Measuring The Effect Of Investors' Private Information On Stock Prices: Implications For Earnings Management And Stock Liquidity

Thu Minh Thi Vu

Doctor of Philosophy

ASTON UNIVERSITY

May 2024

ASTON UNIVERSITY

MEASURING THE EFFECT OF INVESTORS' PRIVATE INFORMATION ON STOCK PRICES: IMPLICATIONS FOR EARNINGS MANAGEMENT AND STOCK LIQUIDITY

Thu Minh Thi Vu

Doctor of Philosophy

2024

Thesis Summary

This thesis has five chapters. Chapter 1 serves as the introduction, Chapters 2, 3, and 4 each present separate empirical studies that focus on the private information investors embed in stock prices and its impact on the real economy, and Chapter 5 concludes the research.

The first article is presented in Chapter 2, where I develop a new measure for investors' private information called the probability of informed trading with size effects (SDPIN) embedded in stock prices. Contrary to the existing measures, SDPIN considers the trade order size (volume) and the trade frequency. I advocate that the existence of private information can be captured through the trade volume. I test the validity of this new measure and examine how it relates to the stock price synchronicity, using various proxies for private information. The findings show that it is more accurate than the existing private information measures. Hence, it can help reduce information asymmetries among market players and enhance both market transparency and managers' awareness of private information, so as to make more optimal decisions.

The second article is presented in Chapter 3 and studies the interplay between financial markets and the real economy, focusing on the effect of the investors' private information on earnings management. I use two econometric models (single-level and multilevel regression models) and three private information measures: the probability of informed trading (PIN), dynamic measure for the probability of informed trading (DPIN), and dynamic measure of the probability of informed trading with size effects (SDPIN). The findings show that managers are less likely to engage in earnings manipulation when investors' private information is higher. I examine the upward and downward earnings management and conclude that private information has a greater effect on the upward earnings management than on downward earnings management. Hence, it is concluded that firms' managers can gain valuable insights from the analysis of the stock price movements. This finding is in alignment with the hypothesis of managerial learning, incentive channels, and the information flow from secondary markets to the economy.

The third article is presented in Chapter 4, where I examine the impact of informed trading on stock liquidity in the context of high-frequency trading, relying on the probability of informed trading (DPIN) developed by Chang et al. (2014) and the probability of informed trading with size effects (SDPIN) proposed in Chapter 2 of this thesis. I analyse daily, weekly, monthly, quarterly, and yearly data from S&P 500 companies covering the period between 2018 and 2021. While the analysis with daily and weekly data reveals that informed trading enhances stock liquidity, those with monthly, quarterly, and yearly data, reveal that such an effect is not evident. Specifically, for daily and weekly data the findings show that informed trading enhances stock liquidity, whereas for monthly, quarterly, and yearly data such effects do not exist. Finally, the above findings hold during the COVID-19 period.

Keywords: Investors' Private Information; Order Size; Earnings Management; Informed Trading; Stock Liquidity.

ACKNOWLEDGEMENT

My PhD experience has been an incredible journey of facing countless challenges and gaining incredible growth. Throughout this time, I have gained invaluable knowledge and development. I would like to express my deepest gratitude to everyone who has supported me along this path.

Firstly, I extend my heartfelt thanks to my exceptional supervisors, Dr. Alcino Azevedo and Dr. Dimitrios Stafylas. Their unwavering guidance and encouragement over the past few years have been my bedrock, empowering me to tackle even the most difficult challenges. Their expertise and mentorship have been instrumental in my academic and personal development, and I am forever grateful for their support.

I would also like to express my appreciation to Dr. Sajid Chaudhry for his support in taking on the role of my second supervisor in my final years.

A special thanks to my friend Minh at Yale University. I will always remember his generous support and unconditional help during the times I struggled with certain aspects of my research. His kindness and willingness to assist meant so much to me.

I also wish to express my sincere appreciation to my beloved family, friends, and colleagues. Their unwavering support and encouragement have made my doctoral journey vibrant and enjoyable. Their belief in me has been a constant source of motivation, especially during the challenging times of the COVID-19 pandemic and lockdowns. Without their love and companionship, this journey would have been a bitter and lonely experience.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1. Introduction

1.1. Research Background and Motivation

Most previous research has focused on the one-way effects of the economy on secondary financial markets. These studies assume that insiders (i.e., managers) possess complete information about the firm, leading to the perception that secondary market prices are passive indicators that simply reflect investor's expectations about the present value of the future cash flows. Consequently, the operations of secondary markets are often viewed as either having no impact on the real economy or affecting it only to the extent that "ex-post liquidity affects the firm's cost of capital in the primary markets" (Bond et al., 2012).

However, market prices are a valuable source of information and can themselves have an impact on the real economy. This process involves two main phases, including the stage of producing knowledge and the reactions of managers to the price changes. In the knowledge production phase, secondary market prices aggregate and reveal dispersed information from a wide array of market participants. This information includes investor sentiment, macroeconomic indicators, and firm-specific news, leading to price movements that encapsulate the market's collective expectations and assessments. In the managerial reaction phase, managers closely monitor these price signals, which provide insights into how the market views their firm's prospects. Significant price changes can prompt managers to revise investment plans, alter financing strategies, or modify operational tactics. This managerial decision-making, influenced by market prices, feeds back into the economy, potentially impacting the firm's performance and thus completing the feedback loop between market perceptions and economic realities. By recognizing the active role of secondary markets in this manner, it becomes evident that they are integral to the information ecosystem of the economy, not merely reflecting economic fundamentals but also shaping them through the continuous interplay of information production and managerial response.

According to Roll (1988), one notable ability of the financial market is to generate price information. Financial markets generate information because they actively uncover, aggregate, and disseminate knowledge that was not previously available. This process goes beyond merely depicting existing data; it involves the dynamic interaction of market participants whose trades reflect diverse expectations, preferences, and private information. Through the mechanism of price discovery, markets aggregate individual decisions into a consensus that captures the collective wisdom of all participants. Informed traders, acting on private knowledge, signal their insights through trading activity, which contributes to the formation of asset prices. These prices, in turn, incorporate and reveal information that helps resolve uncertainty about an asset's value (Glosten and Milgrom, 1985; Kyle, 1985). Additionally, financial markets facilitate risk sharing among participants, and the willingness to buy or sell assets at specific prices reveals important information about risk preferences and attitudes. As new events unfold and market conditions change, prices adjust to reflect updated expectations, continuously generating fresh insights and knowledge about the underlying assets. This dynamic and iterative process distinguishes markets as active generators of information rather than passive depictions of existing data.

Since secondary markets do not directly transfer resources to firms, price movements in these markets have real consequences only if they influence the actions of decision-makers in the real economy. Bond et al. (2012) identify two main channels through which financial markets affect real decision-makers: the learning channel and the incentives channel. The idea of the learning channel goes back to Hayek (1945). He notes that the market price is efficient and comprises information aggregated from various sources, then decision-makers in the real world of business, who are unlikely to be fully informed, will wish to learn from the price. In principle, small pieces of scattered information can be aggregated among numerous participants in the markets, these people have no means of communication with the firm’s managers apart from the trading process (Chen et al., 2007). It is generally understood that a firm’s fundamentals, consisting of both published and confidential news, are reflected in the stock price. Subrahmanyam and Titman (1999) elucidate that through day-to-day operations, traders might, incidentally, discover valuable information about the quality of the firms. This type of information typically named private (or personal or confidential) information can find its way to be incorporated into the stock price via outside investors’ trading activities. Therefore, stock prices can reveal private information held by investors that would otherwise be unavailable to the firm's managers, so managers can learn from stock prices to understand the market's perception of the firm. The current literature provides empirical evidence on how private information can guide managers to make more optimal investment decisions and decisions on mergers, acquisitions, and earnings forecasts (Bakke and Whited, 2010; Chen et al., 2007; Foucault and Fresard, 2012; Luo, 2005; Loureiro and Taboada, 2015). The idea for the incentive channel is that secondary markets can have feedback effects through decision-makers’ incentives to take real actions. That is, directors’ or managers’ decisions, instead of learning from price, are motivated by prices as their contracts are tied to them. This impact was noticed and analytically formalized by Baumol (1965) and Fishman and Hagerty (1989). The roleplay by prices in the two channels is subtly different. In the learning channel, the price incentivizes decision-makers to take real actions by revealing to them what are the efficient ones. In detail, the greater the extent to which the stock price reflects the manager’s actions, the greater his incentives to take desirable steps and avoid undesirable ones (Nagar et al., 2003). There are several critical reasons to facilitate the incentive mechanism. Managers care about the stock price in the short-term because of the executive compensation package, insider trading, takeover threats, or reputation. Kang and Liu (2008) directly examine this mechanism and show that CEO compensation hinges on the market price and is positively associated with the informativeness of the price.

Although in the last 30 years, there has been a branch of microstructure literature that explores the mechanisms through which the secondary market affects the real economy, this topic has not yet been adequately studied, possibly because of the conventional wisdom that relies on one-way effect assumption – i.e., the effect of the economy on stock price, or because of the difficulty in directly identifying and measure the effect of the financial markets on the economy. Edmans et al. (2017) suggest that the impact of stock prices on firms' decisions depends not only on the information contained in stock prices, known as forecasting price efficiency (FPE), but also on the proportion of information that is unknown to managers, referred to as revelatory price efficiency (RPE). Because there is not yet a reliable proxy for RPE, it is difficult to study this topic empirically. On the other hand, different stocks may carry different levels of private information due for instance differences in costs regarding private information production (Grossman and Stiglitz, 1980). Since these costs are difficult to measure directly, it is difficult to assess the level of private information embedded in a stock price.

In the microstructure literature, various measures of private information have been developed, primarily based on stock price fluctuations and trading behaviors. Notable examples include stock price non-synchronicity, introduced by Roll (1988), the probability of informed trading (PIN) proposed by Easley et al. (1997a, b; 2002), and the dynamics measure for the probability of informed trading (DPIN) developed by Chang et al. (2014). While these measures have contributed significantly to our understanding of information asymmetry, they each have limitations that hinder their applicability in today's rapidly evolving financial markets.

Firstly, existing measures often rely on simplistic assumptions about trading dynamics and the nature of private information, failing to capture the multifaceted reality of information flow in modern trading environments. For instance, traditional measures may not adequately account for the complexity introduced by high-frequency trading and the instantaneous nature of information dissemination across various channels, including social media and news platforms. As a result, they may overlook crucial aspects of how private information is generated, shared, and ultimately reflected in stock prices.

Furthermore, many of these measures do not incorporate relevant contextual factors such as trade volume and liquidity, which are vital in understanding the broader implications of private information on market behavior. For example, while PIN and DPIN offer insights into the likelihood of informed trading, they may miss nuances regarding how varying levels of trade volume can influence the intensity and impact of information asymmetry on price formation.

To address these shortcomings, the proposed SDPIN measure introduces a novel approach by integrating additional dimensions, particularly trade volume, into the analysis of private information. By capturing the interplay between trading volume and private information, SDPIN provides a more nuanced and dynamic understanding of how information asymmetry operates in the market. This measure not only enhances the granularity of private information assessment but also offers valuable insights into how managerial decision-making is influenced by fluctuations in private information, leading to more informed and optimal strategic choices.

In essence, SDPIN seeks to fill a critical gap in the literature by providing a more comprehensive and context-aware measure of private information that reflects the complexities of current market dynamics. By advancing our understanding of information asymmetry, SDPIN can significantly contribute to the ongoing discourse on how financial markets operate and their consequential impact on real-world decision-making processes.

Once the proportion of investors' private information embedded in stock prices is accurately estimated, its impact on various economic factors, particularly managerial decisions, becomes a critical area of investigation. This feedback from secondary markets to the real economy is essential for understanding the dynamic relationship between financial markets and the real economy. One promising avenue for exploration lies in the intersection between private information and earnings management—a practice where managers manipulate financial statements to present a desired image of the firm. Managers may be influenced by the information embedded in stock prices to adjust reported earnings in ways that align with market expectations or enhance their personal compensation tied to stock performance. By examining how investors' private information, reflected in stock prices, affects managerial behavior and decisions related to earnings management, researchers can gain insights into the broader implications of information asymmetry in the market.

Moreover, the link between private information and market liquidity is another critical aspect that warrants further study. Liquidity, or the ease with which assets can be bought or sold without impacting the price, plays a vital role in the efficiency of financial markets. When investors possess private information, it can lead to greater uncertainty about a stock's true value, which may, in turn, affect market liquidity. If a significant portion of market participants is trading based on private information, it can result in more volatile and less liquid markets, especially if this information is not fully reflected in the prices. Understanding how private information influences liquidity can offer valuable insights into the stability and functionality of financial markets. Researchers can explore whether the dynamics of private information lead to price distortions or if they contribute to the market's ability to absorb shocks and maintain efficient price formation.

Therefore, the motivation to carry out this research stems from the desire to advance our understanding of how private information—shaped by investors' trading behavior—affects critical aspects of corporate governance and market dynamics. By investigating these relationships, scholars can offer more nuanced perspectives on how financial markets influence decision-making processes in firms and the broader economy. In particular, the ability to measure and understand the role of private information can help regulators, managers, and investors make more informed decisions, leading to more efficient markets and improved economic outcomes. This research can also contribute to the development of better market structures and policies that address information asymmetry, ensuring that financial markets operate in a manner that aligns more closely with real-world economic needs and expectations.

1.2. Research Objectives

One primary objective of this research is to establish and refine an accurate estimate of investors' private information embedded in a stock price. I introduce a new measure for investors' private information which addresses some of the limitations underlying existing ones. Notably, conventional proxies such as PIN and DPIN fail to adequately account for the influence of different size orders, should they exist. I contend that it is imperative to develop measures that encapsulate both key aspects of informed trading: the frequency and volume of informed orders. My proposed private information measure, named SDPIN, builds on the private information measure named DPIN. It encompasses the private information potentially revealed through variations in the trade order sizes.

To contribute to this emerging field of research, the secondary objective is to assess the impact of investors' private information, as reflected in stock prices, on various factors within the economy. Consequently, the second empirical study investigates the influence of stock prices on future earnings manipulation, aiming to determine whether corporate managers consider external investors' private information when making decisions regarding corporate earnings. Specifically, it examines the relationship between the levels of private information embedded in stock prices and the likelihood of earnings management practices by firms. This study seeks to understand whether the dissemination of private information can serve as a monitoring mechanism that potentially deters earnings manipulation.

Furthermore, the third study investigates the dynamics of informed trading and its impact on stock liquidity, considering contemporary conditions of stock trading such as high-frequency trading in our current modern market conditions. This investigation focuses on how the presence of informed traders affects market liquidity, bid-ask spreads, and the overall trading environment. Both the second and third studies rely on the findings and methodology established in the first study, utilizing the new measure of private information (SDPIN) as a foundational component.

1.3. Research Findings and Contributions

1.3.1. First Empirical Study

The first study proposes a new investors' private information measure, which I name SDPIN. The methodology to estimate the SDPIN follows that of Chang et al. (2014) regarding the identification of the informed trade, but it considers trade order sizes, parameters that are so far neglected by the aforementioned measures.

I test the validity of the SPDIN measure using an intraday sample from the U.S. market, which comprises information on 236 firms listed on the NYSE and NASDAQ over the period between 2018 and 2020. The findings show a positive and significant correlation between the SDPIN measure and the stock price non-synchronicity, which means that the new measure can confidently be used to determine the amount of private information that is embedded in a stock price. In general, it exhibits a better fit with both the main model and the robustness models than the DPIN of Chang et al. (2014). It also allows us to measure investors' private information for relatively short periods (e.g., a day or a shorter period), a task that is not possible with the existing private information measures.

This research contributes significantly to the financial markets literature by introducing a more adaptable approach to estimating private information in stock prices compared to existing measures such as PIN and DPIN. The method, termed SDPIN, offers greater flexibility by accommodating a wider range of trade order sizes, thus making it applicable to a broader spectrum of stock trade scenarios. One key advantage of the SDPIN measure is its ease of computation, particularly when dealing with high-frequency data. Unlike the PIN measure, which involves complex and time-consuming calculations, SDPIN provides a simpler alternative, offering researchers a more efficient way to estimate private information in stock prices. Moreover, the approach separates the measurement of investors' private information from the estimation of the probability of informed trading orders. This distinction allows us to precisely identify the content of private information embedded in daily stock prices. Consequently, the methodology opens avenues for exploring the feedback effect of stock prices on the real economy, particularly through an analysis of managers' behaviors and decisions. This has the potential to enrich the existing literature in this field and provide valuable insights into the dynamics feedback effects of financial markets on the real economy.

The new private information measure is a useful tool for both financial markets in general and firms' managers, in the sense that, if used, it reduces information asymmetries among market players, enhancing market transparency, and increases managers' awareness of the existence of private information, so they can act according and make more optimal decisions.

1.3.2. Second Empirical Study

The second study examines the impacts of investors' private information embedded in the stock prices on earnings management. I rely on the work provided in Chapter 2 on a new measure of private information (SDPIN), which builds up on the probability of informed trading (DPIN) measure developed by Chang et al. (2014).

Single-level and multi-level regressions are performed to evaluate the aforementioned relationship. The results show that informed trading negatively affects earnings management. When there is more private information embedded in the stock prices, managers are less inclined to engage in earnings management. In addition, when analyzing upward and downward earnings management, I observe that while investors' private information can influence managers' decisions on firms' earnings in both cases, managers who have consistently inflated earnings exhibit a heightened response to private information. Also, tests employing the multi-level approach show that predictors at the industry level play a significant role in determining earnings management, although they are other factors rather than investors' private information at the industry level.

