.49)	(-2.48)
ETF_SR						0.262***	-0.782***	-0.777***
						(3.67)	(-10.17)	(-10.11)
BTM							-1.552***	-1.554***
							(-34.10)	(-34.14)
Log(MktCap)		-0.594***	-0.603***	-0.598***			-4.148***	-4.146***
		(-18.52)	(-18.77)	(-18.59)			(-74.93)	(-74.82)
IdiosynVol		-98.88***	-93.57***	-92.27***			-12.30***	-10.40***
		(-7.95)	(-7.52)	(-7.40)			(-4.12)	(-3.24)
REV		-4.869***	-4.863***	-4.544***			-2.840***	-2.843***
		(-6.87)	(-6.89)	(-6.30)			(-13.04)	(-13.06)
Turnover			-0.00135***	-0.00135***				-0.00429
			(-3.46)	(-3.46)				(-1.37)
ETF Flow				-1.161***				
				(-2.91)				
Put				-0.438				
				(-1.60)				
ETF-based Flow								0.00000291***
								(4.59)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175,534	173,691	173,691	173,578	912,777	912,777	786,022	786,021
Adjusted R2	0.391	0.396	0.396	0.396	0.263	0.263	0.277	0.277

Table 2.A20: The return predictability of bi-weekly industry ETF shorting activity

This table presents coefficient estimates from panel regressions of industry ETF returns on the industry ETF shorting activity (ΔiETF SI) and the stock-level industry ETF-based shorting activity (iETF-based SA) and controls. Columns (1) to (4) report the coefficient estimates of ETF-level regressions. Columns (5) to (8) report the coefficient estimates of stock-level regressions. For dependent variable(s), Ret_{t+20} is next-20-day returns (in percentage), after skipping one trading day. For independent variable(s), the variables of interest are the industry ETF shorting activity (ΔiETF SI) in Columns (1) to (4) and the industry ETF-base shorting activity(iETF-based SA) in Columns (5) to (8). For control variables, $iETF\_SR$ is the industry ETF-based short ratio from Li and Zhu (2022). Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. IdiosynVol the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. iETF-based Flow is the aggregate flow via industry ETFs scaled by the market capitalization. iETF Flow is the change of industry ETF outstanding shares scaled by the total shares outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) RET_{t+20}	(2) RET_{t+20}	(3) RET_{t+20}	(4) RET_{t+20}	(5) RET_{t+20}	(6) RET_{t+20}	(7) RET_{t+20}	(8) RET_{t+20}
ΔiETF SI	0.0342***	0.0275**	0.0287***	0.0280***				
	(3.73)	(2.18)	(3.16)	(3.02)				
iETF-based SA					0.592***	0.520***	0.377***	0.290***
					(6.16)	(5.68)	(4.16)	(3.76)
$iETF\_SR$						0.158**	0.804***	0.972***
						(2.47)	(12.62)	(14.41)
BTM							-0.0176***	-0.0186***
							(-40.83)	(-43.33)
Log(MktCap)		-0.392***	-0.394***	-0.398***			0.0178***	0.0220***
		(-14.27)	(-14.37)	(-14.49)			(36.34)	(43.23)
IdiosynVol		-21.02**	-17.81*	-16.90*				0.938***
		(-2.17)	(-1.87)	(-1.79)				(22.25)
REV		-1.259**	-1.344**	-1.319**				-0.0508***
		(-2.14)	(-2.29)	(-2.17)				(-26.06)
Turnover			-0.00304***	-0.00309***				-0.000258***
			(-5.29)	(-5.38)				(-9.99)
iETF Flow				-0.0742				
				(-0.25)				
Put				0.00680				
				(0.05)				
iETF-based Flow								-11.04***
								(-3.93)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	No	No	No	No
Date fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Observations	63,566	62,175	62,175	61,808	424,435	424,435	386,279	386,273
Adjusted R2	0.000	0.005	0.005	0.005	0.000	0.000	0.026	0.039

Table 2.A21: Industry ETF-based shorting activity and informational efficiency

This table presents coefficient estimates from panel regressions of the industry ETF-based shorting activity (iETF-based SA) on informational efficiency. For dependent variable(s), Ret_t is quarterly stock returns in percentage from earnings reporting quarter $t-1$ to earnings reporting quarter t. Jump is the jump ratio by Weller(2018). PEAD is the cumulative abnormal return in percentage from the first trading day after the earnings announcements to one month after the earnings announcements. |1-VR| is the absolute difference between one and variance ratio of close-to-close return based on one-day and two-day returns from 21 days before the earnings announcements to two days after the earnings announcements. |AC| is the absolute value of daily close-to-close returns auto-correlation from 21 days before the earnings announcements to two days after the earnings announcements. For independent variable(s), the variables of interest are the interaction term between the seasonally adjusted earnings innovation at t+1 ($Earn_{t+1}$) and iETF-based SA in Columns (1) and (2), which captures the IFERC, and iETF-based SA in Columns (3) to (6). Control variables include seasonally adjusted earnings innovation at t+1, scaled by price ($Earn_{t+1}$), the industry ETF-based shorting activity (iETF-based SA), orthogonal institutional ownership (IO), the logarithm of market capitalization ($Log(MktCap)$), Idiosyncratic volatility ($IdiosynVol$), the book-to-market ratio (BTM), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise ($Loss$), stock-level short interest ratio (SI), lag industry ETF-based shorting activity, lag seasonally adjusted earnings innovation, scaled by price, ($Earn_t$), the interaction term between $Earn_t$ and iETF-based SA, quarterly stock returns at t+1 (Ret_{t+1}), the interaction term between $Earn_t$ and iETF-based SA, the absolute value of seasonally adjusted earnings innovation at t+1, scaled by price, $|Earn_{t+1}|$, the interaction term between $|Earn_{t+1}|$ and iETF-based SA, firm fixed effects and Year-quarter fixed effects. t-statistics from robust standard errors clustered by Year-quarter and firm are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) Ret_t	(2) Ret_t	(3) Jump	(4) |1-VR|	(5) $\|AC\|$	(6) PEAD
$Earn_{t+1}$ × iETF-based SA	2.132***	1.889***				
	(4.66)	(3.67)				
$Earn_{t+1}$	0.236***	0.232***				
	(12.81)	(9.77)				
iETF-based SA	-0.124***	-0.0971***	-7.396	-0.0545***	-0.0302	-6.015***
	(-13.34)	(-6.33)	(-1.59)	(-3.91)	(-1.60)	(-4.37)
IO		0.00566***	-0.170	-0.000328	-0.000552*	0.113***
		(17.38)	(-1.48)	(-1.49)	(-1.77)	(4.66)
Log(MktCap)		0.0769***	-1.739**	-0.000125	-0.00362***	-3.153***
		(35.12)	(-2.40)	(-0.10)	(-2.95)	(-21.05)
IdiosynVol		-0.395*	-522.8***	-0.137	-0.181**	46.16***
		(-1.92)	(-14.71)	(-1.25)	(-2.22)	(4.60)
BTM		-0.000000559	0.000107	3.26e-08	-0.000000116	0.0000110
		(-0.98)	(0.51)	(0.07)	(-0.22)	(0.40)
SI		-0.0104**	2.010	-0.00261	-0.000683	-0.741***
		(-2.33)	(1.38)	(-0.85)	(-0.19)	(-2.66)
Ret_{t+1}		-0.0686***	-3.126	0.0232***	0.0298***	
		(-6.46)	(-1.21)	(2.88)	(2.58)	
$\|Earn_{t+1}\|$		0.167***	2.697	0.0169	0.00251	-1.543
		(4.27)	(0.56)	(1.41)	(0.33)	(-0.73)
$\|Earn_{t+1}\|$ × iETF-based SA		-0.997	78.10	2.163***	1.172**	168.9***
		(-1.44)	(0.43)	(4.35)	(2.15)	(3.42)
Loss		0.00711	2.391*	-0.00601	-0.0154**	-2.540***
		(1.15)	(1.77)	(-1.06)	(-2.39)	(-5.11)
$Earn_{t-1}$		0.00718				
		(0.33)				
$Earn_{t-1}$ × iETF-based SA		0.976*				
		(1.93)				
$Earn_t$		0.0973***				
		(4.54)				
$Earn_t$ × iETF-based SA		0.606				
		(1.21)				
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162,456	108,504	92,101	113,673	113,673	115,453
Adjusted R2	0.012	0.072	0.003	0.003	0.003	0.023

Table 2.A22: Information acquisition and ETF shorting activity

This table presents coefficient estimates from fixed effects OLS regressions of quarterly Information acquisition measured by the logarithm of EDGAR Search Volume (SV) on the stock-level ETF-based shorting activity (ETF-based SA) and controls. For dependent variable(s), $Log(SV)_{t-2}$ is the log of total EDGAR search volume in the previous two months. $Log(SV)_{t+1}$ is the log of total EDGAR search volume in the next month. For independent variable(s), the variables of interest are the ETF-base shorting activity(ETF-based SA), Control variables include the ETF-based short interests(ETF_SR), the logarithm of market capitalization (Log(MktCap)), the inverse of the close price (1/Price), the book-to-market ratio (BTM), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the past monthly stock return (Ret_{t-1}), stock fixed effects and quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively..

	(1) $Log(SV)_{t-2}$	(2) $Log(SV)_{t-2}$	(3) $Log(SV)_{t+1}$	(4) $Log(SV)_{t+1}$
ETF-based SA	-0.0228	0.00990	0.0211	0.0125
	(-0.92)	(0.37)	(0.92)	(0.51)
ETF_SR		-0.0193**		0.00970
		(-2.43)		(1.46)
Log(MktCap)		0.0633***		0.0433***
		(5.72)		(4.60)
IdiosynVol		1.578***		0.565
		(3.54)		(1.52)
1/Price		-0.208***		-0.232***
		(-4.06)		(-4.86)
BTM		0.0760***		0.0697***
		(5.15)		(5.68)
Ret_{t-1}		0.0318***		0.0150***
		(4.66)		(3.21)
Stock fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
Observations	141,300	134,846	141,300	134,846
Adjusted R2	0.000	0.003	0.000	0.003

Table 2.A23: The return predictability of market ETF shorting activity on macroeconomic news period (no time fixed effects)

This table presents coefficient estimates from stock panel regressions of market ETF returns on the market ETF shorting activity (ΔETF SI), and controls. Columns (1) to (2) report the coefficient estimates during the full sample. Columns (3) to (4) report the coefficient estimates on macroeconomic new period. Columns (5) to (6) report the coefficient estimates on non-macroeconomic new period. For dependent variable(s), Ret_{t+20} is next-20-day returns in percentage, after skipping one trading day. For independent variable(s), the variables of interest are the ETF shorting activity(Δ ETF SI). For control variables, Log(MktCap) is the logarithm of market capitalization. IdiosynVol the standard deviation of the residual obtained by fitting Fama-French three-factor model to the time series of daily stock returns. REV is the short-term reversal measure that equals to the lagged monthly return. Put is the put option volume scaled by the total trading volume. ETF Flow is the change of ETF outstanding share scaled by the total share outstanding. t-statistics from robust standard errors clustered by stock and bi-week are reported in parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Market ETF Sample		Macro news periods		Non-Macro news periods	
	(1)	(2)	(3)	(4)	(5)	(6)
	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}	Ret_{t+20}
Δ ETF SI	-0.000624***	-0.000622***	-0.00139***	-0.00132***	-0.000255	-0.000317
	(-3.51)	(-3.49)	(-3.25)	(-3.03)	(-0.55)	(-0.68)
Log(MktCap)	0.00267***	0.00267***	-0.00248***	-0.00195***	-0.00122**	-0.00125**
	(12.00)	(12.00)	(-4.86)	(-3.84)	(-2.26)	(-2.31)
IdiosynVol	0.300	0.301	2.034***	1.959***	-3.435***	-3.388***
	(1.46)	(1.46)	(6.92)	(6.60)	(-8.97)	(-8.91)
REV	-0.0441***	-0.0440***	0.149***	0.0661***	-0.0445***	-0.0393***
	(-7.33)	(-7.33)	(8.07)	(4.01)	(-3.20)	(-2.75)
Turnover	-0.000341***	-0.000340***	-0.00000806***	-0.00000701***	-0.00000668	-0.00000690
	(-5.66)	(-5.65)	(-3.02)	(-2.86)	(-0.46)	(-0.47)
ETF Flow		0.018***		0.135***		-0.0123
		(3.77)		(11.27)		(-1.45)
Put		-0.00166		-0.00373		0.00526*
		(-0.49)		(-0.97)		(1.70)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,755	32,755	9,901	9,901	6,928	6,928
Adjusted R2	0.099	0.117	0.143	0.178	0.071	0.091

Chapter 3

The informativeness of liquidity-demanding and liquidity-supplying shorts

3.1 Introduction

Investors frequently perceive short-selling as a uniform activity, primarily driven by speculation from informed traders.[1] Such an oversimplification may overlook the heterogeneity within short sellers, the diverse motivations behind short-selling and the nuances in the information content carried by different types of short-selling activity. To provide insights into the heterogeneity within short-selling, I study the informativeness of different types of short-selling activity and their respective impact on the financial market. Specifically, I aim to discern whether all short sellers are informed and, if not, which types of short-selling activity are informed and which are not.

To address these questions, I categorize short-selling activity into those that demand liquidity and those that supply liquidity using the public Financial Industry

[1]Existing literature consistently finds that short sellers, collectively, are well-informed investors and possess superior information. Numerous studies corroborate that firms with high levels of short-selling activity tend to experience lower future returns, subsequent financial misconduct, and credit rating downgrades, as well as downgrades by financial analysts. (See Chen et al. (2019a),Boehmer et al. (2008) and Diether et al. (2009), Karpoff and Lou (2010), Henry and Koski (2010), Christophe et al. (2010) and Hou et al. (2019))

Regulatory Authority (FINRA) monthly short-sale file that contains transaction-level data on off-exchange short-selling activity. The existing literature shows that trading activities within dark pools (i.e., off-exchange venues) tend to be less informed compared to those conducted in lit markets, as highlighted by Zhu (2014) and Comerton-Forde and Putniņš (2015). However, this observation does not necessarily imply that all dark trading lacks informativeness or exclude the possibility that a certain portion of dark trading volume may be informed. Empirical studies have shown that some dark trading is informed as evidenced by the increased information acquisition activity associated with dark trading (Brogaard and Pan (2022)). Notably, for short-selling in off-exchange venues, Wang, Yan, and Zheng (2020) and Reed, Samadi, and Sokobin (2020) find that both on-exchange and off-exchange shorts are informed, where on-exchange short sellers typically act on short-lived information and prioritize the immediacy offered by exchanges, whereas off-exchange short sellers are informed by longer-term information, which is gradually reflected in market prices.[2]

Following Comerton-Forde et al. (2016b), I define liquidity-demanding shorts (LD, henceforth) as short volume initiated from the sellers and liquidity-supplying shorts (LS, henceforth) as short volume initiated from the buyers.[3] The price dynamics following short-selling is often used to infer the informativeness of short-selling. Comerton-Forde et al. (2016b) show that LD exhibit no significant price change after 24 hours whereas LS are followed by significant price increases. Such observations appear to diverge from the broader short-selling literature, such as Wang et al. (2020), which consistently observes price declines over more extended horizons. This discrepancy between the short-term and long-term price dynamics necessitates a closer inspection of the long-term impacts of LD/LS on financial markets to better understand the informativeness of these two types of shorts.

[2]In Table 3.A2 of the Appendix, I examine the informativeness using the on-exchange shorts from Nasdaq BX and Nasdaq PSX. I find my results remain qualitatively consistent.

[3]Whether a short is initiated by the seller or the buyer is determined using the Lee and Ready (1991) algorithm. Although the pre-trade transparency is missing in the dark pools, trading in dark pools is often referenced or pegged to the quoted price in the lit markets. Despite the absence of pre-trade transparency in dark pools, dark pool trades typically reference or are pegged to the quoted prices in lit markets. Applying the same logical inference in Lee and Ready (1991) algorithm for lit markets, trades executed at prices above the mid-quote of the lit market are presumed to be initiated by buyers demanding liquidity (LD), while those executed at prices below the mid-quote are likely initiated by sellers providing liquidity (LS).

I begin assessing the informativeness of LD/LS by examining their return predictability. I use portfolio sorting to examine the predictive power of LD/LS over future stock returns. Should LD/LS be informed, I expect that LD/LS predict future stock returns and a long-short portfolio based on these shorts yields an excess return and alpha. My findings reveal that a value-weighted long-short strategy based on LD generates a negative excess return, along with a negative Fama and French (1993) and Carhart (1997) four-factor alpha. Conversely, a value-weighted long-short strategy based on LS fails to yield any statistically significant excess return or Fama and French (1993) and Carhart (1997) four-factor alpha. I also estimate Fama-Macbeth regressions controlling for well-known predictors and find that the LD consistently exhibit a negative association with future stock returns, whereas LS show no such association. This differentiation in return predictability shows that LD are informed about future stock returns, whereas LS are not.

This return predictability alone offers necessary but not sufficient evidence that LD are informed and LS are not. This is because the return predictability could stem either from unknown systematic risks (i.e., risk premium) associated with short-selling activity or from the superior information embedded in short-selling activity. To provide more conclusive evidence, I investigate the predictability of both types of shorts on earnings surprises. If LD possess superior information about earnings, I expect these shorts prior to earnings announcements to forecast earnings surprises. My findings show that LD negatively predict earnings surprises, while LS lack such predictive power. This implies that LD are informed about earnings information prior to the earnings announcements. The distinct predictability of earnings surprises further affirms that LD are informed, whereas LS are uninformed.

To reinforce the distinct predictability between LD and LS surrounding earnings announcements, I examine the association between the market responses to the earnings announcements (measured by Cumulative Abnormal Returns, or CAR) and LD/LS. The market responses to the earnings announcements reflect the proportion of information that is unknown to general market participants before the earnings announcements. If LD/LS are more informed than the average market participants, it follows that pre-announcement LD/LS will be associated with negative post-announcement CAR.

My results confirm a negative association between LD and CAR. No such association, however, emerges for LS. These results support that LD contain negative earnings information but LS do not.

Further, I investigate the association between LD/LS and hedge funds, conventionally regarded as the most informed and sophisticated investors in the market (Brunnermeier and Nagel (2002)). If LD/LS reflect negative information, I hypothesize that hedge funds will abstain from holding the stock with high LD/LS. My results support this hypothesis that hedge fund holdings are negatively associated with LD but not associated with LS. Additionally, non-hedge fund holdings show no association with either type of shorts. The discrepancy implies hedge funds' inclination to avoid holding stocks with high LD but not stocks with high LS, aligning with the notion that LD are negatively informed but LS are not.

In light of the differential informativeness of LD and LS, I study their association with public information acquisition. Typically, the occurrence of more informed trading is preceded by increased public information acquisition. Given the informed nature of LD, I hypothesize that more public information acquisition is followed by higher LD. Using EDGAR search volume as a proxy for public information acquisition, I find that information acquisition is associated with more future LD, whereas there is no such association for LS. This difference confirms the informed trading nature of LD in contrast to uninformed LS.

Last, considering the informed nature of short-selling activity, previous studies observe that short-selling activity is associated with improved informational efficiency (Boehmer and Wu (2013) and Evans et al. (2022)). Consequently, I investigate whether LD/LS are associated with better informational efficiency. My results show that LD enhance informational efficiency around the periods with negative earnings announcements but not around the periods with positive earnings announcements. These distinct effects align with LD being initiated by traders informed of negative information. I also find that LS enhance informational efficiency, especially around the periods with positive earnings announcements. The beneficial effect of LD on informational efficiency aligns with existing literature, which posits that informed short-selling activity aids in correcting mispriced securities and restoring them to their fundamental values

(Diamond and Verrecchia (1987)). The positive effect of LS on informational efficiency is also in accordance with extensive literature that documents a positive association between liquidity and informational efficiency.

The primary contribution of this chapter is to document the different informativeness of LD and LS. Understanding the heterogeneity of information content in short-selling activity is crucial for market participants to develop and implement their trading strategies. Empirical evidence on short-selling shows that short sellers are informed and their shorting activity predicts future returns (Chen et al. (2019a), Boehmer et al. (2008) and Diether et al. (2009)). A typical interpretation of this predictability is that shorting activity is a proxy for the amount of negative information excluded from the market price (Figlewski (1981)). Comerton-Forde et al. (2016b), by decomposing short-sale volume, point out that the short-sale volume is heterogeneous in the intraday horizon. This chapter contributes to the literature by extending Comerton-Forde et al. (2016b) with more comprehensive evidence on the informativeness of LD/LS and their implication for the financial markets. I show that LD and LS are associated with different levels of informativeness, hedge fund holdings, informational efficiency and information acquisition activity. My results provide evidence that not all short-selling activity is informed, helping to better understand the heterogeneity of information content in short-selling activity. This research is also relevant to the regulatory authorities to understand the divergent patterns and information content associated with different types of short-selling activities and to inform the design of more targeted and effective policy interventions conditioning on different types of shorts rather than a one-size-fits-all short-selling regulation.

3.2 Data

I obtain intraday short-sale trades from the FINRA monthly short volume file. This monthly short volume file provides all short-sale trades executed and reported to either a FINRA Trade Reporting Facility (TRF) or FINRA's Alternative Display Facility (ADF). This data is merged with CRSP's daily file, and I filter for U.S. common stock listed in NYSE, NYSE MKT, and NASDAQ (CRSP share codes 10 and 11; exchange codes 1, 2, and 3). To mitigate market microstructure noise, I exclude observations with a monthly closing price of less than \$2.

To construct a measure of LD/LS, I obtain intraday consolidated quotes from Refinitiv Datascope Select and intersect it with the FINRA intraday short-sale trade data. Following the method of Comerton-Forde et al. (2016b), I identify a short-sale trade as a liquidity-demanding short (a liquidity-supplying short) if the trade is seller-initiated (buyer-initiated) as shown in Eq. (3.1), where the trade initiator is determined by the Lee and Ready (1991) algorithm.

$$Short_{i,t,\tau} = \begin{cases} \text{Liquidity-Supplying short}_{i,t,\tau}, & \text{if } P_{i,t,\tau} > \text{NBBO MidQuote}_{i,t,\tau} \\ \text{Liquidity-Demanding short}_{i,t,\tau}, & \text{if } P_{i,t,\tau} < \text{NBBO MidQuote}_{i,t,\tau} \end{cases} \tag{3.1}$$

Where i represents a stock i, t represents a trading day t, and τ represents a trade time τ. P is the trade price of short-sale trade, and NBBO is the National Best Bid and Offer. Subsequently, I aggregate LD/LS volume by day to form the daily LD/LS volume. The daily measures of LD and LS allow me to avoid the mis-classification issues in the intraday LD and LS measures as documented by Chakrabarty, Moulton, and Shkilko (2012). To ensure comparability across stocks, I normalize the daily LD/LS volume by the daily consolidated volume in TRF and ADF during market hours. I combine the daily LD/LS volume with stock returns, consolidated trading volume, market capitalization and share price from the CRSP dataset, financial statement data from Compustat, and institutional holding data from the Thomson Reuters Institutional Holdings (13F) database. Following Shumway (1997), I replace the missing delisting return with -30%. I also gather the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) search volume and Fail-to-Deliver (FTD) volume from the SEC EDGAR database. Finally, I winsorize all variables at the 1% and 99% levels to alleviate the effects of outliers. My sample spans from January, 2010 to March, 2022. My final sample contains 10,313,732 daily observations from 5,237 stocks.

Panel A of Table 3.1 reports stock characteristics at a daily frequency. The mean (median) market capitalization is \$ 5.36b (\$ 0.7b). The mean (median) daily consolidated dollar volume is $ 42.76m ($ 4.48m). The mean (median) short volume is 38.07% (38.61%) of consolidated volume in TRF and ADF during market hours. The mean (median) LD is 11.68% (9.92%) of consolidated volume in TRF and ADF during market hours. The mean (median) LS is 20.09% (30.66%) of consolidated volume in TRF and ADF during market hours. The mean difference between LD and LS is consistent with Chakrabarty et al. (2012).

[Insert Table 3.1 near here]

3.3 The informativeness of LD and LS

The existing literature posits that short sellers are collectively informed.[4] Yet, there is considerable variation in short-sellers and their motivations for taking short positions. Previously, I decomposed short-selling activity into LD and LS following Comerton-Forde et al. (2016b). By construction, LS occur when shorts are initiated from the buy side of the order book. This indicates that LS are passive and contrarian and can occur when market makers provide liquidity without inventory. Such short-selling activity significantly differs from the conventional view of short-selling, where short sellers are perceived as speculating on negative information for profits. This liquidity provision process can result in naked shorts, which, although otherwise illegal, are permissible for market makers.[5] The Securities and Exchange Commission (SEC) Rule 204 permits market makers to engage and postpone their delivery of shorted shares, consequently leading to a "failures-to-deliver" (FTD) occurrence. Consequently, an FTD instance is often an indication of market makers' short-selling activity (Jain and Jain (2015)).

In Table 3.A3 of the Appendix, I show that LS are indeed associated with more FTD volume.[6] This positive association supports that LS are predominantly executed by market makers. Although most investors characterize the liquidity suppliers (i.e., market makers) as uninformed traders, the literature does not necessarily categorize

[4]Boehmer et al. (2008) and Wang et al. (2020) find that short sellers are informed about future stock returns.
[5]A naked short is defined as the sale of a security that the seller does not own and has not borrowed,
[6]I find no such association between LD and FTD volume.

liquidity suppliers as uninformed.[7] Therefore, whether LS are informed remains an empirical question.

Conversely, LD resonate more with the conventional view of short-selling, often characterized by aggressive and speculative trading strategies. These aggressive shorts can cause downward price movements and are typically perceived as informed trades. However, Comerton-Forde et al. (2016b) challenge this consensus by showing that the LD are not fundamentally different from aggressive long-seller order flow. They find that the LD may initially drive downward price pressure but subsequently reverse and have no significant price impact after 24 hours. As a result, the informativeness of LD remains unclear. This uncertainty underscores the need for an extensive examination of the informativeness of LD/LS over an extended horizon.

3.3.1 Contribution to price discovery

To provide preliminary evidence on the informative role of LD/LS, I examine their contribution to the price discovery process. By default, informed shorts should make a more substantial contribution to the price discovery process than uninformed shorts. I measure the contribution to the price discovery process using (i) Hasbrouck (1995) information share (IS) and (ii) Gonzalo and Granger (1995) common factor share (CS). I categorize all trades into three categories: (i) LD, (ii) LS and (iii) other non-short trades. For each stock and each day, I compute the daily IS and CS.[8]

[Insert Table 3.2 near here]

As reported in Panel A of Table 3.1, the mean (median) CS and IS of LD is 73.81% (80.70%). the mean (median) IS of LD is 80.37% (93.24%). The mean (median) CS and IS of LS is 11.99% (6.74%). The mean (median) IS of LS is 7.51% (1.24%). The mean (median) CS of other non-short trades is 14.19% (10.53%). The mean (median) IS of other non-short trades is 12.11% (3.94%). The prominent CS and IS of LD emphasize

[7] Through experiment design, Bloomfield, O'hara, and Saar (2005) find that informed traders use more limit orders than liquidity traders. Collin-Dufresne and Fos (2015) find that 13D activist investors use limit orders, especially for stocks with low liquidity. Brogaard, Hendershott, and Riordan (2019) find that market makers (i.e., high-frequency traders) submit a large number of limit orders, which contribute to most of the price discovery.

[8] The information share varies depending on the order of vector auto-regression (VAR). I calculate the average information share from VARs with different order combinations.

their dominant role in the price discovery process, supporting that LD are informed. Conversely, the small CS and IS of LS indicate that LS are the least informed among the three trade categories, especially given LS accounts for larger proportion of trading volume compared to LD (20.09% v.s. 11.68%) [9] However, these results may vary depending on different stock characteristics. To eliminate this concern, I sort stocks into quintile portfolios based on their market capitalization, book-to-market ratio, idiosyncratic volatility, turnover, order imbalance and relative spread. For each quintile, I compute the average CS and IS of LD, LS and other non-short trades. Table 3.2 depicts the results. I find that LD consistently dominate in price discovery across all classifications, while LS exhibit a lesser role.[10] My investigation into the contribution to price discovery indicates that LD are informed and LS are uninformed.

3.3.2 Return predictability

Previous results of CS and IS describe LD/LS' contribution to the price discovery process at the intraday level. If these types of shorts are genuinely informed, they should possess information beyond the intraday horizon. Therefore, I examine the informativeness of LD/LS by investigating their return predictability in an extended horizon. Following Boehmer et al. (2008), I sort stocks into decile portfolios based on the average LD/LS in the last five trading days. I skip one day to avoid concerns about bid-ask bounce (Kaul and Nimalendran (1990)) and hold these portfolios over the next 20 trading days (i.e. from day 2 through day 21). In addition, I construct a long-short stock portfolio by taking a long position in stocks with the highest LD/LS and a short position in stocks with the lowest LD/LS. I calculate the value-weighted portfolio's excess returns and Fama and French (1993) and Carhart (1997) alphas.

Table 3.3 reports the excess returns and alphas of ten stock portfolios and the long-short portfolio sorted by LD and LS. I adjust the t-statistics based on the Newey-West standard errors with four lags to account for potential autocorrelation. Panel A reports excess returns and alphas for value-weighted portfolios sorted by LD. I find that the long-short stock portfolio based on LD generates a negative and significant

[9] These differences of CS and IS between LD and LS are statistically significant at the 1% level.

[10] For each quintile, the differences of CS and IS between LD and LS are statistically significant at the 1% level.

excess return with a t-stat of -2.67. After controlling for Fama and French (1993) and Carhart (1997) four factors, the long-short portfolio generates a negative alpha with a t-stat of -3.67. The negative return predictability indicates that LD are informed about future stock returns. Panel B reports excess returns and alphas for value-weighted portfolios sorted by LS. I find that the long-short stock portfolio based on LS fails to generate a statistically significant excess return or Fama and French (1993) and Carhart (1997) alpha. The absence of return predictability indicates that LS are uninformed about future stock returns. These results support that LD are informed and LS are uninformed about future stock returns.

[Insert Table 3.3 near here]

Next, I corroborate my results on the predictability of LD/LS using the Fama and MacBeth (1973) cross-sectional regression. The Fama and MacBeth (1973) regression allows me to examine the predictability of shorting activity controlling for other well-known factors. These factors include the size ($Log(MktCap)$), book-to-market ratio (BTM), cumulative return from t-12 to t-2 (MOM), short-term reversal (REV), market beta ($Beta$), Amihud illiquidity ($Illiq$), order imbalance ($OIMB$) and idiosyncratic volatility ($IdiosynVol$).[11] Table 3.4 reports the results of Fama-Macbeth regressions. I first replicate the results of Boehmer et al. (2008) by regressing the next-20-day stock returns on the short ratio (SR) in Column (1).[12] I find that the coefficient estimator of the short ratio (SR) is negative and significant, consistent with Boehmer et al. (2008) that short-selling activity contains negative information and negatively predicts future returns. In Column (2), I decompose the SR into liquidity-demanding short ratio (LD) and liquidity-supplying short ratio (LS), and I find that the coefficient estimator of LD is negative and significant while the coefficient estimator of LS is statistically insignificant. This indicates that LD contain negative information and negatively predict future returns, while LS have no predictive power. In addition, these results also suggest the negative return predictability in Boehmer et al. (2008) stems from LD but not from LS. In Columns (3) and (4), I control for additional predictors and I find my results stay unchanged. Overall, I find that LD negatively predict future

[11] Table 3.A1 defines these variables

[12] The short ratio (SR) is defined as the percentage of short volume relative to total volume over the previous five trading days following Boehmer et al. (2008),

stock returns, while LS do not. The difference in return predictability reinforces that LD are informed and LS are uninformed.

[Insert Table 3.4 near here]

3.3.3 Earnings surprises

Although the difference in return predictability between LD and LS supports that LD are informed and LS are uninformed, alternative explanations might account for this discrepancy in return predictability. For example, LD may contain unobservable or unknown systematic risks that are overlooked in the extant literature, and the return predictability might simply be a manifestation of a risk premium instead of informativeness. Therefore, the return predictability alone is indicative but not definitively conclusive evidence of LD being informed and LS being uninformed. To further investigate the information content of these two shorts, I focus on the information about the firm's fundamentals released in the earnings announcements. Specifically, I examine the predictability of LD/LS on earnings surprises. If LD/LS reflect negative earnings information, it logically follows that high levels of LD/LS would precede negative earnings surprises.

Following Livnat and Mendenhall (2006), I define earnings surprises as the standardized unexpected earnings (SUE):

$$SUE_{i,t} = \frac{EPS_{i,t} - EPSi,t-1}{P_{i,t}} \tag{3.2}$$

Where $EPS_{i,t}$ is the earnings per share of firm i in quarter t, and $P_{i,t}$ is the price of firm i at the end of quarter t. To mitigate the potential noise in the daily short volume measure, I calculate the average LD/LS that is within the five trading days preceding the earnings announcements. I regress average LD/LS on earnings surprises using the regression specification in Eq. (3.3):

$$\text{SUE}_{i,t} = \beta_0 + \beta_1 \text{LS}_{i,t} + \beta_2 \text{LD}_{i,t} + \gamma \text{Controls}_{i,t} + \delta + \epsilon + i, t \tag{3.3}$$

where i represents a firm i, t represents a fiscal quarter t. $\text{LD}_{i,t}$ is the average liquidity-demanding shorts as a percentage of trading volume in TRF and ADF that is within the five trading days prior to the earnings announcements. $LS_{i,t}$ is the average liquidity-supplying shorts as a percentage of trading volume in TRF and ADF that is within the five trading days prior to the earnings announcements. $Controls_{i,t}$ include market capitalization, idiosyncratic volatility, book-to-market ratio and loss indicator. δ are stock fixed effects and year-quarter fixed effects. The standard errors are clustered by stock and year-quarter.

[Insert Table 3.5 near here]

Table 3.5 shows the results of the earnings surprise predictability. Column (1) reports the results of the regression of LD/LS on earnings surprises with stock fixed effects and year-quarter fixed effects. I find that LD are negatively associated with earnings surprises (i.e.,SUE) whereas LS don't exhibit any significant predictive power over SUE. In Columns (2) and (3), I control for additional financial statement characteristics and past returns. I find that LD continue to be negatively associated with earnings surprises, while the association between earnings surprises and LS remains statistically insignificant.[13] Overall, I find that LD are informed about earnings information but LS are not, consistent with my previous findings in return predictability.

One concern in using earnings surprises is that the predictability of earnings surprises might simply capture the information dissemination to the public ahead of earnings announcements. Effectively, when there is earnings information dissemination to all market participants, the predictive power of LD on earnings surprises is only signaling the existence of information dissemination, rather than as evidence of the shorts' informed nature. To further ascertain that LD contain negative information while LS do not, I study the market responses to the earnings announcements. The market responses to the earnings announcements capture the proportion of earnings information that is unknown to the general market participants prior to the earnings announcements. This allows me to examine whether LD/LS are more informed than the average market participants or possess superior information that is unknown by

[13]In Table 3.A4 of the Appendix, I construct the monthly and quarterly average LD/LS that are within one month and one quarter before the earnings announcements, respectively, and re-estimate the regression of earnings surprises on the monthly and quarterly LD/LS. I find that my results on daily and monthly average LD/LS remain consistent with the main results.

other market participants. I use cumulative abnormal return (CAR) in the five trading days following the earnings announcements (earnings announcements date inclusive) to measure market responses to the earnings announcements. Subsequently, I regress CAR on LD/LS using the regression specification in Eq. (3.4).

$$\text{CAR}_{i,t} = \beta_0 + \beta_1\text{LS}_{i,t} + \beta_2\text{LD}_{i,t} + \gamma\text{Controls}_{i,t} + \delta + \epsilon + i, t \tag{3.4}$$

where i represents a firm i, t represents a fiscal quarter t. $\text{LD}_{i,t}$ is the average liquidity-demanding shorts as a percentage of trading volume in TRF and ADF that is within the five trading days prior to the earnings announcements. $\text{LS}_{i,t}$ is the average liquidity-supplying shorts as a percentage of trading volume in TRF and ADF that is within the five trading days prior to the earnings announcements. $Controls_{i,t}$ include market capitalization, idiosyncratic volatility, book-to-market ratio and loss. δ are stock fixed effects and year-quarter fixed effects. The standard errors are clustered by stock and year-quarter. If LD/LS contain negative information that is unknown by general market participants, it logically follows that LD/LS will be negatively associated with market responses to earnings announcements (i.e., a negative CAR).

[Insert Table 3.6 near here]

Table 3.6 shows the results of the market responses to earnings announcements. Column (1) reports the results with stock fixed effects and year-quarter fixed effects. I find that the coefficient of LD is negative and significant, indicating that LD negatively predict market responses to earnings announcements and are more informative than the average market participants. However, the coefficient of LS is statistically insignificant, showing that LS do not possess superior information that is unknown by general market participants. In Columns (2) and (3), I control for additional financial statement characteristics and past returns, and I find my results remain robust.[14] In sum, I find that LD predict market responses to earnings announcements, indicating that LD contain negative earnings information that is not well-known by average market

[14]I construct the daily and monthly LD/LS that are within one day and one month ahead of earnings announcements, respectively, and re-estimate the regression of earnings surprises on the daily and monthly LD/LS. In Table 3.A5 of the Appendix, I find that my results of monthly and quarterly LD/LS remain qualitatively consistent with my main results.

participants, but I fail to find such evidence for LS.

3.3.4 Informed investors: Hedge fund holdings

Given the differential informativeness of LD versus LS, I examine how LD/LS are associated with trading from informed investors. Hedge funds, arguably, are the most informed and sophisticated investors in the stock market. Jiao et al. (2015) show that the simultaneous decrease of hedge fund ownership and increase in short interest indicate a strong informed signal and negatively predict future stock returns. If LD/LS encapsulate the negative information, I hypothesize that hedge funds avoid holding stocks with more negative information (i.e., a high level of LD/LS). I use the change in hedge fund ownership (ΔHIO) to measure hedge fund holding dynamics because the hedge fund ownership is not stationary, leading to spurious results. Specifically, I regress the quarterly change in hedge fund ownership on quarterly average LD/LS using the regression specification in Eq. (3.5).

$$\Delta HIO_{i,t} = \alpha_0 + \alpha_1 \text{LD}_{i,t} + \alpha_2 \text{LS}_{i,t} + \gamma Controls_{i,t} + \sigma + \epsilon_{i,t} \tag{3.5}$$

Where i represents firm i and t represents a fiscal quarter t. $\Delta HIO_{i,t}$ is the quarterly change in hedge fund ownership. $\text{LS}_{i,t}$ is the quarterly average liquidity-supplying shorts as a percentage of consolidated trading volume in TRF and ADF. $\text{LD}_{i,t}$ is the quarterly average liquidity-demanding shorts as a percentage of consolidated trading volume in TRF and ADF. $Controls_{i,t}$ include size, turnover, Book-to-Market ratio, quarterly return, spread, market beta, lag institutional ownership, the inverse of the close price, realized volatility and past 12-month returns. δ are stock fixed effects and year-quarter fixed effects. The standard errors are clustered by stock and year-quarter.

[Insert Table 3.7 near here]

Columns (1) and (2) of Table 3.7 report the results of the change in hedge fund ownership with different controls. I find that the coefficients of LD are negative and significant, indicating that hedge funds avoid holding stocks with high LD. However, the coefficients of LS are statistically insignificant, indicating no discernible adjustment in hedge fund positions in stocks with high LS. The different associations between LD/LS and hedge fund holdings are consistent with LD being informed while LS being uninformed. I also explore how LD/LS are associated with non-hedge fund holdings. Non-hedge fund institutional investors are generally perceived not as informed as hedge funds, and I hypothesize that non-hedge institutional investors do not actively adjust their holdings in stocks that exhibit high LD/LS. I examine this hypothesis by estimating the regression specification in Eq. (3.5) with the replacement of the dependent variable to the change in non-hedge fund ownership (ΔNHIO). Columns (3) and (4) of Table 3.7 report the results of the change in non-hedge fund holdings with different controls. I find that both LD and LS lack a significant association with non-hedge fund holdings. Overall, I observe a negative association between LD and hedge fund holdings, while there is no such association between LS and hedge fund holdings. Additionally, neither form of shorts has a significant association with non-hedge fund holdings. These distinctions collectively corroborate that LD contain negative information, while LS do not.

3.3.5 Information acquisition

Information acquisition must precede informed trading. As more information is acquired, markets are typically paralleled by an escalation in informed trading activities. Taking into consideration that LD are informed and LS are uninformed, one can logically infer that an increase in information acquisition will lead to more LD but not necessarily more LS. Engelberg, Reed, and Ringgenberg (2012) demonstrate that a substantial part of short sellers' information advantage comes from their ability to process public information. In light of these findings, I examine how LD/LS are associated with public information acquisition.

Following Chen, Kelly, and Wu (2020) and Brogaard and Pan (2022), I use EDGAR search volume to proxy the level of public information acquisition.[15] I exclude algorithm-driven search volume following Ryans (2017), and I take the logarithm of the search volume as EDGAR search volume is widely spread (i.e., large kurtosis). I use weekly average short volume and EDGAR search volume to mitigate noise in daily metrics. I

[15] The U.S. Securities and Exchange Commission (SEC) keeps the records of each filling download request in EDGAR log files and provides access to these log files through their website.

regress the future weekly LD/LS on the EDGAR search volume as shown below.

$$\text{Shorts}_{i,t+1} = \beta_0 + \beta_1 \text{Log(Search Volume)}_{i,t} + \gamma \text{Controls}_{i,t} + \delta + \epsilon + i, t \tag{3.6}$$

Where i represents firm i and t represents trading week t. $\text{Shorts}_{i,t}$ is the weekly average LD/LS as a percentage of trading volume in TRF and ADF. $\text{Log(Search Volume)}_{i,t}$ is the logarithm of the weekly EDGAR search volume. $Controls_{i,t}$ include the inverse of the close price, intraday realized volatility and the log of dollar volume. δ are stock fixed effects and week fixed effects. The standard errors are clustered by stock and week.

[Insert Table 3.8 near here]

Table 3.8 shows the results of public information acquisition. Columns (1) and (2) report the results of LD, and Columns (3) and (4) report the results of LS. I find that, in Column (1), the coefficient of EDGAR search volume is positive and significant, indicating that increased public information acquisition is associated with more future LD. Nevertheless, in Column (3), the coefficient of EDGAR search volume is statistically insignificant, indicating no meaningful association between public information acquisition and future LS. When controlling for financial statement characteristics and past returns in Columns (2) and (4), my findings remain robust.[16] Overall, my results show that LD are associated with more public information acquisition, while such an association is absent in the case of LS.

3.4 LD/LS and informational efficiency

The different informativeness of LD/LS implies their different roles in the stock market. In this section, I investigate the implications stemming from the distinct informational content of LD/LS. Specifically, I focus on the association between LD/LS and informational efficiency.

The literature documents that short flows enhance informational efficiency. Such

[16] I construct the monthly average LD/LS and the monthly average EDGAR search volume measure. I re-estimate the regression of the monthly EDGAR search volume measure on monthly LD/LS, and I find that my results remain qualitatively consistent with the main results.

enhancement is largely credited to the informativeness of short flows and their pivotal role in adjusting mispriced securities more closely to the fundamental values, as underscored by Diamond and Verrecchia (1987) and Boehmer and Wu (2013). Given the informativeness of LD, I hypothesize that informed LD enhance informational efficiency. Although LS are uninformed, LS inherently improve liquidity, where liquidity is demonstrated to improve informational efficiency by stimulating arbitrage activity (Chordia, Roll, and Subrahmanyam (2008)). Additionally, when supplying liquidity to market, market makers learn and update their prior about the value of securities from order flows, including informed order flow, which allows market makers to establish more informed and efficient prices, enhancing informational efficiency (Evans et al. (2022)). Therefore, my hypothesis extends that both LD and LS contribute positively to informational efficiency.

I measure informational efficiency using: (i) the absolute value of the 15-second NBBO midpoint return autocorrelation and (ii) the variance ratio. A smaller absolute autocorrelation indicates that the current returns are less predictable by the past returns, indicating a more efficient stock price. The variance ratio refers to the absolute difference between one and the ratio of the 60-second NBBO midpoint return variance over the variance of the 15-second NBBO midpoint variance following O'Hara and Ye (2011), as shown below:

$$\text{Variance Ratio} = |1 - \frac{\sigma^2_{\text{60-sec}}}{4\sigma^2_{\text{15-sec}}}| \tag{3.7}$$

A low variance ratio indicates that prices follow a random walk, indicating a more efficient stock price. Similarly, I average LD, LS and informational efficiency proxies on a weekly basis to mitigate noise in daily metrics. I examine the association between informational efficiency and LD/LS by estimating the following regression specification:

$$\text{InfoEff}_{i,t} = \alpha_0 + \alpha_1 \text{Shorts}_{i,t} + Controls_{i,t} + \sigma + \epsilon_{i,t} \tag{3.8}$$

Where i represents firm i and t represents trading week t. $\text{InfoEff}_{i,t}$ is the weekly average informational efficiency measure. $\text{Shorts}_{i,t}$ are the weekly average LD/LS as a percentage of consolidated trading volume in TRF and ADF. $Controls_{i,t}$ include the inverse of the close price, intraday realized volatility and the log of dollar volume. δ are stock fixed effects and week fixed effects. The standard errors are clustered by stock and week.

Table 3.9 shows the results of informational efficiency. Columns (1) - (3) report the results using autocorrelation as an informational efficiency measure. I find that the coefficients of LD and LS are negative and significant, indicating that LD and LS are associated with better informational efficiency. Columns (4) - (6) report the results using the variance ratio as an informational efficiency measure. I find consistent results that LD and LS remain beneficial to informational efficiency. In summation, my results are consistent with the hypothesis that both LD and LS enhance informational efficiency.

[Insert Table 3.9 near here]

My results reveal that both LD and LS are associated with enhanced informational efficiency, while the channel through which LD/LS improve informational efficiency might not necessarily be the same. LD are typically initiated by traders informed of negative information. As this negative information is progressively incorporated into the stock price through LD, the stock price becomes more efficient. Consequently, I expect the improvement effect of LD on informational efficiency to concentrate on the dissemination of negative news. To investigate this, I focus on periods surrounding earnings announcements characterized by negative earnings surprises.

To ensure that the negative earnings surprises are substantial enough to induce informed short selling, I define an earnings announcement as a negative earning announcement if the standardized unexpected earnings (SUE) falls below the 15th percentile. This definition minimizes the likelihood that informed short sellers would refrain from shorting when the negative earnings surprise is marginal. I subsequently re-examine the association between informational efficiency and LD following the regression specification in Eq. (3.8), focusing on the two-week window before and after the announcements of such significant negative earnings announcements.

Column (1) of Table 3.10 presents the results of informational efficiency proxied by autocorrelation around negative earnings announcements. The coefficient estimate of LD is negative and significant, indicating that more LD are associated with more efficient stock prices. Column (2) reports the results using variance ratio as an informational efficiency proxy. The results consistently shows that more LD are associated with a more close-to-one variance ratio, indicating enhanced informational efficiency during negative earnings announcements.

Columns (3) to (4) of Table 3.10 shift focus to positive earnings announcements. Here, I define a positive earnings news announcement as one where the SUE is above the 85th percentile. However, the coefficient estimates for LD are statistically insignificant during positive earnings announcements, suggesting that the impact of LD on informational efficiency is concentrated during negative earnings announcements. These distinct effects of LS during positive and negative earnings announcements support that LD improves informational efficiency by facilitating the incorporation of negative information when such news is released.

[Insert Table 3.10 near here]

Better liquidity enables different market participants to more effectively and easily price in their information, improving informational efficiency. LS, in particular, improve the liquidity on the buy-side, allowing informed buyers to easily incorporate their positive news into the prices. Furthermore, the influx of information from informed buyers associated with an increase in LS during the positive new periods enables market makers to better assess the value of the securities and provide more efficient and informed quotes. As a result, I expect the effect of LS on improving informational efficiency to concentrate around positive earnings announcements, where LS predominantly provides liquidity to informed buyers. The results in Table 3.10 confirm that the coefficient estimates of LS are negative and significant around positive earnings announcements (Columns (3) and (4)), whereas they remain statistically insignificant around negative earnings announcements (Columns 1 and 2) These distinct effects of LD during positive and negative earnings announcements support that LS improve informational efficiency by enabling informed buyers to incorporate their positive news more effectively during positive earnings periods. These nuanced examinations complement the different roles of LD and LS in enhancing informational efficiency, highlighting their different informational roles in the financial markets.

3.5 Identification strategy

Despite my efforts to control for market and stock characteristics that could affect my results, there remains a possibility that some unobservable factors might be associated with the error term and bias my results. To address this potential omitted variable problem, I use the full repeal of the uptick rule in Regulation SHO (Reg SHO) in 2007 as an instrument to identify the exogenous variation in LD/LS. This approach enables me to draw causal inferences on the effect of LD/LS.

Prior to 2004, short-selling was subject to the uptick rule, which requires short sellers to sell at a price higher than the last trade price. This rule increases short-selling costs and constrains short-selling activity. On July 28, 2004, the SEC launched a pilot program to remove the uptick rule for 1000 randomly selected stocks (treatment stocks) in Russell 3000 constituents, but kept the restrictions on short-selling unchanged for the remaining Russell 3000 constituents (control stocks). On June 13, 2007, the SEC announced plans to remove the uptick rule for all stocks, effectively on July 6, 2007. Following Boehmer, Jones, and Zhang (2020), I refer to this as the “full uptick repeal". The full uptick repeal has no effects on treatment stocks but increases LD/LS for control stocks by reducing their short-selling costs. Therefore, I use the full uptick repeal as an instrumental variable to identify the exogenous variation in LD/LS.

The implementation of the full uptick repeal in 2007 removes the uptick rule for control stocks, lowering the short-selling costs and boosting LD/LS for control stocks. However, this time-series variation in LD/LS may be associated with other unobservable time-series variation. To control for unobservable time-series variation, I use the Reg SHO pilot treatment stocks as a control group, which were not previously constrained by the uptick rule and hence remained unaffected by the full uptick repeal. The time-series variation in treatment stocks will capture the unobservable time-series variation. I formally exploit the exogenous variation in LD/LS by using it as an instrument variable. Specifically, I instrument the LD/LS by the difference in the change in LD/LS between Reg SHO pilot control stocks and treatment stocks after the full uptick repeal. Since the effect of the full uptick repeal is short-lived (Boehmer et al. (2020)), I only focus on the sample spanning from the final quarter preceding the full uptick repeal and the first quarter succeeding the full uptick repeal. I estimate the first-stage regression as shown below.

$$\text{Shorts}_{i,t+1} = \beta_0 + \beta_1 1_{Control,i} \times 1_{Post,t} + \gamma \text{Controls}_{i,t} + \delta + \epsilon + i, t \tag{3.9}$$

Where i represents a firm i and t represents a fiscal quarter t. $1_{Control,i,t}$ is a dummy variable that equals one if stock i is a Reg SHO pilot control stock at time t, and $1_{Post,i,t}$ is a dummy variable that equals one if time t is after the 2007 full uptick repeal. $Controls_{i,t}$ include market capitalization, idiosyncratic volatility, book-to-market ratio and loss indicator. δ are stock fixed effects and year-quarter fixed effects. The standard errors are clustered by stock and year-quarter.

Columns (1) and (2) of Table 3.11 report the first-stage regression results for LD and LS, respectively. I find that the coefficients of the interaction term (β_1) are positive and significant, substantiating that LD (LS) shorts increase by 2.52% (2.04%) for control stocks after the full uptick repeal. Next, I estimate the fitted value of LD/LS ($\widehat{\text{Shorts}_{i,t}}$) from the first-stage regression. To validate the causal association between LD/LS and earnings surprises, I run the second-stage regression by regressing earnings surprises on the fitted value of LD/LS as shown in Eq 3.10.

$$\text{Earnings Surprises}_{i,t+1} = \alpha_0 + \alpha_1 \widehat{\text{Shorts}_{i,t}} + \gamma \text{Controls}_{i,t} + \delta + \epsilon + i, t \tag{3.10}$$

Where Earnings $\text{Surprises}_{i,t+1}$ is SUE or post-earnings CAR. Controls are the same as in the first-stage regression. The standard errors are clustered by stock and year-quarter.

Columns (3) - (6) report the second-stage regression results. Columns (3) and (4) report the results of earnings surprises on the fitted value of LD and LS , respectively. I find consistent results that LD negatively predict earnings surprises, whereas LS remain unassociated with earnings surprises. Similarly, Columns (5) and (6) report the results of the association between the fitted value of LD and LS and market responses to earnings announcements. I find consistent results that LD are negatively associated with post-earnings CAR, while there is no such association for LS. Overall, using the full uptick repeal as an instrument, I confirm that the association between LD and earnings surprises is causal.

[Insert Table 3.11 near here]

I am cautious about re-using the full uptick repeal as an experiment and modify the t-statistics following Heath, Ringgenberg, Samadi, and Werner (2022). Using the modified t-statistics following Heath et al. (2022), my results are only marginally significant. This is because the effect of the full uptick repeal on LD is small. On average, after the full uptick repeal, the control stocks experience a 0.242% and 0.368% higher increase in LS and LD, respectively. The other reason for these marginally significant results is that the effect of the full uptick repeal in increasing short-selling activity is short-lived as documented in Boehmer et al. (2020).

To draw the causal inference on the association between LD/LS and informational efficiency, I again exploit the full uptick repeal as an instrument variable. Specifically, I estimate the regression of informational efficiency measures on the fitted value of LD and LS as detailed below:

$$\text{IE}_{i,t+1} = \beta_0 + \beta_1 \widehat{\text{Shorts}_{i,t}} + \gamma \text{Controls}_{i,t} + \delta + \epsilon + i, t \tag{3.11}$$

Where i represents a firm i and t represents a trading week t. $\text{IE}_{i,t+1}$ is the informational efficiency measure, including daily quoted spread and effective spread. $Controls_{i,t}$ include the log of dollar volume, the inverse of the close price and realized volatility. δ are stock fixed effects and week fixed effects.[17] The standard errors are clustered by stock and week.

Columns (1) and (2) of Table 3.12 report the second-stage regression results of the autocorrelation on the fitted value of LD and LS, respectively. I find consistent results that the coefficients of LD and LS are negative and significant, indicating that more LD and LS enhance informational efficiency. Columns (3) and (4) of Table 3.12 report the second-stage regression results of the variance ratio on the fitted value of LD and

[17] The first-stage and second-stage regression share the same set of controls.

LS, respectively. Similarly, I find that both LD and LS enhance informational efficiency. Overall, using the full uptick repeal as an instrument, I validate that LD/LS improve informational efficiency.

[Insert Table 3.12 near here]

3.6 Conclusion

This study provides a comprehensive perspective on the information content of short-selling volume by decomposing it into LD and LS. My findings have important implications for the financial literature, which has predominantly treated short-sellers as a uniform group and concluded that they are informed. By distinguishing between these two types of short-selling activity, I have demonstrated that LD carry negative information while LS do not.

These results enrich the understanding of the role of short-sellers and their implications in the financial markets and provide some insights to market participants and regulators. My findings suggest that policymakers and market participants should differentiate between LD and LS to better interpret the information content of the short-selling activity and understand their associated impacts.

Table 3.1: Summary Statistics

This table summarizes the number of observations, mean, median, standard deviation, and the 25th and 75th percentiles of variables used in my regression analysis. Panel A reports the summary statistics of stock characteristics at a daily frequency. Panel B reports the summary statistics of financial statement variables at a quarterly frequency. MktCap is the market capitalization in $ 1,000,000,000. DollarVolume is the daily consolidated dollar volume in $ 1,000,000. RET is the daily return including dividends. SR is the short ratio calculated as the daily short-sale volume during market hours scaled by daily consolidated volume in TRF and ADF during market hours. LS is the liquidity-supplying short ratio calculated as the daily liquidity-supplying short volume scaled by daily consolidated volume in TRF and ADF during market hours. LD is the liquidity-demanding short ratio calculated as the daily liquidity-demanding short volume scaled by daily consolidated volume in TRF and ADF during market hours. Price is the daily close price. Search volume is the daily EDGAR view count. CS(LD) is the common factor share of liquidity-demanding shorts in percentage. CS(LS) is the common factor share of liquidity-supplying shorts in percentage. CS(Other) is the common factor share of other non-short trades in percentage. IS(LD) is the information share of liquidity-demanding shorts in percentage. IS(LS) is the information share of liquidity-supplying shorts in percentage. IS(Other) is the information share of other non-short trades in percentage. SUE is the standard unexpected earnings per share. IdiosynVol is the idiosyncratic volatility estimated based on the Fama-French (1993) three-factor model. BTM is the book-to-market ratio. CAR is the five-day cumulative abnormal return following the earnings announcement. Leverage is the leverage ratio calculated as the total debt divided by the total assets. IO is the institutional ownership. The sample period covers from January, 2010 to March, 2022.

Panel A: Summary Statistics of stock characteristics (daily sample)

Variable	Count	Mean	Std	25%	Median	75%
MrkCap ($ billions)	10,313,732	5.36	15.89	0.17	0.7	2.95
DollarVolume ($ millions)	10,313,686	42.76	111.05	0.55	4.48	27.31
Ret (%)	10,313,064	-0.0	2.96	-1.31	0.0	1.3
SR (%)	10,696,417	38.07	22.70	22.59	38.61	53.25
LS (%)	10,696,417	20.09	17.07	3.32	19.33	30.66
LD (%)	10,696,417	11.68	11.68	0.79	9.92	18.26
Price	10,313,732	33.83	45.73	7.0	18.78	41.83
Search Volume	8,989,562	422.06	866.66	0.00	56.00	456.00
CS(LD)(%)	8,842,513	73.81	22.01	60.12	80.70	91.65
CS(LS)(%)	8,842,513	11.99	13.36	2.15	6.74	17.56
CS(Other)(%)	8,842,515	14.19	12.80	3.69	10.53	21.86
IS(LD)(%)	8,562,940	80.37	25.62	70.94	93.24	98.56
IS(LS)(%)	8,562,941	7.51	13.20	0.14	1.24	8.03
IS(Other)(%)	8,562,939	12.11	16.77	0.63	3.94	17.45

Panel B: Summary Statistics of financial statement variables (quarterly observation)

Variable	Count	Mean	Std	25%	Median	75%
SUE (bps)	182,853	6.88	560.17	-59.52	9.95	76.64
IdiosynVol	175,017	0.02	0.01	0.01	0.02	0.03
BTM	172,820	0.66	0.62	0.25	0.49	0.85
CAR (%)	230,266	-0.22	8.09	-2.87	0.0	2.77
Leverage (%)	222,445	0.23	0.22	0.03	0.18	0.37
IO (%)	148,642	66.08	29.23	44.74	73.71	89.45

Table 3.2: Information Share

This table presents the component share and information share of liquidity-demanding/supplying shorts and other non-short-selling trades. CS(LD) is the component share of liquidity-demanding shorts in percentage. CS(LS) is the component share of liquidity-supplying shorts in percentage. CS(Other) is the component share of other non-short-selling trades in percentage. IS(LD) is the information share of liquidity-demanding shorts in percentage. IS(LS) is the information share of liquidity-supplying shorts in percentage. IS(Other) is the information share of other non-short-selling trades in percentage. Stocks are sorted into quintile portfolios based on the market capitalization, book-to-market ratio, idiosyncratic volatility, turnover, order imbalance and relative spread. the order imbalance is calculated using Lee and Ready (1991) and is the share of volume initiated by buyers less the share volume initiated by sellers, normalized by total volume.

	Panel A: Sort by MktCap						**Panel B: Sort by BTM**					
	CS(LD)	CS(LS)	CS(Other)	IS(LD)	IS(LS)	IS(Other)	CS(LD)	CS(LS)	CS(Other)	IS(LD)	IS(LS)	IS(Other)
Low	86.68	6.66	6.66	91.92	3.18	4.9	68.4	14.41	17.19	74.82	9.5	15.68
2	84.95	6.92	8.13	92.05	3.09	4.86	71.36	12.87	15.78	78.34	8.09	13.57
3	80.4	8.49	11.11	89.03	3.99	6.98	75.01	11.24	13.75	82.04	6.72	11.23
4	72.02	12.03	15.95	81.43	6.62	11.95	77.75	10.23	12.02	84.52	6.02	9.46
High	52.92	22.42	24.66	56.38	17.23	26.39	77.75	10.52	11.74	83.84	6.52	9.64
	Panel C: Sort by Idiosyncratic Volatility						**Panel D: Sort by Turnover**					
Low	70.44	13.93	15.63	76.87	9.25	13.88	88.78	5.78	5.44	94.08	2.55	3.37
2	73.22	12.09	14.69	80.15	7.46	12.39	82.62	8.37	9.01	89.26	4.52	6.22
3	74.48	11.35	14.17	81.36	6.83	11.8	76.02	10.88	13.09	83.97	6.18	9.85
4	75.88	10.83	13.29	82.68	6.46	10.86	70.32	13.07	16.61	78.57	7.82	13.61
High	76.02	11.11	12.88	81.93	6.91	11.16	59.64	18.36	22.0	65.03	13.18	21.78
	Panel E: Sort by Order Imbalance						**Panel F: Sort by Spread**					
Low	74.85	10.75	14.4	81.49	6.24	12.28	52.52	22.17	25.31	56.11	17.0	26.89
2	74.43	11.3	14.27	80.87	6.92	12.21	70.52	12.83	16.65	79.76	7.4	12.85
3	73.78	12.37	13.86	80.08	7.97	11.96	80.26	8.68	11.05	89.11	4.08	6.81
4	73.17	12.46	14.37	79.78	7.99	12.23	85.82	6.69	7.49	92.79	2.89	4.32
High	73.85	12.16	13.99	80.87	7.53	11.6	89.15	5.66	5.19	94.45	2.28	3.27

Table 3.3: Stock Portfolio Returns Sorted on liquidity-demanding and liquidity-supplying shorts

This table reports the monthly average excess returns, CAPM alpha, Fama and French (1993) 3-factor alpha, and Fama and French (1993) and Carhart (1997) 4-factor alpha (in percentage) for stock portfolio, as well as for the long-short portfolio (High-Low). Panel A reports ten decile stock portfolios sorted by liquidity-demanding (LD) short ratio. Panel B reports ten decile stock portfolios sorted by liquidity-supplying (LS) short ratio. Following Boehmer et al. (2008), at the end of each day, all securities are sorted into decile portfolios based on the average liquidity-demanding or liquidity-supplying shorts activity over the last five trading days, and a long-short portfolio is formed by buying the highest decile and shorting the lowest decile portfolio. Daily value-weighted returns are calculated with a holding period of 20 trading days (skip the first trading day). t-statistics from errors with Newey-West correction of four lags are reported. The sample runs from January 2010 to March 2022.