This study offers multiple contributions to the existing literature. Firstly, it presents empirical proof supporting the notion of information transmission from secondary markets to the real economy. The findings lend further credibility to the idea that the secondary market is dynamic and can exert feedback effects on the real economy. Outside investors, although not involved in the day-to-day operations of the business, can still either discover secret information hidden behind the accounting picture drawn by the firm managers or generate valuable information that is also new to insiders. Secondly, it addresses a gap in the current body of research by investigating a fresh external element influencing managers' choices regarding earnings management and corporate disclosure. Moreover, the method differs from past studies by adopting a multi-level strategy rather than solely relying on dummy variables for firm sectors to investigate the impact of industry-level factors on earnings management. This approach allows for a deeper analysis of the precise factors that shape earnings management.

The findings aid finance managers in integrating more precise private information into their decision-making processes. They also pose a challenge to regulators, highlighting the importance of disclosing information related to trading orders that could influence market efficiency overall.

1.3.3. Third Empirical Study

This study investigates the impact of informed trading on stock liquidity in the context of high-frequency trading. It extends the work provided in Chapter 2, using SDPIN as the measure of informed trading to investigate its relationship with liquidity. Especially, various timeframes, including daily, weekly, monthly, quarterly, and yearly, are examined to determine whether the results hold across these different data frequencies.

The outcomes show that informed trading enhances stock liquidity when I use daily and weekly, but that such an effect does not hold when I use monthly, quarterly, and annual. The finding holds relevance within the context of the current market dynamics and aligns with the notion that short-term trading activities, driven by informed investors, contribute to market liquidity by enhancing trading volumes and market depth.

The research findings have several contributions. Firstly, the research contributes to the understanding of the link between liquidity and informed trading with the advent of modern markets. It recognizes that modern markets have fundamentally changed the dynamics of liquidity and informed trading compared to traditional ones. This update is crucial as these modern markets with highly frequent trading have become increasingly dominant. This is significant because it suggests that liquidity measures can serve as a valuable tool for market participants and analysts to infer the presence of informed trading, even though informed trading is not directly observable.

Both factors, liquidity and informed trading are crucial for assessing the efficiency and functionality of a stock market. The paper bridges the gap between theoretical models, empirical research, and real-world market dynamics. It provides practical insights that can be used by traders, investors, and regulators to better understand the relationship between liquidity and informed trading in modern markets. Especially, identifying informed trading helps in correcting market inefficiencies. When prices reflect genuine supply and demand rather than manipulated information, market efficiency improves.

1.4. Thesis Structure

The remainder of this thesis is organized as follows. Chapter 2, the first empirical chapter, introduces a new measure for investors' private information (SDPIN) embedded in stock prices. Chapter 3, the second empirical chapter, examines the effect of investors' private information on firms' earnings management. Chapter 4 presents the third empirical chapter, exploring the impact of informed trading on stock liquidity within the context of modern markets and high-frequency trading. Finally, Chapter 5 concludes the thesis by providing an overview of the findings, contributions, and limitations of the three empirical studies.

Chapter 2. A New Measure for Investors' Private Information

2.1. Introduction

Investors who have private information on a listed firm (informed investors) use it to exploit investment opportunities that are not accessible to other investors. The current literature suggests that the actions of informed investors affect not only the stock price but also the managerial decisions (Bakke and Whited, 2010; Zuo, 2016).[1] Therefore, those actions affect the economic development (Bond et al., 2012). There are studies suggesting that investors' private information may also help managers to make more optimal decisions, for instance regarding mergers and acquisitions (Luo, 2005; Kau et al., 2008), investment projects (Chen et al., 2007; Bakke and Whited, 2010), and earnings forecasting (Loureiro and Taboada, 2015; Zuo, 2016). However, the level of private information varies across stocks due to the asymmetric costs related to the acquisition of new information (Grossman and Stiglitz, 1980; Keiber, 2007; Nezafat and Schroder, 2022).

The stock price is the ultimate measure of the shareholders' wealth. Hence, managers should carefully identify the key drivers of the stock price (Rappaport, 1987). There are numerous reasons for a stock price change, being one of them the existence of private information in the hands of one or few investors. Therefore, although challenging, it is important to measure the level of investors' private information embedded in the stock price. Informed investors use various strategies to hide their trade activities. For instance, they break their trades into small trade orders, using the so-called "smokescreen" trading (Keim and Madhavan, 1995; Chordia and Subrahmanyam, 2004), set multiple accounts and rely on various trade platforms to anonymise trading, using the so-called "dark pools" [2] (Yeoh, 2010; Bayona et al., 2023), and trade on unrelated assets (Hsu, 2018) to misguide investors. Besides these strategies, which per se complicate the task of measuring the level of private information, stock prices are also influenced by the so-called market sentiment, irrational behaviors, and the randomness of the macroeconomic and geopolitical evolutions over time.

The first attempt to measure the level of privative information in a stock price was made by Roll (1988) who introduced the concept of "stock price non-synchronicity" and classified "new information" into three types: firm-specific, industry-specific, and market-wide. He advocates that the stock price non-synchronicity is a good proxy for private information and attributes the actual price shifts in stocks primarily to the existence of firm-specific information. This measure of private

[1] In Figure 1 in Appendix 1, I illustrate the overlap between the information owned by the informed investors and the information owned by the firm's managers.

[2] Dark pools are private trading platforms where large blocks of securities can be traded anonymously, offering investors the possibility of trading stocks outside the exchanges.

information has been used in several studies such as those of Morck et al. (2000), Jin and Myers (2006), Hutton et al. (2009), and Zuo (2016).

However, Roll's measure is not effective when firm-specific information is not the main driver of investors' private information. Roll's measure assumes that stock price non-synchronicity is primarily driven by firm-specific private information. According to this perspective, lower synchronicity reflects a higher degree of firm-specific information embedded in stock prices. However, this assumption may not hold in environments where market-wide factors, such as macroeconomic news or sector-specific trends, dominate investors' private information. In such cases, Roll's measure underestimates the role of private information because it is narrowly focused on firm-specific components and does not account for market-wide insights that also influence trading decisions.

Additionally, non-synchronicity can result from various factors unrelated to private information, such as speculative trading, noise, or liquidity effects. This introduces further limitations to Roll's measure, as it may conflate these elements with the informational content it seeks to capture.

Thus, two new measures were introduced later: the probability of informed trading (PIN) and the dynamic measure for the probability of informed trading (DPIN). The former measure was developed by Easley et al. (2002) and estimates the probability of informed trading based on a sequential trade model, whereas the latter was developed by Chang et al. (2014) and it is a dynamic private information measure particularly useful for high-speed trading scenarios. These measures proved to be more accurate than the Roll's price non-synchronicity measure. Contrary to Roll's measure, the PIN and DPIN measures do not only rely on price movements but also on the characteristics of the trading which often convey information not yet reflected in the stock price. Both of these measures differentiate informed trades from uninformed trades and measure the likelihood that a trade is driven by private information. This paper identifies the limitations of the PIN and DPIN measures and proposes a new private information measure (SDPIN) to remedy those limitations.

Estimating the PIN measure using high-frequency trade data can be notably time-consuming due to several intricate steps involved in the process. Specifically, high-frequency trade data requires a high volume of trades to be processed and classified accurately, which requires the use of sophisticated algorithms, such as the Lee-Ready and the EMO, for trade classification across the stock exchanges. Moreover, the estimation of the likelihood function for a single trading day, as proposed by Lin and Ke (2011) and Yan and Zhang (2012), involves complex computations that consider various factors such as the number of *buy* and *sell* orders, as well as dealing with the floating-point exceptions[3] and the

[3] Floating-point exceptions occur during computations involving floating-point numbers when operations lead to mathematically undefined or unrepresentable results within the constraints of the floating-point number system. These exceptions include dividing by zero, reaching values beyond the numerical range, encountering invalid operations like square roots of negative numbers, and dealing with infinite values.

boundary solutions[4]. Additionally, the parameter estimation over multiple days requires significant computational power and time due to the need to handle a large volume of high-frequency data points, to ensure accuracy and statistical significance. The nature of these meticulous computations, along with the sheer volume of data to be processed and analysed, makes the process of computing the PIN measure a complex and time-consuming task.

The DPIN measure, introduced by Chang et al. (2014), was specifically designed as a dynamic tool for capturing investors' private information, particularly suited for the demands of high-frequency trading environments. Its computational simplicity enables aggregation over short time intervals, such as daily or even intra-daily periods, making it highly adaptable to modern trading dynamics. However, a notable limitation of the DPIN measure is its inability to account for private information revealed through trade size. Previous studies have consistently highlighted that the size of trade orders can be a critical indicator of private information. For instance, Easley and O'Hara (1987) demonstrate that informed traders often initiate larger trade orders, which convey greater informational value compared to smaller trades. On the other hand, research by Barclay and Warner (1993), Chakravarty (2001), and Alexander and Peterson (2007) provides compelling evidence that medium-sized trades account for the majority of cumulative stock price changes. This pattern supports the hypothesis that informed trading predominantly occurs in the medium-size category, with such trades carrying the most significant informational content. Consequently, price movements are largely driven by private information embedded within medium-sized orders, underscoring the need for measures that capture this critical dimension of trading behavior.

The new investors' private information measure (SDPIN) overcomes the aforementioned limitations of the PIN and DPIN measures. It builds on the DPIN measure but also considers the trade order size (volume). The DPIN is an effective private information measure when investors (including informed investors) trade stocks using same-size trade orders. However, empirical evidence shows that that is rarely the case, so I advocate that the SDPIN measure is more effective than the DPIN because it captures the private information that might be revealed through the choice of size of the trade order. The methodology to estimate the SDPIN follows that of Chang et al. (2014) regarding the identification of the informed trade, but it considers large, medium, and small trade order sizes.

I study the relationship between stock return synchronicity and investors' private information using high-frequency data and examine the accuracy of the SDPIN measure compared to that of the

[4] Boundary solutions occur in mathematical models when variables or parameters approach the limits of their allowed range. These scenarios arise when values in a model or optimization problem reach their maximum or minimum permissible values. In statistical estimation or optimization, encountering boundary solutions can pose challenges as the values at these limits might affect the accuracy or reliability of computations, potentially leading to biased or less reliable results.

DPIN. I conclude that it is more accurate particularly when there is algorithmic trading.[5] According to the New York Stock Exchange's rule 127.10, a big trade usually involves over 10,000 shares, being the minimum trade size set at 100 shares. However, what counts as a large and a small order differs across stocks because these are influenced by the average daily stock trading volume and its impact on the stock market, for instance. According to the data sample that comprises information on stocks listed in the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations (NASDAQ) Exchange, trading orders vary from 1 to about 15,000 shares, although some investors trade with larger orders that range from hundreds of thousands to millions of shares. Hence, there is a wide range of trade order sizes chosen by investors that the new privative information measure takes into account.

This research contributes to the microstructure literature literature as follows. Firstly, the SDPIN private information measure is more flexible than the existing (PIN and DPIN) measures, since it accommodates a wider spectrum of trade order sizes and applies to a wider range of stock trade scenarios; a characteristic that makes it more reliable than the PIN and DPIN measures for high-frequency trading environments or when there is algorithmic trading. Secondly, the SDPIN measure is easier to compute than the PIN measure, when dealing with high-frequency data. Thirdly, the SDPIN measure can also be used to better understand the feedback loop between stock prices and the economy, since a more accurate measurement of the private information, enables us to better understand the effect of a stock price and a stock price movement on managers' behaviour. Hence, the higher accuracy of the SDPIN measure in measuring the level of private information embedded in a stock price sheds light on how stock prices are shaped by private information and thereby affect managerial decisions and economic development.

The new SDPIN private information measure can have implications on the behaviour of financial market participants. Specifically, it determines the level of private information embedded in a stock price, so investors can make more informed investment and risk management decisions; it helps all market players, including financial markets regulators, in the sense that it reduces information asymmetry; and it is also helpful for the firms' managers in the sense that it helps them to be more aware of the private information embedded in their firm's stock price, so they can make more optimal decisions. Financial regulators also benefit from this work because they are able to more accurately measure the level of private information embedded in a stock price and, therefore, to adopt the necessary measures to combat that to achieve the ultimate goal of reducing asymmetric information among investors and make financial markets as transparent as possible.

[5] In algorithmic trading, big trades, also called block trades or parent orders, are split into smaller ones, called child orders, to reduce their impact on the stock price and suppress the volatility risk.

The remaining of the paper is as follows. Section 2 provides the relevant literature review and states the research hypotheses. Section 3 presents the data sample and research methodology. Section 4 presents the empirical findings. Section 5 concludes the paper.

2.2. Literature Review

This section reviews key concepts and methodologies related to the measurement of private information in financial markets. It begins by discussing the foundational work of Easley et al. (1997a, 1997b), its development, and limitations in identifying and quantifying private information embedded in stock prices. Next, I review advancements in dynamic measures of informed trading, including the Dynamic Probability of Informed Trading (DPIN) introduced by Chang et al. (2014), and the limitations of DPIN when it comes to accounting for the informational content embedded in trade order sizes. To address this gap, I introduce a new measure, the Dynamic Probability of Informed Trading with Size Effects (SDPIN), which combines both trade frequency and order size to provide a more comprehensive assessment of private information. Finally, the section concludes by presenting the research hypotheses that build on the link between informed trading and stock price non-synchronicity, setting the stage for the empirical analysis.

2.2.1. Probability of Informed Trading

The idea of measuring the private information embedded in a stock price is credited to Easley et al. (1997a,b) who analyze the information content revealed by the choice of stock trade orders of different sizes and conclude that the behaviour of uninformed traders is history-dependent. That is, uninformed traders are more likely to copy the stock trade flow compared to informed traders, so there is an ongoing sequence of stock trading that is not very informative. However, if a given pattern of ongoing flow of stock trade orders is suddenly reversed, this event can be very informative.[6] Easley et al. (1997b) present a model showing how the numbers of *buys*, *sells*, and no-trading intervals can be used to estimate the proportion of stock trades that are driven by investors' private information. Easley et al. (2002) named this model as the probabilities of information-based trading (PIN) and it has been widely used for measuring the private information embedded in a stock price (see also Yan and Zhang, 2012; Chakrabarty et al., 2015; and Poppe et al., 2016). Various studies have scrutinized the Easley et al. (1997b) methodology - see Ellis et al. (2000), Lin and Ke (2011), and Yan and Zhang (2012), for instance.

In order to estimate the EHO PIN, first it is used the Lee-Ready Algorithm (Lee and Ready, 1991) to classify the number of buy orders (B) and the number of sell orders (S) in a single trading day.

[6] For instance, when the demand for a stock is high (bullish trend) and it dropts abruptly.

The likelihood function for a stock's single trading day is given as follows:

$$L(\theta|B,S) = (1-\alpha)e^{-\varepsilon_b}\frac{(\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{(\varepsilon_s)^S}{S!} + \alpha\delta e^{-\varepsilon_b}\frac{(\varepsilon_b)^B}{B!}e^{-(\varepsilon_s+\mu)}\frac{(\varepsilon_s+\mu)^S}{S!} +$$

$$+\alpha(1-\delta)e^{-\varepsilon_b+\mu}\frac{(\varepsilon_b+\mu)^B}{B!}e^{-\varepsilon_s}\frac{(\varepsilon_s)^S}{S!} \qquad (1)$$

where δ is the probability of bad news; ε_b and ε_s are the daily arrival rates of noise traders that submit buy and sell orders, respectively; α is the probability that some traders acquire new (private) information about the firm fundamental; and μ is the arrival rate of informed traders, given information, event occurs. Using trading information over J days and assuming cross-trading-day independence to estimate ε_b, ε_s, α and μ main objective is to maximize the likelihood function:

$$V = L(\theta|B,S) = \prod_{j=1}^{j=J} L(\theta|B_j,S_j) \qquad (2)$$

The PIN is calculated as follows:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b} \qquad (3)$$

The estimation of the PIN measure relies on both algorithms that classify trades into *buy* and *sell* orders and a likelihood function for a stock single trading day to estimate the daily arrival rates of noise and informed trades (Easley et al., 1997b; Easley et al., 2002). Specifically, Easley et al. (1997b) classify trades following Lee and Ready (1991), who develop a procedure that combines both the quote rule and the tick rule[7]; a trade that is above or below the midpoint is classified using the quote rule and a trade that is at a midpoint is classified using the tick rule. They also compare trade prices to the trade quotes that exist in the five seconds before the stock transaction. The algorithm used presents an accuracy between 73% and 91%, depending on the stock exchange (Finucane, 2000; Lee and Radhakrishna, 2000; Odders-White, 2000). Ellis et al. (2000) propose a new classification algorithm for the NASDAQ, the EMO algorithm. It is highlighted that the outside-the-spread problem is specific to NASDAQ and that the tick rule is more accurate than the quote rule for trades that are away from the quotes.