Panel A: Performance of Value-weighted stock portfolio sorted by LD

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	Carhart(97)	t-stat
Low	0.12	5.53	0.06	7.80	0.06	9.33	0.06	9.32
2	0.11	5.23	0.05	9.21	0.05	9.60	0.05	9.60
3	0.11	5.11	0.05	8.65	0.05	8.71	0.05	8.71
4	0.10	4.58	0.04	7.47	0.04	7.68	0.04	7.67
5	0.09	4.15	0.03	5.74	0.03	5.78	0.03	5.78
6	0.08	3.59	0.02	3.16	0.02	3.20	0.02	3.22
7	0.07	3.34	0.01	1.95	0.01	1.95	0.01	1.95
8	0.08	3.49	0.02	2.30	0.01	2.31	0.01	2.32
9	0.08	3.17	0.01	1.22	0.01	1.25	0.01	1.25
High	0.09	3.56	0.02	2.20	0.02	2.41	0.02	2.42
High-Low	-0.04	-2.67	-0.04	-3.17	-0.04	-3.67	-0.04	-3.67

Panel B: Performance of Value-weighted stock portfolio sorted by LS

	Exret	t-stat	CAPM	t-stat	FF(93)	t-stat	Carhart(97)	t-stat
Low	0.11	5.13	0.05	7.21	0.05	8.37	0.05	8.37
2	0.09	4.45	0.03	6.50	0.03	7.04	0.03	7.04
3	0.09	4.47	0.03	6.85	0.03	7.10	0.03	7.10
4	0.09	4.36	0.03	6.75	0.03	6.83	0.03	6.85
5	0.08	4.02	0.02	5.55	0.02	5.54	0.02	5.55
6	0.08	3.65	0.02	3.51	0.02	3.50	0.02	3.52
7	0.09	4.01	0.03	4.38	0.03	4.38	0.03	4.38
8	0.08	3.49	0.02	2.06	0.02	2.08	0.02	2.07
9	0.09	3.35	0.02	1.66	0.02	1.76	0.02	1.75
High	0.13	4.48	0.06	3.46	0.06	3.88	0.06	3.90
High-Low	0.02	0.90	0.01	0.39	0.01	0.36	0.01	0.36

Table 3.4: Fama-MacBeth Regression: Liquidity-demanding v.s. liquidity-supplying shorts

This table presents the results from the Fama-Macbeth cross-sectional regressions of stock returns on short ratio, liquidity-supplying shorts, liquidity-demanding shorts and controls. For dependent variable(s), Ret_{t+20} is the cumulative return over the following 20 trading days (skip the first trading day). For independent variable(s), the variables of interest are short ratio (SR) defined as short volume as a percentage of consolidated volume in TRF and ADF over the previous five trading days following Boehmer et al. (2008), liquidity-demanding short ratio (LD) defined as liquidity-demanding shorts as a percentage of consolidated volume in TRF and ADF over the previous five trading days, and liquidity-supplying short ratio (LS) defined as liquidity-supplying shorts as a percentage of consolidated volume in TRF and ADF over the previous five trading days. For control variables, Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. MOM is the cumulative return from t-12 to t-2. REV is the short-term reversal measure that equals the lagged monthly return. IdiosynVol is the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns. ILLIQ is the Amihud illiquidity measure. Turnover is the daily consolidated volume scaled by market capitalization. Beta is the CAPM market beta. ChInvIA is the change in capital investment. MomSeason, MomSeason06yrplus ,MomSeason11yrplus are return seasonality documented in Heston and Sadka (2008). t-statistics from errors with Newey-West correction of four lags are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Return_{t+20}	(2) Return_{t+20}	(3) Return_{t+20}	(4) Return_{t+20}
SR	-0.0215***		-0.0158***	
	(-10.25)		(-8.06)	
LD		-0.0179***		-0.0199***
		(-6.65)		(-7.35)
LS		-0.00245		0.00282
		(-0.79)		(0.95)
Log(MktCap)	-0.117***	-0.110***	-0.111***	-0.105***
	(-12.16)	(-11.19)	(-12.12)	(-11.27)
BTM	-1.017***	-1.016***	-1.199***	-1.199***
	(-33.54)	(-33.54)	(-37.11)	(-37.11)
MOM	-0.0562	-0.0600	-0.242***	-0.246***
	(-1.53)	(-1.64)	(-5.84)	(-5.94)
REV	-1.438***	-1.441***	-2.200***	-2.201***
	(-11.38)	(-11.42)	(-17.14)	(-17.18)
IdiosynVol	28.04***	28.24***	34.68***	34.88***
	(11.55)	(11.65)	(12.67)	(12.78)
ILLIQ	-0.239***	-0.243***	-0.193***	-0.196***
	(-22.91)	(-23.20)	(-18.55)	(-18.83)
Turnover	-0.0149***	-0.0151***	-0.0185***	-0.0186***
	(-22.67)	(-22.90)	(-20.68)	(-20.83)
Beta	-0.0582	-0.0547	-0.0605	-0.0560
	(-1.22)	(-1.16)	(-1.13)	(-1.05)
ChInvIa	-0.0105***	-0.0107***	-0.00339	-0.00401*
	(-7.92)	(-8.07)	(-1.41)	(-1.66)
MomSeason			1.158***	1.108***
			(6.55)	(6.32)
MomSeason06yrplus			1.007***	0.999***
			(7.06)	(7.03)
MomSeason11yrplus			0.318***	0.318***
			(3.15)	(3.15)
Cons	4.542***	4.291***	4.161***	3.980***
	(26.25)	(23.73)	(24.78)	(22.80)
Obs.	7,584,904	7,584,904	5,248,771	5,248,771
R-Square	0.088	0.090	0.104	0.106

Table 3.5: Earnings surprises and liquidity-demanding/supplying shorts

This table presents coefficient estimates from panel regressions of earnings surprises on liquidity-demanding/supplying shorts. For dependent variable(s), $SUE_t t$ is earnings surprises measured as seasonally adjusted earnings innovation at t, scaled by price. The variables of interest are LD and LS, which are weekly liquidity-demanding and liquidity-supplying short ratios before the earnings announcements, respectively. Control variables include the Amihud illiquidity ratio (ILLIQ), the logarithm of market capitalization (Log(MktCap)), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the book-to-market ratio (BTM), quarterly stock returns at t ($Return_t$), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise (Loss), the institutional ownership (IO), the leverage ratio (leverage ratio), the weekly order imbalance one week before the earnings announcements ($OImb\_1$) the past return from 1 year before to 1 month before the earnings announcements ($RET_{t-30,t-225}$), the one-month return before the earnings announcements (RET_{t-30}), the one-week return before the earnings announcements (RET_{t-7}), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) SUE_t	(2) SUE_t	(3) SUE_t
LD	-5.050***	-2.007***	-1.808**
	(-6.20)	(-2.62)	(-2.37)
LS	-0.246	1.020	0.712
	(-0.35)	(1.52)	(1.06)
ILLIQ		325.0***	313.3***
		(4.13)	(4.01)
Log(MktCap)		-110.5***	-137.1***
		(-18.52)	(-22.64)
IdiosynVol		-622.6**	-713.8**
		(-2.00)	(-2.31)
BTM		-243.6***	-193.5***
		(-18.06)	(-13.71)
$Return_t$		179.7***	126.2***
		(14.19)	(8.77)
Loss		-388.7***	-383.2***
		(-57.74)	(-57.34)
IO		-0.875***	-0.295
		(-3.46)	(-1.18)
Leverage		50.25**	66.21***
		(2.40)	(3.18)
OImb_1		0.108***	0.0845**
		(2.76)	(2.14)
$RET_{t-30,t-225}$			167.5***
			(22.38)
RET_{t-30}			135.6***
			(5.01)
RET_{t-7}			-10.10
			(-0.22)
Stock fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes
Observations	132,370	103,160	103,160
Adjusted R2	0.001	0.096	0.107

Table 3.6: Market responses to earnings announcements and liquidity-demanding/suppying shorts

This table presents coefficient estimates from panel regressions of market responses to earnings announcements. For dependent variable(s), CAR_{t+5} is the one-week post-earnings cumulative abnormal return. The variables of interest are LD and LS, which are weekly liquidity-demanding and liquidity-supplying short ratios before the earnings announcements, respectively. Control variables include the Amihud illiquidity ratio (ILLIQ), the logarithm of market capitalization (Log(MktCap)), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the book-to-market ratio (BTM), quarterly stock returns at t ($Return_t$), an indicator variable that takes the value of 1 if $Earn_{i,t}$ is negative, and 0 otherwise (Loss), the institutional ownership (IO), the seasonally adjusted earnings innovation at t, scaled by price ($Earn_t$), the leverage ratio (leverage ratio), the weekly order imbalance one week before the earnings announcements ($OImb\_1$) the past return from 1 year before to 1 month before the earnings announcements ($RET_{t-30,t-225}$), the one-month return before the earnings announcements (RET_{t-30}), the one-week return before the earnings announcements (RET_{t-7}), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) CAR_{t+5}	(2) CAR_{t+5}	(3) CAR_{t+5}
LD	-0.000231**	-0.000344**	-0.000348**
	(-2.10)	(-2.53)	(-2.57)
LS	0.0000842	-0.000111	-0.000103
	(0.99)	(-1.00)	(-0.93)
ILLIQ		0.0152	0.0150
		(1.28)	(1.27)
Log(MktCap)		-0.0261***	-0.0253***
		(-25.82)	(-24.92)
IdiosynVol		0.0175	0.0164
		(0.43)	(0.40)
BTM		0.00318***	0.00190**
		(3.83)	(2.20)
IO		0.000215***	0.000198***
		(5.11)	(4.72)
$Return_t$		-0.00118	0.000723
		(-0.57)	(0.34)
Loss		-0.0271***	-0.0271***
		(-24.53)	(-24.58)
$Earn_t$		0.0000139***	0.0000144***
		(16.39)	(16.78)
Leverage		-0.0212***	-0.0224***
		(-5.76)	(-6.09)
OImb_t+7		0.0000221	0.0000221
		(1.13)	(1.13)
OImb_1		-0.0000158**	-0.0000117
		(-2.23)	(-1.63)
RET_{t-7}			-0.0203***
			(-2.76)
$RET_{t-30,t-225}$			-0.00600***
			(-4.87)
Stock fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes
Observations	154,535	103,155	103,155
Adjusted R2	0.000	0.032	0.032

Table 3.7: The hedge fund holdings and liquidity-demanding/supplying shorts

This table presents coefficient estimates from panel regressions of hedge/non-hedge fund ownership and liquidity-demanding/supplying shorts. For dependent variable(s), ΔHIO is the quarterly change in hedge fund ownership, and ΔNHIO is the quarterly change in non-hedge fund institutional ownership. For independent variable(s), the variables of interest are quarterly average liquidity-demanding short ratio (LD) and liquidity-supplying short ratio (LS). Control variables include the logarithm of market capitalization (Log(MktCap)), turnover ratio (Turnover), book-to-market ratio (BTM), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the inverse of the close price (1/Price), market CAPM beta (Beta), quarterly return (Ret), Relative spread (RelSpread), Lag institutional ownership (Lag IO), the past 12-month return (Mom), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) ΔHIO	(2) ΔHIO	(3) ΔNHIO	(4) ΔNHIO
LD	-0.000549***	-0.000580***	-0.000281	-0.000355
	(-5.66)	(-6.00)	(-0.74)	(-0.91)
LS	-0.00000379	-0.0000244	-0.000188	-0.000135
	(-0.08)	(-0.49)	(-0.97)	(-0.71)
Log(MktCap)	0.0873***	0.0696**	0.386***	0.349**
	(3.19)	(2.57)	(2.90)	(2.57)
Turnover	0.00000591***	0.00000366*	0.000000333	0.00000741*
	(3.70)	(1.92)	(0.08)	(1.68)
BTM	-0.109***	-0.0763***	-0.273***	-0.261***
	(-8.78)	(-6.04)	(-7.64)	(-7.79)
IdiosynVol	-0.00251***	-0.00250***	-0.0121***	-0.0110***
	(-2.98)	(-3.11)	(-3.67)	(-3.47)
1/Price	-0.130***	-0.138***	-0.284***	-0.265***
	(-4.82)	(-5.03)	(-3.21)	(-2.94)
Ret	0.473***	0.472***	1.233***	1.259***
	(17.11)	(16.52)	(19.03)	(18.95)
RelSpread	-1.707***	-2.290***	-8.962***	-8.306***
	(-2.91)	(-3.89)	(-5.47)	(-4.98)
Beta	0.00446	-0.0152	-0.00704	-0.0161
	(0.37)	(-1.11)	(-0.26)	(-0.53)
Lag IO	-0.0128***	-0.0120***	-0.0651***	-0.0635***
	(-4.33)	(-4.23)	(-3.95)	(-3.81)
Mom		0.0818***		0.0358
		(7.54)		(1.33)
Stock fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
Observations	128,485	124,006	128,485	124,006
Adjusted R2	0.030	0.031	0.080	0.079

Table 3.8: Information acquisition and liquidity-demanding/supplying shorts

This table presents coefficient estimates from panel regressions of liquidity-demanding/supplying shorts on Information acquisition. For dependent variable(s), LD_{t+1} is the weekly average liquidity-demanding short ratio at $t+1$. LS_{t+1} is the weekly average liquidity-supplying short ratio at $t+1$. The variable of interest is the log of the EDGAR search volume (Log(Search Volume)). Control variables include the logarithm of market capitalization (Log(MktCap)), the inverse of the close price (1/Price), the book-to-market ratio (BTM), the turnover ratio calculated as the ratio of the total dollar volume over the market capitalization (Turnover), the weekly average order imbalance at $t+1$ ($OIMB_{t+1}$), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the weekly stock return (RET), stock fixed effects and week fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1)	(2)	(3)	(4)
	LD_{t+1}	LD_{t+1}	LS_{t+1}	LS_{t+1}
Log(Search Volume)	0.0340**	0.0292**	0.0120	0.000377
	(2.35)	(2.03)	(0.52)	(0.02)
Log(MktCap)		0.463***		-2.500***
		(10.39)		(-27.55)
1/Price		-3.531***		-8.164***
		(-21.67)		(-21.90)
BTM		0.210***		-0.546***
		(5.48)		(-7.15)
Turnover		0.00636***		0.0103***
		(8.41)		(8.23)
$OIMB_{t+1}$		-1.329**		5.546**
		(-2.30)		(2.44)
IdosynVol		28.60***		55.90***
		(17.22)		(19.07)
RET		0.692***		3.541***
		(3.35)		(9.76)
Stock fixed effects	Yes	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes	Yes
Observations	1,179,096	1,161,683	1,179,096	1,161,683
Adjusted R2	0.000	0.018	0.000	0.043

Table 3.9: The informational efficiency and liquidity-demanding/supplying shorts

This table presents coefficient estimates from panel regressions of informational efficiency on liquidity-demanding/supplying shorts. For dependent variable(s), $|AC|$ is the weekly average of the daily auto-correlation of 15-second NBBO midpoint returns. $|1 - VR|$ is the weekly average of the daily absolute difference between one and the variance ratio based on the 15-second and 1-minute NBBO midpoint return. For independent variable(s), the variables of interest are the weekly average liquidity-demanding short ratio (LD) and the weekly average liquidity-supplying short ratio (LS). Control variables include the logarithm of dollar consolidated volume (Log(Dollar Volume)), the inverse of the close price (1/Price), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\|AC\|$	$\|AC\|$	$\|AC\|$	$\|1-VR\|$	$\|1-VR\|$	$\|1-VR\|$
LS	-0.0000787***		-0.0000517***	-0.000146***		-0.000108***
	(-7.99)		(-4.38)	(-6.78)		(-4.19)
LD		-0.0000603***	-0.0000359**		-0.000104***	-0.0000952**
		(-3.82)	(-2.07)		(-3.04)	(-2.54)
Log(Dvol)	-0.000340***	-0.000476***	-0.000521***	-0.00177***	-0.00207***	-0.00216***
	(-5.32)	(-7.12)	(-7.54)	(-12.67)	(-14.31)	(-14.47)
1/Price	0.00159	-0.000743	-0.00145	-0.00367	-0.00883***	-0.0109***
	(1.42)	(-0.65)	(-1.24)	(-1.52)	(-3.57)	(-4.33)
IdiosynVol	0.000000653***	0.000000672***	0.000000719***	0.00000109***	0.00000113***	0.00000121***
	(80.95)	(78.29)	(79.11)	(65.71)	(63.76)	(64.70)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,090,699	2,008,437	1,947,321	2,090,720	2,008,454	1,947,337
Adjusted R2	0.177	0.176	0.178	0.129	0.129	0.130

Table 3.10: The informational efficiency and liquidity-demanding/supplying shorts around earnings announcements

This table presents coefficient estimates from panel regressions of earnings surprises on liquidity-demanding/supplying shorts with different types of earnings news. Columns (1) and (2) report the coefficient estimates on earnings announcements with negative earnings surprises. Columns (3) and (4) report the coefficient estimates on earnings announcements with positive earnings surprises. The sample spans the periods extending from two weeks before to two weeks after the earnings announcements. For dependent variable(s), $|AC|$ is the weekly average of the daily auto-correlation of 15-second NBBO midpoint returns. $|1 - VR|$ is the weekly average of the daily absolute difference between one and the variance ratio based on the 15-second and 1-minute NBBO midpoint return. For independent variable(s), the variables of interest are the weekly average liquidity-demanding short ratio (LD) and the weekly average liquidity-supplying short ratio (LS). Control variables include the logarithm of dollar consolidated volume (Log(Dollar Volume)), the inverse of the close price (1/Price), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	Negative SUE		Positive SUE	
	(1)	(2)	(3)	(4)
	$\|AC\|$	$\|1 - VR\|$	$\|AC\|$	$\|1 - VR\|$
LS	-0.000000193	-0.000000708	-0.00000168**	-0.00000271**
	(-1.09)	(-1.44)	(-2.55)	(-1.98)
LD	-0.000106**	-0.000152**	0.00000798	-0.0000107
	(-2.05)	(-2.47)	(0.33)	(-0.20)
Log(Dvol)	-0.000560**	-0.00313***	-0.000463*	-0.00265***
	(-1.98)	(-5.52)	(-1.76)	(-4.72)
1/Price	-0.00345	-0.0201**	-0.00560	-0.0160**
	(-0.81)	(-2.48)	(-1.46)	(-2.08)
IdiosynVol	0.000000523***	0.000000885***	0.000000349***	0.000000574***
	(13.67)	(13.85)	(8.76)	(10.50)
Stock fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
Observations	80,594	80,594	80,381	80,381
Adjusted R2	0.024	0.018	0.019	0.012

Table 3.11: Reg SHO pilot program, liquidity-demanding/supplying shorts and earnings announcements

This table presents coefficient estimates from instrument variable OLS regression of Reg SHO pilot program. Columns (1) and (2) report the first-stage instrument variable results of Reg SHO pilot program, and Columns (3) to (6) report the results for the second-stage instrument variable results. For dependent variable(s), LS is the weekly liquidity-supplying short ratio one week before the earnings announcements. LD is the weekly liquidity-demanding short ratio one week before the earnings announcements. $SUE_t t$ is earnings surprises measured as seasonally adjusted earnings innovation at t, scaled by price. CAR_{t+5} is the one-week post-earnings cumulative abnormal return. The variables of interest are the interaction term between treatment and post-Reg SHO pilot program indicator ($1_{Treatment} \times 1_{After}$) in Columns (1) and (2) and the fitted values of LS ($\widehat{LS}$) and the fitted values of LD ($\widehat{LD}$) in Columns (3) - (6). Control variables include the logarithm of market capitalization (Log(MktCap)), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdioVol3F), the book-to-market ratio (BTM), the Amihud illiquidity ratio (ILLIQ), the past return from 1 year before to 1 month before the earnings announcements ($RET_{t-30,t-225}$), the weekly abnormal return of the daily stock return one week before the earnings announcements (RET_t-5), an indicator variable that takes the value of 1 if $Earn$ is negative, and 0 otherwise (Loss), the quarterly earnings per share (Earn), the weekly order imbalance one week before the earnings announcements ($OImb_1$), the weekly liquidity-supplying shorting one week after the earnings announcements (LS_{t+1}), the weekly liquidity-demanding shorting one week after the earnings announcements (LD_{t+1}), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	First-stage		Second-stage			
	(1)	(2)	(3)	(4)	(5)	(6)
	LS	LD	SUE_t	SUE_t	CAR_{t+5}	CAR_{t+5}
$1_{Treatment} \times 1_{After}$	2.037***	2.527***				
	(3.15)	(4.95)				
$\widehat{LS}$			-3.848		-0.0118	
			(-1.60)		(-1.46)	
$\widehat{LD}$				-4.194*		-0.00611**
				(-1.88)		(-2.36)
Log(MktCap)	0.0578	-0.683**	-18.05***	-11.22***	-0.0122*	-0.00148
	(0.13)	(-2.15)	(-3.30)	(-3.91)	(-1.71)	(-1.21)
IdioVol3F	12.47*	4.948	2490.5***	1661.2***	0.867	-0.228
	(1.66)	(0.84)	(3.96)	(3.57)	(0.98)	(-1.17)
BTM	0.246	-0.119	-53.65***	-50.54***	-0.0145*	-0.00747***
	(0.95)	(-0.76)	(-6.95)	(-8.85)	(-1.82)	(-2.97)
ILLIQ	143.4	196.6**	863.9	881.7	4.650	3.472**
	(1.46)	(2.10)	(0.45)	(0.51)	(1.36)	(2.11)
$RET_{t-30,t-225}$	-0.0250	0.479**	43.13**	75.33***	-0.0387	-0.00397
	(-0.09)	(2.22)	(2.11)	(6.26)	(-1.63)	(-0.75)
RET_{t-5}	2.369***	2.928***	198.7***	184.1***	-0.0379	-0.0246
	(2.70)	(4.57)	(2.77)	(2.85)	(-0.92)	(-0.96)
Loss	-0.324*	-0.135	-110.6***	-130.2***	0.0185	-0.00902**
	(-1.73)	(-0.97)	(-6.28)	(-12.60)	(0.89)	(-2.13)
$OImb_1$	0.00994***	-0.00396***	-0.0920	-0.402***	0.000390	-0.000153**
	(9.46)	(-4.11)	(-1.09)	(-3.11)	(1.38)	(-2.43)
EPS	-0.219*	-0.111			0.0121**	0.00625
	(-1.89)	(-1.27)			(2.32)	(1.38)
LS_{t+1}					0.00530	0.000592*
					(1.44)	(1.93)
LD_{t+1}					0.00370	0.00297**
					(1.48)	(2.40)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,341	4,280	4,111	4,067	4,510	4,510

Table 3.12: The informational efficiency and liquidity-demanding/supplying shorts (Reg SHO)

This table presents coefficient estimates from the second-stage instrument variable OLS regression of informational efficiency on liquidity-demanding/supplying shorts during the Reg SHO pilot program. For dependent variable(s), $|AC|$ is the weekly average of the 15-second NBBO midpoint return auto-correlation. $|1-VR|$ is the weekly average of the absolute difference between one and the variance ratio based on the 15-second and 1-minute NBBO midpoint return. For independent variable(s), the variables of interest are the fitted values of liquidity-supplying short ratio ($\hat{LS}$) in Columns (1) and (3), and the fitted values of liquidity-demanding short ratio ($\hat{LD}$) in Columns (2) and (4). Control variables include the logarithm of dollar consolidated volume (Log(Dollar Volume)), the inverse of the close price (1/Price), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), stock fixed effects and year-quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) |AC|	(2) |AC|	(3) |1-VR|	(4) |1-VR|
$\hat{LS}$	-0.00955***		-0.0134***	
	(-8.93)		(-7.34)	
$\hat{LD}$		-0.00734***		-0.0102***
		(-9.26)		(-7.20)
Log(Dollar Volume)	-72.01***	-5.392***	-13.47***	-2.194***
	(-3.69)	(-42.99)	(-3.89)	(-35.29)
1/Price	-17.24	61.77***	39.11***	52.29***
	(-0.66)	(31.66)	(8.17)	(29.06)
IdiosynVol	-0.000232	0.0000819	-0.0000135	0.0000381
	(-1.28)	(1.60)	(-0.62)	(1.53)
Stock fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
Observations	94,057	93,397	94,056	93,396

3.7 Appendix

3.7.1 Informed trading in off-exchange venues

This chapter studies the well-documented informed trading, short-selling, in off-exchange venues, commonly known as dark pools. Despite existing literature indicating that trading in dark pools is generally less informed compared to lit markets (i.e., Zhu (2014)), this does not negate the potential informativeness of dark trading.

A recent study by Wang et al. (2020) shows that both on-exchange and off-exchange short-sellers are informed, Specifically, on-exchanges short sellers tend to trade on short-lived information and, consequently, prioritize the immediacy offered by exchanges despite higher transaction costs. Reed et al. (2020) and Boehmer, Jones, Zhang, and Zhang (2021) further show that off-exchange shorts are informed by long-term information, which takes time to incorporate into prices. This observation provides some insight into the rationale behind short sellers opting to trade in environments with lower execution probabilities, as they are motivated by the prospect of lower liquidity costs associated with trading on long-term information that gradually incorporate into prices.

To enhance the robustness of my findings, I conducted a replication of my results using on-exchange shorts. Specifically, I collect the intraday short transaction data in Nasdaq BX and Nasdaq PSX, as provided by Nasdaq Trader. Using the same methodology as outlined in Comerton-Forde et al. (2016b), I construct the on-exchange LS and LD. The replication of return predictability, using Fama-Macbeth regression on on-exchange LS and LD, is presented in Table 3.A2. I find consistent results that LD negatively predict future stock returns while LS do not exhibit similar return predictability. These findings corroborate the initial results and suggest that my findings are applicable across both off-exchange and on-exchange venues.

3.7.2 Long-term predictability in earnings surprises

Reed et al. (2020) show off-exchange short sellers possess long-lived negative information. To explore the long-term predictability in earning surprises of LD/LS, I examine whether the average LD/LS positioned 1, 2, and 3 months prior to earnings announcements can forecast earnings surprises. Tables 3.A5 and 3.A4 reveal a consistent pattern where LD shorts negatively predict earnings surprises (either measured by SUE or CAR) up to three months before the announcements, whereas LS shorts do not exhibit this predictive capability. I find the qualitatively consistent results that LD that are 1 month, 2 months and 3 months before This distinct long-term predictability in earnings surprises between LD and LS aligns with findings from Reed et al. (2020) and reinforces the earlier conclusion that LD trading is informed and LS is not.

Table 3.A1: Variable definitions

Variable	**Definitions**
CS	Gonzalo and Granger (1995) common factor share based on trades
IS	Hasbrouck (1995) information share based on trades
MktCap	Market value of common shares
BTM	Book value of common shares divided by the market value of equity
MOM	The cumulative returns from month t-12 to t-2
REV	The lagged monthly return
IdiosynVol	The standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock return
ILLIQ	Amihud illiquidity measure based on daily return of last month
Turnover	The daily consolidated volume scaled by the market value of common shares
Beta	CAPM market beta based on daily returns of last month
ChInvIa	The change in capital investment
SUE	The earnings surprises computed as standard unexpected earnings earning (Livnat and Mendenhall (2006))
Loss	An indicator variable that takes the value of 1 if the earnings surprise is negative, and 0 otherwise (Loss)
IO	institutional holdings scaled by the total shares outstanding
Leverage	Total debt divided by total assets
OImb	Order imbalance scaled by the consolidated trading volume
CAR	The cumulative abnormal return from earnings announcements date to n days after earnings announcement
Search Volume	The daily aggregate EDGAR search volume
Dvol	The consolidated dollar trading volume
AC	Daily 15-second NBBO midpoint return auto-correlation
VR	The ratio of the 60-second NBBO midpoint return variance over four times 15-second NBBO midpoint return variance
Spread	The daily time-weighted average of bid-ask spread in basis points of price
EffSpread	The daily time-weighted average of effective spread in basis points of price

Table 3.A2: Fama-MacBeth Regression: shorts on Nasdaq BX and Nasdaq PSX

This table presents the results from the Fama-Macbeth cross-sectional regressions of stock returns on short ratio, liquidity-supplying shorts, liquidity-demanding shorts and controls. The sample consists of on-exchange shorts from Nasdaq BX (formerly the Boston Stock Exchange) and Nasdaq PSX. For dependent variable(s), Ret_{t+1} is the cumulative return over the following 20 trading days (skip the first trading day). For independent variable(s), the variables of interest are short ratio (SR) defined as short volume as a percentage of consolidated volume in TRF and ADF over the previous five trading days following Boehmer et al. (2008), liquidity-supplying short ratio (LS) defined as liquidity-supplying shorts as a percentage of consolidated volume in TRF and ADF over the previous five trading days, and liquidity-demanding short ratio (LD) defined as liquidity-demanding shorts as a percentage of consolidated volume in TRF and ADF over the previous five trading days. For control variables, Log(MktCap) is the logarithm of market capitalization. BTM is the logarithm of the book-to-market ratio. MOM is the cumulative return from t-12 to t-2. REV is the short-term reversal measure that equals the lagged monthly return. IdiosynVol is the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns. ILLIQ is the Amihud illiquidity measure. Turnover is the daily consolidated volume scaled by market capitalization. Beta is the CAPM market beta. ChInvIA is the change in capital investment. MomSeason, MomSeason06yrplus ,MomSeason11yrplus are return seasonality documented in Heston and Sadka (2008). t-statistics from errors with Newey-West correction of four lags are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1)	(2)	(3)	(4)
	Return_{t+1}	Return_{t+1}	Return_{t+1}	Return_{t+1}
SR	-0.0248*		-0.0414***	
	(-1.91)		(-3.11)	
LD		-0.0188***		-0.0275***
		(-3.19)		(-4.06)
LS		-0.00771		-0.00511
		(-1.25)		(-0.61)
Log(MktCap)	-0.119***	-0.118***	-0.119***	-0.137***
	(-10.81)	(-10.85)	(-10.80)	(-11.18)
BTM	-0.903***	-0.904***	-1.043***	-1.206***
	(-31.60)	(-30.92)	(-32.48)	(-26.65)
MOM	-0.0443	-0.0431	-0.0727	-0.174***
	(-0.71)	(-0.69)	(-1.13)	(-2.74)
REV	-1.543***	-1.554***	-2.028***	-1.906***
	(-9.79)	(-9.95)	(-12.85)	(-9.64)
IdiosynVol	29.06***	28.95***	37.58***	37.20***
	(8.44)	(8.38)	(10.64)	(9.53)
ILLIQ	-0.235***	-0.233***	-0.210***	-0.206***
	(-16.98)	(-16.80)	(-15.15)	(-12.28)
Turnover	-0.0157**	-0.00606	-0.0134***	-0.00314
	(-2.18)	(-1.29)	(-3.28)	(-0.44)
Beta	0.0410	0.0463	-0.0119	-0.0624
	(0.80)	(0.91)	(-0.21)	(-0.95)
ChInvIa	0.0252***	0.0148***	0.0462***	0.0446***
	(2.70)	(2.73)	(4.93)	(4.57)
MomSeason			2.174***	1.888***
			(7.95)	(5.87)
MomSeason06yrplus			1.043***	1.104***
			(4.81)	(4.49)
MomSeason11yrplus			1.772***	2.257***
			(11.20)	(5.09)
Cons	4.312***	4.133***	4.310***	4.361***
	(17.93)	(20.23)	(17.39)	(17.52)
Obs.	396,940	396,940	281,523	281,523
R-Square	0.100	0.100	0.116	0.117

Table 3.A3: Failures-to-Deliver and liquidity-demanding/supplying shorts

This table presents coefficient estimates from panel regressions of SEC Failures-to-Deliver volume. For dependent variable(s), FTD_{t+1} is the weekly average fail-to-deliver volume relative to consolidated trading volume $t+1$. The key variable of interest is the weekly average liquidity-demanding short ratio at t (LD_t) and the weekly average liquidity-supplying short ratio at t (LS_t). Control variables include the logarithm of market capitalization (Log(MktCap)), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the book-to-market ratio (BTM), the logarithm of Dollar volume (Log(Dvol)), the inverse of the close price (1/Price), the past 12-month returns (MOM), the weekly average order imbalance (OIMb), the weekly average fail-to-deliver volume relative to consolidated trading volume at t (FTD_t), stock fixed effects and time fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	SEC Fail-to-Deliver		
	(1)	(2)	(3)
	FTD_{t+1}	FTD_{t+1}	FTD_{t+1}
LS	0.0000785***		0.0000666***
	(12.79)		(11.63)
LD		0.00000916	-0.0000102
		(1.34)	(-1.53)
Log(MktCap)	-0.00738***	-0.00771***	-0.00648***
	(-27.79)	(-29.00)	(-28.75)
IdiosynVol	-0.0396***	-0.0364***	-0.0276***
	(-4.97)	(-4.57)	(-4.10)
BTM	-0.000565***	-0.000608***	-0.000571***
	(-3.70)	(-3.98)	(-4.35)
Log(Dvol)	0.00358***	0.00373***	0.00306***
	(26.14)	(27.08)	(26.50)
1/Price	0.0225***	0.0235***	0.0196***
	(12.89)	(13.52)	(13.10)
MOM	-0.00197***	-0.00198***	-0.00187***
	(-14.26)	(-14.32)	(-15.65)
OIMB	0.00939***	0.0118***	0.00654***
	(12.58)	(16.52)	(10.11)
FTD_t	0.329***	0.330***	0.311***
	(88.62)	(88.66)	(91.78)
Stock fixed effects	Yes	Yes	Yes
Week fixed effects	Yes	Yes	Yes
Observations	863,777	863,777	852,079
Adjusted R2	0.086	0.085	0.093

Table 3.A4: Earnings surprises and long-term liquidity-demanding/supplying shorts

TThis table presents coefficient estimates from fixed effects OLS regression of earnings surprises on liquidity-demanding/supplying shorts positioned 1, 2, and 3 months prior to earnings announcements. For dependent variable(s), $SUE_t t$ is earnings surprises measured as seasonally adjusted earnings innovation at t, scaled by price. For independent variable(s), the variables of interest are the average of LD/LS one month before earnings announcements ($\bar{LD}_{t-30,t-1}$, $\bar{LS}_{t-30,t-1}$), the average of LD/LS two month before earnings announcements ($\bar{LD}_{t-60,t-1}$, $\bar{LS}_{t-60,t-1}$) and the average of LD/LS three month before earnings announcements ($\bar{LD}_{t-90,t-1}$, $\bar{LS}_{t-90,t-1}$). Control variables include the logarithm of market capitalization (Log(MktCap)), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the book-to-market ratio (BTM), the Amihud illiquidity ratio (ILLIQ), the past return from 1 year before to 1 month before the earnings announcements ($RET_{t-30,t-225}$), stock fixed effects and quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) SUE_t	(2) SUE_t	(3) SUE_t
$\bar{LD}_{t-30,t-1}$	-2.948***		
	(-2.98)		
$\bar{LS}_{t-30,t-1}$	0.842		
	(1.00)		
$\bar{LD}_{t-60,t-1}$		-3.904***	
		(-3.04)	
$\bar{LS}_{t-60,t-1}$		1.306	
		(1.22)	
$\bar{LD}_{t-90,t-1}$			-5.722***
			(-3.87)
$\bar{LS}_{t-90,t-1}$			1.637
			(1.33)
Log(MktCap)	-40.45***	-40.28***	-40.30***
	(-4.92)	(-4.95)	(-4.96)
IdiosynVol	-1527.1***	-1530.7***	-1525.1***
	(-5.18)	(-5.20)	(-5.18)
BTM	-302.4***	-301.3***	-300.9***
	(-24.13)	(-24.06)	(-24.02)
1/Price	436.2***	432.0***	434.2***
	(3.96)	(3.98)	(3.98)
ILLIQ	16.41	16.48	17.73
	(1.41)	(1.47)	(1.58)
$RET_{t-30,t-225}$	197.8***	197.2***	197.4***
	(15.51)	(15.47)	(15.47)
Stock fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes
Observations	125,276	125,433	125,462
Adjusted R2	0.035	0.034	0.035

Table 3.A5: Market responses to earnings announcements and long-term liquidity-demanding/suppying shorts

This table presents coefficient estimates from fixed effects OLS regression of market responses to earnings announcements. For dependent variable(s), CAR_{t+5} and CAR_{t+30} are the one-week and one-month post-earnings cumulative abnormal return, respectively. For independent variable(s), the variables of interest are the average of LD/LS one month before earnings announcements ($\bar{LD}_{t-30,t-1}$, $\bar{LS}_{t-30,t-1}$), the average of LD/LS two month before earnings announcements ($\bar{LD}_{t-60,t-1}$, $\bar{LS}_{t-60,t-1}$) and the average of LD/LS three month before earnings announcements ($\bar{LD}_{t-90,t-1}$, $\bar{LS}_{t-90,t-1}$). Control variables include the logarithm of market capitalization (Log(MktCap)), the standard deviation of the residual obtained by fitting the Fama-French three-factor model to the time series of daily stock returns (IdiosynVol), the book-to-market ratio (BTM), the Amihud illiquidity ratio (ILLIQ), the past return from 1 year before to 1 month before the earnings announcements ($RET_{t-30,t-225}$), stock fixed effects and quarter fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) CAR_{t+5}	(2) CAR_{t+5}	(3) CAR_{t+5}	(4) CAR_{t+30}	(5) CAR_{t+30}	(6) CAR_{t+30}
$\bar{LD}_{t-30,t-1}$	-0.00110***			-0.00285***		
	(-3.54)			(-6.26)		
$\bar{LS}_{t-30,t-1}$	-0.000315			0.000185		
	(-1.31)			(0.52)		
$\bar{LD}_{t-60,t-1}$		-0.00127***			-0.00379***	
		(-3.24)			(-6.58)	
$\bar{LS}_{t-60,t-1}$		-0.000326			0.000441	
		(-1.09)			(1.00)	
$\bar{LD}_{t-90,t-1}$			-0.00176***			-0.00411***
			(-4.00)			(-6.75)
$\bar{LS}_{t-90,t-1}$			-0.000231			0.000816
			(-0.69)			(1.51)
Log(MktCap)	-0.0281***	-0.0280***	-0.0281***	-0.0558***	-0.0556***	-0.0725***
	(-21.29)	(-21.29)	(-21.31)	(-28.53)	(-28.43)	(-28.42)
IdiosynVol	-0.0822	-0.0791	-0.0771	-0.236**	-0.233**	-0.235**
	(-1.40)	(-1.35)	(-1.31)	(-2.56)	(-2.53)	(-2.54)
BTM	0.00663***	0.00663***	0.00669***	-0.00495	-0.00486	-0.00472
	(3.20)	(3.20)	(3.23)	(-1.58)	(-1.55)	(-1.51)
1/Price	0.00794	0.00882	0.00877	0.0816***	0.0834***	0.0820***
	(0.58)	(0.65)	(0.64)	(3.83)	(3.90)	(3.83)
ILLIQ	0.00105	0.00131	0.00153	0.00972	0.0102	0.0106
	(0.11)	(0.14)	(0.16)	(0.83)	(0.87)	(0.91)
$RET_{t-30,t-225}$	-0.00588**	-0.00587**	-0.00587**	0.00277	0.00253	0.00206
	(-2.13)	(-2.12)	(-2.12)	(0.68)	(0.62)	(0.50)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,780	112,836	112,839	112,780	112,836	112,839
Adjusted R2	0.021	0.021	0.021	0.030	0.030	0.030

Chapter 4

Systematic Internalizers, inter-market competition and market quality

4.1 Introduction

Technology developments and regulatory reforms have increased competition in equity markets over the last twenty years. There has been substantial innovation in the trading mechanisms offered to cater to the needs of investors and to allow market operators to compete for order flow. This innovation spans dark trading mechanisms, periodic auctions and the provision of risk capital. A particularly noteworthy development is the rise in popularity of Single-Dealer Platforms (SDP), risk-taking market actors that supply principal liquidity. SDP are electronic trading platforms managed by broker-dealers who act as principal counterparties when executing their clients' orders. In Europe, SDP have gained market share following the introduction of Markets in Financial Instruments Directive II (MiFID II) and have become a crucial component of the market ecosystem. Despite this growth, little academic research considers SDP.