Later on, Easley et al. (2012) introduced a new algorithm that classifies stock trades, namely the bulk volume classification (BVC) algorithm. This method of classification is characterized by high speeds since it focuses on volume in a bar, which is defined as a short period or trading interval in a

[7] The quote rule classifies a transaction based on the midpoint of the bid and the ask. A trade executed at the price lower (higher) than the midpoint is classified as a *sell* (*buy*). The tick rule is based on price movements relative to previous trades. Accordingly, if the price of a transaction is above (below) the previous price, then it is a *buy* (*sell*).

day. For each bar, the proportion of buyer-initiated volume is estimated. It is shown that the BVC algorithm is effective in estimating the order flow toxicity[8]; for instance, it outperforms the tick rule and Lee-Ready's algorithm in some equities, gold, and oil futures markets. Chakrabarty et al. (2015) test the accuracy of the tick rule, Lee and Ready's algorithm, and the BVC in the equities market, and conclude that the first two methods outperform the third one. Pöppe et al. (2016) add that the BVC is not robust to the choice of classification algorithm while the traditional trade-by-trade classification algorithms, such as that of Lee and Ready (1991) and the EMO algorithm exhibit notably high accuracy of up to 90% in most financial markets.

Specifically regarding the PIN measure, Lin and Ke (2011) argue that the likelihood function under use is biased because large *buys/sells* may trigger the power function embedded in the likelihood to generate a numerical value that exceeds the range of real number values that a computer software program can handle. This biased problem is called the floating-point exception and, according to Lin and Ke (2011), it can influence significantly the estimation of the PIN measure. They suggest a reformulation of the likelihood function to mitigate its shortcomings. Yan and Zhang (2012) raise another concern regarding the boundary values (0 and 1) of the variable representing the probability that traders acquire private information on the firm's fundamentals. If the probability receives a value of 0 or 1, it means that no private information event or uninformed trade ever occurred during a given time, which is unlikely. Hence, they suggest the use of a modified factorized likelihood function (the LK factorization) to solve the problem.

2.2.2. The Dynamic Measure for Probability of Informed Trading

Another important limitation of the PIN measure is that it is very challenging to use it with high-frequency data. This problem is exacerbated by the fact that high-frequency trading has become more popular in recent times, and I note that the estimation of informed trading relying on daily or weekly data does not fully capture the information that is possibly associated with stock trades. Following this view, Chang et al. (2014) present a dynamic intraday measure for the probability of informed trading (DPIN). They use a new methodology, following Schwert (1990) and Avramov et al. (2006), to compute the proportion of informed trades; this new measure also has the advantage of being much easier to compute since it does not require, like the PIN measure does, the estimation of a function for the numerical optimization.

The DPIN measure is determined using the Lee-Ready algorithm, to delineate informed and uninformed (herding) trades. Chang et al. (2005) isolate the unexpected components of returns

[8] Order flow toxicity is the measure of a trader's exposure to the risk that a counterparty has private information or other informational advantages (Easley et al., 2012).

(unexpected returns) based on the residual of the autoregressive model developed by Schwert (1990), which was then modified by Jones et al. (1994) and Avramov et al. (2006) as follows: [9]

$$R_{i,j} = \gamma_0 + \sum_{k=1}^{5} \gamma_{1i,k} D_k^{day} + \sum_{k=1}^{26} \gamma_{2i,k} D_k^{Int} + \sum_{1}^{12} \gamma_{3i,k} R_{i,j-k} + \varepsilon_{i,j} \tag{4}$$

where $R_{i,j}$ is the dependent variable and represents the return on stock *i* at the intraday interval *j,* with $j = 1, \ldots, 26$; D_k^{Day} represents the day-of-week dummy variables, from Monday to Friday, and D^{Int} represents dummy variables corresponding to each 15-minute interval over the day *t*; $\varepsilon_{i,j}$ is a measure of the unexpected return.

A *buy* order that generates a negative (positive) unexpected return is classified as an informed (uninformed) trade, whereas a *sell* order with a positive (negative) unexpected return is classified as an informed (uninformed) trade. The DPIN measure is, therefore, constructed, as follows:

$$DPIN_{BASE_{i,j}} = \frac{NB_{i,j}}{NT_{i,j}}(\varepsilon_{i,j} < 0) + \frac{NS_{i,j}}{NT_{i,j}}(\varepsilon_{i,j} > 0) \tag{5}$$

where $NB_{i,j}$ and $NS_{i,j}$ are the number of *buy* and *sell* orders, respectively, and $NT_{i,j}$ is the total number of trades of stock *i* on day *j*.

However, the DPIN measure relies on the number of trades only, therefore, it neglects the information that might be embedded in the investors' choice of trading stocks using different order sizes, so I advocate that it should be improved. Chang et al. (2014) were the first to highlight this problem and propose two new measures, the DPIN_SIZE and DPIN_SMALL, which estimate private information based on either large-size trade orders or small-size trade orders, respectively. However, this is still insufficient since informed and uninformed investors can strategically use small and large trade order sizes to hide informed trade. To remedy this problem, I propose a new dynamic intraday measure with size order effects (SDPIN) that considers both the number and the order size of the informed trades.

2.2.3. SDPIN Estimation

In this subsection, I introduce the new investors' private information measure, which is based on the Chang et al. (2014) private information measure. I proceed as follows: firstly, I use the Lee and Ready (1991) algorithm to classify the orders; secondly, I isolate the unexpected component of the returns (the residuals) from the following regression: [10]

---

[9] Chang et al. (2014) follow Jones et al. (1994) and Avramov et al. (2006) to regress the daily return of individual stocks on its own 12 lags (covering a time period of about 2 weeks) instead of on the 22 lags (covering a time period of about one-month) as in Schwert (1990) model. The DPIN is calculated for each 15-minute interval of the day, which means that its average is computed using 26 DPIN values in a day.

[10] I estimate SDPIN directly for each day of trading, replacing dummy variables corresponding to the particular 15-minute intervals by dummy variables for each trading day in the week and including 22 lags of daily return.

$$R_{i,j} = \gamma_0 + \sum_{k=1}^{5} \gamma_{1i,k} D_k^{day} + \sum_{k=1}^{22} \gamma_{3i,k} R_{i,j-k} + \varepsilon_{i,j} \quad (6)$$

where $R_{i,j}$ is the stock i return on day k (j=1, …, 4), D_k^{Day} is the day-of-week dummy variable (from Monday to Friday).

To identify informed trades, I use the residual $\varepsilon_{i,j}$ as a proxy for unexpected returns and classify buying trades in the presence of negative (positive) unexpected returns as informed (uninformed) trades and selling trades in the presence of positive (negative) unexpected returns as informed (uninformed) trades.

The level of investors' private information is given by the ratio total volume of informed transactions (including both buy-informed and sell-informed trades) over the total trading volume:

$$SDPIN_{i,j} = \frac{\sum V_{i,j}^{B}}{TV_{i,j}} \left(\varepsilon_{i,j} < 0\right) + \frac{\sum V_{i,j}^{S}}{TV_{i,j}} \left(\varepsilon_{i,k} > 0\right) \quad (7)$$

where $V_{i,J}^{B}$ is the volume of buy-informed orders and $V_{i,j}^{S}$ the volume of sell-informed orders; and $TV_{i,j}$ is the total trading volume.

2.2.4. Developing Research Hypotheses

The current microstructure literature acknowledges the link between stock price asynchrony and the investors' private information. Stock return asynchrony arises from firm-specific information - this happens when the stock trade is driven by firm-specific information only. According to Roll (1998), private information explains a given proportion of stock price asynchrony and indicates to what extent changes in the stock return are affected by investors' private information.

Roll (1988), Morck et al. (2000), Durnev et al. (2004), and Zuo (2016) show that there is a positive relationship between stock price non-synchronicity and investors' private information; and Morck et al. (2000) and Fernandes and Ferreira (2008) advocate that investors actively seek for not yet public firm-specific information to obtain abnormal returns – this happens when some investors have access to not yet public information about the firm and take advantage from this private information. This privileged status of some (informed) investors affects the stock price and increases the stock return non-synchronicity.

The above literature assumes that there is a positive relationship between investors' private information and the stock price non-synchronicity.[11] Therefore, I set the following research hypotheses:

[11] In Table 1 in Appendix 2, I summarize the results from previous studies on the relationship between stock price non-synchronicity and multiple other factors, including investors' private information, the idiosyncratic volatility, firm size, stock trading volume, Amihud (2002) illiquidity measure, and stock return.

Hypothesis 1 (H1): *Dynamic Probability of Informed Trading (DPIN) is positively related to stock price non-synchronicity.*

Hypothesis 2 (H2): *Dynamic Probability of Informed Trading with Size Effect (SDPIN) is positively related to stock price non-synchronicity.*

2.3. Data Sample and Methodology

This section is designed to test the hypotheses concerning the relationship between DPIN, SDPIN, and stock price non-synchronicity (SYNCH). First, the section outlines the rationale behind the choice of data and its relevance to examining the connection between private information measures (DPIN and SDPIN) and SYNCH. The description includes the sample characteristics, data sources, and criteria for inclusion, ensuring the robustness of the dataset. Next, it provides a detailed discussion of the models used to test the hypotheses. This includes the construction of the regression framework, the incorporation of control variables, and the rationale for selecting specific estimation techniques.

2.3.1. Data Sample

The data sample comprises daily and intraday information on energy sector firms listed on the NYSE or the NASDAQ, covering the period between January 2018 and December 2020. Due to a lack of computational power,[12] I could not use a data sample comprising information on all the U.S. public firms. Hence, I decide to choose one industry sector to sector to focus on. The energy sector seemed to us as a good choice for the research because energy prices are often regulated and subsidized, so there is technological uncertainty since new and more efficient renewable energy technologies are being released and adopted and there is also regulation policy uncertainty (with new renewable energy policies being introduced and often suddenly stopped), so this is the ideal market environment for the existence of private information. Moreover, as the recent Russia-Ukraine war shows, it is a sector that is more prone to be affected by geopolitical tensions among countries (Elder and Serletis, 2010; Yazdi et al., 2022).

The data about firms was collected from "Bloomberg" and "Wharton Research Data Services" (WRDS), whereas the intraday trading data on the stock prices was collected from the "Trade and Quote" (TAQ). The initial data sample comprises information on 236 stocks; I note that for a stock to be included in the sample, the firm has to be listed on the exchanges for at least a year. Overall, the data sample consists of 236 firms and comprises 154,797 observations. Table 2.1 provides further details on the data sample.

---

[12] "Lack of computational power" refers to the insufficient capability of our computing resources to handle and process large volumes of data efficiently. In the context of our research, analyzing the intraday data for entire dataset of all U.S. public firms would require significant processing speed, memory, and storage, which our current computational setup cannot support. Handling such extensive data involves running complex algorithms and computations that demand high-performance hardware and considerable time, both of which are beyond our present capacity. Consequently, I decided to narrow our focus to a specific industry sector, choosing the energy sector due to its unique characteristics and relevance, to ensure that our analysis is both feasible and effective within our computational constraints.

Table 2.1: This table presents the number of stocks of the sample listed in the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations (NASDAQ) per energy industry sub-sectors, following the global industry classification standard (GICS).

Energy sub-sectors		NYSE	NASDAQ	Total
Energy Equipment & Services	Oil & gas drilling	6	3	9
	Oil & gas equipment & services	17	8	25
Oil, Gas & Consumable Fuels	Integrated Oil & Gas	19	5	24
	Oil & gas exploration & production	89	28	117
	Oil & gas refining & marketing	9	1	10
	Oil & gas storage & transportation	32	9	41
	Coal & consumable fuels	7	3	10
	Total	179	57	236

In Table 2.2, I present the descriptive statistics on the variables used in the regression models; in panel A, the descriptive statistics are provided per year, whereas in panel B these are provided for the whole sample period; panel C shows the number of observations per year.

Specifically, panel A reveals that the mean value of the SDPIN is lower than that of the DPIN for all years; the mean value of the SYNCH consistently decreased from 2018 to 2020, whereas the mean values of both the DPIN and the SDPIN increased. For instance, between 2019 and 2020, the DPIN and the SDPIN increased by 62.56% and 107.06% respectively, and the SYNCH decreased by 13%. This drop of 13% indicates that the stock prices have a higher level of co-movement with the market in 2020 than in 2019. This could be attributed to the effect of the COVID-19 crisis that adversely impacted the whole global market and the energy industry was among the most severely affected.[13] Panel B provides further empirical evidence of the effects on the U.S. market of the COVID-19 crisis; there is a significant decrease in market liquidity in 2020, compared to 2019. Specifically, the mean value of the illiquidity variables IVOL and ILLIQ, increased by 85.49 % and 75.44% respectively; the trade volume (VOL) increased by 56.48%, and market capitalization (SIZE) decreased by 31.14%. All these results are in alignment with those from Chung and Chuwonganant (2023).

Table 2.2: This table presents the statistical descriptions of regression variables: in panel A are the descriptive statistics per year, in panel B are the descriptive statistics for the sample period (2018-20), and in Panel C are the number of the sample observations per year. SYNCH is the proxy for stock price non-synchronicity. DPIN and SDPIN are the measures for investors' private information, calculated according to Eq.(6) and Eq.(8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1.

Panel A	Mean			Min			Max		
Variable	*2018*	*2019*	*2020*	*2018*	*2019*	*2020*	*2018*	*2019*	*2020*
SYNCH	2.591	1.893	1.647	0.321	-0.174	-0.807	13.816	13.816	11.870
DPIN	0.167	0.175	0.317	0.000	0.000	0.000	1.000	1.000	0.910
SDPIN	0.188	0.184	0.381	0.000	0.000	0.000	1.000	1.000	0.996

[13] For instance, during the Covid-19 period, on the 20 April, 2020 expiration date, the price of oil contracts with a delivery date of May 2020 dropped to negative values -$37.63/barrel (Corbet et al., 2021).

IVOL	0.024	0.026	0.049	0.008	0.002	0.008	0.072	0.103	0.156
SIZE	2,540	2,052	1,413	3.220	2.950	2.770	31,874	32,108	35,522
SPREAD	0.844	0.658	0.715	0.000	0.000	0.000	14.700	27.300	25.900
VOL	0.760	0.765	1.202	0.000	0.000	0.000	30.300	27.900	146.500
ILLIQ	0.035	0.073	0.128	0.000	0.000	0.000	15.700	63.400	62.200
RETURN	-0.001	0.000	0.001	-0.412	-0.295	-0.568	0.471	1.005	1.604

Panel B

Variable	Mean	Min	Max	St Dev	Skewness	Kurtosis
SYNCH	2.032	-0.807	13.820	1.355	3.115	1.182
DPIN	0.221	0.000	1.000	0.171	0.422	-0.627
SDPIN	0.253	0.000	1.000	0.183	0.380	0.512
IVOL	0.033	0.002	0.156	0.021	2.074	5.332
SIZE	1,989	2.770	35,522	4,330	1.427	2.218
SPREAD	0.737	0.000	27.270	0.970	4.245	3.541
VOL	0.914	0.000	146.500	2.000	14.672	40.009
ILLIQ	0.080	0.000	63.420	0.630	16.540	12.748
RETURN	0.000	-0.568	1.604	0.049	2.769	16.352

Panel C

Number Obs. per year	2018	2019	2020
	49,435	51,617	53,745

2.3.2. Regression Models

2.3.2.1. Main model

I use the following regression model (Model 1) to estimate the relationship between the stock return synchronicity and the PIN measure of Roll (1988), the DPIN measure of Chang et al. (2014), and the SDPIN measure. The dependent variable, $SYNCH_{i,t}$, is a proxy for the stock price non-synchronicity; a higher $SYNCH_{i,t}$ means that the stock price movements are less synchronous with the market movements.

$$SYNCH_{i,t} = \beta_0 + \beta_1 PI_{i,t}^{DPIN} + \beta_2 PI_{i,t}^{SDPIN} + \beta_3 CONTROL_{i,t} + \beta_4 CONTROL_{i,t-1} + \varepsilon_{it} \quad (8)$$

where $i = 1, \ldots, n$ and $t = 1, \ldots, m$, with i denoting the stock and *t* the day; $PI_{i,t}^{DPIN}$ represents the investors' private information measured by the DPIN, and $PI_{i,t}^{SDPIN}$ represents the investors' private information measured by the SDPIN; the CONTROL variable accounts for the control variables, a vector that includes the idiosyncratic volatility (IVOL) measured by the three-factor model Fama-French (1993), firm size (SIZE) given by market capitalization divided by 10^6, volume (VOL) given by the stock trade volume divided by 10^6, the bid-ask spread (SPREAD) given by the difference between the highest ask price and the lowest bid price, illiquidity (ILLIQ) measured through Amihud (2002) illiquidity measure, and the stock return at day (RETURN) given by the difference between the closing price at day *t* and the closing price at day *t-1* divided by the closing price at day *t-1*. Finally, to account for lag effects, the level-one lag of each control variable is also added to the vector of controls (*CONTROL*$_{i,t-1}$*);* and ε_{it} is a zero-mean residual.[14]

Model (1) uses three measures for stock illiquidity: the idiosyncratic volatility (IVOL), the bid-ask spread (SPREAD), and the Amihud's (2002) illiquidity measure (ILLIQ); the bid-ask spread variable measures the market liquidity. Transparency and high-quality information disclosure enhance stock liquidity, so I conjecture that it is positively related to SYNCH. I use daily volume (VOL) and daily return (RETURN) as control variables, following Chang et al. (2014).

2.3.2.2 The estimation of Stock Price Non-Synchronicity

As stated previously, Roll (1998) decomposes the variation of a stock return into three different sources: firm-related, industry-related, and market-related. Following their methodology, a stock co-movement/synchronicity is measured by the coefficient of determination (R^2) of the following regression:

$$IR_{i,j,t} = \alpha_{i,0} + \alpha_{i,m} MR_{m,t} + \alpha_{i,j} SR_{j,t} + \varepsilon_{i,t} \tag{9}$$

where $IR_{i,j,t}$ is the return of stock of a firm *i* that operates in sector *j*; $MR_{m,t}$ is the market return, and $SR_{j,t}$ is the return of the industry *j*; the stock price synchronicity is given by the R^2, so 1-R^2 measures the stock price non-synchronicity.

Morck et al. (2000) argue that it is difficult to distinguish stock price movements driven by changes in the industry from stock price movements driven by changes in the economy as a whole. Additionally, the industry returns are often driven disproportionally by a small group of firms, so the stock price movements may not be a good representative of the overall industry. Therefore, adding the industry return to the model (Eq. (9)) can yield spurious results. Finally, as R^2 values are bounded within the unit interval [0, 1], it might not serve as an appropriate dependent variable. Consequently, instead of using the original non-synchronicity Roll (1988) model (Model 2 above), I use an adjusted model proposed by Morck et al. (2000) according to which I determine first the coefficient of determination (R^2) of the following regression:

$$IR_{jt} = \alpha_{i0} + \beta_{im} MR_{mt} + \varepsilon_{it} \tag{10}$$

where $IR_{i,t}$ is the return of stock *i*; $MR_{m,t}$ is the market return; ε_{it} is the error term and, then, I use the *SYNCH* variable used in Model 1 as the stock price non-synchronicity measure (or the inverse measure of price synchronicity), which is estimated as follows:

$$SYNCH_{it} = Log\frac{1-R^2}{R^2} \tag{11}$$

[14] Appendix C.3 describes the types of data that need to be collected for each variable.

where R^2 is the stock price synchronicity.

2.4. Empirical Results

2.4.1. Descriptive Analysis

Table 2.3 presents the correlation coefficient matrix; in general, it shows that the correlations are below 0.5, the exceptions being for the following pairs of regression-independent variables: IVOL vs. LagIVOL, SIZE vs. LagSIZE, SPREAD vs. LagSPREAD, and VOL vs. LagVOL. Regarding these variables, there is a clear indication of a multicollinearity problem. Therefore, I use the "variance inflation factors" (VIF) measure to determine the level of collinearity between the regressors of the model(s), and the results are provided in Table 2.4.[15] Within the correlation matrix, the two measures for private information DPIN and SDPIN are strongly correlated. However, the SDPIN measure has a slightly stronger correlation with SYNCH than it has with the DPIN measure, which favors the use of the SDPIN measure.

[15] The VIF measures the magnitude of the variance of the coefficient estimations of the regressors that have been inflated due to collinearity with the other regressors. In Table 2.4, I provide the VIFs results and conclude that there is a multicollinearity problem with two pairs of variables: SIZE vs. LagSIZE (285.03 and 285.03) and IVOL vs. LagIVOL (182.82 and 181.85). Consequently, the LagSIZE and LagIVOL independent variables were removed from the regression model.

Table 2.3: This table presents the correlation coefficients between the different pairs of the regression variables. SYNCH is the proxy for stock price non-synchronicity, measured by Equation (4). DPIN and SDPIN are the measures for investors' private information, calculated according to Equations (6) and (8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest ask price and the lowest bid price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

Variables	SYNCH	DPIN	SDPIN	IVOL	SIZE	SPREAD	VOL	ILLIQ	RETURN	LagDPIN	LagSDPIN	LagIVOL	LagSIZE	LagSPREAD	LagVOL	LagILLIQ	LagRETURN
SYNCH	1																
DPIN	0.116	1															
SDPIN	0.148	0.896	1														
IVOL	-0.168	-0.397	-0.369	1													
SIZE	-0.261	-0.205	-0.208	-0.275	1												
SPREAD	-0.203	-0.011	-0.006	0.190	0.359	1											
VOL	-0.135	0.125	0.089	0.154	0.138	-0.042	1										
ILLIQ	-0.164	-0.105	-0.115	0.085	-0.052	0.058	-0.042	1									
RETURN	-0.004	0.010	0.013	0.046	-0.004	0.021	0.137	0.009	1								
LagDPIN	0.115	0.408	0.384	0.400	-0.205	0.050	0.111	0.094	0.022	1							
LagSDPIN	0.147	0.384	0.375	0.372	-0.208	0.058	0.080	0.093	0.026	0.896	1						
LagIVOL	-0.169	-0.393	-0.365	0.997	0.274	0.194	0.147	0.084	-0.035	-0.398	-0.370	1					
LagSIZE	-0.260	-0.205	-0.208	-0.275	0.998	-0.359	0.137	-0.052	-0.006	-0.205	-0.208	-0.275	1				
LagSPREAD	-0.202	-0.043	-0.050	0.186	0.352	0.757	0.016	0.053	-0.004	-0.018	-0.001	0.188	0.353	1			
LagVOL	-0.134	0.109	0.080	0.155	0.137	-0.014	0.737	-0.041	0.001	0.128	0.091	0.153	0.137	-0.043	1		
LagILLIQ	-0.167	-0.089	-0.095	0.087	0.052	0.051	0.039	0.190	-0.009	-0.108	-0.119	0.086	0.052	0.056	0.042	1	
LagRETURN	-0.004	0.015	0.015	0.051	0.001	0.012	0.092	-0.007	-0.013	0.009	0.011	0.050	-0.004	0.018	0.137	0.008	1

Table 2.4: This table presents the variance inflation factors (VIF) of all the regression variables. Panel A shows the VIF and the 1/VIF before the removal of the LagSIZE and LagIVOL variables (in bolt below) and Panel B shows the VIF and the 1/VIF after the LagSIZE and LagIVOL variables are removed from the regression equation. SYNCH is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according to Eq.(6) and Eq.(8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

	Panel A		**Panel B**	
Variables	VIF	1/VIF	VIF	1/VIF
SIZE	285.030	0.004	1.330	0.755
LagSIZE	285.030	0.004	-	-
IVOL	182.820	0.005	1.410	0.710
LagIVOL	181.850	0.005	-	-
SPREAD	2.490	0.401	2.490	0.402
LagSPREAD	2.480	0.404	2.470	0.404
VOL	2.350	0.426	2.330	0.428
LagVOL	2.320	0.431	2.320	0.432
DPIN	1.360	0.737	1.360	0.737
LagDPIN	1.370	0.732	1.370	0.732
RETURN	1.070	0.938	1.040	0.957
LagRETURN	1.030	0.970	1.020	0.979
ILLIQ	1.050	0.949	1.050	0.949
LagILLIQ	1.050	0.949	1.050	0.949

2.4.2. Main Findings

Tables 2.5 and 2.6 present the results for the main model (Model 1); these show the relationship between the stock return variations (measured by SYNCH) and the investors' private information (measured by the DPIN and SDPIN measures). Specifically, while Table 2.5 shows the results for the entire sample, Table 2.6 shows the results per year (2018-20).

From Table 2.5, it is evident that the SDPIN measure performs slightly better than the DPIN measure across all years. The SDPIN model has a higher R^2 of 0.144, compared to 0.139 for the DPIN model, suggesting that SDPIN explains a marginally greater proportion of the variation in stock return synchronicity. All the regression coefficients are statistically significant at the 1% level, for both the DPIN and the SDPIN measures, except for those of RETURN and LagRETURN. It means that the investors' private information reduces the synchronicity of the stock price movements with market movements. This finding corroborates the research hypotheses *H1* and *H2* and is also in line with the findings of previous literature (Roll, 1988; Durnev et al., 2004).

Also, from the correlation matrix in Table 2.3, we observe that the correlation between SDPIN and SYNCH is higher than the correlation between DPIN and SYNCH, suggesting that the SDPIN measure is more strongly associated with stock price synchronicity (SYNCH) than the DPIN measure. In other words, SDPIN may be a more effective and reliable indicator of how investors' private information impacts stock price movements relative to market movements. This implies that SDPIN captures private information in a manner that is more closely linked to the degree of stock price synchronicity compared to DPIN. Consequently, SDPIN could offer a better representation of how private information influences the synchronization of stock prices with the broader market. These results further reinforce the idea that SDPIN is potentially a stronger model for measuring private information. Based on this analysis, we can confidently conclude that the SDPIN measure is a more reliable indicator of investors' private information and slightly outperforms the DPIN measure in this data sample. This enhanced performance highlights the robustness of SDPIN as the preferred model for capturing private information in financial markets.

The R^2 of financial markets of developed countries tend to be low (Morck et al., 2000; Jin and Myers, 2006), meaning that the movements of the stock prices are mainly due to firm-specific information (including private information). The findings show that the stock price non-synchronicity and the investors' private information, although distinct (private information) measures, lead to very similar results. I use the same control variables in the regression models for DPIN and SDPIN and conclude that the coefficient of the SDPIN is 0.429 (with a standard error of 0.033) and the coefficient for the SYNCH of DPIN is 0.254 (with a standard error of 0.036). These findings are also favourable to the SDPIN measure since they show that it has a stronger (positive) relationship to SYNCH than the DPIN has.

In Table 2.6, I show the results per year for the DPIN and the SDPIN regression models. Again, the data demonstrates that SDPIN outperforms DPIN as a measure of private information in several ways. First, SDPIN exhibits slightly higher R-squared values in 2019 (0.356 vs. 0.348) and 2020 (0.209 vs. 0.206), indicating that models using SDPIN explain more variation in the dependent variable, thereby offering better predictive reliability over time. Moreover, the coefficients for SDPIN are consistently positive and statistically significant across all years, with values of 0.083 (2018), 0.796 (2019), and 0.509 (2020). In contrast, DPIN shows insignificant effects in 2018 ($\beta=-0.066$) and only becomes significant in 2019 and 2020. These results highlight SDPIN's ability to consistently capture private information effects, even in earlier periods where DPIN struggles. Additionally, lagged coefficients for SDPIN are significant across all years, emphasizing its persistence and stronger predictive power, while DPIN's lagged effects are insignificant in 2018. F-test values for SDPIN models are also slightly higher in 2019 and 2020, suggesting greater model robustness and better handling of market variations. Furthermore, control variables like IVOL, SIZE, SPREAD, and ILLIQ maintain significance and expected signs across both measures, but SDPIN retains these relationships more consistently across years, reinforcing its reliability.

Overall, SDPIN's stronger performance can be attributed to its enhanced formulation, which incorporates trade volume alongside order imbalances. This additional sensitivity allows SDPIN to capture more nuanced market dynamics, making it a superior measure of private information, particularly in diverse and active trading environments.

The coefficient of the SDPIN is statistically significant and positively related to SYNCH, for all the years. The relation between the DPIN and SYNCH in 2018 is negative but insignificant. The R^2 and the coefficient of DPIN and SDPIN was lower in 2020 than in 2018 and 2019. This might be because of the negative impact on the US of the Covid-19 crisis. Furthermore, the three liquidity measures (ILLIQ, SPREAD, and IVOL) as well as their Lags (LagILLIQ, LagSPREAD, and Lag IVOL) are negative and statistically significant at the 1% level, as expected. These findings are in line with those of Morck et al. (2000) and Zuo (2016) who show that the stocks with low R^2 are those with the least trade and with the greatest impediments to informed trades. According to Morck et al. (2000), markets with low R^2 stocks mean that they provide investors with a more efficient information environment. For instance, stronger investor protection helps improve market liquidity, leading to a higher level of firm-specific information embedded in the stock price and, therefore, lower stock return synchronicity with the market (Morck et al., 2000).

Moreover, the coefficient of SIZE is negative and statistically significant at the 1% level. According to Chan and Hameed (2006), when the number of stocks within an index is small, a few large companies dominate the market movements. Consequently, if R^2 estimation is based on the value-weighted index, it is expected to have a positive relationship between the market capitalization of assets with stock return synchronicity (or a negative relationship with SYNCH). Previous literature shows that larger companies with high trading volumes of shares tend to attract a larger number of analysts (see Alford and Berger, 1999). Furthermore, Bhushan (1989) argues that the supply of analyst services is also affected by the correlation between the stock return and the market return. For a given level of information acquisition cost related to macro variables, a higher correlation (higher stock return synchronicity) leads to a lower information acquisition marginal cost, and this enhances the supply of analyst services.

Notice that the level of a trading volume affects the stock return synchronicity because it influences the speed of price adjustments. I find that the coefficients for VOL and the LagVOL are negative and statistically significant at a 1% level. This is because stocks traded very frequently react to market information on a timely basis, so their individual price movements are more synchronous with market movement, whereas infrequently traded stocks experience a greater delay in their price reactions which results in a lower stock return synchronicity.

Table 2.5: This table presents results for the main regression model. In Panel A are the regression coefficients (Coef.) and the standard errors (SE) for both the DPIN and the SDPIN measures, whereas in Panel B are the R-squared and F-test for the DPIN and SDPIN measures. SYNCH is the proxy for stock price non-synchronicity and is a dependent variable and is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according to the Eq.(6) and Eq.(8),

respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

Panel A	DPIN	SDPIN
Variables	*Coef.* *(SE)*	*Coef.* *(SE)*
R-square	0.139	0.144
F-Test	823.96***	856.8***
Cons.	2.022***	1.987***
	(0.013)	(0.015)
DPIN	0.254***	-
	(0.036)	-
SDPIN	-	0.429***
	-	(0.033)
VOL	-4.33E-08***	-4.35E-08***
	(3.62E-09)	(3.61E-09)
ILLIQ	-0.235***	-0.228***
	(0.009)	(0.008)
IVOL	-4.880***	-3.968***
	(0.293)	(0.288)
SIZE	-5.00E-05***	-4.00E-05***
	(1.40E-06)	(1.39E-06)
SPREAD	-0.095***	-0.097***
	(0.009)	(0.009)
RETURN	-0.004	-0.014
	(0.115)	(0.114)
LagDPIN	0.251***	-
	(0.036)	-
LagSDPIN	-	0.427***
	-	(0.033)
LagSPREAD	-0.096***	-0.100***
	(0.009)	(0.009)
LagVOL	-4.15E-08***	-4.20E-08***
	(3.58E-09)	(3.57E-09)
LagILLIQ	-0.239***	-0.231***
	(0.009)	(0.009)
LagRETURN	0.253**	0.260**
	(0.109)	(0.109)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

The coefficients for RETURN and LagRETURN have opposite signs when regressed on SYNCH. However, the coefficients for RETURN are insignificant in the models for both DPIN and PIN. In contrast, Table 2.6 indicates that in 2019, the coefficients for RETURN and LagRETURN are both positive and statistically significant. These findings align with the results of Chan and Chan (2014).

Table 2.6: This table presents results for the main regression model across the years of the data sample time period: 2018, 2019, and 2020. SYNCH is the proxy for stock price non-synchronicity and is a dependent variable and is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according to the Eq.(6) and Eq.(8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided

by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

Variable	DPIN			SDPIN		
	2018	2019	2020	2018	2019	2020
R-Square	0.446	0.348	0.206	0.446	0.356	0.209
F-Test	1,236.07***	832.74***	405.16***	1236.57***	863.17***	413.26***
_cons	1.572***	0.825***	1.204***	1.552***	0.812***	1.223***
	(0.023)	(0.023)	(0.022)	(0.023)	(0.022)	(0.021)
DPIN	-0.066	0.658***	0.501***	-	-	-
	(0.052)	(0.055)	(0.052)	-	-	-
SDPIN	-	-	-	0.083*	0.796***	0.509***
	-	-	-	(0.050)	(0.051)	(0.045)
IVOL	-59.860***	-38.260***	-6.710***	-58.780***	-37.090***	-6.640***
	0.786	0.732	0.320	(0.782)	(0.721)	(0.316)
SIZE	-1.23E-05***	-3.1E-05***	-1.27E-05***	-1.09E-05***	-2.94E-05***	-1.2E-05***
	(1.93E-06)	(2.12E-06)	(2.24E-06)	(1.93E-06)	(2.11E-06)	(2.23E-06)
SPREAD	-0.123***	-0.006	-0.142***	-0.125***	-0.005	-0.142***
	(0.013)	(0.014)	(0.012)	(0.013)	(0.014)	(0.012)
VOL	-0.148***	-0.102***	-0.015***	-0.150***	-0.100***	-0.015***
	(0.011)	(0.010)	(0.003)	(0.011)	(0.010)	(0.003)
ILLIQ	-0.912***	-0.223***	-0.203***	-0.889***	-0.220***	-0.200***
	(0.053)	(0.012)	(0.010)	(0.053)	(0.011)	(0.010)
RETURN	0.299	0.517**	-0.071	0.342	0.515**	-0.088
	(0.257)	(0.240)	(0.114)	(0.257)	(0.239)	(0.113)
LagDPIN	-0.053	0.619***	0.490***	0.080*	0.774***	0.496***
	(0.052)	(0.055)	(0.052)	(0.050)	(0.051)	(0.044)
LagSPREAD	-0.111***	-0.008	-0.120***	-0.113***	-0.011	-0.120***
	(0.013)	(0.014)	(0.011)	(0.013)	(0.014)	(0.011)
LagVOL	-0.148***	-0.102***	-0.015***	-0.150***	-0.101***	-0.014***
	(0.011)	(0.010)	(0.003)	(0.011)	(0.010)	(0.003)
LagILLIQ	-0.365***	-0.236***	-0.199***	-0.358***	-0.233***	-0.194***
	(0.035)	(0.011)	(0.010)	(0.035)	(0.011)	(0.010)
LagRETURN	0.246	0.711***	0.016	0.311	0.709***	0.012
	(0.256)	(0.240)	(0.106)	(0.256)	(0.238)	(0.106)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

2.4.3. Robustness Tests

For the robustness tests, as in Chang et al. (2014), I follow the Fama and MacBeth (1973) framework[16] and apply cross-sectional regressions to obtain estimates for the regression parameters/coefficients, after which I use the time-series average across all days to arrive at parameter estimates. Specifically, I perform two robustness tests based on the same data sample used in Model 1 and the first-order differentiated data (the difference in value between data for the current day and the preceding day) for the sensitivity test. The two Fama–MacBeth models are specified as follows:

[16] The Fama and MacBeth (1973) model is a widely used approach in finance for estimating the relationship between asset returns and their underlying risk factors. It is a two-stage procedure that helps analyze how different variables or factors influence asset returns. The first step involves estimation of *N* cross-sectional regressions, and the second step involves *T* time-series averages of the coefficients of the *N*-cross-sectional regressions.