The implementation of MiFID II in Europe led to the proliferation of SDP in the form of Systematic Internalizers (SI).[1] According to the MiFID II, SI are investment

[1] According to the European Securities and Markets Authority (ESMA), SI market share surged from a trivial amount in 2017 to 25% in the first nine months after the MiFID II implementation. See ESMA consultation paper: MiFID II/Markets in Financial Instruments Regulation(MiFIR) review report on the transparency regime for equity and equity-like instruments, the double volume cap mechanism

firms that, on an organized, frequent, systematic and substantial basis, deal on their own account when executing their client orders. One primary objective of the SI regime under MiFID II is to increase the transparency of off-venue trading. MiFID II mandates SI to provide public quotes with a prescribed minimum quote size, making SI more transparent than other Over-the-Counter(OTC) markets. While this pre-trade transparency on SI remains relatively limited compared to the pre-trade transparency on lit venues (i.e., exchanges),[2] which mandate the publication of all bid and offer prices with corresponding depth. The minimum quote size for SI public quotes only amounts to €1,000 or €2,000 for most liquid stocks, and SI may also offer bilateral quotes without pre-trade transparency when dealing in sizes exceeding Standard Market Size (SMS).[3] In addition, SI are characterized by their ability to trade at sub-tick increments, unlike trading venues constrained by the tick-size regime (TSR). This allows SI to compete with trading venues without being constrained by the TSR.

This chapter investigates the impact of (i) SI sub-tick trading and (ii) the SI minimum quote size on market quality. We ask five questions. First, how does SI's ability to engage in sub-tick trading reshape inter-market competition for liquidity provision and affect the liquidity on SI and lit markets? Second, what is the impact of SI sub-tick trading on SI's capacity to offer price improvement? Third, can SI sub-tick trading enhance informational efficiency? Lastly, does an increased SI minimum quote size attract more trading towards SI, and what are the consequences of this redirected trading on SI?

We begin our empirical analysis by studying the impact of SI sub-tick trading. We study a regulatory change that bans SI sub-tick trading. We separately exploit the effects of banning sub-tick trading on stocks that are constrained by the tick size (referred to as “constrained stocks") and stocks that are not constrained by the tick size (referred to as “unconstrained stocks"). This distinction is relevant as sub-tick trading

and the trading obligations for shares.

[2]Nevertheless, pre-trade transparency requirements in trading venues and SI can be waived if a trade falls into one of four pre-trade transparency waivers categories: (i) the reference price waiver, (ii) the negotiated trade waiver, (iii) the Large-In-Scale (LIS) waiver and (iv) the order management facility waiver.

[3]The SI pre-trade transparency regime requires SI to make public quotes with the size of at least 10% SMS for liquid stocks. The SI pre-trade transparency regime is only applicable to the sizes up to SMS, meaning that SI can offer bilateral quotes without pre-trade transparency for sizes above SMS.

plays a more significant role in constrained stocks than in unconstrained stocks (Kwan, Masulis, and McInish (2015), Buti, Consonni, Rindi, Wen, and Werner (2015) and Comerton-Forde, Grégoire, and Zhong (2019)).

We assess the impact of banning SI sub-tick trading on competition for liquidity provision between lit markets and SI. SI's ability to trade at sub-tick increments allows SI to easily undercut limit orders by offering marginal price improvement, reducing lit market traders' incentive to post limit orders. Banning SI sub-tick trading prevents SI from undercutting limit orders, fostering competition for liquidity provision on lit markets and enhancing lit liquidity. Our results confirm this enhancement in lit liquidity evidenced by narrower European Best Bid and Offer (EBBO) spreads and larger EBBO depth following the ban. In light of the improved lit market liquidity, we also examine the effect of this ban on SI public quotes. We observe an improvement in SI public quotes as well: the spread of SI public quotes narrows after the ban, though without economically meaningful changes in the depth of SI public quotes.

We examine the impact of banning SI sub-tick trading on price improvement. Mechanically, the SI sub-tick trading ban prevents SI from offering sub-tick price improvement, compelling SI either to provide on-tick price improvement or to cease offering price improvement (i.e., execute trades at the EBBO instead). Our results reveal that banning SI sub-tick trading causes SI to offer price-improvement less frequently and to offer price improvement at EBBO more frequently. We find that this effect is stronger in constrained stocks, aligning with sub-tick trading being more important in constrained stocks.

Apart from the frequency of receiving price improvement, traders are also interested in the magnitude of price improvement per trade. Our findings indicate that banning SI sub-tick trading increases the magnitude of price improvement per trade for constrained stocks (from 0.79 ticks per trade to 1 tick per trade) and decreases the magnitude of price improvement per trade for unconstrained stocks (from 2.43 ticks per trade to 2.16 ticks per trade). We also study the overall price-improvement benefits traders receive by combining the frequency and the magnitude of price improvement. The decrease in price-improvement benefits for unconstrained stocks is evident because both the frequency and the magnitude of price improvement decrease post-ban.

Regarding constrained stocks, our results show that the price-improvement benefits decrease post-ban. This decrease implies that the decline in the frequency of price improvement outweighs the positive effect of the increased magnitude of price improvement. Overall, we show that banning sub-tick trading reduces the price-improvement benefits for both constrained and unconstrained stocks.

We also investigate the effect of the SI sub-tick trading ban on informational efficiency. Sub-tick trading allows traders to bypass tick constraints and incorporate their information into sub-tick prices when their valuations lie within one tick, improving informational efficiency (Comerton-Forde et al. (2019)). Conversely, banning sub-tick trading obstructs traders from capturing within-tick valuations, undermining informational efficiency. Our results confirm that banning SI sub-tick trading harms price efficiency for constrained stocks, but not for unconstrained stocks, consistent with Comerton-Forde et al. (2019).

We next investigate the effect of the SI minimum quote size on market liquidity. When providing public quotes, SI can choose the size of their public quotes but the minimum quote size is prescribed at the 10% SMS. These SI public quotes can be viewed as “free options" (Copeland and Galai (1983) and Easley and O'Hara (1987)), and the increase in the SI minimum quote size obliges SI to provide more “free options", consequently attracting more trading to SI. Our results support that an increase in the SI minimum quote sizes leads to more SI trading. We also find that this increased SI trading primarily comes from informed traders rather than uninformed traders. We demonstrate that SI contribute more to the price discovery process when the SI minimum quote size increases. This positive association aligns with the increased SI trading mainly coming from informed traders. We also notice that an increase in the SI minimum quote size widens SI quoted spread. This observation is consistent with SI protecting themselves from the increased exposure when encountering more informed trading. Our results also show that the increase in the SI minimum quote size lowers lit market trading costs documented by narrower EBBO spreads and smaller effective spread and price impact of lit trades. These results collectively suggest that a larger SI minimum quote size attracts informed traders from lit markets to SI,[4] thereby rendering

[4]Zhu (2014) demonstrate that informed traders primarily reside in lit markets.

lit markets less toxic while simultaneously making SI more toxic.

Lastly, since some SI are aware of their clients' (i.e., brokers) identity, SI can infer, albeit imperfectly, the trader type based on the broker identity. This allows SI to better capture the information from informed traders compared to what lit market makers can do. Hence, the incoming informed trading on SI will facilitate the price discovery process, improving informational efficiency. We find the increase in the SI minimum quote size improves the informational efficiency for liquid stocks but not for illiquid stocks. These results resonate with the ESMA's recent proposal and industry consensus advocating for an increase in the SI minimum quote size for liquid stocks but not for illiquid stocks.[5]

This chapter contributes to the literature on sub-tick trading and competition between markets. Buti et al. (2015) show that sub-tick trading is more frequent for liquid stocks (i.e., constrained stocks), and sub-tick trading improves market quality for liquid stocks. Kwan et al. (2015) find that dark pools offer sub-tick trading to attract order flow when the National Best Bid and Offer (NBBO) is constrained. Comerton-Forde et al. (2019) study sub-tick trading on inverted venues and find that the increase in inverted venue trading activity has a positive impact on informational efficiency and liquidity. Our research complements their findings by providing additional evidence on sub-tick trading on SI. Our research provides evidence on how banning SI sub-tick trading changes inter-market competition and affects SI's ability to provide price improvement. Our results are relevant to regulators to understand the consequence of the recent European Commission proposal of further restricting sub-tick trading.[6] We find that banning SI sub-tick trading lowers price-improvement benefits received by traders but prohibits sub-tick undercutting, fostering competition for liquidity provision and improving liquidity.

[5]ESMA found that the current SI minimum quote size is small: (i) the SI minimum quote size for most of stocks only amounts to €1,000, leading to very limited transparency (ii) a large proportion of SI trading activity has a trade size larger than the current minimum quote size. Therefore, ESMA proposed to raise the SI minimum quote size from 10% SMS to 100% SMS. Source: ESMA proposal to raise SI minimum quote size

[6]The European Commission proposed to prohibit SI from matching at the midpoint when trading below twice the standard market size. The European Commission proposed to modify transparency by that SI should no longer be allowed to match at the midpoint below twice the standard market size. .See The European Commission proposal

Our studies provide some of the first evidence on SDP trading and quoting behavior. Aramian and Norden (2021) show that SDP trades have a lower price impact than lit market trades, suggesting that SDP in Europe mainly consist of uninformed traders. Aramian and Norden (2020) find that high-frequency SDP manage inventory imbalances by reducing liquidity provision on exchanges, harming the liquidity on exchanges. Battalio, Hatch, and Saglam (2022) find that routing orders to SDP leads to lower net effective spreads for child orders but higher implementation shortfall costs for parent orders. Our research contributes to the burgeoning literature on SDP by examining two SDP features: sub-tick trading and pre-trade transparency.

Last, this chapter contributes to the literature on pre-trade transparency. Most studies in this literature focus on the effect of pre-trade transparency on liquidity, where mixed evidence is found.[7] We complement the previous literature on pre-trade transparency for lit markets and dark pools with results for SI. Our results show that the increase in the SI minimum quote size improves the lit market liquidity but increases adverse selection costs on SI. We also show that a larger SI minimum quote size improves informational efficiency for liquid stocks, supporting the recent ESMA's proposal to raise the SI minimum quote size.

4.2 Institutional details

Like U.S. equity markets, European equity markets are highly fragmented. Within the MiFID II framework, the European equity trading landscape includes four types of trading mechanisms: Regulated Markets (RM), Multilateral Trading Facilities (MTF), SI and OTC. RM refer to the heavily regulated exchanges (i.e., Euronext), where stocks are listed. MTF, analogous to Alternative Trading Systems (ATS) in the U.S., are alternative venues that match multiple third-party buying and selling interests (i.e., Cboe

[7]Pagano and Röell (1996), Barclay, Christie, Harris, Kandel, and Schultz (1999), Boehmer, Saar, and Yu (2005) and Eom, Ok, and Park (2007) show that higher pre-trade transparency reduces quoted and effective spread and increase depth. However, Madhavan, Porter, and Weaver (2005) and Bloomfield and O'Hara (1999) reach the opposite conclusion that higher pre-trade transparency widen quoted and effective spread. More recently, on the one hand, Buti, Rindi, and Werner (2011) and Buti, Rindi, and Werner (2017) study dark pools and find a higher dark pool activity helps reduce quoted spread and increase depth. On the other hand, Degryse, De Jong, and Kervel (2015) show that competition among dark venues has a detrimental effect on liquidity. In the area of hidden order, Boulatov and George (2013), Bessembinder, Panayides, and Venkataraman (2009) and Bloomfield, O'Hara, and Saar (2015) show that hidden orders improve liquidity.

Chi-X). MTF offer both lit and dark trading mechanisms. SI differ from these trading venues as SI execute their clients' orders on a bilateral basis using their own capital while trading venues match multiple third-party orders. According to ESMA, as of 2020, there were 41 RM, 124 MTF and 124 SI authorized in European equity markets. SI represent 19% of consolidated volume, with RM and MTF accounting for 56%, and the remainder goes to the pure OTC market during 2020.[8]

Unlike U.S. equity markets, European equity markets operate without a trade-through rule or a consolidated tape. Instead, MiFID II imposes the best execution obligations that require brokers to take all reasonable steps to obtain the best possible outcomes for their clients. Nevertheless, brokers can choose the markets they connect to, based on their best execution policy and transaction cost analysis, without an obligation to connect to all markets. The European TSR is more complex than the one in U.S. equity markets. Under MiFID II, the tick size for a stock is jointly determined by the stock's turnover and price level,[9] and stocks with lower turnover and higher prices have larger tick sizes.

The scope of SI was initially introduced in MiFID I and has been extended to a broader range of asset classes in MiFID II. The objectives of the SI regime in MiFID II are mainly twofold: (i) to increase the transparency in off-venue trading and (ii) to level the playing field between trading venues and SI. The SI regime under MiFID II is different from the regime for SDP in U.S. markets. Specifically, SDP in the U.S. are not required to disseminate quotes, whereas MiFID II requires SI to provide public quotes throughout continuous trading hours.

4.2.1 The SI Regime: Sub-tick trading

Under MiFID II, RM and MTF must always comply with the TSR. However, SI enjoy an exemption from the TSR until 26 June, 2020. This exemption grants SI the distinctive capability to trade at sub-tick prices and offer sub-tick price improvement that is not otherwise available on trading venues.

As the market share of SI grew rapidly following MiFID II, ESMA voiced concerns

[8] Source: ESMA Annual Statistical Report 2021.
[9] We provide the detailed Tick-Size table in Table 4.A2 of the Appendix.

about the unlevel playing field between trading venues and SI due to SI's ability to trade at sub-tick increments. Consequently, ESMA proposed to extend the TSR to SI, culminating in the Investment Firms Regulation (EU) No 2019/2033 (IFR, henceforth).[10] Initially scheduled to take effect on 26 March, 2020 but delayed to 26 June, 2020 due to the Covid-19 pandemic, IFR mandates SI to comply with the TSR just as trading venues do, effectively banning SI sub-tick trading.

4.2.2 The SI Regime: Pre-trade transparency

Liquidity assessment in MiFID II is critical to the SI pre-trade transparency. A stock is considered to be a liquid stock if the stock has (i) a free float on RM above €100 million, (ii) an Average Daily Number of Transactions (ADNTE) above 250 and (iii) an Average Daily Turnover (ADT) above €1 million. Any stock failing to satisfy these criteria is classified as illiquid and is exempt from the SI pre-trade transparency requirements.

The cornerstone of the SI pre-trade transparency regime is the minimum quote size requirement that mandates SI to make public quotes at a minimum of 10% SMS.[11] SI need to provide their public quotes consistently during continuous trading hours and execute the incoming orders at the quoted prices. Furthermore, SI can restrict the number of transactions from the same client to manage their exposure to a single client.[12] This allows SI to reduce the adverse selection risk, giving them an advantage in liquidity provision relative to liquidity they might provide on venues.

The above SI pre-trade transparency requirements apply when dealing in sizes up to SMS. In situations where SI are dealing in sizes exceeding SMS, they may offer bilateral quotes to their clients without any pre-trade transparency. SI can decide which clients can access their bilateral quotes. However, we do not have access to SI bilateral quotes, and we focus on the analysis of SI public quotes and the effect of SI minimum quote size, which applies to SI public quotes but does not extend to SI bilateral quotes.

---

[10]Before submitting the proposal to the European Commission for endorsement, ESMA undertook a public consultation on the proposed amendments to the SI's TSR exemption. This proposal received overwhelmingly supportive feedback, and we provide the detailed industry feedback in the Appendix

[11]According to MiFID II, SMS is a discontinuous function of the Average Value of Transactions (AVT). For example, SMS equals €10,000 when the AVT is below €20,000 and equals €30,000 when the AVT is between €20,000 and €40,000. We provide the detailed calculation of SMS in Table 4.6.

[12]See SI pre-trade transparency requirements specified in Articles 14, 15, 16 and 17 of MiFIR.

4.3 Data

We construct our sample by collecting the Annual Transparency Calculation (ATC) data from the ESMA registry. We use the 2020 ATC to identify 782 liquid stocks listed in Euronext Paris, Xetra, Euronext Amsterdam and the London Stock Exchange (LSE). We obtain trade and quote data from the BMLL Technologies dataset. For each stock, we obtain intraday trade and quote data from its primary listing, all available MTF, and all available Approved Publication Arrangements (APA).[13] As for SI public quotes, we obtain the daily SI public quote message data from all available APA for each stock and each SI. We measure the competitiveness of SI public quotes for each SI based on (i) public quoted spreads, (ii) the frequency of public quotes at the prevailing EBBO and (iii) the frequency of public quotes better than the prevailing EBBO. We find that not all SI provide competitive public quotes. For example, the average SI public quoted spread provided by Credit Suisse is ten times wider than the EBBO spread of stocks for which Credit Suisse acts as SI,[14] whereas Hudson River Trading, on average, provides a narrower SI quoted spread than the EBBO spread of stocks for which Hudson River Trading acts as SI. To avoid the uncompetitive SI public quotes contaminating our results, we exclude them from our analysis, focusing only on SI public quotes provided by the eighteen most competitive SI.[15] We do not have access to the identity of SI trades. Hence, we cannot link SI trades to a specific SI.

[13] All available MTF in our sample include Cboe Europe, Turquoise, Aquis exchange, SIGMA X, UBS MTF, POSIT Dark (ITG), Blockmatch MTF (Instinet Europe), and Liquidnet Europe. All available APAs are Cboe Europe's suite of Trade Reporting Services and LSE TRADEcho

[14] Under the MiFID II SI regime, SI are assessed on a stock basis: investment firms are required to assess whether they are SI in a specific stock. This means that a SI can be SI for stock a but not for stock b.

[15] We measure the competitiveness of SI based on its public quotes from three dimensions: (i) SI average spread, (ii) the average proportion of time providing SI public quotes at the EBBO and (iii) the average proportion of time providing SI public quotes better than the EBBO. A SI is classified as a competitive SI if the average spread of SI public quote is less than 1.2 times the EBBO spread, the average proportion of time providing SI public quotes at the EBBO is greater than 30%, and the average proportion of time providing SI public quotes better than EBBO is greater than 10%. The eighteen competitive SI include XTX Markets, XTX Markets SAS, Citadel Connect Europe - Systematic Internalizer, BOFA Securities Europe - Systematic Internalizer, Merrill Lynch International - Systematic Internalizer, Hrteu Limited - Systematic Internalizer, J.P. Morgan Securities PLC, J.P. Morgan AG - Systematic Internalizer, EXANE BNP Paribas - Systematic Internalizer, Morgan Stanley AND CO. International PLC - Systematic Internalizer, Morgan Stanley Europe S.E. - Systematic Internalizer, Deutsche Bank AG - Systematic Internalizer, UBS Europe SE - Systematic Internalizer, UBS AG London Branch - Systematic Internalizer, Goldman Sachs International - Systematic Internalizer, VIRTU Financial Ireland Limited - Systematic Internalizer, Jane Street Financial Ltd - Systematic Internalizer, Tower Research Capital Europe Ltd. We report the summary statistics of public quotes for each SI in Table 4.A3 of the Appendix.

Additionally, we obtain daily market capitalization, price, and industry classification from Refinitiv Datastream. We convert all prices to euros using the intraday foreign exchange rate from Refinitiv DataScope Select. Our sample spans from April 1, 2020 to September 30, 2020.[16] We exclude stocks with zero consolidated euro volume over the sample period, resulting in a final sample comprising 747 liquid stocks with 112,213 stock-day observations and 1,532,499 stock-day-SI observations.

Panel A of Table 4.1 reports stock characteristics. The mean (median) market capitalization is €9.02b (2.29b). The mean (median) daily consolidated euro volume is €61.04m (10.32m).[17] The mean (median) euro volume on SI is 12.8% (10.3%) of consolidated euro volume. SI sub-tick trading, on average, accounts for around 10% of SI euro volume before the SI sub-tick trading ban.[18] For each stock and each day, we calculate the time-weighted spread and depth of SI public quotes for each SI and subsequently average across all competitive SI to obtain SI average time-weighted spread and depth.[19] The mean (median) SI average time-weighted spread is 34.8 basis points (20.0 basis points). The mean (median) SI average time-weighted depth is €4,300 (€4,150). We also aggregate all SI public quotes to form a "SI consolidated tape", and compute the corresponding time-weighted spread and depth of this "consolidated tape". We compare SI time-weighted consolidated spread and depth with the EBBO time-weight spread and depth. We show that SI time-weighted consolidated spread (24.1 basis points) is narrower than the EBBO time-weighted spread (30.3 basis points) and SI time-weighted consolidated depth (€44,920) is larger than the EBBO time-weighted depth (€20,250). SI average trade size is also larger than the average lit market trade size. The mean SI trade size is €11,620, significantly larger than the average lit market trade size of €2,590.

To understand the extent to which SI contribute to the price discovery process, we estimate the SI relative contribution to the price discovery processing using (i) Gonzalo and Granger (1995) common factor share (CS) and (ii) Hasbrouck (1995)

[16] Xetra went through an outrage on July 1, 2020, leading to a 3-hour trading halt on this date: there are two-hour trading halts and one hour for pre-open auction and intraday auction. To avoid this disruption contaminating our results, we drop all stocks listed in Xetra on July 1, 2020.

[17] The daily consolidated euro volume includes all euro volume from all trading mechanisms throughout the day (including trading outside continuous trading hours)

[18] A sub-tick trade refers to a trade whose price is not an integer multiple of the tick size.

[19] For each SI, we calculate SI depth as the euro depth at the best bid and offer of SI public quotes.

information share (IS). We estimate CS and IS by running the Vector Error Correction Model (VECM) regressions using SI trade prices and lit trade prices. These price discovery shares (CS and IS) allow us to analyze the relative contribution to the price discovery process by SI trades compared to lit trades. As shown in Panel A of Table 4.1, on average, the CS of SI trades is 24.6% and the IS of SI trades is 14.2% for liquid stocks, both of which are lower than the percentage of SI euro volume relative to the total lit and SI euro volume (28.4%). This indicates that, on average, SI trades contribute less to the price discovery process than lit markets do and are less informed than lit trades, consistent with Aramian and Norden (2021).

Panel B of Table 4.1 reports the descriptive statistics of SI public quotes for all stock-day-SI observations.[20] The mean (median) time-weighted spread of SI public quotes is 32.9 basis points (17.6 basis points). The mean (median) time-weighted depth of SI public quotes is €4,368 (€2,040). We also calculate the percentage of SI euro depth relative to SMS. The median and 75th percentile of the ratio of SI euro depth to SMS is 20.3% and 24.8%, both of which exceed the SI minimum quote size of 10% SMS. SI, on average, post public quotes at the EBBO for 34.9% of continuous trading hours and post public quotes better than the EBBO for 27.4% of continuous trading hours.

[Insert Table 4.1 near here]

4.3.1 Determinants of SI market share

We assess the determinants of cross-sectional variation in SI market share. For each stock, we compute the daily SI market share, the log daily consolidated euro volume, the time-weighted average EBBO spread, the percentage of continuous trading hours the spread is constrained by the tick size (referred to as the tick-constraint time), and the log market capitalization over our sample period. Then, we estimate an ordinary least squares (OLS) regression by regressing the daily SI market share on the above potential determinants, as shown in Eq. (4.1).

$$\text{SI shr}_{i,t} = \beta_0 + \beta_1 1/Price_{i,t} + \beta_2 \text{RealVol}_{i,t} + \delta + \lambda X_{i,t} + \epsilon_{i,t} \tag{4.1}$$

[20]This summary statistics are based on SI public quotes from competitive SI.

Where i represents stock i, and t represents trading day t. SI shr is the percentage of consolidated euro volume executed on SI. $1/Price$ is the inverse of the volume-weighted trade price. RealVol is the realized volatility of 15-second EBBO midpoint returns. δ are controls, including date fixed effects, industry fixed effects, and country fixed effects. X are other potential determinants of SI market share, including market capitalization, the consolidated euro volume and the tick-constraint time. $\epsilon_{i,t}$ is the standard error clustered by stock and date.

Table 4.2 depicts the coefficient estimates following the regression specification of Eq. (4.1). We find that stocks with larger market capitalization, higher prices and higher consolidated euro volume have more SI trading. We notice that the EBBO spread is positively associated with SI market share (Column (2)), indicating that stocks with wider EBBO spread have more SI trading. This positive association aligns with the competitive relationship between SI and lit markets. We find that the tick-constraint time is positively associated with the SI market share (Column (3)), indicating that stocks that are more frequently constrained by the tick size have more SI trading. We also find a negative association between SI trading activities and volatility, showing that SI trading is relatively low in volatile periods. This negative association is consistent with the negative association between SDP trading activity and volatility found by Battalio et al. (2022) and the negative association between dark trading activity and volatility documented by Menkveld, Yueshen, and Zhu (2017) and Buti et al. (2011).

[Insert Table 4.2 near here]

4.4 SI sub-tick trading and market quality

Our primary goal in this section is to investigate the effects of sub-tick trading on market quality. The empirical challenge is that some unobservable factors that affect both SI sub-tick trading and market quality might exist, leading to an omitted variable problem. To alleviate this concern, we study a regulatory change that bans SI sub-tick trading activity, enabling us to identify exogenous variation in SI sub-tick trading and examine its causal impact.

Although the SI sub-tick trading ban applies to all stocks simultaneously, the ban's effect differs between constrained and unconstrained stocks. Considering the importance of sub-tick trading in constrained stocks (Kwan et al. (2015), Buti et al. (2015) and Comerton-Forde et al. (2019)), we hypothesize that SI sub-tick trading ban has more pronounced effects on constrained stocks than on unconstrained stocks. Our empirical approach exploits this cross-sectional difference in the effect of the SI sub-tick trading ban between constrained and unconstrained stocks by using the difference-in-difference (*diff-in-diff*) analysis. Specifically, we compare the changes in market quality between constrained and unconstrained stocks after the SI sub-tick trading ban to identify the causal effects of the SI sub-tick trading ban on market quality. In our *diff-in-diff* specification, we identify constrained and unconstrained stocks based on the pre-ban average tick-constraint time. Specifically, we classify stocks as constrained stocks if their pre-ban average tick-constraint time is above the 75th percentile, and as unconstrained stocks if it is below the 25th percentile.[21] We define the pre-ban period (post-ban period) as one month before (after) the implementation date of the SI sub-tick trading ban.[22] The pre-ban period starts on 26 May, 2020 and ends on 25 June, 2020, and the post-ban period starts on 26 June, 2020 and ends on 26 July, 2020. Our final sample comprises 180 constrained stocks and 180 unconstrained stocks with 15,922 stock-day observations and 197,843 stock-day-SI observations.

[Insert Table 4.3 near here]

We report the pre-ban summary statistics of constrained and unconstrained stocks in Panels A and B of Table 4.3. On average, constrained stocks are constrained by the tick size for 37% of continuous trading hours, whereas unconstrained stocks are constrained by the tick size for less than 2% of continuous trading hours. On average, constrained stocks have larger market capitalization and higher consolidated euro volume than unconstrained stocks.

[21]Note that in the following regression analyses, our sample consists solely of constrained and unconstrained stocks. Stocks with pre-ban average tick-constraint times between the 25th and 75th percentiles are excluded from the analysis. However, in the robustness section, we conduct an analysis using the full sample.

[22]Alternatively, we define the pre-ban period (post-ban period) as two months before (after) the implementation date of the SI sub-tick trading ban, and our results remain qualitatively consistent. However, this alternatively pre-ban period is close to the covid eruption around March, 2020, which might contaminate our results. We present these results in Table 4.A6 and 4.A7 of the Appendix.

4.4.1 Sub-tick trading and Inter-market competition

According to ESMA, the SI sub-tick trading ban aims to level the playing field between trading venues and SI. SI's ability to trade at sub-tick increments gives them an unfair advantage over trading venues, because SI can easily undercut lit market limit orders using sub-tick trading and divert orders towards SI. Effectively, this advantage discourages liquidity provision on lit markets. In contrast, banning SI sub-tick trading incentivizes liquidity provision on lit markets because traders who provide limit orders will no longer face the risk of being undercut by SI sub-tick trading. Therefore, we hypothesize that banning SI sub-tick trading intensifies competition for liquidity provision on lit markets, leading to enhanced lit liquidity. We measure the lit liquidity using (i) the EBBO time-weighted spread and (ii) the EBBO time-weighted depth. To test this hypothesis, we conduct the *diff-in-diff* analysis based on the SI sub-tick trading ban to estimate the impact of banning SI sub-tick trading on lit market liquidity with the following OLS regression specification:

$$Y_{i,t} = \beta_0 + \beta_1 1_{[Constrained]i} \times 1_{[post]t} + \beta_2 1_{[post]t} + \lambda X_{i,t} + \epsilon_{i,t} \tag{4.2}$$

Where i represents stock i, and t represents trading day t. $Y_{i,t}$ is the EBBO time-weighted spread and depth. $1_{[Constrained]i}$ is the constrained group indicator that equals 1 if stock i is in the constrained group and 0 otherwise. $1_{[post]t}$ is the post-ban indicator that equals 1 if date t is in the post-ban period and 0 otherwise. $X_{i,t}$ are controls, including daily realized volatility of 15-second EBBO midpoint returns, the inverse of the share price, the consolidated euro volume and stock fixed effects.[23] $\epsilon_{i,t}$ is the standard error clustered by stock and date.

In our following empirical analyses, we focus on the coefficients of the interaction term (β_1) and the post-ban indicator (β_2). We interpret the post-ban indicator coefficient (β_2) as the average change in the dependent variable after banning SI sub-tick trading for unconstrained stocks.[24] We interpret the sum of coefficients of the

[23]We use these controls following the previous studies by O'Hara and Ye (2011), Buti et al. (2015) and Comerton-Forde et al. (2019)

[24]We are aware that the post-ban indicator coefficient might also capture the time-series variation in dependent variables driven by unobservable or unmeasurable factors. To alleviate this concern, we

interaction term and the post-ban indicator ($\beta_1 + \beta_2$) as the average change in the dependent variable after banning SI sub-tick trading for constrained stocks.[25]

[Insert Table 4.A12 near here]

Column (1) of Table 4.A12 reports the coefficient estimates for the EBBO time-weighted spread regression. The post-ban indicator coefficient (β_2) and the sum of the coefficients of the interaction term and the post-ban indicator ($\beta_1+\beta_2$) are both negative and significant,[26] indicating that the EBBO time-weighted spread narrows after the SI sub-tick trading ban for both constrained and unconstrained stocks. On average, banning SI sub-tick trading leads to decreases in the EBBO time-weighted spread of 0.7 basis points for constrained stocks and 3.5 basis points for unconstrained stocks, compared to the pre-ban level. This translates to the EBBO spread narrowing by 5.1% (6.4%) for constrained stocks (unconstrained stocks) of its pre-ban average. The reason behind the smaller decrease in the EBBO time-weighted spread in constrained stocks is two-fold: (i) compared to constrained stocks, unconstrained stocks are less liquid and have a larger pre-ban EBBO time-weighted spread, and (ii) it is more challenging to narrow the EBBO spread for constrained stocks whose spread is more frequently constrained.

Column (2) of Table 4.A12 shows the coefficient estimates for the EBBO time-weighted depth regression. Both the post-ban indicator (β_2) and the interaction term coefficient estimates (β_1) are positive and significant, signaling an increase in the EBBO time-weighted depth post-ban. On average, banning SI sub-tick trading increases the EBBO time-weighted depth by €2,879.4 for constrained stocks and €862.3 for unconstrained stocks, corresponding to 7.6% and 6.8% of its pre-ban mean, respectively. The stronger effect in constrained stocks is consistent with O'Hara, Saar, and Zhong (2019) that traders choose to compete on the depth dimension when the spread is constrained. The

study the time-series variation in dependent variables in stocks that have zero SI euro volume before IFR by estimating time-series OLS regressions with the same set of control variables. We do not find statistically significant differences in dependent variables before and after the implementation of IFR, suggesting that , on average, those unobservable or unmeasurable factors do not drive the time-series variation in dependent variables.

[25]We report the statistical significant of the sum of coefficients of the interaction term and the post-ban indicator ($\beta_1 + \beta_2$) in the Appendix.

[26]In Column (1) of Table 4.A4 in the Appendix, we show that the sum of the coefficients of the interaction term and the post-ban indicator is negative and significant.

improvement in lit liquidity reinforces our hypothesis that banning SI sub-tick trading incentivizes liquidity provision on lit markets and improves lit liquidity.

Following the ban, SI lose their sub-tick trading advantage, and intensified competition for liquidity provision on lit markets may further incentivize SI to provide more or cheaper liquidity. We examine the effect of banning sub-tick trading on SI public quotes. We estimate the OLS regression of Eq. (4.2) with the same set of control variables. The dependent variables are replaced with SI time-weighted spread and SI time-weighted depth, and we additionally include SI fixed effects to control for within-SI variation.