Fama–MacBeth Robust Test 1:

$$SYNCH_{i,t} = \beta_0 + \beta_{1,t}PI_{i,t} + \beta_{2,t}PI_{i,t-1} + \beta_3 CONTROL_{i,t} + \beta_4 CONTROL_{i,t-1} + \varepsilon_{i,t} \quad (12)$$

Fama–MacBeth Robust Test 2:

$$\Delta SYNCH_{i,t} = \beta_0 + \beta_{1,t}\Delta PI_{i,t} + \beta_{2,t}\Delta PI_{i,t-1} + \beta_3 \Delta CONTROL_{i,t} + \beta_4 \Delta CONTROL_{i,t-1} + \varepsilon_{i,t} \quad (13)$$

where, Δ represents the difference (change) in the value of the corresponding variable in relation to the previous day.

The variables in the first robust test are the same as those in the original Model 1 for the independent and the control variables. As in Chang et al. (2014), I include contemporaneous and lagged returns (RET) but not their daily difference. The difference in firm size (SIZE) and change for its lag are not included in Test 2 as these are highly correlated with the stock return.

In Tables 2.7 and 2.8, I present the results for the two Fama–MacBeth robustness tests. Within the first robustness test, the parameters are estimated from the time-series average of cross-sectional regressions. I conclude that the coefficients of the DPIN and SDPIN and the LagDPIN and LagSDPIN are positive and statistically significant at the 5% and 10% level, respectively; this corroborates the idea that the SPDIN is a reliable measure for investors' private information. Regarding the control variables, IVOL, SIZE, and SPREAD are significant at 5%, 10%, and 10% levels, respectively, for both the DPIN and SDPIN robustness models.

For the second robustness check, the measures for the daily change of DPIN (ΔDPIN) and of SDPIN (ΔSDPIN) show that there is a negative relation to the daily change of SYNCH (ΔSYNCH), but these coefficients are insignificant. The coefficient of ΔIVOL is negative and significant at the level of 10%.

Table 2.7: This table presents the results for the robustness test 1. SYNCH is the proxy for stock price non-synchronicity and is a dependent variable; it is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according to the Eq.(6) and Eq.(8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

Variables	**DPIN** Coef. (SE)	**SDPIN** Coef. (SE)
DPIN	0.794*	-
	(0.418)	-
SDPIN	-	0.739**
	-	(0.324)
IVOL	-152.110**	-126.050**
	(72.339)	(64.557)
SIZE	-3.01E-06**	-2.87E-06**
	(1.44-06)	(1.29E-06)
VOL	-1.05E-07	-1.05E-07
	(2.81E-07)	(2.46E-07)
ILLIQ	1.933	1.911
	(1.636)	(1.704)
SPREAD	-0.070**	-0.069**
	(0.034)	(0.028)
RETURN	0.612	0.649
	(4.118)	(4.723)
LagDPIN	0.249*	-
	(0.135)	-
LagSDPIN	-	0.186*
	-	(0.099)
LagVOL	-9.26E-08	-9.05E-08
	(2.15E-07)	(2.46E-07)
LagILLIQ	1.847	1.822
	(1.750)	(1.696)
LagSPREAD	-0.112	-0.114
	(0.291)	(0.284)
LagRETURN	0.733	0.776
	(2.068)	(4.255)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

Table 2.8: This table presents the results of the robustness test 2. SYNCH is the proxy for stock price non-synchronicity and is a dependent variable; it is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according to Eq.(6) and Eq.(8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day t compared to day t-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

Variables	**DPIN Model**	**SDPIN Model**
	Coef. (SE)	Coef. (SE)
ΔDPIN	-0.002	-
	(0.069)	-
ΔSDPIN	-	-0.005
	-	(0.011)
ΔIVOL	-62.080*	-62.460*
	(35.227)	(33.048)
ΔVOL	1.79E-09	1.72E-09*
	(2.05E-09)	(1.24E-09)
ΔILLIQ	-0.008	-0.009
	(0.035)	(0.047)
ΔSPREAD	2.00E-04	5.00E-04
	(2.74E-04)	(4.13E-04)
RETURN	-0.012	-0.015
	(0.056)	(0.059)
ΔLagDPIN	-0.002	-
	(0.094)	-
ΔLagSDPIN	-	-0.006
	-	(0.039)
ΔLagVOL	0.14E-09	1.40E-09
	(0.35E-09)	(1.19E-09)
ΔLagILLIQ	-0.015	-0.015
	(0.017)	(0.008)
ΔLagSPREAD	5.01E-04	6.00E-04
	(0.029)	(0.001)
LagRETURN	-0.030	-0.037
	(0.026)	(0.061)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

2.5. Conclusion

There are currently two main measures for investors' private information, the probability of informed trading (PIN) developed by Roll (1998) and the high-frequency measure for the probability of informed trading (DPIN) proposed by Chang et al. (2014). This study presents a new investors' private information measure, which I name SDPIN. It builds on the DPIN measure of Chang et al. (2014) but considers the size of the stock trade orders (trade volume), a parameter that is so far neglected by the aforementioned measures. I advocate that private information can be revealed through the size of the trade orders.

I test the accuracy of the SPDIN measure and compare it to those of the PIN and DPIN measures, using intraday data collected from the U.S. energy sector that comprises information on 236 firms listed on the NYSE and the NASDAQ over the period between 2018 and 2020. I find that there is a positive and significant correlation between the SDPIN measure and the stock price non-synchronicity, which attests to the reliability of this new private information measure. It proved to be more effective than DPIN measures particularly when there is algorithm trading. Since algorithm trading is relatively popular these days and is expected to grow in the future, the development of the SDPIN measure can be deemed as a contribution to the literature.

The use of the SDPIN measure can have positive implications for investors and the firm’s managers. Specifically, it enables market participants to more accurately measure the level of private information embedded in a stock price, reducing therefore the information asymmetries among investors and enhancing financial market transparency. The Firms’ managers can more easily determine the level of private information embedded in the stock price and, the use of this information enables them to make more informed (optimal) decisions.

Despite its advantages, the study faces certain limitations. First, it does not account for trading fees, such as brokers' commissions and spreads, which can significantly influence the process of investors' private information acquisition and overall trading behaviour. These fees can alter the cost-benefit analysis for investors, potentially affecting their decisions and the market outcomes observed.

Second, the study focuses on a single sector, characterized by its unique idiosyncratic characteristics. This sector-specific focus may limit the generalizability of the findings, as different sectors can exhibit varied responses to the same strategies due to differences in market dynamics, regulatory environments, and competitive landscapes.

Third, due to time constraints, the study was unable to incorporate high-frequency trading variables and advanced measures such as tick-based relative and effective spreads, which could have provided a more nuanced understanding of trading dynamics and market microstructure.As future avenues for research, I suggest the test of the SDPIN measure in other sectors and financial markets, including those of developing countries. It would be interesting to see how the SDPIN measure performs, compared to the PIN and DPIN measures, in these new markets and microstructure contexts. It also would be interesting to study the relationship between the SDPIN measure and the stock liquidity, and firms’ earnings management. Such studies could also incorporate trading fees to better understand their impact on investors’ behavior and market efficiency.

Chapter 3. Market Signal to Managers: Effect of Investors' Private Information on Earnings Management

3.1. Introduction

Prior literature, studies primarily the one-way effect of the economy on the stock market. It is assumed that the insiders (managers) have complete information about the firm and that stock prices reflect the investors' expectations about the ability of the firm to generate cash flows in the future. Therefore, the only way that stock markets affect the economy is through the stock trade liquidity, since it can impact negatively the cost of capital, especially during the initial public offering (Bond et al., 2012). However, the notion of secondary financial markets as a mere sideshow is subject to debate, as market prices also serve as a crucial source of information capable of influencing economic dynamics. Previous studies delineate a two-stage mechanism by which financial markets can affect the economy. One mechanism entails the generation of knowledge leading to price movements, and another considers the managerial reaction to stock price movements (Glosten and Milgrom, 1985; Kyle, 1985).

There are theoretical and empirical works on the feedback impact of the financial markets on managerial decisions and thereby on the economy. For instance, Boot and Thakor (1997) use the feedback effects to rationalize the firms' choice of issuing publicly traded securities over receiving finance from private sources, Luo (2005) find that decision-makers may learn from prices while evaluating mergers opportunities, Bakke and Whited (2010) show that the sensitivity of the investment to price is stronger when there is more confidential news embedded in the stock price, and Durnev et al. (2004) reveals that price informativeness is positively correlated to the effectiveness of investment.

Another strand of literature examines the impact of financial markets on firm value, management forecasts, discretionary disclosure, CEO compensation, and board dependence (see, e.g., Kang and Liu, 2008; Bharath et al., 2013; and Zuo, 2016). Rappaport (1987) advocates that managers should not worry about what the stock market says but, instead, learn about what the stock price unveils on the investors' expectations about the firm performance. They also show how managers can "read" market expectations about hurdle rates and are guided by the market via the payment price, concluding that financial markets can have a feedback effect on the economy. In general, there is evidence suggesting that managers are better positioned to evaluate operational and financial restructuring alternatives for their firms if they carefully interpret market signals (Bond et al., 2010; Peress, 2014; Sletten, 2012). However, the mechanisms through which the stock market affects the economy are not yet fully understood. This gap in understanding may stem from the conventional belief that the economy influences stock prices rather than the reverse. Additionally, studying the effect of stock prices on the economy is technically challenging due to the difficulty in identifying such effects. The goal of this paper is to identify such mechanisms for which I use various econometric approaches. The goal is to complement existing studies on the effect of stock prices on earnings management.

This study relies on Chapter 2 for a new measure of private information (SDPIN), which builds up on the dynamic measure for the probability of informed trading (DPIN) developed by Change et al. (2014), and also considers the size and volume of the stock trade. The aim is to study the effect of aggregate stock market dynamics on a firm's earning management, that is on the firm's tendency to meet its earnings thresholds. Specifically, this study investigates the impact of investors' private information embedded in the stock price on earnings manipulation and examines whether managers do consider investors' private information in stock prices while making managerial decisions with effects on earnings.

I use single-level and multi-level regressions to evaluate the aforementioned relationship and obtain important results. First, the results show that the feedback effect is more pronounced in firms whose stock prices are more affected by private information. This decreases the earnings manipulation by the firm's managers, that is, managers are less inclined to engage in earnings management if stock prices are more affected by investors' private information. Second, when analyzing upward and downward earnings management, I observe that while investors' private information can influence managers' decisions on firms' earnings in both cases, managers who have consistently inflated earnings exhibit a heightened response to private information.

This paper makes several contributions to the literature. Firstly, it provides empirical evidence supporting the presence of information transmission from secondary markets to the economy. This substantiates the notion that financial markets play a pivotal role beyond being merely peripheral, showcasing their potential for feedback effects on the economy. Secondly, it addresses a gap in the existing literature by exploring one new external factor that impacts managers' decisions on earnings management and corporate disclosure. By doing so, it expands upon the currently limited understanding of these crucial aspects of corporate behavior. Previous studies suggest that informed traders are drawn to less transparent firms, since opacity increases the profitability of private information acquisition (Brown and Hillegeist, 2007; Gao and Liang, 2013; Verrecchia, 1982;). Tests conducted by Zuo (2016) yield similar results: smaller-sized companies with limited analyst coverage and low institutional holdings, experience more informed trading. Consequently, managers in these firms with privately informed trading are held accountable by investors if they disregard market reactions and persistently manipulate company accounting profits. Thirdly, the methodology diverges from previous literature by employing a multi-level approach instead of simply using dummy variables for firm sectors to analyze industry-level effects on earnings management. This enables a more in-depth examination of specific factors influencing earnings management, particularly industry-level investors' private information. The outcomes show that intrinsic firm characteristics primarily drive earnings management variation, with industry-level variables also playing a significant role, although I find that investors' private information at the industry level does not significantly impact earnings management. Additionally, the use of SDPIN measure in the study provides significant advantages for both managers and non-manager traders. For managers, it enhances decision-making by helping them assess the level of informed trading in the market, which can reduce the likelihood of earnings manipulation when private information strongly influences stock prices. It also promotes transparency, as managers can better align their actions with market expectations, and aids in risk management by helping them understand when the market is more sensitive to private information. For non-manager traders, SDPIN offers insights into the level of private information embedded in stock prices, allowing them to identify trading opportunities and time their trades more effectively. It also helps assess the risk of adverse selection, enabling traders to make more informed decisions and avoid trading with informed investors. Overall, SDPIN equips both managers and traders with a valuable tool to navigate the complexities of informed trading and improve market efficiency.

The findings are relevant to firms and financial market players. I concluded that the feedback effect from informed trading enhances firms' financial management efficiency by reducing discretionary accruals in accounting profits and promoting more transparent corporate information disclosures, and the decrease in information asymmetry helps to reduce the investors' costs and duplicative efforts to look for information in prices on their own. While there may be concerns that this discourages informed investors, it is essential to note that personal information for trading includes both managers' secrets and investor-discovered insights that are yet unknown to businesses. Furthermore, speculators will continue to pursue novel private information, provided the benefits outweigh acquisition costs, especially when abnormal returns are at stake.

The remaining sections of the paper are organized as follows. Section 3.2 reviews the literature and develops the research hypotheses. Section 3.3 presents the data sample and methodology. Section 3.4 provides the main findings and discusses the results. Section 3.5 concludes the paper.

3.2. Literature Review

This section reviews existing literature on investors' private information, earnings management, and the relationship between the two in corporate decision-making. It begins by exploring how private information influences stock prices and managerial decisions, followed by a discussion of earnings management and its various forms and motivations. The section concludes by presenting the hypothesis that links private information to earnings management performance.

3.2.1. Investors' Private Information

3.2.1.1. The concept of investors' private information and its impact on the economy

Generating information is one remarkable ability of the financial markets. According to Roll (1988). It occurs through two mechanisms: one mechanism is the frequent revaluations of information such as the share price following the public disclosure of relevant information, such as the country's GDP and unemployment rate, the central bank's interest rate, and the firm's quarterly results, another is the stock trading. This private information can be reflected in stock prices via outside investors' trading (Glosten and Milgrom, 1985; Kyle, 1985). Private information, distinct from publicly available data, often holds significant value because it drives trading decisions. Investors leverage this information to place buy or sell orders, aiming to capitalize on discrepancies between a stock's current price and its perceived intrinsic value. This process not only reflects the informational efficiency of financial markets but also creates incentives for market participants to uncover and act on proprietary insights.

Stock prices, therefore, serve as repositories of both public and private information. And private information, by its nature, is not uniformly accessible to all market participants. The private component may range from "secret" (unknown to managers) to "not secret" (known by managers) or "partially secret" (partially known by managers). This asymmetry means that private information can influence decision-making in ways that extend beyond the market itself, potentially impacting the real economy. The interplay between private information and market dynamics can be conceptualized in two key phases. First, the generation and dissemination of new information lead to price movements as markets incorporate this information into stock prices. Second, these price changes elicit responses from other market participants—such as firm managers, institutional investors, and policymakers—who adjust their strategies based on updated price signals. These informational nuances underscore the important role of stock prices in guiding managerial decisions. For instance, when firm managers rely on stock prices to gauge market sentiment or assess the cost of equity capital, their decisions—whether related to investment, financing, or operations—can be directly shaped by the information embedded in prices.

In this sense, the stock market does not merely reflect the state of the economy; it also influences it. When stock prices fluctuate due to private information, the resulting managerial actions and investor reactions can ripple through the broader economy, impacting corporate investment decisions, resource allocation, and even economic growth. This highlights the intricate feedback loop between financial markets and the real economy, with private information serving as a critical link in this dynamic process. Bond et al. (2012) identify two primary channels through which the stock market may affect managerial decisions: the learning and incentives channels.[17]

[17] The idea of the learning channel goes back to Hayek (1945), who argued that the "price is a useful source of information". The market price is efficient and comprises of information aggregated from various sources' the decision-makers in the real world, who are unlikely to be fully informed, will wish to learn from the price.