Column (3) of Table 4.A12 describes the results of SI spread. The effect of banning SI sub-tick trading on SI spread is similar to the effect on the EBBO time-weighted spread. The post-ban indicator coefficient estimate and the sum of the coefficients of the interaction term and the post-ban indicator are negative and significant,[27] indicating a narrower SI spread following the SI sub-tick trading ban. After banning SI sub-tick trading, SI spread narrows by 0.7 basis points in constrained stocks and 3.0 basis points in unconstrained stocks compared to the pre-ban level, representing about 5.2% and 5.1% of the pre-ban average SI spread, respectively.

Column (4) of Table 4.A12 reports the regression results of SI depth. The post-ban indicator is positive and marginally significant, indicating that SI provide larger depth after the SI sub-tick trading ban for unconstrained stocks. The interaction term coefficient estimate is negative and is smaller than the post-ban indicator coefficient estimate, indicating an increase in SI depth for constrained stocks. However, the effect of SI depth is economically negligible (around 0.1% - 0.2% of the pre-ban average level). Overall, our results underscore enhanced lit liquidity and SI liquidity post-ban, demonstrating that the SI sub-tick trading ban intensifies competition for liquidity provision.

[27] In Column (2) of Table 4.A4 in the Appendix, we show that the sum of the coefficients of the interaction term and the post-ban indicator is negative and significant.

4.4.2 Price improvement and the SI sub-tick trading ban

After banning sub-tick trading, SI are constrained to trade on the specified tick schedule and are no longer able to provide sub-tick price improvement.[28] For each trade that would have benefited from sub-tick price improvement, SI must decide whether to stop or continue providing price improvement at on-tick prices. The decision to stop providing sub-tick price-improvement trades will lower the frequency of price-improvement trades, but where the price improvement is offered it will be at a higher magnitude per trade. Consequently, we hypothesize that SI will provide price-improvement trades less often after banning SI sub-tick trading. We measure the frequency of SI price-improvement trades as the proportion of SI trades whose price lies within the prevailing EBBO (denoted as $f(\text{PI})$). We estimate the OLS regression of the frequency of SI price-improvement trades following the specification of Eq. (4.2) with the replacement of the dependent variable to $f(\text{PI})$.

Column (1) of Table 4.A13 reports the results of the frequency of price-improvement trades. The coefficient estimates on both the interaction term and the post-ban indicator are negative and significant, showing that banning SI sub-tick trading lowers the frequency of price-improvement trades. The negative coefficient on the interaction term further shows that the decrease in the frequency of price-improvement trades is more pronounced in constrained stocks. After banning sub-tick trading, the frequency of SI price-improvement trades for unconstrained stocks decreases by 2.8%, whereas the frequency of SI price-improvement trades for constrained stocks decreases by 8.3%. This decrease is economically meaningful; the decrease in the frequency of SI price-improvement trades in constrained stocks corresponds to about 39% of the pre-ban average level.

As SI cannot execute trades at sub-tick prices following the ban, it prompts the question: at what price will SI execute trades that previously would have received price improvement? Assuming that SI choose to execute these trades at the closest on-tick prices, SI will end up executing these trades at the EBBO prices. We hypothesize that, after banning SI sub-tick trading, SI will execute trades at the EBBO more frequently.

[28]SI cannot trade at sub-tick prices unless they meet the criteria for one of three waivers: (i) the reference price waivers, (ii) the negotiated trade waiver and LIS waiver.

We measure the frequency of SI trades at EBBO as the proportion of SI trades at EBBO (denoted as $f(\text{EBBO})$). We assess the effect of banning SI sub-tick trading on the frequency of SI trades at EBBO by estimating OLS regression using the same specification of Eq. (4.2) with the replacement of the dependent variable to $f(\text{EBBO})$.

[Insert Table 4.A13 near here]

Column (2) of Table 4.A13 reports the results of the frequency of SI trades at EBBO. The coefficient estimates on both the interaction term and the post-ban indicator are positive and significant, indicating that banning SI sub-tick trading increases the frequency of SI trades at EBBO. The positive coefficient on the interaction term indicates that this increasing effect is stronger in constrained stocks than in unconstrained stocks. We find that the increase in the frequency of SI trades at EBBO is about the same magnitude as the decrease in the frequency of SI price-improvement trades. After banning SI sub-tick trading, the frequency of SI trades executed at EBBO increases by 2.9% for unconstrained stocks and 8.9% for constrained stocks.

Traders care about not only the frequency of receiving price improvement but also the magnitude of price improvement per trade. Banning sub-tick trading affects the magnitude of price improvement per trade in two ways. First, banning sub-tick trading forces SI to provide price-improvement trades at a minimum of one tick improvement, increasing the magnitude of price improvement per trade. Second, the SI sub-tick trading ban forces SI to round to the closest on-tick prices. Depending on the pre-ban level, this rounding decision can either increase or decrease the magnitude of price improvement per trade. For example, SI i provides one and a half ticks of price improvement for stock j before the SI sub-tick trading ban. SI i can choose to provide either one tick or two ticks of price improvement post-ban.

We focus on SI trades that receive price improvement, and we measure the intraday magnitude of price improvement per trade: (i) in basis points of the EBBO midpoint

$(RelPI_{i,t,\tau})$ and (ii) by the number of tick sizes $(TickPI_{i,t,\tau})$ as shown below:

$$PI_{i,t,\tau} = \begin{cases} P_{i,t,\tau} - \text{EBO}_{i,t,\tau} & \text{for buyer-initiated trades} \\ P_{i,t,\tau} - \text{EBB}_{i,t,\tau} & \text{for seller-initiated trades} \end{cases} \tag{4.3}$$

$$RelPI_{i,t,\tau} = \frac{PI_{i,t,\tau}}{\text{the EBBO Midpoint}_{i,t,\tau}} * 10000 \tag{4.4}$$

$$TickPI_{i,t,\tau} = \frac{PI_{i,t,\tau}}{\text{Tick Size}_{i,t,\tau}} \tag{4.5}$$

Where the subscript i,t denotes the stock i at trading day t. $P_{i,t,\tau}$ is the price of the trade at time τ, $EBO_{i,t,\tau}$ is the contemporaneous European Best Offer at time τ, and $EBB_{i,t,\tau}$ is the contemporaneous European Best Bid at time τ. Trade sign is determined using the Lee and Ready (1991) algorithm. We then calculate the daily measures of the magnitude of price improvement per trade $(RelPI_{i,t}$ and $TickPI_{i,t})$ by taking the volume-weighted average of the intraday magnitude of price improvement per trade $(RelPI_{i,t,\tau}$ and $TickPI_{i,t,\tau})$ over the trading day.

We estimate the regressions using the same specification of Eq. (4.2) with an extra control of the tick size because, without sub-tick trading, the minimum magnitude of price improvement per trade is one tick. Column (3) of Table 4.A13 reports the results of the magnitude of price improvement per trade. The post-ban indicator coefficient estimate is negative and significant, indicating a decrease in the magnitude of price improvement per trade for unconstrained stocks after the ban. The sum of coefficients of the interaction term and the post-ban indicator is positive and significant,[29] showing an increase in the magnitude of price improvement per trade for constrained stocks post-ban. Column (4) of Table 4.A13 reports consistent results when measuring the magnitude of price improvement per trade by the number of ticks. The magnitude of price improvement per trade increases from 0.79 ticks to 1 tick for constrained stocks and decreases from 2.43 ticks to 2.16 ticks for unconstrained stocks. These results

[29] In Columns (3) and (4) of Table 4.A4 in the Appendix, we show that the sum of the coefficients of the interaction term and the post-ban indicator is positive and significant.

suggest that SI simply round the sub-tick trading price to the closest on-tick price.

To understand whether traders benefit from the ban, we compute the price-improvement benefits by combining the frequency and the magnitude of price improvement as shown below:

$$\text{PI benefits}_{i,t} = \underbrace{f(\text{PI})_{i,t} \times \text{SI share volume}_{i,t}}_{\substack{\text{the number of trades that} \\ \text{receive price improvement}}} \times \underbrace{\text{PI}_{i,t}}_{\substack{\text{the magnitude of price} \\ \text{improvement per trade}}} \tag{4.6}$$

This "price-improvement benefits" variable captures the daily price-improvement benefits in euros traders receive when trading with SI. We hypothesize the ban to lower price-improvement benefits for unconstrained stocks due to the simultaneous decline in both frequency and magnitude. However, the implication of price-improvement benefits for constrained stocks is less clear because the magnitude increases while the frequency decreases after banning sub-tick trading. Therefore, it is an empirical question whether, for constrained stocks, the benefits from the increased magnitude of price improvement outweigh the costs from the decreased frequency of price improvement.

We estimate the regression using the specification of Eq. (4.2) with the replacement of the dependent variable with the above price-improvement benefit measure. We include an additional control of SI euro volume because the size of price-improvement benefits is expected to be larger when the consolidated SI volume is larger. Column (5) of Table 4.A13 reports the results. Both the interaction term and the post-ban indicator coefficient estimators are negative and significant, showing that banning SI sub-tick trading lowers the price-improvement benefits and this effect is even stronger in constrained stocks. The effect of banning sub-tick trading is economically large. After banning sub-tick trading, the price-improvement benefits for constrained stocks decrease by €161.04 per day, which accounts for around a quarter of the pre-ban average level. The decrease in the price-improvement benefits in constrained stocks indicates that the negative effect on the decreased frequency of price improvement dominates the positive effect on the increased magnitude of price improvement per trade.

4.4.3 Sub-tick trading and informational efficiency

The TSR mandates traders to execute their orders at specific and discrete prices. When informed traders' valuations lie within the prevailing best quotes and the next tick, sub-tick trading allows informed traders to bypass the tick-size constraint and incorporate their information at sub-tick prices, improving informational efficiency. Conversely, banning sub-tick trading impedes traders from incorporating their information into prices, undermining informational efficiency. Since banning sub-tick trading is equivalent to banning all trades within the spread when the spread is constrained by a tick size, we hypothesize this detrimental effect on informational efficiency to concentrate on constrained stocks. Following O'Hara and Ye (2011), we measure informational efficiency using (i) the absolute value of the 30-second EBBO midpoint return auto-correlation and (ii) the absolute difference between one and the variance ratio based on 2-minute and 30-second EBBO midpoint returns.[30] An efficient stock price (or return) should be unpredictable by its past-day return. Therefore, when the stock price is more efficient, the auto-correlation of daily returns is closer to zero, and a smaller absolute value of autocorrelation indicates a more efficient stock price. We compute the variance ratio:

$$\text{Variance Ratio} = |1 - \frac{\sigma_{2\text{min}}}{4\sigma_{30\text{sec}}}| \tag{4.7}$$

Where $\sigma_{2\text{min}}$ and $\sigma_{30\text{sec}}$ are the daily variances of 2-minute and 30-second EBBO midpoint returns, respectively. If prices are efficient, prices follow a random walk with stable variance, and the variance ratio should be closer to zero, and a smaller variance ratio indicates a more efficient price. We run the regression using the specification of Eq. (4.2) with the replacement of the dependent variable with two informational efficiency measures. We additionally control for the consolidated euro volume to account for the effect of liquidity on informational efficiency.

Columns (6) and (7) of Table 4.A13 present the coefficient estimates of OLS regressions of two informational efficiency measures. Both post-ban indicator coefficient estimates are statistically insignificant, indicating that banning sub-tick trading does not significantly alter informational efficiency for unconstrained stocks. Both sums of post-

[30] We do not use the common 60-second/15-second variance ratio measure because the median quote frequency of EBBO is around every 30 seconds.

ban indicator and interaction term coefficient estimates are positive and significant,[31] indicating that the SI sub-tick trading ban undermines informational efficiency for constrained stocks. The deterioration of informational efficiency in constrained stocks aligns with our hypothesis that banning sub-tick trading prohibits price discovery at sub-tick prices and harms informational efficiency primarily for constrained stocks. This result is also consistent with Comerton-Forde et al. (2019) that sub-tick trading helps improve informational efficiency for constrained stocks but not for unconstrained stocks. The negative effect on informational efficiency in constrained stocks implies that constrained stocks' tick size is excessively large, preventing traders from incorporating their information into the market and making SI sub-tick trading especially helpful in incorporating information at sub-tick prices.

Since the level of constraint and consolidated euro volume are positively associated, constrained stocks tend to be more liquid and have large ADNTE. To alleviate the negative impact on informational efficiency, regulators might consider reducing the tick size for stocks with a large ADNTE, most of which are constrained stocks, and make these stocks less constrained. Specifically, regulators can adjust the tick size schedule by reducing the tick size for ADNTE categories above 3,000 while maintaining the current tick size of other small ADNTE categories. In our sample stocks, the above adjustment leads to a reduction of tick size for 50% of constrained stocks but a tick size reduction for only 12% of unconstrained stocks. By making half of constrained stocks less constrained, this adjustment of the tick size schedule should help mitigate the negative effects on informational efficiency concentrated in constrained stocks without affecting the majority of unconstrained stocks.

4.4.4 Robustness checks on SI sub-tick trading

In this section, we complement our results with three robustness checks on identification strategy, covariate imbalance and parallel trend assumption. To be concise, we highlight the main results in this section and leave the detailed description of the empirical implementation in the Appendix.

[31] In Columns (5) and (6) of Table 4.A4 in the Appendix, we show that the sum of the coefficients of the interaction term and the post-ban indicator is positive and significant.

4.4.4.1 Identification strategy

We assess the sensitivity of our results to the choice of the threshold for identifying stocks as constrained or unconstrained. Previously in our identification strategy, we assign a stock to the constrained (unconstrained) group if the pre-ban average tick-constraint time is above (below) the 75th percentile (25th percentile). In this section, we explore two alternative thresholds: (i) median and (ii) tercile. Specifically, we assign a stock to the constrained (unconstrained) group if the pre-ban average tick-constraint time of the stock is above (below): (i) the median or (ii) the top tercile (the bottom tercile). Using these alternative thresholds leads to a larger alternative sample compared to our main sample (i.e., the previous sample in *diff-in-diff* analysis), with the main sample now being a subset of these two alternative samples.

One key difference between the two alternative samples and the main sample is the smaller variation in tick-constraint time in the alternative samples than in the main sample. In Table 4.A4 of the Appendix, we show that the pre-ban average tick-constraint time difference between constrained stocks and unconstrained stocks are 23.3% and 30.1% in the alternative samples using median and tercile as thresholds, respectively, whereas this number is 35.4% in the main sample. We use the alternative samples to re-estimate our results and report the re-estimated results from Tables 4.A6 to 4.A9 of the Appendix. We find consistent results with our earlier findings, albeit with smaller interaction term coefficient estimates. These smaller coefficient estimates are attributed to the smaller variation in tick-constraint time between constrained and unconstrained stocks in the alternative samples compared to the main sample. Overall, we find qualitatively consistent results using the alternative samples, showing that our results are not driven by the choice of threshold in identifying constrained and unconstrained stocks.

4.4.4.2 Covariate Imbalance

We are aware that constrained and unconstrained stocks differ in market capitalization, consolidated euro volume and price. To alleviate the concern that these covariate differences drive our results, we adopt a Propensity Score Matching (PSM) approach to control the effects of these covariates. First, we use the median pre-ban average tick-constraint time as the threshold to identify constrained and unconstrained stocks.[32] We estimate the logit regression, where the dependent variable is an indicator variable that equals one for a constrained stock and zero for an unconstrained stock, with independent variables comprising market capitalization, consolidated euro volume, and price. We subsequently estimate the propensity score for each stock and match each constrained stock with the nearest unconstrained stock based on propensity score.[33] To enhance matching quality, we exclude pairs with propensity score distances greater than the median distance.[34] We compare stock characteristics between constrained stocks and matched unconstrained stocks and do not find significant differences, except for the pre-ban average tick-constraint time. We find that the variation in the level of constraint is 18.3% in the PSM sample, which is smaller than the variation in the main sample. We use this PSM sample to re-estimate our results and report the results in Tables 4.A10 and 4.A11 of the Appendix, and all our *diff-in-diff* regression results are consistent with the main results. Similarly, we find that the interaction term coefficient estimates are smaller than the estimates in our main results due to the smaller variation in tick-constraint time between constrained and unconstrained stocks in the PSM sample. These consistent results using the PSM approach allay concerns that our results are driven by the covariate imbalance between constrained and unconstrained stocks.

4.4.4.3 Parallel Trend Assumption

We examine the "parallel trend" assumption in the diff-in-diff analysis, which posits that constrained stocks and unconstrained stocks should exhibit similar trends prior to the implementation of the SI sub-tick trading ban. To validate this assumption, we estimate the effect of the regulation on constrained stocks over different time periods surrounding the introduction of the SI sub-tick trading ban. The results are reported in Tables 4.A12 and 4.A13. The results indicate that, prior to the SI sub-tick trading ban, there are no significant differences in the dependent variables used in the main analysis between constrained and unconstrained stocks. This observation supports the

[32]We use the median as a threshold to identify constrained and unconstrained stocks so that the sample is large enough to ensure adequate variation in the above covariates to implement PSM.

[33]When matching stocks with the nearest propensity score, we allow matching one unconstrained stock to several constrained stocks. (i.e., PSM with replacement)

[34]This is because large constrained stocks generally do not match unconstrained stocks well. We calculate propensity score distance as the absolute difference in propensity score between the constrained stock and the matched unconstrained stock.

validity of the parallel trend assumption.

Specifically, we divided the sample period into four bi-weekly (i.e.,two-week) intervals: two prior to the SI sub-tick trading ban and two following the ban. For each bi-week period, we generated an indicator variable that equals one if a trading day falls within that period. We then re-estimated the regressions by replacing the original interaction terms with interaction terms of each bi-weekly indicator and constrained stock group indicator ($1[Constrained]i$).

The results show that the interaction terms corresponding to the first two bi-week periods are statistically insignificant,[35] suggesting no significant differences in the dependent variables between constrained and unconstrained stocks prior to the SI sub-tick trading ban. More critically, these differences become statistically significant only after the ban's implementation, showing that the effects observed in the diff-in-diff analyses are driven by the introduction of the SI sub-tick trading ban and not by underlying time trends.

4.5 SI Pre-trade transparency: the SI minimum quote size

This section studies another important feature of SI: pre-trade transparency. Our goal is to investigate the impact of the SI minimum quote size on market quality. However, despite our effort to control for observable market characteristics, some unobservable market characteristics that drive the cross-sectional results on market quality may exist. To address the potential omitted variable problem, we use the Sharpe Regression Discontinuity Design (RDD) to identify exogenous variation in the SI minimum quote size in liquid stocks.

Our identification strategy relies on the discontinuity of the SI minimum quote size as determined by the Average Value of Transactions (AVT). Under the SI pre-trade transparency regime, SI are required to provide public quotes for a minimum quote size of 10% SMS for liquid stocks. As depicted in Table 4.6, SMS is a discontinuous function of the AVT, with several cutoffs occurring when the AVT crosses specific thresholds.

[35]There are some coefficient estimates are marginally signficant, while the size of these coefficient estimates is relatively small compared to the coefficient estimates of interaction terms corresponding to the last two bi-week periods.

This leads to a discontinuity in the SI minimum quote size with respect to the AVT. Specifically, we focus on the cutoff at €20,000 AVT because the stock population around other cutoffs is sparse.[36] The SI minimum quote size increases from €1,000 to €3,000 whenever the AVT exceeds the €20,000 AVT threshold. Hence, we exploit this discontinuity by adopting the Sharp RDD with the AVT as the forcing variable to identify exogenous variation in the SI minimum quote size.

[Insert Table 4.6 near here]

To make sure that the Sharp RDD is valid, we select stocks that are close to the AVT threshold. Following Imbens and Kalyanaraman (2012), we pick the optimal bandwidth of €5,000 AVT around the €20,000 AVT threshold.[37] This bandwidth selection restricts our sample to stocks with an AVT within the range from €15,000 to €25,000. With this bandwidth, we can obtain stocks that are sufficiently similar in AVTs, and assume that these stocks are randomly assigned to either side of the AVT threshold. Under this assumption, the observed differences in market quality between stocks with AVTs above or below the threshold can be attributed to the causal association between the SI minimum quote size and market quality.

We collect the stock sample using the 2020 ATC.[38] Our sample consists of stocks whose AVT lies between €15,000 and €25,000. Our sample spans from April 1, 2020 to October 1, 2020. Our final sample comprises 23 liquid stocks with €3,000 SI minimum quote size and 23 liquid stocks with €1,000 SI minimum quote size, totaling 5,897 stock-day observations and 94,307 stock-day-SI observations.

[Insert Table 4.7 near here]

Table 4.7 reports the summary statistics for the study on SI pre-trade transparency. Panels A and B of Table 4.7 report stock characteristics for stocks with the SI minimum

[36]There are less than 15 liquid stocks around all other cutoffs (with the bandwidth of AVT of €5,000). See Table 4.6.

[37]The optimal bandwidth is estimated by minimizing the asymptotic mean squared error (MSE) following Imbens and Kalyanaraman (2012). Using the sum of the one-sided estimators MSE expansion, we obtain the optimal bandwidth of around €5,000.

[38]This is because (i) the 2019 ATC is the first time publishing the ATC and was amended multiple times during the year. and (ii) in 2021 ATC, most sample stocks falling within the range from €15,000 to €25,000 AVT are American Depositary Shares (ADS) or Global Depositary Shares (GDS).

quote size of €3,000 and €1,000, respectively. Stocks with €3,000 SI minimum quote size have larger market capitalization and higher consolidated euro volume (€62.2b and €371.0m, respectively) than stocks with €1,000 SI minimum quote size (€35.7b and €227.9m, respectively). The mean euro volume on SI is 17.1% and 15.6% of consolidated euro volume for stocks with the SI minimum quote size of €3,000 and €1,000, respectively. The difference in SI market share is consistent with our previous results that larger and more liquid stocks tend to have higher SI market share.

We compare the characteristics of SI public quotes between stocks with €3,000 SI minimum quote size and stocks with €1,000 SI minimum quote size. We find that the number of SI providing public quotes in these two groups of stocks is not significantly different. However, SI consolidated depth for stocks with €3,000 SI minimum quote size is close to three times SI consolidated depth for stocks with €1,000 SI minimum quote size. We show that the EBBO time-weighted spreads and SI consolidated spreads for the two groups of stocks are not significantly different. While stocks with €3,000 SI minimum quote size have larger EBBO time-weighted depth (€62.2k) than stocks with €1,000 SI minimum quote size (€44.5k), this difference in the EBBO time-weighted depth (€17.7k) is relatively small compared to the difference in SI consolidated depth (€81.0k).

4.5.1 The SI minimum quote size and the trade migration

SI public quote can be likened to a "free option" to traders because SI are obligated to execute incoming orders against their public quotes (Copeland and Galai (1983) and Easley and O'Hara (1987)). When the SI minimum quote size increases, SI are compelled to provide larger volume of SI public quotes. We hypothesize that larger volume on SI public quotes (or more "free options") attracts traders to SI, consequently boosting SI trading. To explore this, we examine how the SI minimum quote size affects the SI market share. We estimate the impact of the SI minimum quote size using the Sharpe RDD regression following Roberts and Whited (2013):

$$Y_{i,t} = \beta_0 + \beta_1 1_{[3,000]i} + \beta_2 (AVT_{i,t} - AVT_i) + \beta_3 (AVT_{i,t} - AVT_i) \times 1_{[3,000]i} + \lambda X_{i,t} + \sigma_t + \epsilon_{i,t} \tag{4.8}$$

Where the subscript i,t denotes the stock i at trading day t. $Y_{i,t}$ is the percentage of consolidated euro volume executed on SI. $1_{[3,000]i}$ is the treatment indicator variable that equals 1 when the stock has €3,000 SI minimum quote size (i.e., AVT above €20,000) and 0 otherwise. AVT_i is the €20,000 AVT threshold. $X_{i,t}$ are controls, including daily realized volatility level, the market capitalization and the inverse of the share price. σ_t are the date fixed effects. The standard errors are clustered by stock and date. We focus on the treatment indicator coefficient (β_1). We interpret this coefficient as the change in SI market share when the SI minimum quote size increases from €1,000 to €3,000. We also control for the lag SI market share as SI market share is likely to follow an autoregressive process.

[Insert Table 4.8 near here]

Column (1) of Table 4.8 reports the results of SI market share. The treatment indicator coefficient estimate is positive and significant, indicating a positive association between the SI minimum quote size and SI market share. When the SI minimum quote size increases from €1000 to €3000, SI market share grows by 1.1%. The growth corroborates the hypothesis that the increase in the SI minimum quote size attracts traders to SI.

Next, we investigate whether the incoming trades come primarily from informed or uninformed traders. If the increase in the SI minimum quote size mainly attracts informed traders, we would hypothesize that this migration makes SI trades more informed and contribute more to the price discovery process. To test this hypothesis, we study how the SI minimum quote size affects the SI's relative contribution to the price discovery process. We measure SI relative contribution to the price discovery processing using (i) Gonzalo and Granger (1995) CS and (ii) Hasbrouck (1995) IS.[39] We examine the impact of the SI minimum quote size on price discovery shares using the OLS regression specification of Eq. (4.8) with the relative contribution to the price discovery process as the dependent variable. We additionally control for the proportion of SI trades relative to the total SI and lit market euro volume because more trading

[39]Given that dark pools are hardly informed (Zhu (2014)), we focus on the price discovery process in SI and lit markets. Specifically, we estimate daily CS and IS by running the Vector Error Correction Model (VECM) regressions using intraday SI trade prices and lit trade prices.

naturally leads to larger contribution to price discovery.

Columns (2) and (3) of Table 4.8 report coefficient estimates on SI relative contribution to the price discovery process. The treatment indicator coefficient estimates are both positive and significant. Specifically, when the SI minimum quote size increases from €1000 to €3000, the CS of SI trades increases by 3.2%, and the IS of SI trades increases by 1.4%. The increase in SI price discovery share indicates that SI become more informed than relative to lit markets. Such a positive association between the SI minimum quote size and SI's relative contribution to price discovery suggests that the increase in the SI minimum quote size largely attracts informed traders to migrate to SI and makes SI more informed.

The migration of informed traders to SI increases SI's exposure to informed traders. Nonetheless, SI can widen their quoted spreads to compensate their losses to informed traders. We then examine how SI react to this increased exposure to informed traders associated with the increased SI minimum quote size. Specifically, we estimate the OLS regression specification of Eq. (4.8) with the replacement of the dependent variables to SI time-weighted spread.

Column (4) of Table 4.8 reports the results of SI spread. We find that the coefficient estimate for the treatment indicator is positive and significant, confirming that a larger SI minimum quote size indeed increases SI spread. SI spread widens by 0.5 basis points when the SI minimum quote size increases from €1000 to €3000, corresponding to 8.0% of the sample average of SI spread. These results suggest that SI protect themselves by widening SI spreads in response to the higher exposure to informed traders induced by a larger SI minimum quote size.

The wider SI spreads imply higher trading costs of SI trades. The elevated trading costs on SI are consistent with the notion that SI demand more compensation when encountering more informed traders. These lead us to investigate the impact of the SI minimum quote size on trading costs on SI. We measure trading costs on SI using two proxies: effective spread and price impact of SI trades, which are estimated as follows:

$$EffSpread_{i,t,\tau} = \frac{2 \times |P_{i,t,\tau} - m_{i,t,\tau}|}{m_{i,t,\tau}} \tag{4.9}$$

$$PrcImpact_{i,t,\tau} = \begin{cases} \frac{2*(m_{i,t,\tau+1min} - m_{i,t,\tau})}{m_{i,t,\tau}} & \text{for buyer-initiated trades} \\ \frac{2*(m_{i,t,\tau} - m_{i,t,\tau+1min})}{m_{i,t,\tau}} & \text{for seller-initiated trades} \end{cases} \quad (4.10)$$

Where the subscript i,t denotes the stock i at trading day t. $P_{i,t,\tau}$ is the SI trade price of the trade at time τ, $m_{i,t,\tau}$ is the contemporaneous EBBO midpoint at time τ. Trade sign is determined using the Lee and Ready (1991) algorithm. We calculate the daily effective spread and price impact ($EffSpread_{i,t}$ and $PrcImpact_{i,t}$) of SI trades by taking the volume-weighted average of intraday effective spread and price impact ($EffSpread_{i,t,\tau}$ and $PrcImpact_{i,t,\tau}$) of SI trades over the trading day. We then estimate OLS regressions of the effective spread and price impact of SI trades using the regression specification of Eq. (4.8).

Columns (5) and (6) of Table 4.8 report the results. The coefficient estimates for the treatment indicator are positive and significant, confirming that a larger SI minimum quote size increases trading costs on SI. On average, the effective spread and price impact of SI trades increase by 0.7 and 0.2 basis points, respectively, when the SI minimum quote size increases from €1000 to €3000. This increase corresponds to 7.9% and 8.6% of the sample average of each measure. Since the price impact captures the proportion of informed traders (Bessembinder and Venkataraman (2010)), the increase in price impact of SI trades also aligns with the notion that the increase in the SI minimum quote size attracts informed traders.

Zhu (2014) illustrate that informed traders are concentrated in lit markets. Consequently, their migration away from lit markets to SI should make lit markets less toxic as traders on lit markets are less likely to be adversely picked off by informed traders, lowering adverse selection risks for liquidity provision on lit markets. Hence, we hypothesize the trading costs on lit markets to go down when informed traders are attracted to SI due to the larger SI public quote volume and migrate away from lit markets. We measure the trading costs in lit markets in the effective spread and price impact of lit trades and the EBBO time-weighted spread. We estimate the effective spread and price impact of lit trades using Eq. (4.9) and (4.10) with the modification where $P_{i,t,\tau}$ refers to the lit trade price rather than the SI trade price. We estimate OLS regressions on lit market trading costs using the regression specification of Eq. (4.8) with the replacement of the dependent variable to the lit market trading cost measures.

[Insert Table 4.9 near here]

Columns (1) to (3) of Table 4.9 present the coefficient estimates of regressions on lit market trading costs. The treatment indicator coefficient estimates are all negative and significant, supporting a larger SI minimum quote size lowers lit market trading costs. The effective spread and price impact of lit trades decrease by 0.7 and 0.1 basis points, respectively, and the EBBO time-weighted spread decreases by 1.0 basis points when the SI minimum quote size increases from €1000 to €3000. These decreases translate to 11.4%, 3.2% and 11.8% reduction in the sample average of effective spread, price impact and the EBBO spread. The negative association between trading costs and the SI minimum quote size is consistent with our hypothesis that increased the SI pre-trade transparency attracts informed traders to migrate away from lit markets and make lit markets less toxic, lowering lit market trading costs and improving lit liquidity.

Finally, we examine the effect of the SI minimum quote size on informational efficiency. In lit markets, market makers do not know the trader's identity and adjust their quotes based on the ex-ante density of informed traders (Glosten and Milgrom (1985)). However, some SI can observe their clients' (i.e., brokers) identity and infer, albeit imperfectly, the underlying trader type based on the broker identity. Knowing their clients' identity allows these SI to better grasp, learn and extract the information from the informed trades, and update their quotes accordingly, rendering their quotes more informed. Therefore, we hypothesize that the increased informed trading on SI allows SI to learn more information and provide more informed quotes on SI and lit markets, improving informational efficiency. Following O'Hara and Ye (2011), we measure informational efficiency using (i) the absolute value of the 15-second SI Best Bid and Offer (BBO) midpoint return auto-correlation (|AC (SI)|), (ii) the absolute difference between one and the variance ratio based on 1-minute and 15-second SI BBO midpoint returns (|1-VR (SI)|), (iii) the absolute value of the 15-second EBBO midpoint return auto-correlation (|AC|) and (iv) the absolute difference between one and the variance ratio based on 1-minute and 15-second EBBO midpoint returns (|1-VR|). We estimate OLS regressions of informational efficiency using the regression specification of Eq. (4.8) with the replacement of the dependent variable to the informational efficiency measures.

Columns (4) to (7) of Table 4.9 present the results of informational efficiency. The treatment indicator coefficient estimates are all negative and significant, affirming that a larger SI minimum quote size enhances informational efficiency. When the SI minimum quote size increases from €1000 to €3000, the |AC (SI)| decreases by 0.008, |1-VR (SI)| decreases by 0.01, |AC| decreases by 0.005 and |1-VR| decreases by 0.01, indicating an improvement in informational efficiency. This improvement in informational efficiency is economically large and corresponds to 16.4%, 11.0%, 12.7% and 12.2% of the sample average of each informational efficiency measure. The positive effect of SI minimum quote size on informational efficiency aligns with our hypothesis and suggests that, from the perspective of improving informational efficiency, the current level of the SI minimum quote size is relatively small. These results also lend support to the ESMA's proposal to increase the SI minimum quote size.

4.5.2 Robustness Check: SI pre-trade transparency for illiquid stocks

Our previous results are based on the liquid stock sample and might not necessarily apply to illiquid stocks. This section performs a robustness check using an alternative identification strategy to extend our examination of the SI pre-trade transparency to illiquid stocks. We report the detailed description of the empirical implementation and regression results in the Appendix.

The alternative identification strategy exploits the variation in the SI minimum quote size that occurs when a stock's liquidity status changes.[40] According to MiFID II, the SI minimum quote size requirement only applies to liquid stocks but not to illiquid stocks. This means that, when an illiquid stock is re-identified as a liquid stock, SI are obligated to provide public quotes for this stock, leading to an increase in the SI minimum quote size. We subsequently identify the exogenous variation in the SI minimum quote size by exploiting the time-series variation in the liquidity status.

Stock liquidity fluctuates periodically, resulting in switches in liquidity status and

[40] A liquidity status is a flag indicating whether the stock is liquid or illiquid.

the SI minimum quote size. We assign stocks that experience a liquidity status switch from illiquid to liquid into the treatment group. We use PSM to match each treatment stock with a control stock that remains an illiquid stock. This match is conducted based on market capitalization and consolidated euro volume. Subsequently, we compare the market quality changes in treatment stocks with the market quality changes in control stocks using the *diff-in-diff* analysis to recover the causal impact in the SI minimum quote size.

We are aware that the re-classification of the liquidity status is not random and is driven by the annual changes in three liquidity measures that determine the liquidity status, leading to a reverse causality problem. We avoid this potential endogeneity issue by segregating our sample period from the ATC calculation period. The liquidity status is set annually based on the data from the previous calendar year. Every March, ESMA announces the liquidity status for application from April of the current year to March of the subsequent year.[41] For illustration, the 2021 ATC was revealed on March 1, 2021, calculated using data from January 1, 2020 to December 31, 2020, and is applicable from April 1, 2021, to March 31, 2022.[42]

We restrict our sample period within a two-month time window surrounding the ATC application date such that our sample period does not overlap with the ATC calculation period. We demonstrate the separation of our sample period and the ATC calculation period using the following example and Figure 4.1. Suppose that stock i was identified as an illiquid stock in the 2019 ATC and re-identified as a liquid stock in the 2020 ATC. This means that, based on data from 1 January, 2018 to 31 December, 2018, stock i is classified as an illiquid stock from 1 April, 2019 to 31 March, 2020; and, based on data from 1 January, 2019 to 31 December, 2019, stock i is classified as a liquid stock from 1 April, 2020 to 31 March, 2021. However, our sample period for stock i focuses on the two-month window surrounding the ATC effective date (1 February, 2020 to 1 June, 2020), not coinciding with either the year 2018 or the year

[41] ESMA's annual transparency calculation is available on ESMA Financial Instruments Transparency System.

[42] Although the transparency calculation is published on an annual basis, ESMA might update the transparency calculation before the next annual publication date. Nevertheless, those updates are not applicable unless ESMA further notifies the application of the updated version of transparency calculation.

2019. Therefore, there is no clear channel suggesting that stock i has experienced a liquidity improvement during our sample period.

[Insert Figure 4.1 near here]

In the Appendix, we verify that SI generally do not voluntarily provide public quotes for illiquid stocks. As a result, we cannot compute SI spread, and we cannot estimate the effect of the SI minimum quote size on SI spread for illiquid stocks. Table A10 reports the results of market quality on the SI minimum quote size using the switch of liquidity status as an identification strategy. All our previous results remain unchanged except for SI price discovery share and informational efficiency, where we do not find any significant impacts on these measures. We provide three potential justifications for these insignificant results. First, sample stocks in this subsection are smaller and illiquid and have fewer trades and quotes on each date. This means that, when estimating price discovery share and informational efficiency metrics, we might not have adequate variation in daily VECM regression to generate meaningful measures. Second, the variation in the SI minimum quote size in the liquidity status switch is relatively smaller than the variation previously identified in the Sharpe RDD strategy. The switch of liquidity status from illiquid to liquid leads to an increase in the SI minimum quote size of €1000, smaller than the change in the SI minimum quote size of €2000 in the Sharpe RDD strategy. This leads to a less significant effect. Last, SI is more informed in less liquid stocks. We show that, in this less liquid stock sample, the average SI price discovery share measures are 37.6% for the CS of SI trades and 26.6% for the IS of SI trades, where SI account for only 8% of the total consolidated euro volume on average. Given SI's prominent role in the price discovery process, the additional improvement in informativeness associated with the increased SI minimum quote size is marginal. Nevertheless, the insignificant results of informational efficiency indicate that forcing SI to provide more public quotes on illiquid stocks does not necessarily improve informational efficiency. This finding aligns with ESMA's view and industry response of not requiring SI to provide public quotes for illiquid stocks. In summation, we find that the increase in the SI minimum quote size improves lit liquidity but lowers the number of SI price-improvement trades. This indicates that our results are widely applicable in both liquid and illiquid stocks.

4.6 Conclusion

The implementation of MiFID II increased the number of SI and their market share. Since then, ESMA and the European Commission have initiated several consultations and proposals to modify SI regulations. From June 2020, the implementation of IFR banned SI sub-tick trading. The following month, ESMA proposed a dramatic increase in SI pre-trade transparency requirements from 10% of SMS to 100% of SMS and a revision in the method of calculating SMS. In late 2021, the European Commission proposed a change to SI that would further limit SI's ability to trade at the midpoint.

This chapter examines the two characteristic aspects of SI: (i) sub-tick trading and (ii) pre-trade transparency. Using a regulatory change as a natural experiment, we study the effect of banning SI sub-tick trading on inter-market competition and market quality. We show the pros and cons of sub-tick trading that banning SI sub-tick trading curtails price-improvement benefits received by market participants but enhances lit liquidity and SI liquidity. Our results provide insights into the benefits and costs of banning SI sub-tick trading on different types of stocks, which is relevant to the recent proposal made by the European Commission to further restrict SI's ability to trade at the midpoint. We also find that banning SI sub-tick trading is detrimental to informational efficiency for constrained stocks but not for unconstrained stocks, consistent with Comerton-Forde et al. (2019). This negative effect can be alleviated by lowering the tick size in large ADNTE categories (i.e., the ADNTE of €3,000 and above), as outlined in the MiFID II TSR.

Last, we study the effect of the SI pre-trade transparency on market quality using the Sharpe RDD as an identification strategy. We show that an increase in SI displayed liquidity unfavorably positions SI in comparison to trading venues. Considering the recent proposal to increase the SI minimum quote size on SI from both ESMA and Financial Conduct Authority (FCA). Our results hold significant relevance for policymakers to comprehend the consequences and implications of raising the SI minimum quote size.

Figure 4.1: The timeline of ATC calculation and application period This figure demonstrates the timeline of ATC calculation period, ATC application period, and our sample period, which is separated from the ATC calculation period.

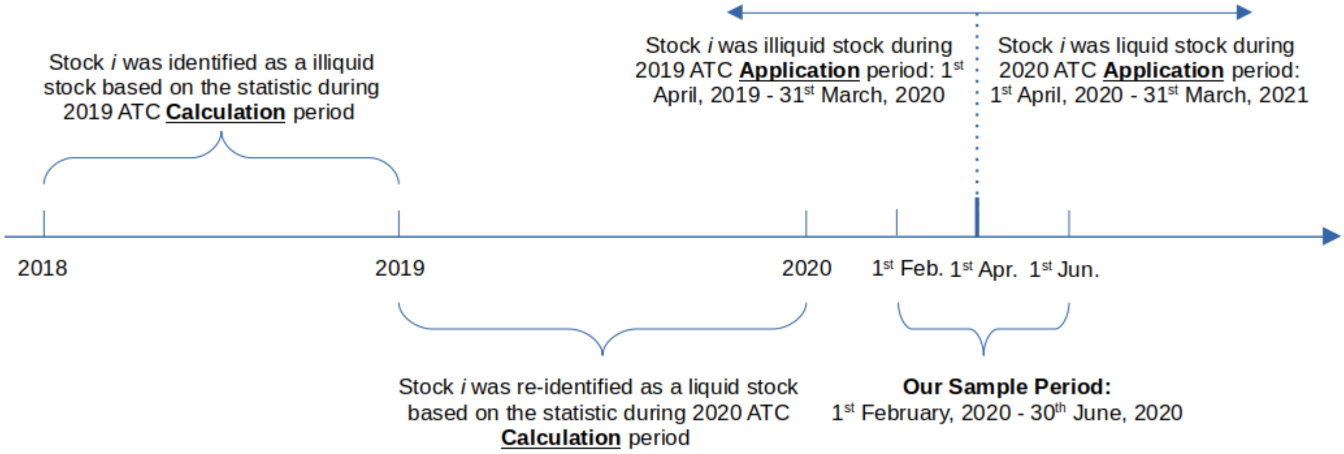

Table 4.1: Summary Statistics

This table summarizes the number of observations, mean, median, standard deviation, minimum and maximum values and the 25th and 75th percentiles of variables used in our regression analysis. The sample covers 747 liquid stocks in Euronext Paris, Xetra, Euronext Amsterdam and the LSE from 1 April, 2020 to 1 October, 2020. Panel A and Panel B report stock characteristics at the stock-day level and stock-day-SI level, respectively. MktCap is the market capitalization in €1,000,000,000. SI market share is the percentage of consolidated euro volume executed on SI. Euro Volume is the consolidated euro volume in €1,000,000 in all trading venues, SI and OTC (including trades outside continuous trading hours). Price is the volume-weighted average price in euros. Pre-ban Subtick trading is the proportion of SI sub-tick trading in the pre-ban perod. EBBO Spread is EBBO time-weighted spread in basis points. EBBO Depth is EBBO time-weighted depth measured in €1,000. TrdSize is the average trading size on lit markets in €1,000. SI AvgSpread is the average of time-weighted spread of SI public quotes across eighteen competitive SI in basis points. SI AggSpread is the time-weighted consolidated spread of SI public quotes for eighteen competitive SI in basis points. SI AvgDepth is the average of time-weighted depth of SI public quotes across eighteen competitive SI measured in €1,000. SI AggDepth is the time-weighted consolidated depth of SI public quotes for eighteen competitive SI measured in €1,000. SI TrdSize is the average SI euro trading size in €1,000. CS is the component share of SI trades in percentage. IS is the information share of SI trades in percentage. SI/(SI+Lit) is the percentage of SI euro volume relative to the sum of SI and lit market euro volume. SI Spread is SI time-weighted spread of SI public quotes in basis points. SI Depth is SI time-weighted depth of SI public quotes measured in €1,000. AtEBBO is the average proportion of time that SI provide quotes at EBB and EBO prices. BetterEBBO is the average proportion of time that SI provide quotes better than EBB and EBO prices. SI EuroDepth/SMS is the percentage of SI time-weighted euro depth to the Standard Market Size.

Panel A: Summary Statistics for stock characteristics (Stock-Day Observations)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	91,455	9.33	20.06	0.01	0.93	2.40	7.66	215.10
SI market share (%)	94,994	12.82	10.87	0.0	6.33	10.34	15.9	100.0
Euro Volume (€millions)	94,994	53.11	127.41	0.0	2.71	10.72	45.79	7,235.01
Price (€)	94,988	35.6	82.15	0.01	3.32	12.46	35.95	1,605.0
Pre-ban Sub-tick trading (%)	43,205	9.64	14.11	0.0	0.0	3.79	14.46	100.0
EBBO Spread (bps)	94,678	30.33	41.33	1.67	8.36	18.42	38.12	910.65
EBBO Depth (€thousands)	94,678	20.27	59.82	0.49	7.52	12.77	21.59	3,621.81
TrdSize (€thousands)	94,637	2.51	1.58	0.0	1.45	2.16	3.14	49.35
SI AvgSpread (bps)	93,140	34.79	51.74	2.05	9.19	19.92	41.0	3,152.24
SI AggSpread (bps)	93,140	24.13	29.75	0.03	5.55	13.5	30.63	175.65
SI AvgDepth (€thousands)	93,140	4.3	1.67	1.12	3.93	4.15	4.38	19.92
SI AggDepth (€thousands)	93,140	44.93	29.52	1.1	32.9	42.41	52.04	4,099.77
SI TrdSize (€thousands)	92,710	10.75	55.44	0.0	2.18	4.02	8.21	7,080.08
CS (%)	87,547	24.61	18.71	0.0	10.4	20.55	34.21	100.0
IS (%)	87,544	14.21	17.34	0.0	2.66	7.77	18.58	100.0
SI/(Lit + SI) (%)	92,674	28.4	20.34	0.0	13.95	23.65	37.12	100.0

Panel B: Summary Statistics for SI quote (Stock-Day-SI Observations)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
SI Spread (bps)	1,519,961	32.85	62.67	1.29	8.44	17.59	37.36	8,018.17
SI Depth (€thousands)	1,522,887	4.36	6.58	1.00	1.96	2.03	2.83	81.59
AtEBBO (%)	1,522,887	34.89	36.72	0.0	0.0	16.4	71.23	100.0
BetterEBBO (%)	1,522,887	27.35	20.38	0.0	7.3	22.72	49.97	99.99
SI Depth/SMS (%)	1,522,887	40.39	54.5	10.01	19.66	20.34	24.78	507.66

Table 4.2: SI market share

This table presents coefficient estimates from OLS regressions of SI market share. For dependent variable(s), SI market share is the percentage of consolidated euro volume executed in SI. Independent variables include realized volatility of 15-second EBBO midpoint returns, the lag SI market share, the inverse of the share price (1/Price), the logarithm of market capitalization (Log(MktCap)), the logarithm of consolidated euro volume (Log(Euro Volume)), the time-weighted average EBBO spread (Spread), the proportion of time when the quoted spread is at one-tick spread on lit markets (Tick-Constraint), date fixed effects, industry fixed effects and country fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) SI Shr	(2) SI Shr	(3) SI Shr
1/Price	-3.436***	-0.424***	-3.371***
	(-5.62)	(-3.34)	(-5.53)
RealVol	-0.366***	-0.737***	-0.339***
	(-3.08)	(-5.57)	(-2.88)
Lag SI Shr	0.252***	0.229***	0.248***
	(27.72)	(25.74)	(27.11)
Spread	0.00282	0.0242***	0.00491
	(0.94)	(7.62)	(1.62)
Log(MktCap)	0.688***		0.517***
	(10.55)		(7.68)
Log(Euro Volume)		1.410***	
		(30.42)	
Tick-Constraint			0.0405***
			(8.98)
Date Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes
Observations	31,607	32,631	31,607
Adjusted R2	0.087	0.119	0.089

Table 4.3: Summary Statistics: SI sub-tick trading study

This table summarizes the number of observations, mean, median, standard deviation, minimum and maximum values and the 25th and 75th percentiles for the daily stock-level observations of variables used in our sub-tick trading study. The sample covers 195 constrained stocks and 195 unconstrained stocks in Euronext Paris, Xetra, Euronext Amsterdam and the LSE during the pre-ban period (from 26 May, 2020 to 26 June, 2020). Panel A and Panel B report stock characteristics at the stock-day level for constrained and unconstrained stocks, respectively. MktCap is the market capitalization in €1,000,000,000. SI market share is the percentage of consolidated euro volume executed on SI. Euro Volume is the consolidated euro volume in €1,000,000 in all trading venues, SI and OTC (including trades outside continuous trading hours). Tick-Constraint is the proportion of time when the quoted spread is at one-tick spread on lit markets. Sub-tick PI is the percentage of SI euro volume that receives sub-tick price improvement relative to SI euro volume that receives price improvement.

Panel A: Summary Statistics for Constrained Stocks (Pre-ban period)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	4,602	21.68	30.68	0.36	3.62	9.88	25.61	204.16
SI market share (%)	4,674	15.0	12.18	0.0	7.94	12.2	18.36	100.0
Euro Volume (€millions)	4,674	152.22	224.36	0.0	18.62	71.24	189.82	2,651.75
Tick-Constraint (%)	4,633	36.98	17.92	0.0	23.52	34.02	48.54	97.8
Sub-tick PI (%)	4,408	47.34	27.13	0.0	28.58	51.22	67.11	100.0

Panel B: Summary Statistics for Unconstrained Stocks (Pre-ban period)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	4,489	10.27	25.92	0.02	0.55	1.16	2.9	204.16
SI market share (%)	4,600	10.58	12.05	0.0	3.8	7.79	13.09	100.0
Euro Volume (€millions)	4,600	14.31	39.84	0.0	2.04	4.7	11.53	1,071.72
Tick-Constraint (%)	4,588	1.63	2.01	0.0	0.34	1.0	2.15	22.67
Sub-tick PI (%)	3,825	24.63	25.91	0.0	0.0	17.17	41.33	100.0

Table 4.4: Inter-market competition and the SI sub-tick trading ban

This table presents coefficient estimates from OLS regressions of liquidity on lit markets and SI. For dependent variable(s), EBBO Spread is time-weighted EBBO spread in basis points. EBBO depth is the time-weighted EBBO depth measured in euros. SI Spread is the time-weighted spread of SI public quotes in basis points. SI Depth is the time-weighted depth of SI public quotes measured in euros. For independent variable(s), the variable of interest is the interaction between $1_{[Constrained]}$ and $1_{[post]}$, where $1_{[Constrained]}$ is an indicator that equals 1 if the stock is in the constrained group and 0 otherwise. $1_{[post]}$ is a post-ban indicator that equals 1 if the date is in post-ban period and 0 otherwise. Control variables include realized volatility of 15-second EBBO midpoint returns (RealVol), the inverse of the share price (1/Price), the logarithm of consolidated euro volume (Log(Euro Volume)), stock fixed effects and SI fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	SI sub-tick trading study			
	(1)	(2)	(3)	(4)
	EBBO Spread	EBBO Depth	SI Spread	SI Depth
$1_{[Constrained]} \times 1_{[post]}$	2.795***	2017.1***	2.309***	-6.003
	(8.42)	(9.15)	(7.79)	(-1.36)
$1_{[post]}$	-3.487***	862.3***	-2.984***	9.196***
	(-10.69)	(9.59)	(-10.31)	(2.65)
1/Price	3.965***	-102.1	1.162	-79.18***
	(3.34)	(-0.56)	(0.89)	(-4.75)
RealVol	1.516***	-444.4***	1.398***	-1.154
	(5.84)	(-4.94)	(5.76)	(-0.25)
Log(Euro Volume)	-1.641***	2695.9***	-1.351***	21.00***
	(-7.23)	(17.52)	(-6.84)	(7.76)
Stock Fixed Effects	Yes	Yes	Yes	Yes
SI Fixed Effects	No	No	Yes	Yes
Observations	15,798	15,798	197,924	198,235
Adjusted R2	0.023	0.045	0.012	0.000

Table 4.5: SI price improvement, informational efficiency and the SI sub-tick trading ban

This table presents coefficient estimates from OLS regressions of SI price improvement. For dependent variable(s), f(PI) is the frequency in percentage of SI trades whose price lies within the prevailing EBBO. f(EBBO) is the frequency in percentage of SI trades at EBBO. RelPI is the magnitude of price improvement per trade in basis points. TickPI is the magnitude of price improvement per trade measured by the number of ticks. PI Benefit is SI price-improvement benefits in euros. |AC| is the absolute value of the 30-second EBBO midpoint return auto-correlation. |1-VR| is the absolute difference between one and the variance ratio based on 2-minute and 30-second EBBO midpoint returns. For independent variable(s), the variable of interest is the interaction between $1_{[Constrained]}$ and $1_{[post]}$, where $1_{[Constrained]}$ is an indicator that equals 1 if the stock is in the constrained group and 0 otherwise. $1_{[post]}$ is a post-ban indicator that equals 1 if the date is in post-ban period and 0 otherwise. Control variables include realized volatility of 15-second EBBO midpoint returns (RealVol), the inverse of the share price (1/Price), the tick size in basis points of price (TickSize), SI euro volume (SI EuroVolume), the logarithm of consolidated euro volume (Log(Euro Volume)) and stock fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	SI sub-tick trading study						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	f(PI)	f(EBBO)	RelPI	TickPI	PI Benefit	|AC|	|1-VR|
$1_{[Constrained]} \times 1_{[post]}$	-5.502***	6.072***	1.772***	0.380***	-101.5***	0.00327**	0.00652**
	(-10.04)	(8.11)	(12.73)	(18.59)	(-8.04)	(2.33)	(2.30)
$1_{[post]}$	-2.834***	2.896***	-0.918***	-0.169***	-59.54***	-0.00112	-0.000950
	(-6.16)	(5.18)	(-6.89)	(-8.43)	(-7.04)	(-0.96)	(-0.41)
RealVol	0.598**	-1.326***	0.194***	0.0361***	37.81***	0.00170*	-0.00842***
	(2.40)	(-2.96)	(2.80)	(3.11)	(5.88)	(1.78)	(-8.98)
1/Price	-0.634	0.485	0.497	0.175	-25.33	0.000829	-0.00365
	(-0.47)	(0.28)	(0.44)	(0.96)	(-1.22)	(0.22)	(-0.46)
TickSize			19.62***				
			(9.71)				
SI EuroVolume					1.339***		
					(4.99)		
Log(Euro Volume)						0.00229***	0.00217
						(2.92)	(1.40)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,539	15,539	14,781	14,781	14,781	15,796	15,796
Adjusted R2	0.038	0.024	0.015	0.028	0.045	0.002	0.003

Table 4.6: Standard Market Size thresholds

Average value of transactions (AVT) in EUR	Standard Market Size	SI minimum quote size	No. of Stocks
AVT <20,000	10,000	1,000	2914
20,000 $\leq$ AVT <40,000	30,000	3,000	112
40,000 $\leq$ AVT <60,000	50,000	5,000	19
60,000 $\leq$ AVT <80,000	70,000	7,000	17
80,000 $\leq$ AVT <100,000	90,000	9,000	6
100,000 $\leq$ AVT <120,000	110,000	11,000	3
120,000 $\leq$ AVT <140,000	130,000	13,000	3

Source: table 3 of ANNEX II in COMMISSION DELEGATED REGULATION (EU) 2017/587

Table 4.7: Summary Statistics: SI pre-trade transparency study

This table summarizes the number of observations, mean, median, standard deviation, minimum and maximum values and the 25th and 75th percentiles for the daily stock-level level observations of variables used in our SI pre-trade transparency study. The sample covers 46 liquid stocks in Euronext Paris, Xetra, Euronext Amsterdam and the LSE from 1 April, 2020 to 31 October, 2020. Panel A and Panel B report stock characteristics at the stock-day level for stocks with €3,000 and €1,000 SI minimum quote size, respectively. MktCap is the market capitalization in €1,000,000,000. SI market share is the percentage of consolidated euro volume executed on SI. Euro Volume is the consolidated euro volume in €1,000,000 in all trading venues. Price is the volume-weighted average price in euros. EBBO Spread is EBBO time-weighted spread in basis points. EBBO Depth is EBBO time-weighted depth measured in €1,000. SI EffSpread is the volume-weighted average effective spread of SI trades in basis points. SI PrcImpact is the volume-weighted average price impact of SI trades in basis points. Lit EffSpread is the volume-weighted average effective spread of lit trades in basis points. Lit PrcImpact is the volume-weighted average price impact of lit trades in basis points. CS is the component share of SI trades in percentage. IS is the information share of SI trades in percentage. No. SI is the number of competitive SI. SI AggSpread is the time-weighted consolidated spread of SI public quotes for eighteen competitive SI in basis points. SI AggDepth is the time-weighted consolidated depth of SI public quotes for all competitive SI measured in €1,000.

Panel A: Summary Statistics for stocks with €3,000 SI minimum quote size

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	2,938	62.24	28.66	22.49	42.03	53.56	72.63	150.47
SI market share (%)	2,938	17.05	12.55	0.0	9.37	13.97	21.44	100.0
Euro Volume (€millions)	2,938	371.04	340.75	0.15	187.63	289.94	445.78	7,252.56
Price (€)	2,930	82.98	71.5	10.42	32.84	63.25	110.09	351.21
EBBO Spread (bps)	2,930	8.92	23.47	1.67	2.96	4.14	5.31	303.92
EBBO Depth (€thousands)	2,930	62.21	37.49	11.26	32.47	52.96	82.11	208.74
SI EffSpread (bps)	2,802	9.89	15.62	2.09	4.1	5.37	7.79	97.27
SI PrcImpact (bps)	2,802	2.0	1.34	-1.18	1.16	1.88	2.63	7.03
Lit EffSpread (bps) (bps)	2,827	3.99	1.94	1.79	2.83	3.73	4.59	16.31
Lit PrcImpact (bps)	2,827	3.31	1.39	1.19	2.34	3.11	3.92	9.44
CS (%)	2,686	12.42	10.31	0.01	4.28	10.01	18.53	86.94
IS (%)	2,686	5.68	7.36	0.0	1.14	3.25	7.32	95.12
SI AggSpread (bps)	2,881	6.43	20.07	-8.57	1.81	2.71	3.67	126.49
SI AggDepth (€thousands)	2,881	130.68	40.72	15.17	106.41	141.89	162.22	201.30

Panel B: Summary Statistics for stocks with €1,000 SI minimum quote size

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	2,831	35.73	22.69	1.52	19.29	29.06	46.27	100.08
SI market share (%)	2,959	15.59	10.87	0.0	9.28	13.37	19.21	100.0
Euro Volume (€millions)	2,959	227.86	166.6	0.0	119.53	206.33	303.78	2,267.72
Price (€)	2,931	69.82	102.43	3.1	11.64	35.63	92.73	597.0
EBBO Spread (bps)	2,931	8.73	15.7	1.76	3.08	4.29	5.99	117.57
EBBO Depth (€thousands)	2,931	44.5	29.56	5.2	24.36	32.49	60.69	389.89
SI EffSpread (bps)	2,800	7.59	7.01	2.02	3.67	5.35	8.08	43.69
SI PrcImpact (bps)	2,800	2.57	2.25	-1.03	1.36	2.12	3.1	14.69
Lit EffSpread (bps)	2,931	7.63	13.74	1.81	2.8	3.97	5.33	85.13
Lit PrcImpact (bps)	2,931	5.24	5.86	1.18	2.45	3.54	4.83	34.63
CS (%)	2,798	12.29	10.01	0.0	4.61	10.03	17.52	78.55
IS (%)	2,798	5.17	6.39	0.0	1.1	3.08	7.11	73.07
SI AggSpread (bps)	2,886	6.12	12.80	-2.70	1.71	2.56	4.05	69.57
SI AggDepth (€thousands)	2,886	49.72	13.93	10.07	43.10	52.11	59.09	87.63

Table 4.8: SI minimum quote size, SI market share and SI adverse selection costs

This table presents coefficient estimates from OLS regressions of SI transparency on SI market share and SI trading costs. For dependent variable(s), SI shr is the percentage of consolidated euro volume executed in SI. CS is the component share of SI trades in percentage. IS is the information share of SI trades in percentage. SI RelSpread is SI public quoted spread in basis points. EffSpread is the effective spread of SI trades in the basis points. PrcImpact is the price impact of SI trades in the basis points. For independent variable(s), the variable of interest is $1_{[3,000]}$: a treatment indicator that equals 1 if the stock has SI minimum quote size of €3,000 and 0 otherwise. Control variables include the difference between stock's AVT and AVT threshold $(A\bar{V}T - AVT)$, the interaction term between the difference between stock's AVT and AVT threshold and the treatment indicator, realized volatility of 15-second EBBO midpoint returns (RealVol), the inverse of the share price (1/Price), the logarithm of market capitalization (Log(MktCap)), the lag SI market share (Lag SI Shr), the percentage of SI euro volume relative to the total SI and lit market euro volume (SI/(SI+Lit)), date fixed effects, country fixed effects and SI fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	SI pre-trade transparency					
	(1)	(2)	(3)	(4)	(5)	(6)
	SI shr	CS	IS	SI Spread	EffSpread	PrcImpact
$1_{[3,000]\,i}$	1.175**	2.815***	1.300***	0.409***	1.096***	0.229**
	(2.29)	(5.21)	(4.04)	(3.48)	(2.93)	(2.52)
1/Price	3.530	22.83***	13.47***	1.465**	-40.71***	1.367***
	(1.32)	(7.30)	(7.37)	(2.08)	(-9.30)	(2.97)
RealVol	-1.031**	-1.644***	-0.837***	0.00543	1.459***	0.618***
	(-2.32)	(-3.04)	(-3.01)	(0.06)	(2.60)	(3.25)
$A\bar{V}T - AVT$	-10.37***	-3.499**	-1.485	5.780***	21.54***	-0.183
	(-6.32)	(-1.99)	(-1.47)	(11.55)	(10.41)	(-0.52)
$(A\bar{V}T - AVT) \times 1_{[3,000]\,i}$	11.05***	-7.234**	-2.831	-0.882	-12.28***	0.907*
	(3.50)	(-2.12)	(-1.39)	(-1.28)	(-3.62)	(1.74)
Log(MktCap)	1.117***	-1.560***	-0.766***	-4.377***	-5.350***	-1.052***
	(4.64)	(-5.79)	(-4.61)	(-41.87)	(-18.42)	(-13.96)
Lag SI Shr	0.355***					
	(18.13)					
SI/(SI+lit)		-0.0264**	0.0304***			
		(-2.52)	(4.80)			
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
SI Fixed Effects	No	No	No	Yes	No	No
Observations	5,562	5,357	5,357	94,307	5,475	5,475
Adjusted R2	0.170	0.038	0.035	0.462	0.124	0.195

Table 4.9: SI minimum quote size, lit market trading costs and informational efficiency

This table presents coefficient estimates from OLS regressions of SI minimum quote size on lit market trading costs and informational efficiency. For dependent variable(s), EffSpread (lit) is the effective spread of lit trades in the basis points. PrcImpact (lit) is the price impact of lit trades in the basis points. EBBO Spread is the EBBO time-weighted spread in basis points. |AC (SI)| is the absolute value of the 15-second SI BBO midpoint return auto-correlation. |1-VR (SI)| is the absolute difference between one and the variance ratio based on 1-minute and 15-second SI BBO midpoint returns. |AC| is the absolute value of the 15-second EBBO midpoint return auto-correlation. |1-VR| is the absolute difference between one and the variance ratio based on 1-minute and 15-second EBBO midpoint returns. For independent variable(s), the variable of interest is $1_{[3,000]}$: a treatment indicator that equals to 1 if the stock has SI minimum quote size of €3,000 and 0 otherwise. Control variables include the difference between stock's AVT and AVT threshold ($A\bar{V}T - AVT$), the interaction term between the difference between stock's AVT and AVT threshold and the treatment indicator, the inverse of the share price (1/Price), realized volatility of 15-second EBBO midpoint returns (RealVol), log market capitalization (Log(MktCap)), the logarithm of lit market Euro volume (Log(Lit Euro Volume)), the logarithm of consolidated euro volume (Log(Euro Volume)), and date fixed effects and country fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	SI pre-trade transparency study						
	(1) EffSpread (Lit)	(2) PrcImpact (Lit)	(3) EBBO Spread	(4) |AC (SI)|	(5) |1-VR (SI)|	(6) |AC|	(7) |1-VR|
$1_{[3,000]}i$	-0.993***	-0.214**	-1.668***	-0.00519***	-0.0105***	-0.00526***	-0.0113***
	(-5.91)	(-2.25)	(-7.43)	(-3.02)	(-2.58)	(-3.05)	(-2.79)
RealVol	1.278***	0.731***	2.925***	0.00258*	0.00443	0.00119	-0.00391
	(3.29)	(2.88)	(3.22)	(1.84)	(1.23)	(0.81)	(-1.15)
$A\bar{V}T - AVT$	8.137***	5.498***	10.17***	0.00687	0.0323**	0.00660	0.0326**
	(14.61)	(14.20)	(11.72)	(1.22)	(2.49)	(1.18)	(2.53)
$(A\bar{V}T - AVT) \times 1_{[3,000]}i$	0.166	-1.072*	3.701***	0.0316***	0.0341	0.0336***	0.0403*
	(0.20)	(-1.91)	(2.74)	(3.08)	(1.41)	(3.27)	(1.66)
1/Price	6.601***	2.890***	11.56***	0.00194	-0.00608	0.00237	-0.00679
	(6.71)	(5.28)	(8.02)	(0.20)	(-0.28)	(0.24)	(-0.31)
Log(MktCap)	-0.111	-2.206***	1.668***				
	(-0.35)	(-15.15)	(6.62)				
Log(Lit Euro Volume)	-4.320***	-1.633***	-7.240***				
	(-12.19)	(-12.08)	(-33.50)				
Log(Euro Volume)				-0.00704***	-0.0123***	-0.00714***	-0.0125***
				(-9.58)	(-7.27)	(-9.70)	(-7.41)
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,631	5,631	5,734	5,861	5,861	5,861	5,861
Adjusted R2	0.611	0.652	0.833	0.038	0.020	0.037	0.020

4.7 Appendix

4.7.1 SI sub-tick trading study

ESMA consultation: Industry response to banning SI sub-tick trading

SI's Tick-Size Regime (TSR) exemption allows SI to compete with trading venues without constrained by a TSR, allowing SI to undercut the limit orders at the sub-tick prices that are not available in trading venues. The London Stock Exchange Group argued that SI's TSR exemption leads to unfair advantages for SI over trading venues, diverting the routing of client orders towards SI.[43] Aquis exchange contends that the different tick size between SI and trading venues contradicts the general MiFID II objective in leveling the playing field between trading venues and SI, resulting in trading volume migration from trading venues to SI, where not all participants have access.[44] Meanwhile, regulators (the European Securities and Markets Authority (ESMA)) also share their concerns in SI's advantage in trading at sub-tick prices and opened a public consultation on proposed amendments to extend TSR to SI.[45]

The proposal to extend TSR to SI received diversified responses from different industry participants. On the one hand, the proponents support that the proposal facilitates the process of leveling the playing field between trading venues and SI and avoids further market fragmentation. On the other hand, the opponents stress that the extension can benefit some trading venues from the lighter tick size regime because trading venues are not always bound by TSR.[46] Overall, most respondents support the proposal to extend TSR to SI. Based on the overwhelmingly supportive feedback, ESMA submitted the proposal to extend TSR to SI for endorsement by the European Commission.

[43]See London Stock Exchange Group's response to ESMA's Consultation Paper on Amendments to Commission Delegated Regulation (EU) 2017/587 (RTS 1).

[44]See Aquis Exchange Limited Response to Consultation Paper - Amendments to Commission Delegated Regulation (EU) 2017/587 (RTS 1).

[45]Source: The Consultation Paper: Amendments to Commission Delegated Regulation (EU) 2017/587 (RTS 1)

[46]For example, trading facilities operating under a reference price waiver and matching orders at midpoint

Cross-SI Summary of SI public quotes

As discussed in the Data section, there are considerable cross-SI variations in SI public quotes. This section provides the characteristics of public quotes provided by each SI. We compute the SI time-weighted quoted spread, the SI time-weighted Depth, the percentage of time in providing SI public quotes at European Best Bid (EBB) and European Best Offer (EBO) prices, the percentage of time in providing SI public quotes better than EBB and EBO. For each SI, we calculate the average of the above measures across stocks for which SI provide public quotes. These measures capture the competitiveness of SI public quotes relative to lit market quotes. Table 4.A3 shows the results sorted by average SI quoted spread in ascending order. We find that SI spread widens dramatically from the 19th SI (HSBC SI) and onwards: The 18th SI (Goldman Sachs international SI) has an average SI spread of 20.6 basis points while all SI ranked from 19 have an average SI spread of more than 30 basis points. Moreover, the average time of SI providing quotes at EBBO or better EBBO are close to 0 for SI ranked from 19, whereas the 18th SI has an average time of SI providing quotes at EBBO and better EBBO at 57.0% and 10.5%, respectively. The cutoff in the competitiveness measures around 18th and 19th SI suggests that SI ranked above this cutoff provides more competitive public quotes than public quotes provided by SI ranked below this cutoff. Hence, we use this cutoff and identify SI ranked from 0 to 18 as competitive SI and SI ranked from 19 to 32 as uncompetitive SI. Specifically, we identify a SI as a competitive SI if (i) the average SI time-weighted quoted spread is less than 30 basis points, the average time of SI providing quotes at EBBO or better EBBO is greater than 30% and identify the rest of SI as uncompetitive SI. By construction, we end up with 18 competitive SI and 14 uncompetitive SI.