In the learning channel, it is assumed that managers' decisions are affected by the investors' private information embedded in the stock price, but this is not known by the managers. For instance, suppose that a manager has private information *A* that is not yet known by investors, and outside investors have private information *B* that is not yet known by managers. According to conventional wisdom, information *A* is more impactful on managerial decisions than information *B*, since it is known by the managers. However, according to Chen et al. (2007), as long as information *B* exists, the manager can learn from carefully observing the stock price fluctuations, if *B* is embedded in it. Subrahmanyam and Titman (1999) explain that in the course of daily trading, traders may inadvertently come across valuable insights about firms that are not known by the managers. Consequently, managers can learn about the firm's future prospects by observing the evolution over time of the stock price, and, by doing so, their management decisions might be different from those they would made otherwise. The literature documents several crucial instances where private information affects managers' decisions. For instance, in the evaluation of mergers and acquisitions, investment projects, and disclosure of earnings forecasts (Bakke and Whited; 2010; Foucault and Fresard, 2012; Luo, 2005; and Loureiro and Taboada, 2015). Furthermore, Chen et al. (2007) show that stock prices contain information that managers do not know and that managers can learn from the (not yet known) information and take it into account in their decisions, and Zuo (2016) reveals that investors' private information helps managers improve the accuracy of their earnings forecast.

In the incentives channel, managers do not learn new information from observing stock price fluctuations. Incentives are driven by the stock price due to the attachment of stock options to their employment contracts. This channel was initially identified by Baumol (1965), and an early formalization can be found in Fishman and Hagerty (1989). It is noted that the role played by the stock price in the aforementioned two channels is, however, slightly different (Bond et al., 2012). While in the learning channel, the price reveals new information to the managers, so they can consider it in their decisions, in the incentives channel, the stock price affects the manager's incentive to take real actions. In detail, the greater the extent to which the stock price reflects the manager's decisions, the higher the incentive for the manager to make more optimal decisions (Nagar et al., 2003).[18] Several reasons justify the incentive mechanisms. Executives care about short-term stock prices due to CEO compensations, bonuses, takeover threats, or reputation. Kang and Liu (2008) examine this mechanism and their findings are in line with the theory that CEO compensation hinges on the market price and is positively associated with the informativeness of the price.

[18] For example, the manager's primary goal is to maximize the firm's stock price. However, since the market cannot directly observe the manager's decisions, stock prices do not fully reflect the expected future cash flows. For instance, it leads to underinvestment if investments that increase cash flows by \$1 does not raise the share price by \$1. When price efficiency improves, stock prices better reflect the firm's true value, including benefits from the manager's investments, reducing underinvestment as stock prices align more closely with the actual value generated by the manager's decisions.

3.2.1.2. Investors' private information measurement

In Chapter 2, I provide a critical review of the existing measures in literature for investors' private information, including the Probability of Informed Trading (PIN) by Easley et al. (1997a, 1997b), Dynamic Probability of Informed Trading (DPIN) introduced by Chang et al. (2014), and the newly developed measure named the Dynamic Probability of Informed Trading with Size Effects (SDPIN).

Probability of informed trading (PIN)

The PIN measure builds on Easley et al.'s (2002) EHO PIN measure, incorporating enhancements from Lee and Ready (1991), Ellis et al. (2000), Lin and Ke (2011), and Yan and Zhang (2012). The PIN measure is computed as follows:

1. Classify the number of buy orders (B) and the number of sell orders (S) in a single trading day, using the Lee-Ready Algorithm and EMO algorithm to classify trades of stock listed on the NYSE exchange.
2. Estimate the likelihood function for a single trading day of a stock, using Lin and Ke (2011) and Yan and Zhang (2012) methods:

$$L(\theta|B_i, S_i) = \mathrm{Log}[\alpha\delta e^{e_{1i}-e_{\max i}} + \alpha(1-\delta)e^{e_{2i}-\varepsilon_{\max i}} + (1-\alpha)e^{e_{3i}-e_{\max i}}] + \\ + B_i \log(\varepsilon_b + \mu) + S_i \log(\varepsilon_s + \mu) - (\varepsilon_b + \varepsilon_s) + e_{\max i} - \log(S_i!\,B_i!) \quad (14)$$

where $e_{1i} = -\mu - B_i \log\left(1 + \frac{\mu}{\varepsilon_b}\right)$, $e_{2i} = -\mu - S_i \log\left(1 + \frac{\mu}{\varepsilon_s}\right)$, and $e_{3i} = -B_i \log\left(1 + \frac{\mu}{\varepsilon_s}\right) - S_i \log\left(1 + \frac{\mu}{\varepsilon_s}\right)$, $e_{\max i} = \max(e_{1i}, e_{2i}, e_{3i})$, δ is the probability of bad news, ε_b and ε_s are the daily arrival rates of noise traders that submit buy and sell orders, respectively; α is the probability that some traders acquire new (private) information about the firm fundamental, and μ is the arrival rate of informed traders, given the information, the event occurs.

3. Using trading information over J days and assuming cross-trading-day independence to estimate *(ε_b, ε_s, α, μ)* by maximizing the following likelihood function:

$$V = L(\theta|B,S) = \prod_{j=1}^{j=J} L(\theta|B_j, S_j) \quad (15)$$

4. Then PIN is calculated as follows:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b} \quad (16)$$

Dynamic probability of informed trading and dynamic probability of informed trading with the size order effects

The DPIN measure is calculated following both Chang et al. (2014) and Chang and Wang (2015). It isolates the unexpected components of returns from the residual of the autoregressive model at a daily frequency. The SDPIN is a new investors' private information measure developed in Chapter 2 that also takes into account the trading volume of different trade order sizes. To determine SDPIN, I employ the same method used by Chang et al. (2014) and Chang and Wang (2015) for obtaining the unexpected returns and to recognize the informed and uninformed orders. However, in the last steps, SDPIN is defined by the ratio of the total volume of all informed transactions over the total trading volume, which includes both buy and sell-initiated trades).

3.2.2. Earnings Management

3.2.2.1. The concept of earnings management

While Healy and Wahlen (1999) define earnings management as the use of subjective accounting estimates in financial statements or structure transactions to alter financial reports by the managers, Schipper (1989) defines earnings management as the interference of managers in the information disclosure process to gain private benefits. This phenomenon is driven by the belief that investors may pay more attention to the accounting profits than to the cash flows. Graham et al. (2005) show that executives see earnings as the key metric for public disclosure and that optimistic views by analysts on the firm's short-term profit help build credibility with the market participants. The existing literature also shows that there is a positive relationship between earnings and stock returns (Chan et al., 2006; Demirtas and Zirek, 2011).

Other studies such as that of Ronen and Yaari (2007), argue that earnings management is not necessarily a bad practice, although it is difficult to distinguish between earnings manipulation (as a fraud) and reconciling profit through firm revenues and expenses; and Fatemeh and Narjes (2013) argue that earnings management should not be confused with an illegal manipulation of financial results to distort reality ("cooking the book"). Despite these opposite views, most of the literature considers earnings management as one form of manipulation.

Earnings management can be divided into upward and downward earnings management. Upward earnings management refers to the practice where managers intentionally manipulate financial reporting to inflate a company's earnings. This strategy is typically employed to meet or exceed market expectations, achieve financial targets, secure bonuses, or present a more favorable financial position to investors and stakeholders. Techniques used for upward earnings management include prematurely recognizing revenue, delaying expense recognition, or reclassifying expenses to reduce their impact on earnings. While this practice may provide short-term benefits, it can undermine long-term transparency and credibility. Downward earnings management involves intentionally reducing reported earnings below their actual levels. Managers may use this strategy to "save" earnings for future periods (also known as "income smoothing") or to create a lower earnings benchmark for easier outperformance in subsequent periods. Other motivations might include reducing tax liabilities, avoiding regulatory scrutiny, or attributing poor performance to temporary factors, such as an economic downturn. Downward earnings management techniques often involve accelerating expense recognition, deferring revenue recognition, or writing off assets. Both forms of earnings management are often viewed critically as they can mislead stakeholders and undermine the integrity of financial reporting.

The manipulation of earnings by the firms has been extensively studied by the existing literature (Moardi et al., 2020; Park, 2017; Watts and Zimmerman, 1978). It focuses mainly on the subjective and self-interested aspects related to the managers. The most common motivations for firms to manipulate earnings are to attract outside investments, enhance bonuses, apply for governmental grants, and buyback stocks. Apart from the firm and the managers, which have a direct interest in earnings manipulation, other factors are also examined in the literature (see, e.g., Chen et al., 2007; Cheng and Warfield, 2005; Fakhfakh and Nasti, 2012; Fathi, 2013; Goh et al., 2013; Healy, 1985). Specifically, Charfeddine et al. (2013) and Watts and Jimmerman (1990) classify the aformentioned factors into two categories: incentive factors, such as the debt of the firm, and the firm size, performance, and growth, as well as constraint factors such as the characteristics of the board of directors, ownership structure, and auditing quality, and regarding the effect of external factors, there is a growing attention to the effect of investors' sentiment on the information disclosure policies and earnings management; for instance, Hurwitz (2018) examines the relationship between investors' sentiment and managers' biased behaviours in forecasting firm accounting performances. It shows that, during periods of high sentiment, earnings estimations are more optimistic, and vice-versa.[19] Simpson (2013) indicates that discretionary accruals are used by firms to inflate their earnings in times of predominantly optimistic sentiment but disclose more conservative results during times of investor sentiment pessimism.

3.2.2.2. Earnings management Measurement

Marai and Vladan (2014) and McNichols (2000) review different methods widely used in the literature to identify earnings management, namely aggregate accruals, specific accruals, and the statistical distribution of earnings. Although all of these methods are centered on various ideas and assumptions to provide a solution to the previously highlighted problems, there is no sole technique with the ability to completely answer the mean questions about magnitude, and the techniques of earnings management.

Aggregate Accruals Method

A significant body of literature focuses on identifying discretionary accruals by examining the relationship between total accruals and explanatory factors. Earnings management is carried out through the accrual basis of accounting, where revenues and expenses are recorded when earned or incurred, as

[19] Their reasoning is that financial executives in response to investors' sentiment tend to overstate firm accounting profit by adjusting the accruals, influencing the market's ability to price firm shares efficiently.

opposed to the cash basis, which recognizes them when cash is received or paid. The difference between operating income and net cash flow from operations represents non-cash accounting profits or accruals.

Total accrual (TAit) = Earnings for operations - Net cash flow from operations (17)

Total accruals are comprised of non-discretionary accruals and discretionary accruals. Non-discretionary accruals (NDA) are accrual values that appear or be eliminated spontaneously over periods or business cycle length. This process complies with accounting principles. Meanwhile, the discretionary accrual value (DAC) is the adjustment to change the cash flow by choice of managers. According to Kothari (2001), the use of discretionary accrual models, where discretionary accrual is often used as a synonym for earnings management, is very common. When analyzing earnings adjustments, DAC is considered an index measuring the quality of financial information disclosed (Francis et al, 2005).

Total accrual = Non-discretionary accruals + Discretionary accruals (18)

This literature began with Healy (1985) and DeAngelo (1986), who used total accruals and change in total accruals, respectively, as measures of management's discretion over earnings. Jones (1991) introduced a regression approach to control for nondiscretionary factors influencing accruals, specifying a linear relation between total accruals and change in sales and property, plant and equipment.

Jones' (1991) model is the first to predict total accrual changes using explanatory variables, distinguishing between non-discretionary accruals (arising from the organization's economic position) and discretionary accruals (resulting from earnings management). Non-discretionary accruals, such as those linked to revenue changes, reflect economic conditions, while discretionary accruals indicate manipulation. To estimate discretionary accruals or earnings management, the model controls for the impact of non-discretionary changes, isolating the discretionary component. The method to estimate the discretionary accruals or earnings management by Jones (1991) is as follows:

1. Values of β_0, β_1, β_2 and β_3 are estimated by Equation (19):

$$\frac{TAC_t}{TA_{t-1}} = \beta_0 + \beta_1 \frac{1}{TA_{t-1}} + \beta_2 \frac{\Delta Sales_t}{TA_{t-1}} + \beta_3 \frac{TFA_t}{TA_{t-1}} + \varepsilon_t \quad (19)$$

2. Non-discretionary accruals are then calculated using Equation (20) with the above-estimated values of coefficients β_0, β_1, β_2, and β_3:

$$NDAC_t = \beta_0 + \beta_1 \frac{1}{TA_{t-1}} + \beta_2 \frac{\Delta Sales_t}{TA_{t-1}} + \beta_3 \frac{TFA_t}{TA_{t-1}} + \varepsilon_t \quad (20)$$

3. Discretionary accruals (operating DAC) can be achieved using Equation (21):

$$\mathrm{DAC} = \frac{\mathrm{TAC_t}}{\mathrm{TA_{t-1}}} - \mathrm{NDAC_t} \tag{21}$$

where DAC is the discretionary accruals (earnings management), $\mathrm{TAC_t}$ is the total accruals (or net operating accruals) estimated by taking net income minus cash flow from operations, TAC_{t-1} is the total assets at the last fiscal year-end; *△Sales* is the change in sales from operation, *TFA* is total fixed assets, $NDAC_t$ is non-discretionary accruals.

Several variants of Jones' (1991) model have been developed to improve the detection of earnings management. The modified Jones Model (Dechow et al., 1995) adjusts for changes in receivables to better capture revenue-based manipulation, under the assumption that credit sales are more easily manipulated than cash sales. This modification improves the model's ability to identify discretionary accruals during periods of suspected earnings management. The margin Model (Peasnell et al., 2000) shifts focus from total accruals to working capital accruals (WCA), excluding depreciation due to its lower susceptibility to systematic manipulation. It highlights abnormal accruals, particularly those related to non-bad debt expense manipulations, by examining changes in creditors, debtors, and inventory. The performance matching Model (Kothari et al., 2005) adds return on assets (ROA) as a control for firm performance, helping to reduce bias from heteroskedasticity and omitted variables. This approach also improves the symmetry of discretionary accruals estimation. Lastly, the reversal-Based Model (Dechow et al., 2012) posits that earnings management through accruals is likely to reverse in subsequent periods. By accounting for these reversals, the model enhances the detection of earnings management and reduces model misspecifications, though it relies on researchers to define when these reversals occur.

Specific accruals Method

The specific accruals approach focuses on analyzing individual accruals that require significant managerial judgment, such as bad debt provisions or claim loss reserves, to identify earnings management. This method isolates discretionary accruals by modeling expected values for these accruals and comparing them to actual values, thereby highlighting managerial discretion. Studies like McNichols and Wilson (1988), Beneish (1997), and Cecchini et al. (2012) use this approach to examine whether earnings management occurs in specific sectors, such as banking and insurance, by focusing on particular accrual accounts and controlling for non-discretionary factors.

Statistic distribution of earnings Method

The statistical distribution approach focuses on the idea that managers often manipulate earnings to meet specific benchmarks or goals, resulting in distribution anomalies around certain thresholds. The approach examines the frequency of earnings reports at key points, such as avoiding losses, maintaining positive earnings, or meeting analyst expectations. A discontinuity in earnings distribution—such as an unusually high frequency of small positive earnings changes or an unusually low frequency of small losses—can indicate earnings management. Key studies, like Burgstahler and Dichev (1997), Degeorge et al. (1999), and Gore et al. (2007), have used this approach to identify earnings management patterns linked to these thresholds, often demonstrating that managers adjust accruals to meet earnings targets.

3.2.3. Hypotheses Development

The examination of external factors influencing future earnings management is yet notably limited. The concept of informational feedback from secondary stock markets to earnings management is relatively new within this strand of literature. In terms of information flow, studies conducted by Richardson (2000), Dai et al. (2013), and Fatemeh and Narjes (2013) validate the influence of information asymmetry on firm earnings management, although these studies focus on Chinese and Iranian stock exchanges only. Nevertheless, these articles measure asymmetric information with the premise that managers hold complete information about firms and hence, secondary markets as essentially a sideshow.

Additionally, the existing literature studying the feedback effect of speculators' personal information tends to follow the managerial learning hypothesis, which assumes that managers can learn about the firm by observing the stock prices. For instance, Chen et al. (2007) argue that some of the investors' private information embedded in a stock price might be new to the manager. Zuo (2016), following Chen et al. (2007), concludes that the learning channel helps managers improve the earnings forecast and that the existence of investors' private information can potentially lead to higher abnormal returns at the expense of insiders. This can deter managers from persistently inflating accruals and it increases the likelihood that the truth will eventually come to light. Thus, when managers are aware of investors' private information, they are less likely to manipulate earnings.

Therefore, I hypothesize that:

Hypothesis 1 (H1): *There is a negative relationship between the investors' private information and earnings management performance.*

3.3. Data Sample and Methodology

To investigate the hypothesized negative relationship between investors' private information and earnings management performance (H1), this study employs both single-level and multi-level regression models, each offering unique insights into the dynamics of private information and managerial behavior. The single-level model, rooted in prior methodologies such as Chen et al. (2007) and Luo (2016), focuses on the direct relationship between private information proxies (PIN, DPIN, and SDPIN) and discretionary accruals (DAC), with interaction terms and control variables ensuring robustness. Complementing this, the multi-level model explores the hierarchical structure of earnings management decisions by accounting for clustering effects at the observation, firm, and industry levels. Recognizing that industry-specific characteristics and profitability dynamics influence earnings manipulation, the multi-level approach captures variations across sectors, providing a richer understanding of how private information and earnings management interplay at different levels of aggregation. This combined approach not only validates the hypothesis but also extends the literature by integrating cross-sectional and hierarchical dimensions in earnings management research.