The significance of the effect of banning SI sub-tick trading on constrained stocks

In section IV, we claim the effects of banning SI sub-tick trading (i) EBBO spread, (ii) SI spread and (iii) the magnitude of price improvement per trade in constrained stocks without showing the significant statistics. To show the significant level of the change of the above measures for constrained stocks, we modify the regression specification used in section IV by replacing the original interaction term with the interaction term between the post-ban indicator and the unconstrained group indicator, as shown below:

$$Y_{i,t} = \beta_0 + \beta_3 1_{[unconstrained]i} \times 1_{[post]t} + +\beta_4 1_{[post]t} + \lambda X_{i,t} + \epsilon_{i,t} \qquad (4.11)$$

where $1_{[unconstrained]i}$ is the unconstrained group indicator that takes values of 1 if stock i is in the unconstrained group and 0 otherwise. The coefficient estimate of the post-ban indicator (β_4) in Eq. (4.11) captures the average changes in the dependent variable after banning SI sub-tick trading in constrained stocks.

Table 4.A4 reports the results. Columns (1) and (2) show that the EBBO spread and SI spread of constrained stocks become narrower after banning SI sub-tick trading. Columns (3) and (4) show that banning SI sub-tick trading increases the magnitude of price improvement per trade for constrained stocks.

Robustness to identification strategy

In section IV, we document the importance of SI sub-tick trading to SI's ability to provide price improvement and in the inter-venue competition in liquidity provision. The regression analysis relies on the identification of constrained and unconstrained groups, and we identify a stock as a constrained (unconstrained) stock if its average tick-constraint time is above (below) the 75th (25th) percentile. We consider an alternative threshold to identify constrained and unconstrained stocks to make sure that our results are robust to the choice of the threshold of identification strategy. In particular, we choose the median as the new threshold: we assign a stock to the constrained group if tick-constraint time of the stock is above the median tick-constraint time, and we assign the rest of stocks to the unconstrained group. This dichotomy suggests that a stock will be either constrained or unconstrained stocks, and the new data sample will cover the full sample. For simplicity, we refer to the new sample as the full sample in the rest of the section.

One key difference between the full sample and our main sample is that the variations in tick-constraint time between constrained and unconstrained stocks are smaller in the full sample than in our main results. Panels A and B of Table 4.A5 compare the stock characteristics of constrained and unconstrained stocks in the full sample. We show that the average tick-constraint time is 25.56% for constrained stocks and 3.68% for unconstrained stocks, whereas these two numbers are 36.96% and 1.53% in our main sample. This is because the constrained (unconstrained) group in the full sample contains all stocks whose tick-constraint time is above (below) the median, while the constrained group in our main sample contains all stocks whose tick-constraint time is above the 75th (below the 25th) percentile. Previously literature asserts that sub-tick trading is more important in constrained stocks than in unconstrained stocks. When the variations in tick-constraint time between constrained and unconstrained stocks become smaller, we expect that the differences in the effect of banning SI sub-tick trading between constrained and unconstrained stocks will also become smaller. We re-estimate regression results in Tables IV and V in Chapter 4 using this full sample. Tables 4.A6 and 4.A7 report the results using the full sample.

All our regression results in Tables 4.A6 and 4.A7 remain unchanged.[47] As we expect, the differences in the effect of banning SI sub-tick trading on SI price improvement between constrained and unconstrained stocks is smaller than that in our main results. Overall, we find consistent results using the full data sample, suggesting our results are not driven by the choice of threshold in identifying constrained and unconstrained stocks.

Robustness to Covariates Balance

In the summary statistics of SI sub-tick trading study, we show that constrained and unconstrained stocks are different in market capitalization, euro trading volume and price. To alleviate the concern of covariate imbalance, we use the Propensity Score Matching (PSM) approach to match each constrained stock with unconstrained stock based on market capitalization, euro trading volume and price.

First, to ensure adequate variation in the above covariates to implement PSM, we again use the median pre-ban average tick-constraint time as the threshold to identify constrained and unconstrained stocks as we did in the previous section. We estimate the logit regression, where the dependent variable is an indicator variable that equals one for a constrained stock and zero for an unconstrained stock, and independent variables are log market capitalization, log euro trading volume and price. We

[47]Similarly, we re-estimate regression results in Tables IV and V in Chapter 4 using sample tercile as the threshold to identify constrained and unconstrained stocks. Tables 4.A8 and 4.A9 report the results using the full sample. All our regression results in Tables 4.A8 and 4.A9 remain unchanged

compute the propensity score for each stock and match each constrained stock with an unconstrained stock with the nearest propensity score.[48] We are aware that large (small) stocks are generally more (less) liquid and have a narrower (wider) spread and, therefore, are more likely to be constrained (unconstrained). This positive correlation between stock size and the level of constraint results in a low-quality matching for large constrained stocks. Hence, we define the matching quality as the negative absolute difference of propensity score in between matching pairs, and we filter out the matching pairs whose matching quality is above the median. Panels C and D of Table 4.A5 compare the summary statistics between constrained stocks and matched unconstrained stocks in our PSM sample. We find that stock characteristics of constrained stocks are not significantly different from these of the matched unconstrained stocks, except for tick-constraint time. We find that the difference in tick-constraint time between constrained stocks and match unconstrained stocks is smaller than the difference in tick-constraint time in our main data sample. Similarly, we expect the different effects of banning SI sub-tick trading between constrained and unconstrained stocks will be smaller in the PSM sample than the difference in our main results.

We use this PSM data sample to re-estimate our results in Tables 4.A10 and 4.A11 , and all our *diff-in-diff* regression results remain the same. We find that the size of interaction term coefficient estimates is smaller compared to that in our main results since the variations in tick-constraint time between constrained stocks and match unconstrained stocks is smaller in the PSM sample. These consistent results using the PSM approach suggest that our results are not driven by the covariate imbalance between constrained and unconstrained stocks.

4.7.2 SI pre-trade transparency

We study the impact of SI minimum quoting size among liquid stocks, while the results for liquid stocks might necessarily apply to the illiquid stocks. Hence, we study the change in liquidity status associated with the annual transparency calculation (ATC) to exploit the exogenous variations in SI minimum quoting size among illiquid stocks.

[48]When matching with the stock with the nearest propensity score, we allow matching one unconstrained stocks to several constrained stocks. (i.e. PSM with replacement)

Each March, ESMA publishes the liquidity status for stocks, indicating whether each stock is liquid or illiquid. According to MiFID II, SI minimum quoting size requirements apply to liquid stocks only, and there are no such requirements for illiquid stocks. When an illiquid stock is re-identified as a liquid stock, SI are required to provide public quotes for this stock, leading to an increase in SI minimum quoting size. We then identify the exogenous variations in SI minimum quoting size by exploiting the time-series variations in the liquidity status.

To be identified as a liquid stock, a stock must satisfy the minimum thresholds of three annual liquidity measurements: 1. free float, 2. the Average Daily Number of Transactions (ADNTE) and 3. the Average Daily Turnover (ADT). A stock is identified as a liquid stock when all three annual liquidity measurements meet their corresponding thresholds. Since three annual liquidity measurements are calculated on an annual basis, the change of stock liquidity will change these three annual liquidity measures, causing the liquidity status to change. For example, Nostrum Oil & Gas PLC was identified as an illiquid stock in 2019 and, hence, SI minimum quoting size requirements do not apply to Nostrum Oil & Gas PLC in 2019. Nonetheless, in the 2020 ATC, Nostrum Oil & Gas PLC was identified as a liquid stock, meaning that SI need to provide public quotes for Nostrum Oil & Gas PLC from 1st Apr 2020. As a result, SI are forced to provide public quotes for Nostrum Oil & Gas PLC when Nostrum Oil & Gas PLC switches its liquidity status from illiquid to liquid, increasing SI minimum quoting size. However, for the illiquid stock whose liquidity status remains illiquid after the ATC publication, their SI minimum quoting size will stay the same. We can then compare the market quality changes among stocks whose liquidity status switches from illiquid to liquid with the market quality changes among stocks whose liquidity status remains illiquid to recover the causal impacts in SI minimum quoting size. Building on that, we formalize this comparison using the *diff-in-diff* analysis to identify the impact of SI minimum quoting size on market quality.

We start with all stocks that are classified as illiquid stocks before the ATC publication. We assign stocks whose liquidity status switches from illiquid to liquid to the treatment group and stocks whose liquidity status remains illiquid after ATC publication to the control group. We are aware that the re-classification of the liquidity status is not random and is driven by the changes in three annual liquidity measures that determine the liquidity status. Precisely, treatment stocks are reclassified as liquid stocks because treatment stocks have experienced increases in liquidity and size (i.e., free float), leading to two potential problems: 1. reverse causality and 2. covariate imbalance.

First, this association between liquidity and the liquidity status might lead to reverse causality between liquidity and SI minimum quoting size. Nevertheless, we avoid this potential reverse causality issue by separating our sample period from the calculation period of ATC. To be specific, we restrict our sample period to a two-month time window around the ATC application date such that our sample period does not overlap with the ATC calculation period. We demonstrate the separation of our sample period and the ATC calculation period using the following example. Suppose that stock i was identified as an illiquid stock in the 2019 ATC and re-identified as a liquid stock in the 2020 ATC. This means that, based on data from 1 January, 2018 to 31 December, 2018, stock i is classified as an illiquid stock from 1 April, 2019 to 31 March, 2019; and, based on data from 1 January, 2019 to 31 December, 2019, stock i is classified as a liquid stock from 1 April, 2020 to 31 March, 2021. The change of liquidity status might indicate that stock i has experienced an enhancement in liquidity from the year 2018 to the year 2019. Nevertheless, our sample period for stock i focus two-month window around the ATC effective date (1 February, 2020 to 1 June, 2020), which does not overlap with either the year 2018 or the year 2019. Therefore, there is no clear channel suggesting that stock i has experienced a liquidity improvement over our sample period. The second potential problem is that treatment stocks tend to have a larger market capitalization and higher liquidity than control stocks because treatment stocks have experienced increases in liquidity and size. To ensure the co-variate balance between treatment and control stocks, we use the PSM to match each treatment stock with a control stock based on the stock's liquidity and market capitalization, and we only keep the matched control stocks in our sample. Additionally, we control for ADNTE and ADT that determine the liquidity status to capture the effect led by the differences between treatment and controls.

Altogether, we estimate the effect of SI minimum quoting size by running the following generalized diff-in-diff regression:

$$Y_{i,t} = \beta_0 + \beta_1 1_{[Switch]i} \times 1_{[post]t} + \lambda X_{i,t} + \epsilon_{i,t} \tag{4.12}$$

Where the subscript i, t denotes the stock i at time t, respectively. $Y_{i,t}$ is the measurement of market quality. $1_{[Switch]i}$ is an indicator that takes values of 1 if stock i switches its liquidity status from illiquid to liquid and 0 otherwise. $1_{[post]t}$ is an indicator that takes values of 1 if date t is in the post ATC application date and 0 otherwise. $X_{i,t}$ is controls that depend on the dependent variables $Y_{i,t}$, including daily realized volatility level, the log total euro volume, the log SI euro volume , ADNTE, ADT and the inverse of the share price, stock fixed effect, date fixed effect and SI fixed effect depending on the dependent variable $Y_{i,t}$. The standard errors are clustered at stock-day level. we focus on the coefficient (β_1) of the interaction term, which recovers the average treatment effect under the *diff-in-diff* setting.

Table 4.A14 reports the results. Columns (1)-(3) show that the increase in the SI minimum quoting size increases SI trading costs evidenced by the lower frequency of providing price improvement and larger effective spreads. Columns (4)-(6) indicate that the increase in the SI minimum quoting size lowers the lit market trading costs demonstrated by a small price impact and narrower effective spread and EBBO spread. The positive association between SI trading costs and SI minimum quoting size and the negative association between lit market trading costs and SI minimum quoting size are consistent with our hypothesis that the increase in SI minimum quoting size attracts informed traders from lit markets to SI, making SI more toxic and lit markets less toxic. Columns (7) - (10) report the coefficient estimates of regression of SI price discovery share and informational efficiency. However, all the interaction term coefficient estimates are statistically insignificant, indicating that SI minimum quoting size does not significantly affect the price discovery process.

[Insert Table 4.A14 near here]

We provide three potential explanations for these insignificant results. First, sample stocks using the liquidity status switch are relatively small and illiquid. This means that we might not have adequate variations in trade prices to generate meaningful price discovery share estimates. Second, the variations in SI minimum quoting size in the alternative identification strategy are relatively smaller than the variations identified in the Sharpe RDD strategy. The switch of liquidity status from illiquid to liquid leads to an increase in SI minimum quoting size of €1000, smaller than the increase in SI minimum quoting size of €2000 in the Sharpe RDD strategy. The shock identified in the alternative identification strategy is relatively small, leading to a less significant effect. Last, SI is more informative in illiquid stocks. We show that the average SI price discovery share measures are for CS of SI trades and for IS of SI trades, where SI, on average, accounts for only 8% of total consolidated volume. Given SI's prominent role in the price discovery process, the additional increase in informativeness associated with the increase in SI minimum quoting size is marginal.

Table 4.A1: Abbreviation definitions

Abbreviation	**Definition**
AC	AutoCorrelation
ADNTE	Average Daily Number of Transactions
ADT	Average Daily Turnover
APA	Approved Publication Arrangements
ATC	Annual Transparency Calculation
ATS	Alternative Trading Systems
AVT	Average Value of Transactions
CS	Gonzalo and Granger (1995) Common factor Share
EBBO	European Best Bid and Offer
ESMA	European Securities and Markets Authority
IFR	Investment Firms Regulation (EU) No 2019/2033
IS	Hasbrouck (1995) Information Share
LIS	Large in Scale
MiFID II	Markets in Financial Instruments Directive II
MiFIR	Markets in Financial Instruments Regulation
MTF	Multilateral Trading Facilities
NBBO	National Best Bid and Offer
OLS	Ordinary Least Squares
OTC	Over-the-Counter
PSM	Propensity Score Matching
RDD	Regression Discontinuity Design
RM	Regulated Markets
SDP	Single-Dealer Platforms
SI	Systematic Internalizers
SI BBO	SI Best Bid and Offer (SI public quote)
SMS	Standard Market Size
TSR	Tick-Size Regime
VECM	Vector Error Correction Model
VR	Variance Ratio

Table 4.A2: Tick size table

This table consists of six liquidity bands and 19 price ranges and specifies the minimum tick size regime for shares, depositary receipts and ETFs. This tick size regime has taken into effect since 2018.

	Liquidity bands					
Price ranges	$0 \leq Average$ daily number of transactions <10	$10 \leq Average$ daily number of transactions <80	$80 \leq Average$ daily number of transactions <600	$600 \leq Average$ daily number of transactions <2000	$2000 \leq Average$ daily number of transactions <9000	$9000 \leq Average$ daily number of transactions
$0 \leq price < 0.1$	0.0005	0.0002	0.0001	0.0001	0.0001	0.0001
$0.1 \leq price < 0.2$	0.001	0.0005	0.0002	0.0001	0.0001	0.0001
$0.2 \leq price < 0.5$	0.002	0.001	0.0005	0.0002	0.0001	0.0001
$0.5 \leq price < 1$	0.005	0.002	0.001	0.0005	0.0002	0.0001
$1 \leq price < 2$	0.01	0.005	0.002	0.001	0.0005	0.0002
$2 \leq price < 5$	0.02	0.01	0.005	0.002	0.001	0.0005
$5 \leq price < 10$	0.05	0.02	0.01	0.005	0.002	0.001
$10 \leq price < 20$	0.1	0.05	0.02	0.01	0.005	0.002
$20 \leq price < 50$	0.2	0.1	0.05	0.02	0.01	0.005
$50 \leq price < 100$	0.5	0.2	0.1	0.05	0.02	0.01
$100 \leq price < 200$	1	0.5	0.2	0.1	0.05	0.02
$200 \leq price < 500$	2	1	0.5	0.2	0.1	0.05
$500 \leq price < 1000$	5	2	1	0.5	0.2	0.1
$1000 \leq price < 2000$	10	5	2	1	0.5	0.2
$2000 \leq price < 5000$	20	10	5	2	1	0.5
$5000 \leq price < 10000$	50	20	10	5	2	1
$10000 \leq price < 20000$	100	50	20	10	5	2
$20000 \leq price < 50000$	200	100	50	20	10	5
$50000 \leq$ price	500	200	100	50	20	10

Source: Tick size table from Annex of RTS 1

Table 4.A3: Summary statistics of the SI public quotes

This table presents the average of the SI public quote characteristics for each SI across all the stocks, to which SI make public quotes available. SI Spread is the average SI time-weighted spread in the basis points of price. SI Depth is the average SI time-weighted depth in shares. AtEBBO is the average percentage of time that SI provide quotes at European Best Bid(EBB) and European Best Offer(EBO) prices. BetterEBBO is the average percentage of time that SI provide quotes better than EBB and EBO prices. AtSIBBO is the a average percentage of time that SI provide quotes at Best Bid and Best Offer prices of SI consolidated tape. EBBO Spread is the average time-weighted EBBO spread.

SI Name	SI Spread	SI Depth	At EBBO	Better EBBO	At SIBBO	EBBO Spread	Rank by SI Spread
HRTEU Ltd	14.44	64.14	77.87	6.14	66.91	16.47	1
VIRTU FINANCIAL IRELAND Ltd	14.75	75.4	72.5	7.1	58.45	15.64	2
MORGAN STANLEY EU S.E.	16.38	71.61	73.7	12.22	66.55	19.77	3
MORGAN STANLEY AND CO. INTERNATIONAL PLC	16.38	71.61	73.55	12.29	66.54	19.77	4
BOFA SECURITIES EU	16.44	105.89	74.79	13.98	66.46	19.34	5
MERRILL LYNCH INTERNATIONAL	16.44	105.89	74.8	13.97	66.46	19.34	6
CITADEL CONNECT EU	16.69	82.95	85.84	4.2	63.6	18.36	7
J.P. MORGAN AG	16.73	68.94	90.94	3.16	65.45	18.5	8
J.P. MORGAN SECURITIES PLC	16.73	68.94	90.94	3.15	65.44	18.5	9
EXANE BNP PARIBAS	17.26	71.13	75.46	10.66	64.32	19.32	10
DEUTSCHE BANK AG	17.44	71.02	64.89	14.85	63.65	19.4	11
XTX MARKETS SAS	18.23	711.1	48.9	23.74	68.02	18.6	12
XTX MARKETS	18.23	711.1	48.87	23.75	67.34	18.6	13
TOWER RESEARCH CAPITAL EU LTD	19.93	106.0	32.13	0.72	23.3	11.77	14
JANE STREET FINANCIAL LTD	20.21	80.35	55.46	6.27	37.4	18.96	15
UBS EU SE	21.0	72.09	65.4	7.32	58.73	18.83	16
UBS AG LONDON BRANCH	21.0	72.09	65.41	7.31	58.73	18.83	17
GOLDMAN SACHS INTERNATIONAL	21.11	70.88	57.09	10.47	55.01	19.43	18
HSBC	33.0	69.84	0.34	0.25	0.14	19.42	19
HSBC FRANCE	33.01	69.84	0.34	0.26	0.14	19.42	20
DNB BANK ASA	44.88	441.83	0.73	0.43	0.1	9.36	21
SOCIETE GENERALE	46.38	97.5	9.96	1.63	8.69	17.71	22
INVESTEC BANK PLC	97.18	85.94	0.9	0.72	4.75	15.49	23
Citi Match - HK	102.89	69.43	0.4	0.53	0.29	19.18	24
CITIGROUP GLOBAL MARKETS Ltd	102.89	69.43	0.4	0.53	0.29	19.18	25
CREDIT SUISSE INTERNATIONAL	123.04	70.53	0.0	0.0	0.0	19.0	26
CREDIT SUISSE SECURITIES (EU) Ltd	123.13	79.67	0.0	0.01	0.01	19.13	27
CREDIT SUISSE EU	123.13	79.01	0.0	0.01	0.01	19.13	28
BERENBERG	172.33	104.78	0.0	0.04	0.02	19.29	29
RBC EU Ltd	497.79	91.11	0.0	0.0	0.0	12.29	30
JEFFERIES INTERNATIONAL	996.84	146.78	0.0	0.0	0.0	18.65	31
JEFFERIES EU	1,000.0	147.64	0.0	0.0	0.0	18.65	32

Table 4.A4: Inter-market competition and SI sub-tick trading ban

This table presents coefficient estimates from panel regressions of Inter-market competition. For dependent variable(s), EBBO Spread is the EBBO time-weighted Spread in basis points of price. SI Spread is SI time-weighted spread in basis points of price. RelPI is the magnitude of price improvement per trade in basis points of price. PI/Tick is the magnitude of price improvement per trade measured by the number of ticks. |AC| is the absolute value of the 30-second EBBO midpoint return auto-correlation. |1-VR| is the absolute difference between one and the variance ratio based on 2-minute and 30-second EBBO midpoint returns. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Unconstrained]\,i}$ and $1_{[post]\,t}$, where$1_{[Unconstrained]\,i}$ is an indicator that equals to 1 if stock i is in unconstrained stock group and 0 otherwise.$1_{[post]\,t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include realized volatility on lit market, the log of Euro trading volume ,the inverse of share price, the tick size in euros, stock fixed effects and SI fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Signficance level of Main results					
	(1) EBBO Spread	(2) SI Spread	(3) RelPI	(4) TickPI	(5) |AC|	(6) |1-VR|
$1_{[Unconstrained]\,i} \times 1_{[post]\,t}$	-2.795***	-2.309***	-1.772***	-0.380***	-0.00327**	-0.00652**
	(-8.42)	(-7.79)	(-12.73)	(-18.59)	(-2.33)	(-2.30)
$1_{[post]\,t}$	-0.693***	-0.675***	0.853***	0.211***	0.00215**	0.00557***
	(-5.43)	(-6.84)	(20.87)	(50.48)	(2.45)	(3.10)
1/Price	3.965***	1.162	0.497	0.175	0.000829	-0.00365
	(3.34)	(0.89)	(0.44)	(0.96)	(0.22)	(-0.46)
RealVol	1.516***	1.398***	0.194***	0.0361***	0.00170*	-0.00842***
	(5.84)	(5.76)	(2.80)	(3.11)	(1.78)	(-8.98)
Log(Euro Volume)	-1.641***	-1.351***			0.00229***	0.00217
	(-7.23)	(-6.84)			(2.92)	(1.40)
TickSize			19.62***			
			(9.71)			
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
SI Fixed Effects	No	Yes	No	No	No	No
Observations	15,798	197,924	14,781	14,781	15,796	15,796
Adjusted R2	0.023	0.012	0.015	0.028	0.002	0.003

Table 4.A5: Summary Statistics: SI sub-tick trading study robustness check

This table summarizes the number of observations, mean, median, standard deviation, minimum and maximum values and the 25th and 75th percentiles for the daily stock-level and stock-SI level observations of variables used in our regression analysis. The sample covers 747 liquid stocks in Euronext Paris, Xetra, Euronext Amsterdam and the LSE from 26 May, 2020 to 26 July, 2020. MktCap is the market capitalization in €1,000,000,000.. SI market share is the percentage of total euro volume executed on SI. Euro Volume is the of Euro volume in €1,000,000. Tick-Constrain is the percentage of time at one-tick spread in lit market. Sub-tick PI is the percentage of SI price improvement trades executed at sub-tick prices.

Panel A: Summary Statistics for Control Stocks (Full sample)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	9,014	15.2	25.22	0.28	1.93	6.11	14.85	204.16
SI market share (%)	9,271	14.04	11.8	0.0	7.14	11.37	17.27	100.0
Euro Volume (€millions)	9,271	108.74	186.63	0.0	7.18	37.7	118.45	2,651.75
Tick-Constrain (%)	9,216	25.62	18.16	0.0	11.67	20.77	35.29	97.8
Sub-tick PI (%)	8,731	40.99	27.73	0.0	18.06	43.1	61.43	100.0

Panel B: Summary Statistics for Control Stocks (Full sample)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	8,866	6.83	19.31	0.02	0.58	1.27	3.08	204.16
SI market share (%)	9,217	11.0	11.56	0.0	4.46	8.23	13.68	100.0
Euro Volume (€millions)	9,217	20.96	55.89	0.0	1.96	5.6	17.23	2,429.85
Tick-Constrain (%)	9,202	3.78	4.51	0.0	0.8	2.27	5.2	65.44
Sub-tick PI (%)	7,949	26.59	26.48	0.0	0.0	21.78	43.94	100.0

Panel C: Summary Statistics for Constrained Stocks (Tercile sample)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	6,058	19.33	28.74	0.36	2.93	7.97	21.77	204.16
SI market share (%)	6,201	14.59	12.07	0.0	7.58	11.84	17.96	100.0
Euro Volume (€millions)	6,201	134.77	208.44	0.0	13.03	54.9	163.2	2,651.75
Tick-Constrain (%)	6,155	32.25	18.22	0.0	18.5	28.5	43.13	97.8
Sub-tick PI (%)	5,868	44.9	27.36	0.0	24.49	48.11	65.08	100.0

Panel D: Summary Statistics for Unconstrained Stocks (Pre-ban period)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	5,889	8.85	23.28	0.02	0.57	1.19	3.05	204.16
SI market share (%)	6,120	10.71	11.79	0.0	4.04	7.88	13.3	100.0
Euro Volume (€millions)	6,120	18.33	57.91	0.0	1.92	4.98	13.67	2,429.85
Tick-Constrain (%)	6,108	2.21	2.79	0.0	0.49	1.33	2.87	38.56
Sub-tick PI (%)	5,157	25.32	26.23	0.0	0.0	18.68	42.48	100.0

Panel E: Summary Statistics for Constrained Stocks (PSM sample)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	4,691	4.55	6.23	0.28	1.31	2.43	5.23	57.26
SI market share (%)	4,691	13.12	11.65	0.0	6.38	10.33	16.07	100.0
Euro Volume (€millions)	4,691	40.02	83.97	0.02	3.34	13.33	43.02	2,004.51
Tick-Constrain (%)	4,681	20.58	14.74	0.0	9.86	17.0	27.36	97.8
Sub-tick PI (%)	4,395	35.93	28.9	0.0	5.74	35.98	56.68	100.0

Panel F: Summary Statistics for Unconstrained Stocks (PSM sample)

Variable	Count	Mean	Std	Min	25 Pct	Median	75 Pct	Max
MktCap (€billions)	3,070	4.05	6.46	0.25	1.14	2.19	3.85	64.22
SI market share (%)	3,070	12.17	10.39	0.0	6.14	9.38	15.15	100.0
Euro Volume (€millions)	3,070	34.3	62.94	0.05	4.0	12.01	34.52	751.72
Tick-Constrain (%)	3,067	4.52	4.65	0.0	1.32	3.06	6.18	50.34
Sub-tick PI (%)	2,896	30.63	25.61	0.0	6.51	28.54	46.73	100.0

Table 4.A6: SI trades allocation and the SI sub-tick trading ban (Full Sample)

This table presents coefficient estimates from panel regressions of SI price improvement. For dependent variable(s), f(PI) is the frequency in percentage of SI trades whose price lies within the prevailing EBBO. f(EBBO) is the frequency in percentage of SI trades at EBBO. RelPI is the size of price improvement in basis points of price. PI/Tick is the size of price improvement measured by the number of tick. PI Benefit is SI price-improvement benefits received in euros. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Constrained]i}$ and $1_{[post]t}$, where$1_{[Constrained]i}$ is an indicator that equals to 1 if stock i is in constrained stock group and 0 otherwise.$1_{[post]t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include Realized volatility on lit market, the inverse of price, the tick size in euros, the diff-in-diff interaction term using control group indicator, SI Euro trading volume. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Full Sample						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	f(PI)	f(EBBO)	RelPI	RelPI	TickPI	TickPI	PI Benefit
$1_{[Constrained]i} \times 1_{[post]t}$	-3.441***	3.698***	1.227***		0.213***		-61.57***
	(-8.77)	(6.99)	(12.47)		(18.44)		(-7.64)
$1_{[post]t}$	-3.818***	4.701***	-0.521***	0.706***	-0.0669***	0.146***	-67.89***
	(-12.12)	(12.00)	(-5.83)	(16.79)	(-6.09)	(37.36)	(-12.15)
$1_{[Unconstrained]i} \times 1_{[post]t}$				-1.227***		-0.213***	
				(-12.47)		(-18.44)	
RealVol	0.348	-1.159***	0.374***	0.374***	0.0650***	0.0650***	41.33***
	(1.56)	(-3.81)	(5.53)	(5.53)	(6.32)	(6.32)	(7.36)
TickSize	-53.50***	76.40***	37.32***	37.32***	-2.416***	-2.416***	
	(-5.83)	(4.91)	(11.94)	(11.94)	(-8.05)	(-8.05)	
1/Price	-0.270	1.204	1.194	1.194	0.407**	0.407**	-117.2***
	(-0.19)	(0.67)	(1.14)	(1.14)	(2.31)	(2.31)	(-3.07)
SI dvol							1.699***
							(5.85)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,849	31,849	29,872	29,872	29,872	29,872	29,872
Adjusted R2	0.030	0.024	0.012	0.012	0.021	0.021	0.045

Table 4.A7: Inter-market competition and the SI sub-tick trading ban (Full Sample)

This table presents coefficient estimates from panel regressions of SI and lit liquidity. For dependent variable(s), RelSpread is EBBO Spread in basis points of price. EuroDepth is EBBO depth measured in euros. SI RelSpread is SI spread in basis points of price. SI EuroDepth is SI depth on the top of the book measured in euros. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Constrained]\,i}$ and $1_{[post]\,t}$, where$1_{[Constrained]\,i}$ is an indicator that equals to 1 if stock i is in constrained stock group and 0 otherwise.$1_{[post]\,t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include the diff-in-diff interaction term using control group indicator, Realized volatility on lit market, the inverse of price, the logarithm of Euro volume. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Full sample				
	(1)	(2)	(3)	(4)	(5)
	EBBO Spread	EBBO Depth	SI Spread	SI Depth	SI Depth
$1_{[Constrained]\,i} \times 1_{[post]\,t}$		898.4***		0.346	
		(7.17)		(0.11)	
$1_{[post]\,t}$	-0.208*	1103.9***	-0.486***	5.138**	5.484**
	(-1.94)	(13.38)	(-5.28)	(2.08)	(2.54)
$1_{[Unconstrained]\,i} \times 1_{[post]\,t}$	-1.677***		-1.333***		-0.346
	(-7.78)		(-6.62)		(-0.11)
1/Price	5.448***	-1399.6***	3.935***	-92.52***	-92.52***
	(4.06)	(-4.40)	(3.22)	(-5.00)	(-5.00)
RealVol	1.918***	-516.4***	2.201***	-14.74*	-14.74*
	(7.86)	(-6.17)	(7.24)	(-1.80)	(-1.80)
Log(Euro Volume)		2429.7***		16.35***	16.35***
		(26.06)		(7.75)	(7.75)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes
SI Fixed Effects	No	No	No	Yes	Yes
Observations	32,652	32,652	408,322	408,945	408,945
Adjusted R2	0.015	0.056	0.009	0.000	0.000

Table 4.A8: SI trades allocation and the SI sub-tick trading ban (Tercile Sample)

This table presents coefficient estimates from panel regressions of SI price improvement. For dependent variable(s), f(PI) is the frequency in percentage of SI trades whose price lies within the prevailing EBBO. f(EBBO) is the frequency in percentage of SI trades at EBBO. RelPI is the size of price improvement in basis points of price. PI/Tick is the size of price improvement measured by the number of tick. PI Benefit is SI price-improvement benefits received in euros. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Constrained]\,i}$ and $1_{[post]\,t}$, where$1_{[Constrained]\,i}$ is an indicator that equals to 1 if stock i is in constrained stock group and 0 otherwise.$1_{[post]\,t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include Realized volatility on lit market, the inverse of price, the tick size in euros, the diff-in-diff interaction term using control group indicator, SI Euro trading volume. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Tercile Sample						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	f(PI)	f(EBBO)	RelPI	RelPI	TickPI	TickPI	PI Benefit
$1_{[Constrained]\,i} \times 1_{[post]\,t}$	-5.069***	5.171***	1.534***	0.289***	-82.82***	0.00257**	0.00654**
	(-10.36)	(7.91)	(11.88)	(16.51)	(-6.98)	(2.19)	(2.39)
$1_{[post]\,t}$	-2.882***	3.438***	-0.681***	-0.101***	-64.07***	-0.000809	0.0000581
	(-7.12)	(7.05)	(-5.62)	(-5.93)	(-7.96)	(-0.87)	(0.03)
RealVol	0.241	-1.142***	0.291***	0.0685***	45.40***	0.00130	-0.00962***
	(0.96)	(-3.41)	(3.81)	(4.68)	(5.60)	(1.59)	(-9.40)
TickSize			25.18***				
			(8.74)				
1/Price	-0.617	1.528	0.645	0.490**	-140.9***	0.00292	-0.00347
	(-0.43)	(0.84)	(0.59)	(2.36)	(-2.99)	(0.81)	(-0.32)
SI dvol					1.662***		
					(5.25)		
Log(Euro Volume)						0.00139**	0.00469***
						(2.23)	(3.21)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,527	20,527	19,444	19,444	19,444	22,030	22,030
Adjusted R2	0.032	0.022	0.011	0.023	0.041	0.001	0.003

Table 4.A9: Inter-market competition and the SI sub-tick trading ban (Tercile Sample)