3.3.1. Data Sample

The data sample comprises quarterly information on the stock prices of firms that belong to the S&P 500 index.[20] Specifically, I collect data on 475 U.S. stocks listed on the NYSE and NASDAQ over the period between January 2018 and December 2021. To be included in the sample, companies must have been listed in the S&P 500 for at least two years and remain in the S&P 500 during the research period. Table 3.1 provides further information on the data sample.

Table 3.1: This table describes the data sample. Investors' private information is measured by PIN, DPIN, and SDPIN calculated according to Eq.(13), Eq.(6), and Eq.(8), respectively.

	Firms	Obs.	Mean			Skewness			Kurtosis		
			PIN	DPIN	SDPIN	PIN	DPIN	SDPIN	PIN	DPIN	SDPIN
Industrials	71	1,156	0.136	0.188	0.132	0.525	-0.128	0.413	-1.427	-1.042	-0.338
Health Care	63	940	0.228	0.201	0.238	0.001	0.538	-0.593	0.729	1.323	0.342
Information Technology	61	952	0.142	0.224	0.267	0.062	0.862	0.030	0.844	-0.260	0.026
Communication Services	27	320	0.062	0.144	0.102	-0.399	0.599	0.138	0.538	-0.804	1.068
Consumer Staples	32	456	0.081	0.775	0.063	-0.728	1.545	0.588	0.613	1.739	0.003
Consumer Discretionary	60	880	0.106	0.093	0.089	0.363	-0.210	0.353	0.508	-0.977	-0.556
Utilities	28	384	0.198	0.182	0.134	-0.194	-0.089	-0.061	0.539	-1.206	-0.179
Financials	61	888	0.137	0.169	0.171	-0.793	0.405	0.057	0.626	-1.034	0.782
Materials	26	376	0.275	0.156	0.232	0.289	0.744	0.857	0.024	-0.317	0.026
Real Estate	25	384	0.176	0.181	0.198	0.777	0.586	0.381	0.625	-0.532	0.017
Energy	21	303	0.293	0.201	0.236	0.483	0.581	0.118	0.532	-0.919	0.637
Total	475	7,039	0.158	0.204	0.235	0.149	0.289	0.115	-0.867	-1.189	-0.351

[20] The S&P 500 is a stock market index tracking the stock performance of 500 of the largest companies listed on the U.S. stock exchanges. According to Standard and Poor's, the index represents about 80 percent of the total value of all stocks traded in the U.S. It includes stocks across 11 industry sectors, as defined by the Global Industry Classification Standard (GICS) classification system.

3.3.2. Regression Models

3.3.2.1. Single-level regression model

I regress earnings management (DAC) on informed investors' private information (PI), where DAC is the dependent variable estimated using the modified Jones model (Raman and Shahrur, 2008). The model follows the methodology used by Chen et al. (2007) and Luo (2016). The independent and control variables are estimated for the same period (quarterly) as the dependent variable. The control variables used are the same as those used in the previous literature – the stock return (RETURN) and the institutional investors (INST) – see Chen et al. (2007), La Porta et al. (2000), and Zuo (2016).

$$DAC_{i,t} = \beta_0 + \beta_1 PI_{i,t}^m + \beta_2 RETURN_{i,t} + \beta_3 INST_{i,t} + \beta_4 RETURN_{i,t}\ x\ PI_{i,t}^m + \beta_5 PI_{i,t}^m\ x\ INST_{i,t-1} + \gamma CONTROL_{i,t} + INDUSTRY_{i,t} + YEAR_{i,t} + \varepsilon_t \qquad (22)$$

where *i* denotes the stock and *t* denotes the period. *PI* is the measure for the investors' private information, measured by the PIN, DPIN, and the SDPIN for m = 1, 2, and 3, respectively; *PI* is one major independent variable in the model, denoting the direct impact of confidential information in the stock price on earnings management. The study uses three proxies for investors' private information: probability of informed trading (PIN) developed by Easley et al. (2002), dynamic measure for probability of informed trading (DPIN) presented by Chang et al. (2014) and Chang and Wang (2015), and probability of informed trading with size effects (SDPIN) proposed in Chapter 2.

DAC is earnings management (also called discretionary accruals) and is computed through the modified Jone's (1991) model. *RETURN* represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarter). *INST* denotes the institutional investors variable, defined by the proportion of firm shares owned by the institutional investors. (*RETURN x PI)* is the interaction term between stock return and investors' private information.[21] I use lagged *PI* to capture some firm characteristic that results in return containing more private information. (*PI x INST)* is the interaction term between institutional investors and investors' private information.[22] Other control variables are coded *CONTROL*, including *DUAL* which is *CEO* duality and is equal to 1 if the *CEO* of the company also serves as the chairman of the board of directors and equal to 0 otherwise; firm size (*SIZE*) is given

[21] Zuo (2016) investigates the relationship between private information and management forecasts, suggesting that the complex nature of private information and its acquisition process in stock price makes its relationship with earnings management difficult to observe; the role of its interaction with other dependent variables should not be ruled out. S/he also observed the effect of stock return and its interaction with the variable private information on management forecasts.

[22] I use institutional investors as a factor that can control the quality of the information disclosed by firms especially information presented in the financial reports. Zuo (2016) documents that a higher percentage of shares owned by institutional investors decreases the amount of private information contained in stock prices. I expect that institutional investors can play both direct and mediating roles in our model.

by quarterly market capitalization divided by 10^6; firm growth opportunity (*GROWTH*) measured as the market value of equity plus book value of assets minus book value of equity, scaled by the book value of the total assets; and profitability (*ROE*).

Measure for Earnings Management

I run different models to estimate this variable and choose the performance-adjusted modified Jone (1991) model according to Raman and Shahrur (2008), given that its coefficient of determination (R-squared) is the highest (65.29%). The discretionary accruals or earnings management is calculated as follows:

2. Values of β_0, β_1, β_2 and β_3 are estimated by Equation (23):

$$\frac{\mathrm{TAC_t}}{\mathrm{TA_{t-1}}} = \beta_0 + \beta_1 \frac{1}{\mathrm{TA_{t-1}}} + \beta_2 \left(\frac{\Delta \mathrm{Sales_t} - \Delta \mathrm{Rec_t}}{\mathrm{TA_{t-1}}}\right) + \beta_3 \frac{\mathrm{TFA_t}}{\mathrm{TA_{t-1}}} + \beta_4 \mathrm{ROA_t} + \beta_5 \mathrm{MB_t} + \varepsilon_t \quad (23)$$

2. Non-discretionary accruals are then calculated using Equation (24) with the above-estimated values of coefficients β_0, β_1, β_2, and β_3:

$$\mathrm{NDAC_t} = \beta_0 + \beta_1 \frac{1}{\mathrm{TA_{t-1}}} + \beta_2 \left(\frac{\Delta \mathrm{Sales_t} - \Delta \mathrm{Rec_t}}{\mathrm{TA_{t-1}}}\right) + \beta_3 \frac{\mathrm{TFA_t}}{\mathrm{TA_{t-1}}} + \beta_4 \mathrm{ROA_t} + \beta_5 \mathrm{MB_t} + \varepsilon_t \quad (24)$$

3. Discretionary accruals (operating DAC) can be achieved using Equation (25):

$$\mathrm{DAC} = \frac{\mathrm{TAC_t}}{\mathrm{TA_{t-1}}} - \mathrm{NDAC_t} \quad (25)$$

where, DAC is the discretionary accruals (earnings management). A positive value of DAC indicates upward earnings management, where managers inflate earnings by increasing income or deferring expenses. Conversely, a negative value of DAC signifies downward earnings management, where managers reduce reported earnings by accelerating expenses or deferring income.

$\mathrm{TAC_t}$ is the total accruals (or net operating accruals) estimated by taking net income minus cash flow from operations, TAC_{t-1} is the total assets at the last fiscal year-end; *△Sales* is the change in sales from operation, *△Rec* is the change in accounts receivables, *TFA* is total fixed assets, $NDAC_t$ is non-discretionary accruals, *ROA* is the ratio between total net income and total assets, and *MB* is the ratio between the market value and the book value of equity.

3.3.2.2. Multi-level model

There are various factors underlying the managerial decision of managing earnings. We can study earnings management by categorizing the level of factors involved in such a decision, for instance, the use of firm-level predictors or industry-level predictors (Fan and Jahan-Parvar, 2012; Mackay and Phillips, 2005). It is argued that firms are influenced by factors that are specific to their industries which in turn affect their profitability. The substantial variations in profitability observed among different industries can partially be attributed to the different levels of earnings management practiced among industries. For instance, Beneish (2001) notes that there are industries (e.g., financial and insurance industries) that have greater incentives for earnings manipulation. This is because factors such as "loan loss reserves for banks and property casualty claim loss reserves" heavily rely on management's discretion (Healy and Wahlen, 1999), so there may be a greater propensity for earnings management due to the flexibility and subjectivity in estimating these reserves.

On the other hand, each industry sector has distinct characteristics and information dynamics, making it a valuable resource for speculators seeking to uncover new or private insights about specific firms. Additionally, while earnings management is a strategy used by managers across various industries, there can be significant disparities in the extent of earnings manipulation among industry sectors. To capture the industry-related patterns regarding private information and earnings adjustment behaviors, I employ a multi-level model to examine the level of clustering within the research dataset. This approach allows us to identify and explain variations in private information and earnings management across industry sectors. The use of multi-level modeling in this study is a contribution to the literature.

I define three determinant levels of earnings management. The first level is observation, the second is the firm level, and the third is the industry level. The multi-level model is extended gradually from the empty model (Eq. 26) to the model with the random intercepts and random coefficients (Eq. 34). Then, control variables are added. The mean of private information of each firm (FPI) is a level-two (firm-level) variable. At level-three analysis, the mean of private information of all firms in each industry (IPI) will be employed.

3.2.2.1 The empty model

As a first step, I use the empty model to determine whether there is evidence of clustering in the data with respect to the dependent variable DAC_{ijk}.

Level 1:

$$\mathrm{DAC}_{\mathrm{ijk}} = \beta_{0\mathrm{jk}} + \varepsilon_{\mathrm{ijk}} \tag{26}$$

where DAC is earnings management (also called discretionary accruals) and is computed through the modified Jone’s (1991) model.

Level 2:

$$\beta_{0jk} = \gamma_{00k} + \mu_{0jk} \tag{27}$$

Level 3:

$$\gamma_{00k} = \delta_{000} + r_{00k} \tag{28}$$

Combined empty model: $$DAC_{ijk} = \delta_{000} + r_{00k} + \mu_{0jk} + \varepsilon_{ijk} \tag{29}$$

Where DAC is earnings management (also called discretionary accruals) and is computed through the modified Jone's (1991) model.

3.2.2.2 Random models with covariates

The combined model (see below Eq.34) is the combination of Eq.302 to Eq.335 and is a more consolidated mixed-effect model which assumes the intercepts and slopes of some firm-level variables are random and affected by firm and industry variables. In other words, this model helps to analyse the indirect influence of sector characteristics levels on earnings management.

Level 1 equation:

$$\mathrm{DAC_{ijk}} = \beta_{0jk} + \beta_{1jk}\, PI^{m}_{ijk} \text{ x } \mathrm{PI_{ijk}} + \varepsilon_{ijk} \tag{30}$$

where DAC is earnings management (also called discretionary accruals) and is computed through the modified Jone's (1991) model. *PI* is the measure for the investors' private information, measured by the PIN, DPIN, and the SDPIN for *m* = 1, 2, and 3, respectively.

Level 2 equation:

$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k}\, \mathrm{FPI_{0jk}} + \mu_{0jk} \tag{31}$$

Where FPI is the mean of private information of each firm (level-two (firm-level) variable).

Level 3 equation:

$$\gamma_{00k} = \delta_{000} + \delta_{001}\, \mathrm{IPI_{00k}} + \mathrm{r_{00k}} \tag{32}$$

$$\beta_{1jk} = \delta_{110} + \delta_{111}\, \mathrm{IPI_{11k}} + \mathrm{r_{11k}} \tag{33}$$

Where IPI is the mean of private information of all firms in each industry (level-three (industry-level) variable).

The combined model is given by:

$$DAC_{\mathrm{ijk}} = \delta_{000} + \delta_{001}\, \mathrm{IPI_{00k}} + \gamma_{01k}\, \mathrm{FPI_{0jk}} + \delta_{110} PI^{m}_{ijk} + \delta_{111}\, \mathrm{IPI_{11k}}\, x\, PI^{m}_{ijk} + \mathrm{r_{11k}} PI^{m}_{ijk} + \gamma\, \mathrm{CONTROL_{ijk}} + \mathrm{YEAR} + \varepsilon_{ijk} + \mu_{0jk} + \mathrm{r_{00k}} \tag{34}$$

where DAC is earnings management (also called discretionary accruals) *PI* is the measure for the investors' private information, measured by the PIN, DPIN, and the SDPIN for *m* = 1, 2, and 3, respectively. FPI is the mean of private information of each firm (level-two (firm-level) variable). IPI is the mean of private information of all firms in each industry (level-three (industry-level) variable). Other control variables are coded CONTROL, including DUAL which is CEO duality and is equal to 1 if the CEO of the company also serves as the chairman of the board of directors and equal to 0 otherwise; firm size (SIZE) is given by quarterly market capitalization divided by 10^6; firm growth opportunity (GROWTH) measured as the market value of equity plus book value of assets minus book value of equity, scaled by the book value of the total assets; and profitability (ROE). YEAR is a dummy variable, representing the time effect.

3.4. Results and Discussion

3.4.1 Main Model

Table 3.2 provides some statistics on the data sample. The mean values of investors' private information proxies PIN, DPIN, and SDPIN are 0.158, 0.204, and 0.235, respectively, all with standard deviations around 0.1. The values of these variables are relevant to what they reflect because of the nature of private information - it is not publicly announced and is only possessed by a minority of outsiders. Additionally, the SDPIN has the largest amount of private information in stock prices, with a mean of 0.695 compared to the other proxies. The R^2 measure indicates the level to which the price of a stock co-moves with the market. A higher R^2 means that the stock price is more synchronous with the market. The mean of R^2 is 0.418 which is higher than the 0.223 and 0.250 reported by Chan et al. (2013) and Hutton et al. (2009), respectively. The higher value of R^2 is in line with the notion that stock prices of larger firms reflect more industry and market-specific factors in the stock prices (Roll, 1988) than smaller firms; the firm sample is based on the S&P 500 index which is a good proxy for the market as a whole. However, the mean value of R^2 is below 0.5, which indicates that besides the effect of the macroeconomic factors, individual stocks have their own unsystematic risk component, and so they move asynchronously with the market.

Table 3.2: This table presents the statistical descriptions of regression variables. PIN, DPIN, and SDPIN are the measures for investors' private information, calculated according to Eq.(13), Eq.(6), and Eq.(8), respectively. NSYN is the stock price non-synchronicity measured by Roll's (1988) market model. DAC is earnings management (also called discretionary accruals) and is computed through the modified Jones model (Raman and Shahrur, 2008); CONTROL variables include: RETURN represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarterly); INST denotes institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; DUAL is CEO duality and is equal 1 if the CEO of the company also serves as the chairman of the board of directors and equal 0 otherwise; firm size (SIZE) is given by quarterly market capitalization divided by 10^6; GROWTH is firm growth opportunity; ROE presents profitability.

	N	Min	Max	Mean	SD	Skewness	Kurtosis
PIN	7,039	0.002	0.519	0.158	0.124	0.149	-0.867
DPIN	7,039	0.005	0.637	0.204	0.148	0.289	-1.189
SDPIN	7,039	0.006	0.695	0.235	0.121	0.115	-0.351
R-square	7,039	0.213	0.906	0.418	0.357	0.26	-0.984
NSYN	7,039	-0.984	0.568	0.144	0.601	6.545	-21.489

DAC	7,039	-0.127	0.119	-0.005	0.021	0.289	-1.189
RETURN	7,039	-0.819	0.63	-0.131	0.133	4.149	-9.867
SIZE*	7,039	6,070	92,914	67,300	0.955	12.016	-36.410
ISNT	7,039	0.461	0.932	0.793	0.174	1.445	3.917
DUAL	7,039	0.000	1.000	0.493	0.502	0.218	-1.146
GROWTH	7,039	0.081	12.304	2.437	1.312	12.462	23.180
ROE	7,039	-39.331	37.037	0.265	4.706	2.368	3.612

* Firm size (SIZE) in million US dollars.

Earnings management (DAC) is estimated based on the performance-adjusted modified Jone (1991) model by Raman and Shahrur (2008). The max, min, and mean of earnings management (DAC) are 0.119, -0.127, and -0.005, respectively. DAC can be negative or positive which identifies two ways of earnings management. If positive, it means that there is an upward earnings management behavior, if it is negative there is a downward earnings management behavior. According to Degeorge et al. (1999), earnings management is a managers' game of information disclosure that outsiders must play. Firms can inflate and deflate their profits following a strategy that can vary over time. For example, managers exaggerate earnings to sustain recent performance, that is, make at least last year's earnings or to meet shareholders' and investors' expectations as it is related to the company's reward scheme. On the other hand, Holthausen et al. (1995) show that firms' profits are manipulated downward when managers are at the upper bounds of their bonus contracts.