This table presents coefficient estimates from panel regressions of SI and lit liquidity. For dependent variable(s), RelSpread is EBBO Spread in basis points of price. EuroDepth is EBBO depth measured in euros. SI RelSpread is SI spread in basis points of price. SI EuroDepth is SI depth on the top of the book measured in euros. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Constrained]\,i}$ and $1_{[post]\,t}$, where$1_{[Constrained]\,i}$ is an indicator that equals to 1 if stock i is in constrained stock group and 0 otherwise.$1_{[post]\,t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include the diff-in-diff interaction term using control group indicator, Realized volatility on lit market, the inverse of price, the logarithm of Euro volume. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	Tercile sample			
	(1)	(2)	(3)	(4)
	EBBO Spread	EBBO Depth	SI Spread	SI Depth
$1_{[Constrained]} \times 1_{[post]}$	2.353***	2032.2***	1.897***	-4.922
	(7.94)	(11.69)	(6.80)	(-1.28)
$1_{[post]}$	-3.004***	447.3***	-2.696***	9.018***
	(-10.49)	(5.93)	(-10.19)	(2.92)
1/Price	4.955***	-1084.6***	3.115**	-84.13***
	(3.81)	(-3.59)	(2.37)	(-4.79)
RealVol	1.992***	-440.9***	2.527***	-17.23*
	(6.13)	(-4.90)	(5.90)	(-1.76)
Log(Euro Volume)	-2.153***	2205.6***	-2.185***	20.18***
	(-9.80)	(20.06)	(-11.11)	(7.07)
Stock Fixed Effects	Yes	Yes	Yes	Yes
SI Fixed Effects	No	No	Yes	Yes
Observations	22,032	22,032	270,406	270,982
Adjusted R2	0.024	0.045	0.017	0.001

Table 4.A10: SI trades allocation and the SI sub-tick trading ban (PSM)

This table presents coefficient estimates from panel regressions of SI price improvement. For dependent variable(s), f(PI) is the frequency in percentage of SI trades whose price lies within the prevailing EBBO. f(EBBO) is the frequency in percentage of SI trades at EBBO. RelPI is the size of price improvement in basis points of price. PI/Tick is the size of price improvement measured by the number of tick. PI Benefit is SI price-improvement benefits received in euros. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Constrained]i}$ and $1_{[post]t}$, where$1_{[Constrained]i}$ is an indicator that equals to 1 if stock i is in constrained stock group and 0 otherwise.$1_{[post]t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include Realized volatility on lit market, the inverse of price, the tick size in euros, the diff-in-diff interaction term using control group indicator, SI Euro trading volume. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	PSM Sample						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	f(PI)	f(EBBO)	RelPI	RelPI	TickPI	TickPI	PI Benefit
$1_{[Constrained]i} \times 1_{[post]t}$	-1.940***	1.567*	0.998***		0.158***		-12.05
	(-3.21)	(1.94)	(9.12)		(12.12)		(-1.09)
$1_{[post]t}$	-4.951***	6.400***	-0.312***	0.686***	-0.0539***	0.104***	-77.70***
	(-10.11)	(10.20)	(-3.50)	(10.47)	(-4.64)	(17.21)	(-9.04)
$1_{[Unconstrained]i} \times 1_{[post]t}$				-0.998***		-0.158***	
				(-9.12)		(-12.12)	
RealVol	0.534	-1.189**	0.267***	0.267***	0.0394***	0.0394***	47.14***
	(1.28)	(-2.03)	(3.49)	(3.49)	(4.09)	(4.09)	(4.79)
TickSize	-45.09***	39.05**	30.22***	30.22***			
	(-4.68)	(2.49)	(9.38)	(9.38)			
1/Price	-0.682	66.89**	7.065	7.065	-1.397*	-1.397*	-2385.6***
	(-0.03)	(1.99)	(1.62)	(1.62)	(-1.91)	(-1.91)	(-3.64)
SI dvol							2.571***
							(3.44)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,016	14,016	13,110	13,110	13,110	13,110	13,110
Adjusted R2	0.034	0.027	0.018	0.018	0.019	0.019	0.051

Table 4.A11: Inter-market competition and the SI sub-tick trading ban (PSM)

This table presents coefficient estimates from panel regressions of SI and lit liquidity. For dependent variable(s), RelSpread is EBBO Spread in basis points of price. EuroDepth is EBBO depth measured in euros. SI RelSpread is SI spread in basis points of price. SI EuroDepth is SI depth on the top of the book measured in euros. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[Constrained]\,i}$ and $1_{[post]\,t}$, where$1_{[Constrained]\,i}$ is an indicator that equals to 1 if stock i is in constrained stock group and 0 otherwise.$1_{[post]\,t}$ is an indicator that equals to 1 if date t is in post-ban period and 0 otherwise. Control variables include the diff-in-diff interaction term using control group indicator, Realized volatility on lit market, the inverse of price, the logarithm of Euro volume. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	PSM sample				
	(1)	(2)	(3)	(4)	(5)
	EBBO Spread	EBBO Depth	SI Spread	SI Depth	SI Depth
$1_{[Constrained]\,i} \times 1_{[post]\,t}$		176.4		-15.87***	
		(1.41)		(-3.53)	
$1_{[post]\,t}$	-0.662***	1149.4***	-0.849***	17.95***	2.079
	(-4.09)	(12.00)	(-6.21)	(5.29)	(0.66)
$1_{[Unconstrained]\,i} \times 1_{[post]\,t}$	-0.843***		-0.615***		15.87***
	(-3.21)		(-2.63)		(3.53)
1/Price	7.225	-29007.3***	23.69	-2012.1***	-2012.1***
	(0.56)	(-6.74)	(1.55)	(-9.79)	(-9.79)
RealVol	1.734***	-611.3***	1.717***	-4.542	-4.542
	(5.21)	(-5.05)	(6.70)	(-1.55)	(-1.55)
Log(Euro Volume)		2248.4***		14.89***	14.89***
		(23.27)		(5.44)	(5.44)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes
SI Fixed Effects	No	No	Yes	Yes	Yes
Observations	14,267	14,267	181,846	181,959	181,959
Adjusted R2	0.014	0.090	0.009	0.001	0.001

Table 4.A12: Inter-market competition and the SI sub-tick trading ban: parallel trend assumption

This table presents the dynamics of average treatment effects of the SI sub-tick trading ban. I generate four event time indicators (Biweek(j)), where j takes values of (-2, -1, 1, 2) to indicate the j biweeks (i.e., two weeks) after the SI sub-tick trading ban. For dependent variable(s), EBBO Spread is time-weighted EBBO spread in basis points. EBBO depth is the time-weighted EBBO depth measured in euros. SI Spread is the time-weighted spread of SI public quotes in basis points. For independent variable(s), the variable of interest is the interaction between $1_{[Constrained]}$ and Biweek(j), where $1_{[Constrained]}$ is an indicator that equals 1 if the stock is in the constrained group and 0 otherwise. Control variables include realized volatility of 15-second EBBO midpoint returns (RealVol), the inverse of the share price (1/Price), the logarithm of consolidated euro volume (Log(Euro Volume)), stock fixed effects and SI fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) EBBO Spread	(2) EBBO Depth	(3) SI Spread
$1_{[Constrained]} \times$ Biweek(-2)	-0.561	468.0	0.280
	(-0.67)	(0.70)	(0.38)
$1_{[Constrained]} \times$ Biweek(-1)	0.999	-153.4	1.353*
	(1.22)	(-0.23)	(1.85)
$1_{[Constrained]} \times$ Biweek$(+1)$	2.960***	1507.7**	3.579***
	(3.45)	(2.28)	(4.81)
$1_{[Constrained]} \times$ Biweek$(+2)$	3.035***	2844.4***	4.185***
	(3.69)	(4.12)	(5.63)
Biweek(-2)	0.807	-471.7***	-0.112
	(1.02)	(-2.93)	(-0.16)
Biweek(-1)	-0.906	-6.730	-1.070
	(-1.17)	(-0.04)	(-1.51)
Biweek$(+1)$	-2.914***	328.9**	-3.965***
	(-3.64)	(2.17)	(-5.50)
Biweek$(+2)$	-4.235***	1022.0***	-4.767***
	(-5.43)	(5.86)	(-6.60)
1/Price	0.201	-31.87***	0.199
	(0.98)	(-2.72)	(1.01)
RealVol	1.468***	-440.1***	1.430***
	(5.78)	(-4.93)	(6.56)
Log(Euro Volume)	-1.809***	2785.8***	-1.586***
	(-8.15)	(18.15)	(-8.63)
Stock Fixed Effects	Yes	Yes	Yes
SI Fixed Effects	No	No	Yes
Observations	15,990	15,990	250,639
Adjusted R2	0.027	0.052	0.017

Table 4.A13: SI price improvement, informational efficiency and the SI sub-tick trading ban: parallel trend assumption

This table presents the dynamics of average treatment effects of the SI sub-tick trading ban. I generate four event time indicators (Biweek(j)), where j takes values of (-2, -1, 1, 2) to indicate the j biweeks (i.e., two weeks) after the SI sub-tick trading ban. For dependent variable(s), f(PI) is the frequency in percentage of SI trades whose price lies within the prevailing EBBO. f(EBBO) is the frequency in percentage of SI trades at EBBO. RelPI is the magnitude of price improvement per trade in basis points. TickPI is the magnitude of price improvement per trade measured by the number of ticks. PI Benefit is SI price-improvement benefits in euros. |AC| is the absolute value of the 30-second EBBO midpoint return auto-correlation. |1-VR| is the absolute difference between one and the variance ratio based on 2-minute and 30-second EBBO midpoint returns. For independent variable(s), the variable of interest is the interaction between $1_{[Constrained]}$ and Biweek(j), where $1_{[Constrained]}$ is an indicator that equals 1 if the stock is in the constrained group and 0 otherwise. Control variables include realized volatility of 15-second EBBO midpoint returns (RealVol), the inverse of the share price (1/Price), the tick size in basis points of price (TickSize), SI euro volume (SI EuroVolume), the logarithm of consolidated euro volume (Log(Euro Volume)) and stock fixed effects. t-statistics from robust standard errors clustered by date and stock are reported in parentheses, and *,** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

	(1) Q(PI)	(2) Q(EBBO)	(3) RelPI	(4) TickPI	(5) PI Benefit	(6) |AC|	(7) 1-VR
$1_{[Constrained]} \times$ Biweek(-2)	0.261	0.910	0.0384	-0.00399	-4.658	0.00244	0.00293
	(0.24)	(0.66)	(0.13)	(-0.10)	(-0.17)	(0.92)	(0.89)
$1_{[Constrained]} \times$ Biweek(-1)	0.330	-0.841	0.413	0.0580	10.83	0.00291	0.00302
	(0.31)	(-0.64)	(1.52)	(1.50)	(0.41)	(1.09)	(0.35)
$1_{[Constrained]} \times$ Biweek($+1$)	-7.002***	7.528***	2.109***	0.400***	-107.0***	0.00646**	0.0104**
	(-6.61)	(5.53)	(7.70)	(10.40)	(-4.07)	(2.43)	(2.00)
$1_{[Constrained]} \times$ Biweek($+2$)	-3.454***	4.181***	1.819***	0.404***	-76.42***	0.00467*	0.00800
	(-3.24)	(3.00)	(6.63)	(10.27)	(-2.83)	(1.75)	(1.55)
Biweek(-2)	-1.798**	0.633	-0.416	-0.0370	-32.80	-0.00457**	-0.00461
	(-2.00)	(0.62)	(-1.52)	(-0.95)	(-1.59)	(-2.06)	(-1.10)
Biweek(-1)	-4.183***	3.825***	-0.492*	-0.0406	-96.17***	0.0000179	-0.000456
	(-4.89)	(3.90)	(-1.90)	(-1.07)	(-5.07)	(0.01)	(-0.11)
Biweek($+1$)	-3.380***	3.021***	-1.308***	-0.186***	-108.4***	-0.00314	-0.00179
	(-3.85)	(2.98)	(-5.01)	(-4.93)	(-5.67)	(-1.45)	(-0.43)
Biweek($+2$)	-7.711***	6.954***	-1.254***	-0.208***	-126.7***	-0.00200	-0.00408
	(-8.71)	(6.66)	(-4.78)	(-5.38)	(-6.52)	(-0.90)	(-0.97)
RealVol	0.425*	-1.162***	0.187***	0.0356***	36.19***	0.00196**	-0.00861***
	(1.76)	(-2.67)	(2.73)	(3.10)	(5.77)	(2.02)	(-9.28)
1/Price	0.194**	-0.0682	0.00136	0.00556	-0.0898	0.00216***	0.00220***
	(2.22)	(-0.61)	(0.03)	(0.63)	(-0.05)	(8.93)	(3.84)
TickSize			20.62***				
			(10.16)				
SI EuroVolume					1.374***		
					(5.03)		
Log(Euro Volume)						0.00250***	0.00233
						(3.24)	(1.52)
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,707	15,707	14,932	14,932	14932	15,970	15,970
Adjusted R2	0.047	0.027	0.016	0.029	0.050	0.003	0.003
SE Type	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Table 4.A14: SI transparency and market quality (Illiquid stocks)

This table presents coefficient estimates from panel regressions of SI transparency. For dependent variable(s), f(PI) is the frequency in percentage of SI trades whose price lies within the prevailing EBBO. EffSpread is the effective spread in the basis point of price in SI market. PrcImpact is the price impact in the basis point of price in SI market. EffSpread(lit) is the effective spread in the basis point of price in lit market. PrcImpact(lit) is the price impact in the basis point of price in lit market. RelSpread is EBBO Spread in basis points of price. CS is the component share of SI in percentage. IS is the information share of SI in percentage. |AC| is the absolute value of 15-seconds EBBO midquote return auto-correlation. $\|1-VR\|SI$ is the absolute difference between one and the variance ratio based on EBBO midquote return]. For independent variable(s), the variable of interest (the interaction term) is the interaction between $1_{[switch]}i$ and $1_{[post]}t$, where$1_{[switch]}i$ is an indicator that equals to 1 if stock i switches its liquidity status from illiquid to liquid and 0 otherwise. $1_{[post]}t$ is an indicator that equals to 1 if date t is in post-treatment period and 0 otherwise. Control variables include Realized volatility on lit market, the inverse of price, the tick size in euros, the average daily number of transcation, the average daliy turnover, the logarithm of Euro volume, SI market share of all euro-volume trading (including trades in non-continuous trading hours), the proportion of SI trades in euros that executed within EBBO. t-statistics from robust standard errors clustered by date and stock are reported in the parentheses, and *,** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

	(1) f(PI)	(2) EffSpread	(3) PrcImpact	(4) EffSpread (lit)	(5) PrcImpact (lit)	(6) EBBO Spread	(7) CS	(8) IS	(9) |AC|	(10) | 1-VR |
$1_{[1_{[switch]i}]i} \times 1_{[post]t}$	-3.482***	3.642***	0.277	-1.231***	-0.748**	0.630	-1.580	-1.640	0.00132	-0.00122
	(-3.53)	(3.73)	(0.55)	(-2.75)	(-2.12)	(1.27)	(-1.51)	(-1.45)	(0.71)	(-0.30)
1/Price	-2.505	6.023*	3.789**				0.653	-0.115	-0.00146	-0.0149
	(-1.13)	(1.94)	(2.55)				(0.27)	(-0.05)	(-0.39)	(-1.54)
RealVol	0.766	25.15***	8.603***	10.76***	12.46***	8.682***	-6.391***	-7.047***	0.0146***	0.0260***
	(0.83)	(11.21)	(6.43)	(5.55)	(4.69)	(6.02)	(-4.76)	(-5.10)	(4.70)	(4.24)
TickSize	12.70	12.30	7.057	21.88***	13.64***	19.49***	22.87*	19.34		
	(1.22)	(1.03)	(0.94)	(4.08)	(3.37)	(4.86)	(1.82)	(1.47)		
ADNTE	-0.257	0.209	-0.0741	-0.294*	-0.776***	0.351*	0.404	0.292	-0.000797	-0.00326**
	(-0.77)	(0.65)	(-0.39)	(-1.81)	(-5.89)	(1.83)	(1.14)	(0.76)	(-1.22)	(-2.23)
ADT	-0.584	-0.215	0.262	0.0627	0.0409	-0.402**	-0.636	-0.515	-0.000289	0.000474
	(-1.52)	(-0.60)	(1.00)	(0.38)	(0.31)	(-2.40)	(-1.49)	(-1.11)	(-0.42)	(0.28)
Log(Euro Volume)		-3.625***	-0.0815							
		(-7.17)	(-0.41)							
SI Shr							-0.0220	0.0587***	-0.0000504	-0.000118
							(-1.22)	(2.93)	(-1.54)	(-1.62)
Trddvol withinshr							0.0404***	0.0202		
							(3.39)	(1.57)		
Stock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,463	8,880	8,888	17,956	17,956	18,199	7,549	7543	14,435	14,437
Adjusted R2	0.002	0.057	0.028	0.051	0.098	0.028	0.007	0.007	0.006	0.005

Chapter 5

Conclusion

In financial economics research, the efficient market hypothesis (EMH) stands as a cornerstone, underscoring the critical role of information in the valuation of assets within financial markets. EMH posits that markets should ideally reflect all available information in stock prices promptly and accurately. However, the reality diverges, with market frictions in trading and some agents such as corporate insiders and analysts often accessing crucial information ahead of the public. Furthermore, technological advancements and regulatory changes have significantly reshaped market structures, highlighting the need to assess what influences informational efficiency today and how it can be enhanced in this evolved landscape.

Informed trading emerges as a pivotal element affecting informational efficiency. Investors who actively gather and analyze information engage in informed trading to leverage and capitalize on their insights, thereby incorporating more information into asset prices and improving informational efficiency. This thesis starts with focusing on shorting activity, a prevalent form of informed trading. I find that ETF shorting activity not only contains market-wide information but also positively influences the informational efficiency, as detailed in Chapter 2. Further exploration into the nuances of short-selling highlights the differential impact of liquidity-demanding and liquidity-supplying shorts (Chapter 3). Specifically, liquidity-demanding shorts demonstrate informed trading through their ability to predict future returns and earnings surprises negatively, a characteristic not observed with liquidity-supplying shorts. I also delve into further understanding the heterogeneity in informativeness among shorting activity.

Beyond informed trading, market structure and regulatory frameworks play a significant role in shaping informational efficiency. The market structure and regulatory frameworks can influence how information is incorporated into asset prices. This thesis extends to the practices of sub-tick trading and pre-trade transparency of Systematic Internalizers (SI) in European equity markets. Sub-tick trading, by permitting traders to engage in transactions that reflect valuations within tick sizes, enhances informational efficiency. I find supporting evidence that banning SI sub-tick trading, on the contrary, impairs informational efficiency (Chapter 4). Furthermore, imposing greater pre-trade transparency on SI diverts informed traders from lit markets to SI. This shift improves informational efficiency as some SI, aware of their clients' identities, can make more informative quotes based on the trader types inferred from the broker identity, facilitating the incorporation of information into asset prices.

Collectively, this thesis weaves a comprehensive narrative that traverses the complexities of information dissemination and its implications for informational efficiency and market quality. This thesis contributes to the understanding of contemporary market development and challenges that affect informational efficiency. The insights derived from this research not only contribute to academic scholarship but also offer practical guidance for enhancing market structures and regulatory policies. In doing so, this thesis lays the groundwork for future inquiries into the evolving landscape of financial markets, encouraging a continued exploration of the delicate interplay among information, trading activities, and informational efficiency.

Bibliography

Appel, Ian R, Todd A Gormley, and Donald B Keim, 2016, Passive investors, not passive owners, *Journal of Financial Economics* 121, 111–141.

Aramian, Fatemeh, and Lars L Norden, 2020, High-Frequency Traders and Single-Dealer Platforms, *Available at SSRN 3738608* .

Aramian, Fatemeh, and Lars L Norden, 2021, Costs and Benefits of Trading with Stock Dealers: The Case of Systematic Internalizers, *Available at SSRN 3409878* .

Barclay, Michael J, William G Christie, Jeffrey H Harris, Eugene Kandel, and Paul H Schultz, 1999, Effects of market reform on the trading costs and depths of Nasdaq stocks, *The Journal of Finance* 54, 1–34.

Battalio, Robert H, Brian C Hatch, and Mehmet Saglam, 2022, The cost of exposing large institutional orders to electronic liquidity providers, in *Sixth Annual Conference on Financial Market Regulation*.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do etfs increase volatility?, *The Journal of Finance* 73, 2471–2535.

Bernile, Gennaro, Jianfeng Hu, and Yuehua Tang, 2016, Can information be locked up? informed trading ahead of macro-news announcements, *Journal of Financial Economics* 121, 496–520.

Bessembinder, Hendrik, Marios Panayides, and Kumar Venkataraman, 2009, Hidden liquidity: an analysis of order exposure strategies in electronic stock markets, *Journal of Financial Economics* 94, 361–383.

Bessembinder, Hendrik, and Kumar Venkataraman, 2010, Bid-ask spreads: Measuring trade execution costs in financial markets, *Encyclopedia of Quantitative Finance* 184–190.

Bhojraj, Sanjeev, Partha Mohanram, and Suning Zhang, 2020, Etfs and information transfer across firms, *Journal of Accounting and Economics* 70, 101336.

Bhojraj, Sanjeev, and Wuyang Zhao, 2021, Supply constraints and limits to arbitrage in the etf loan market, *Available at SSRN 3862603* .

Bloomfield, Robert, and Maureen O'Hara, 1999, Market transparency: who wins and who loses?, *The Review of Financial Studies* 12, 5–35.

Bloomfield, Robert, Maureen O'hara, and Gideon Saar, 2005, The "make or take" decision in an electronic market: Evidence on the evolution of liquidity, *Journal of Financial Economics* 75, 165–199.

Bloomfield, Robert, MAUREEN O'Hara, and Gideon Saar, 2015, Hidden liquidity: Some new light on dark trading, *The Journal of Finance* 70, 2227–2274.

Boehmer, Ekkehart, Zsuzsa R Huszar, and Bradford D Jordan, 2010, The good news in short interest, *Journal of Financial Economics* 96, 80–97.

Boehmer, Ekkehart, Zsuzsa R Huszár, Yanchu Wang, Xiaoyan Zhang, and Xinran Zhang, 2022, Can shorts predict returns? a global perspective, *The Review of Financial Studies* 35, 2428–2463.

Boehmer, Ekkehart, Charles M Jones, and Xiaoyan Zhang, 2008, Which shorts are informed?, *The Journal of Finance* 63, 491–527.

Boehmer, Ekkehart, Charles M Jones, and Xiaoyan Zhang, 2020, Potential pilot problems: Treatment spillovers in financial regulatory experiments, *Journal of Financial Economics* 135, 68–87.

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *The Journal of Finance* 76, 2249–2305.

Boehmer, Ekkehart, and Eric K Kelley, 2009, Institutional investors and the informational efficiency of prices, *The Review of Financial Studies* 22, 3563–3594.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2005, Lifting the veil: An analysis of pre-trade transparency at the NYSE, *The Journal of Finance* 60, 783–815.

Boehmer, Ekkehart, and Juan Wu, 2013, Short selling and the price discovery process, *The Review of Financial Studies* 26, 287–322.

Boulatov, Alex, and Thomas J George, 2013, Hidden and displayed liquidity in securities markets with informed liquidity providers, *The Review of Financial Studies* 26, 2096–2137.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2019, Price discovery without trading: Evidence from limit orders, *The Journal of Finance* 74, 1621–1658.

Brogaard, Jonathan, and Jing Pan, 2022, Dark pool trading and information acquisition, *The Review of Financial Studies* 35, 2625–2666.

Brunnermeier, Marcus K, and Stefan Nagel, 2002, Arbitrage at its limits: Hedge funds and the technology bubble, *London University Hedge Fund Center Working Paper* .

Busse, Jeffrey A, 1999, Volatility timing in mutual funds: Evidence from daily returns, *The Review of Financial Studies* 12, 1009–1041.

Buti, Sabrina, Francesco Consonni, Barbara Rindi, Yuanji Wen, and Ingrid M Werner, 2015, Sub-penny and queue-jumping, *Charles A. Dice Center Working Paper* .

Buti, Sabrina, Barbara Rindi, and Ingrid M Werner, 2011, Diving into dark pools, *Charles A. Dice Center Working Paper* .

Buti, Sabrina, Barbara Rindi, and Ingrid M Werner, 2017, Dark pool trading strategies, market quality and welfare, *Journal of Financial Economics* 124, 244–265.

Cao, Charles, Yong Chen, Bing Liang, and Andrew W Lo, 2013, Can hedge funds time market liquidity?, *Journal of Financial Economics* 109, 493–516.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of*

Finance 52, 57–82.

Chakrabarty, Bidisha, Pamela C Moulton, and Andriy Shkilko, 2012, Short sales, long sales, and the lee–ready trade classification algorithm revisited, *Journal of Financial Markets* 15, 467–491.

Chen, Hang, Yushu Zhu, and Liang Chang, 2019a, Short-selling constraints and corporate payout policy, *Accounting & Finance* 59, 2273–2305.

Chen, Yong, Zhi Da, and Dayong Huang, 2019b, Arbitrage trading: The long and the short of it, *The Review of Financial Studies* 32, 1608–1646.

Chen, Yong, Bryan Kelly, and Wei Wu, 2020, Sophisticated investors and market efficiency: Evidence from a natural experiment, *Journal of Financial Economics* 138, 316–341.

Chen, Yong, and Bing Liang, 2007, Do market timing hedge funds time the market?, *Journal of Financial and Quantitative Analysis* 42, 827–856.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2008, Liquidity and market efficiency, *Journal of financial Economics* 87, 249–268.

Chordia, Tarun, and Lakshmanan Shivakumar, 2006, Earnings and price momentum, *Journal of Financial Economics* 80, 627–656.

Christophe, Stephen E, Michael G Ferri, and Jim Hsieh, 2010, Informed trading before analyst downgrades: Evidence from short sellers, *Journal of Financial Economics* 95, 85–106.

Collin-Dufresne, Pierre, and Vyacheslav Fos, 2015, Do prices reveal the presence of informed trading?, *The Journal of Finance* 70, 1555–1582.

Comerton-Forde, Carole, Binh Huu Do, Philip Gray, and Tom Manton, 2016a, Assessing the information content of short-selling metrics using daily disclosures, *Journal of Banking & Finance* 64, 188–204.

Comerton-Forde, Carole, Vincent Grégoire, and Zhuo Zhong, 2019, Inverted fee structures, tick size, and market quality, *Journal of Financial Economics* 134, 141–164.

Comerton-Forde, Carole, Charles M Jones, and Tālis J Putniņš, 2016b, Shorting at close range: a tale of two types, *Journal of Financial Economics* 121, 546–568.

Comerton-Forde, Carole, and Tālis J Putniņš, 2015, Dark trading and price discovery, *Journal of Financial Economics* 118, 70–92.

Copeland, Thomas E, and Dan Galai, 1983, Information effects on the bid-ask spread, *the Journal of Finance* 38, 1457–1469.

Crane, Alan D, Sébastien Michenaud, and James P Weston, 2016, The effect of institutional ownership on payout policy: Evidence from index thresholds, *The Review of Financial Studies* 29, 1377–1408.

Da, Zhi, and Sophie Shive, 2018, Exchange traded funds and asset return correlations, *European Financial Management* 24, 136–168.

Degryse, Hans, Frank De Jong, and Vincent van Kervel, 2015, The impact of dark trading and visible fragmentation on market quality, *Review of Finance* 19, 1587–1622.

Diamond, Douglas W, and Robert E Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277–311.

Diether, Karl B, Kuan-Hui Lee, and Ingrid M Werner, 2009, Short-sale strategies and return predictability, *The Review of Financial Studies* 22, 575–607.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Eglīte, Elza, Vinay Patel, and Tālis J Putniņš, 2023, Using etfs to conceal insider trading, *Available at SSRN 4343579* .

Engelberg, Joseph E, Adam V Reed, and Matthew C Ringgenberg, 2012, How are shorts informed?: Short sellers, news, and information processing, *Journal of Financial Economics* 105, 260–278.

Eom, Kyong Shik, Jinho Ok, and Jong-Ho Park, 2007, Pre-trade transparency and market quality, *Journal of Financial Markets* 10, 319–341.

Ernst, Thomas, 2021, Stock-specific price discovery from etfs, Technical report, Working Paper, University of Maryland, College Park.

Evans, Richard B, Rabih Moussawi, Michael S Pagano, and John Sedunov, 2022, Operational shorting and etf liquidity provision, *Available at SSRN* .

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Fang, Yanhao, and Gary C Sanger, 2011, Index price discovery in the cash market.

Figlewski, Stephen, 1981, The informational effects of restrictions on short sales: Some empirical evidence, *Journal of Financial and Quantitative Analysis* 16, 463–476.

Glosten, Lawrence, Suresh Nallareddy, and Yuan Zou, 2021, Etf activity and informational efficiency of underlying securities, *Management Science* 67, 22–47.

Glosten, Lawrence R, and Paul R Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Gonzalo, Jesus, and Clive Granger, 1995, Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics* 13, 27–35.

Hasbrouck, Joel, 1995, One security, many markets: Determining the contributions to price discovery, *The Journal of Finance* 50, 1175–1199.

Hasbrouck, Joel, 2003, Intraday price formation in us equity index markets, *The Journal of Finance* 58, 2375–2400.

Heath, Davidson, Matthew C Ringgenberg, Mehrdad Samadi, and Ingrid M Werner, 2022, Reusing natural experiments, *Fisher College of Business Working Paper* 021.

Henry, Tyler R, and Jennifer L Koski, 2010, Short selling around seasoned equity offerings, *The Review of Financial Studies* 23, 4389–4418.

Heston, Steven L, and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Hou, Deshuai, Qingbin Meng, Kai Zhang, and Kam C Chan, 2019, Motives for corporate philanthropy propensity: Does short selling matter?, *International Review of Economics & Finance* 63, 24–36.

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhong, 2021, Innovation and informed trading: Evidence from industry etfs, *The Review of Financial Studies* 34, 1280–1316.

Imbens, Guido, and Karthik Kalyanaraman, 2012, Optimal bandwidth choice for the regression discontinuity estimator, *The Review of Economic Studies* 79, 933–959.

Israeli, Doron, Charles Lee, and Suhas A Sridharan, 2017, Is there a dark side to exchange traded funds? an information perspective, *Review of Accounting Studies* 22, 1048–1083.

Ivanov, Stoyu I, Frank J Jones, and Janis K Zaima, 2013, Analysis of djia, s&p 500, s&p 400, nasdaq 100 and russell 2000 etfs and their influence on price discovery, *Global Finance Journal* 24, 171–187.

Jain, Archana, and Chinmay Jain, 2015, Fails-to-deliver before and after the implementation of rule 203 and rule 204, *Financial Review* 50, 611–636.

Jiao, Yawen, Massimo Massa, and Hong Zhang, 2015, Short selling meets hedge fund 12f: An anatomy of informed demand, *Journal of Financial Economics* 121, 543–567.

Karpoff, Jonathan M, and Xiaoxia Lou, 2010, Short sellers and financial misconduct, *The Journal of Finance* 65, 1879–1913.

Kaul, Gautam, and Mahendrarajah Nimalendran, 1990, Price reversals: Bid-ask errors or market overreaction?, *Journal of Financial Economics* 28, 67–93.

Kwan, Amy, Ronald Masulis, and Thomas H McInish, 2015, Trading rules, competition for order flow and market fragmentation, *Journal of Financial Economics* 115, 330–348.

Lee, Charles MC, and Mark J Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Li, Frank Weikai, and Qifei Zhu, 2022, Short selling etfs, *Review of Asset Pricing Studies* 1.

Livnat, Joshua, and Richard R Mendenhall, 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts, *Journal of accounting research* 44, 177–205.

Lou, Dong, 2012, A flow-based explanation for return predictability, *The Review of Financial Studies* 25, 3457–3489.

Madhavan, Ananth, David Porter, and Daniel Weaver, 2005, Should securities markets be transparent?, *Journal of Financial Markets* 8, 265–287.

Menkveld, Albert J, Bart Zhou Yueshen, and Haoxiang Zhu, 2017, Shades of darkness: A pecking order of trading venues, *Journal of Financial Economics* 124, 503–534.

O'Hara, Maureen, Gideon Saar, and Zhuo Zhong, 2019, Relative tick size and the trading environment, *The Review of Asset Pricing Studies* 9, 47–90.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics* 100, 459–474.

Pagano, Marco, and Ailsa Röell, 1996, Transparency and liquidity: a comparison of auction and dealer markets with informed trading, *The Journal of Finance* 51, 579–611.

Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Reed, Adam V, Mehrdad Samadi, and Jonathan S Sokobin, 2020, Shorting in broad daylight: Short sales and venue choice, *Journal of Financial and Quantitative Analysis* 55, 2246–2269.

Roberts, Michael R, and Toni M Whited, 2013, Endogeneity in empirical corporate finance1, in *Handbook of the Economics of Finance*, volume 2, 493–572 (Elsevier).

Ryans, James, 2017, Using the edgar log file data set, *Available at SSRN 2913612* .

Saffi, Pedro AC, and Kari Sigurdsson, 2011, Price efficiency and short selling, *The Review of Financial Studies* 24, 821–852.

Shumway, Tyler, 1997, The delisting bias in crsp data, *The Journal of Finance* 52, 327–340.

Wang, Xue, Xuemin Sterling Yan, and Lingling Zheng, 2020, Shorting flows, public disclosure, and market efficiency, *Journal of Financial Economics* 135, 191–212.

Weller, Brian M, 2018, Does algorithmic trading reduce information acquisition?, *The Review of Financial Studies* 31, 2184–2226.

Yu, Lei, 2005, Basket securities, price formation, and informational efficiency, *Price Formation, and Informational Efficiency (March 25, 2005)* .

Zhu, Haoxiang, 2014, Do dark pools harm price discovery?, *The Review of Financial Studies* 27, 747–789.

Zhu, Qifei, 2020, The missing new funds, *Management Science* 66, 1193–1204.