Table 3.3 shows the correlation matrix of all variables. The correlations of the three pairs formed by PIN, DPIN, and SDPIN have the strongest Pearson correlation coefficients, 0.531, 0.694, and 0.672 respectively. However, these results do not affect the accuracy of the regression models. Instead, these figures may signal the close relationship among the three proxies for investors' private information in stock prices. In addition, the pair of firm size (SIZE) and institutional ownership (INST) also exhibits a strong correlation of 0.573. However, the VIF-index check shows normality with no multicollinearity issues. Overall, the coefficients of correlations are generally below 0.5, meaning that there are no issues of serious multi-collinearity among variables.

Table 3.3: This table presents the variables correlation matrix. PIN, DPIN, and SDPIN are the measures for investors' private information, calculated according to Eq.(13), Eq.(6), and Eq.(8), respectively. NSYN is the stock price non-synchronicity measured by Roll's (1988) market model. DAC is earnings management (also called discretionary accruals) and is computed through the modified Jones model (Raman and Shahrur, 2008); CONTROL variables include the firm size (SIZE) given by market capitalization divided by 10^6; RETURN represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarterly); INST denotes institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; DUAL is CEO duality and is equal 1 if the CEO of the company also serves as the chairman of the board of directors and equal 0 otherwise; GROWTH is firm growth opportunity; ROE presents profitability.

	NSYN	DAC	PIN	DPIN	SDPIN	RETURN	INST	DUAL	GROWTH	ROE	SIZE	VIF
NSYN	1											1.18
DAC	-0.07*	1										3.56
PIN	0.325***	-0.055**	1									2.09
DPIN	0.101**	-0.069**	0.531**	1								1.75
SDPIN	0.363***	-0.032***	0.694**	0.672*	1							4.84
RETURN	-0.218	0.037*	-0.122	-0.105**	-0.033	1						2.66
INST	-0.221**	0.045**	-0.091**	-0.104	-0.037**	0.883	1					1.73
DUAL	0.261**	0.181*	0.134*	0.071**	-0.048	-0.411	-0.383**	1				1.25
GROWTH	0.685	-0.372	0.124	-0.065	-0.019**	-0.218*	-0.221	-0.288	1			1.19
ROE	0.339*	0.744**	0.003	0.04	-0.017	-0.031	-0.014*	0.175***	0.34	1		2.67
SIZE	0.284***	-0.406*	0.019***	-0.274**	-0.238	-0.086	0.573**	-0.194***	-0.072	-0.116**	1	2.36

Note: *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively

One issue with the measure of earnings management in Raman and Shahrur's (2008) model is that the proxy can be negative or positive. The negative or positive values of earnings management do not indicate the degree and magnitude but the direction of earnings manipulation. To specify, a negative value means downward earnings manipulation and a positive value of earnings management means upward earnings manipulation. To overcome this issue, instead of using the original values of earnings management, I conduct tests for the absolute value of earnings management. The use of the absolute value of discretionary accruals when measuring earnings management (Warfield et al., 1995; Hribar and Nichols, 2007) enables a more accurate evaluation of the influence of different factors on earnings management. Along with statistical tests, I perform pooled OLS, Fixed Effects, and Random Effects for the main model to choose the best fit with the dataset. The Fixed Effects model assumes individual-specific characteristics (e.g., company factors) are correlated with the explanatory variables, allowing for unit-specific intercepts and controlling for unobserved heterogeneity. In contrast, the Random Effects model assumes these characteristics are uncorrelated with the explanatory variables, estimating a common intercept for all units. The Hausman test is used to compare Fixed Effects (FE) and Random Effects (RE) models in panel data analysis. If the Hausman test rejects the null hypothesis, it indicates that the Fixed Effects model is more suitable, as it suggests that individual effects are correlated with the regressors. If the null hypothesis is not rejected, the Random Effects model is preferred.

Table 3.4 presents *F* tests and Hausman test results. For each model of PIN, DPIN, and SDPIN, the *F* test has values of 1028.14, 758.23, and 179.12, respectively, with P values = 0.000; thus, the Fixed Effects regression is more suitable than pooled OLS for all models. Then Hausman test is performed to compare the Fixed Effects and the Random Effects technique. The Chi-square is 92.06, 15.88, and 104.52 for the PIN, DPIN, and SDPIN models respectively, with *p*-values equal to 0.000. Thus, we can reject the null hypothesis of the Hausman test, indicating that the Fixed Effects is the best fit for the data sample. Also, I conduct further tests to check for heteroscedasticity and autocorrelation issues. However, as the *p*-values of heteroscedasticity and autocorrelation are equal to 0.000 for all models, we can reject the null hypothesis of non-heteroscedasticity non-autocorrelation. To deal with heteroscedasticity, I run GLS regressions for all the models (the outcomes for GLS regressions are presented in Table 3.4 for PIN, DPIN, and SDPIN models).

Table 3.4 also provides estimates for the impact of private information contained in share prices on earnings management. The SDPIN shows the best fit with the Fixed Effects regression as its model has the highest R-Square (0.1931), indicating that it explains the most variation in the dependent variable (earnings management) among the three models. The DPIN model follows with 0.1802, and the PIN model has the lowest R-Square at 0.1669.

SDPIN shows a significant negative correlation to absolute values of DAC at the 1% level. Models of PIN and DPIN also reveal similar patterns as they negatively impact earnings management but with a 5% and 1% significance level, respectively. This suggests that, across all three models, higher levels of investors' private information (PI) are associated with lower levels of earnings management, with SDPIN and PIN showing the most significant effect. The greater amount of confidential information in prices means the lower possibility of earnings management. Therefore, the hypothesis *H1* is confirmed. Through trading, informed investors inject their private news into stock prices, causing their movements. As this type of private information can be known or not by managers, I argue that it can affect managers' earnings manipulation through either the learning channel or the incentive channel although the first channel is rarer. When informed traders own confidential information about firms and incorporate this information in stock prices through their tradings, managers will have a lesser chance to manipulate the firm's earnings, leading to a higher quality of financial statements released by firms. This finding may help to motivate outside investors to put more effort into studying firms and making transactions that can benefit them in many ways.

Table 3.4: This table presents results for the relationship between investors' private information and earnings management. PI is the investors' private information, measured by the PIN, DPIN, and SDPIN. DAC is earnings management (also called discretionary accruals) and is computed through the modified Jones model (Raman and Shahrur, 2008); CONTROL variables include the firm size (SIZE) given by market capitalization divided by 10^6; RETURN represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarterly); INST denotes institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; DUAL is CEO duality and is equal 1 if the CEO of the company also serves as the chairman of the board of directors and equal 0 otherwise; GROWTH is firm growth opportunity; ROE presents profitability. Standard errors are in parentheses.

	PIN Model	**DPIN Model**	**SDPIN Model**
R-Square	0.1669	0.1802	0.1931
F Test	1028.14***	758.23***	179.12***
	Coef. (SE)	**Coef. (SE)**	**Coef. (SE)**
_Cons	0.709	1.023	1.314
PI	-1.148***	-0.557**	-1.202***
	(0.361)	(0.259)	(0.213)
RETURN	0.838***	0.497***	0.614***
	(0.034)	(0.039)	(0.049)
INST	-0.059***	-0.016**	-0.040***
	(0.003)	(0.008)	(0.006)
RETURN x PI	2.263***	0.557**	1.192***
	(0.187)	(0.247)	(0.047)
PI x INST	-0.524*	-0.269**	-0.635**
	(0.318)	(0.149)	(0.300)
SIZE	-7.2E-07***	-1.3E-07***	-8.5E-07***
	(2.1E-07)	(1.1E-08)	(3.6E-08)
DUAL	0.025**	0.099**	0.099**
	(0.012)	(0.046)	(0.040)
GROWTH	-0.004	-0.003	-0.008
	(0.013)	(0.013)	(0.016)
ROE	-0.162**	-0.109**	-0.041***
	(0.080)	(0.053)	(0.007)
Year Dummies	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes

Hausman Test	92.06***	15.88***	104.52***
Heteroskedasticity	5.83***	11.64***	7.08***
Autocorrelation	1.22E+05***	3.78E+05***	2.28E+06***

Note: *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively

DAC is positively associated with stock return in the PIN, DPIN, and SDPIN models. Firm managers with higher stock returns tend to increase the level of accounting profit manipulation which is consistent with the findings of past studies. For instance, the tests of Zuo (2016) also show positive and significant coefficients on the relationship between stock return and earnings forecasts. This study finds that "the positive coefficient on return is consistent with managers learning from prices, but may not be solely attributed to this channel since it can be partly driven by public information".[23]

The joint effects of stock return (RETURN) and private information (PI) are positive and statistically significant, with the SDPIN and PIN models showing stronger effects at the 1% significance level, compared to the DPIN model, which shows significance at the 5% level. Firm managers are more sensitive to responding to stock returns when stock prices contain more investor information that is new to them. This result is in alignment with the results regarding the nature of private information. More personal information incorporated in stock price strengthens the relationship between stock return and earnings management by restraining managers from distorting corporate profit (urging them to unfold the truths) or encouraging managers to acquire new knowledge and incorporate it into their decisions.

The coefficients of institutional investors (INST) and the interaction terms between institutional investors and private information of investors in stock prices are both negative statistically significant. Specifically, all three models (PIN, DPIN, and SDPIN) demonstrate a significant impact of private information (PI) on institutional investors (INST) at the 1% level. However, the DPIN and SDPIN models show stronger significance for the interaction terms between institutional investors and private information in stock prices, with significance at the 5% level, compared to the PIN model, which shows significance at the 10% level. There is evidence that institutional investors constrain earnings distortions. Previous studies show that foreign and institutional ownership can effectively monitor and improve a firm's corporate governance (Chung and Zhang, 2011; He et al. 2013). Especially, these kinds of outside investors are more interested in firm stocks with better managerial performance and better disclosure (Giannetti and Simonov, 2006). Being usually the most diversified investors, institutional investors minimize the risk of firm-specific information and expect to experience only wide-market risk (Farooq and Ahmed, 2014). The findings support this view and add that under the

[23] There are studies examining the effect of investor sentiment on earnings management that also agrees that in a period of optimism with a high stock return, companies are used to inflate their accounting profits to respond to high expectations from speculators (see, e.g., Hurwitz, 2018; Santana et al., 2020).

moderating role of institutional investors, the sensitivity of earnings management to private information seems stronger.

Similar to institutional investors, CEO duality (DUAL) variable proxies for corporate governance characteristics that influence earnings management as documented by previous studies (Halioui et al., 2016; Nuanpradit, 2019). DUAL represents the independence of the company's board. Normally, it is considered poor practice when a firm combines the roles of the CEO and the chairman. The test provides additional support for this viewpoint since the results show positive and significant coefficients for the impact of CEO duality on earnings manipulation. There are more chances for profit manipulation when a firm has the same director chair as the chief executive.

The factor of firm size (SIZE) and profitability (ROE) have statistically significant coefficients at the 1% and 5% levels, thus, having a negative impact on earnings management for all models. The results are consistent with those of Dechow and Dichev (2002) and Warfield et al. (1995). Large S&P 500 firms have fewer incentives to distort earnings because they are in the public interest and followed by a larger number of analysts and investors. Regarding profitability, well-performing firms can use and manage their assets effectively, limiting the adjustment of profits in the business. Firm growth (GROWTH) has no impact on earnings management as all coefficients are statistically insignificant.

3.4.2 Multi-level Model

Tables 3.5 and 3.6 present the results for covariance parameters and estimates for the Fixed Effects, respectively, using multi-level models. Specifically, Table 3.5 shows that all estimated parameters for the 3 levels are statistically significant. This means we reject the null hypothesis of the Wald Z test, inferring that the variation in the level 1 outcome and the intercepts at the firm level and industry level is significantly greater than zero. There is evidence of non-trivial clustering of observation units within firm-level clusters within industry-level clusters. The intraclass correlation coefficients (ICC) are also computed in Table 3.5. ICC (%) is also considered an indicator of whether there is evidence of clustered observations within level 2 and 3 units. Overall, in all the models, ICCs for the firm level account for around 68% to 76% of all three levels, meaning that more than 60% of the variation of earnings management activity occurs between firms. In other words, intrinsic firm characteristics are responsible for the largest proportion of earnings management variation. The following is the observation-unit level with around 20% of earnings management variance, and industry-related attributes have a 5% to 6% effect. Heck et al. (2014) noted that "5% is often considered a "rough cut-off" of evidence of substantial clustering". Based on the results, we can conclude that substantial clustering is found at both between-firm and industry levels. In this case, multi-level models generally show a better fit than the traditional regression (single-level) model.

Table 3.5: This table presents estimates of covariance parameters for multi-level models. PIN, DPIN, and SDPIN are three measures for investors' private information, calculated according to Eq.(13), Eq.(6), and Eq.(8), respectively.

	Null model	Random- intercept and random-coefficient model		
		PIN	DPIN	SDPIN
Parameter Estimation				
Individual Obs.	2.072***	1.992***	7.244***	7.085***
Firm Level	5.260***	8.054***	24.637***	25.926***
Industry Level	0.417*	0.588*	2.049*	2.142*
Parameter Estimation (%)				
Individual Obs.	26.74%	18.73%	21.35%	20.15%
Firm Level	67.88%	75.74%	72.61%	73.75%
Industry Level	5.38%	5.53%	6.04%	6.09%

***Note:** *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

Table 3.6 presents the estimates for the Fixed Effects of the multi-level model. Overall, there is a negative and statistically significant relationship between earnings management and PIN, DPIN, and SDPIN as well when using multi-level models. The multi-level models' outcomes still confirm the hypothesis. In addition, the impact of investors' private information at the firm level (FPI) is also recognized as their coefficients are negative and significant. However, investors' private information at the industry level (IPI) has no significant results under the three models. The results confirm the essential role of industry-level variables in determining earnings management but there are other factors rather than investors' private information at the industry level (IPI).

Table 3.6: This table presents the estimates of the Fixed Effects for multi-level models. PI is the investors' private information, measured by the PIN, DPIN, and SDPIN. DAC is earnings management (also called discretionary accruals) and is computed through the modified Jones model (Raman and Shahrur, 2008); CONTROL variables include the firm size (SIZE) given by market capitalization divided by 10^6; RETURN represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarterly); INST denotes institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; DUAL is CEO duality and is equal 1 if the CEO of the company also serves as the chairman of the board of directors and equal 0 otherwise; GROWTH is firm growth opportunity; ROE presents profitability. Standard errors are in parentheses.

Parameter	Null model	Random-intercept and random-coefficient model		
		PIN	DPIN	SDPIN
Intercept	-0.011 (0.005)	0.448 (0.106)	0.830 (0.291)	0.110 (0.049)
Year Fixed Effects	No	Yes	Yes	Yes
PI		-2.962** (1.507)	-1.084* (0.633)	-0.969** (0.447)
FPI		-1.125* (0.633)	-1.001** (0.475)	-1.267** (0.642)
IPI		0.540 (0.779)	0.692 (1.284)	-1.100 (1.14)
IPI x PI		-3.306 (3.512)	-1.648 (1.525)	-1.019 (1.137)
RETURN		0.185** (0.100)	0.606* (0.304)	0.953* (0.355)
INST		-0.008** (0.0037)	-0.072** (0.036)	-0.059** (0.029)
SIZE		-8.45E-07***	-5.11-06***	-5.64E-06***

	(3.12E-07)	(2.04E-06)	(1.19E-06)
DUAL	0.051**	0.138**	0.147**
	(0.023)	(0.064)	(0.073)
GROWTH	0.004*	0.009*	0.007*
	(0.002)	(0.005)	(0.003)
ROE	0.276**	0.095*	0.114*
	(0.126)	(0.054)	(0.062)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

3.4.3 Robustness

3.4.3.1 Upward and downward earnings management

The concepts of upward and downward earnings management carry distinct meanings and implications, each potentially influenced by a variable in different ways. The combination of them in one regression model can suppress some statistical features and relationships. I carry out additional tests to investigate the effects of investors' private information on upward and downward earnings manipulation separately. Table 3.7 presents the Fix Effects estimates for PIN, DPIN, and SDPIN models during upward and downward earnings management. Overall, DPIN and SDPIN influence both downward and upward earnings management as there are statistically significant negative coefficients at the 1% level. However, PIN exhibits its negative impact only on upward earnings management. No significant correlation is found between PIN and downward earnings manipulation. This outcome demonstrates the better fit of DPIN and SDPIN, indicating their advantage over PIN in calculating the private information of outsiders in higher trading frequencies. In addition, the coefficients for DPIN (-0.739) and SDPIN (-0.212) on upward earnings management are statistically significant at the 5% level; those coefficients for the downward earnings management are -0.184 and -0.133 respectively, and are statistically significant at the 5% and 1% levels, respectively. The impact of the three measures for private information in stock prices on earnings management seems to manifest stronger in the model of upward earnings management than downward earnings management. Although private information of investors can affect managers' decisions on firms' earnings in both cases, managers who systematically inflated earnings in the past may react more strongly to private information.

The factor CEO duality (DUAL) and the joint effect between stock return (RETURN) and private information (PIN and SDPIN) impact upward earnings management in the same pattern as the absolute values of earnings management (DAC). Nevertheless, these control factors show no significant associations with downward earnings management in all three models of PIN, DPIN, and SDPIN for the S&P 500. I find that firm growth (GROWTH) is positively correlated to upward earnings management (using the SDPIN model) but negatively connected to downward earnings management (using the DPIN model). High market-to-book ratios can limit downward earnings management, but the opposite is valid with upward earnings management. These contrasting results on the relationship between market-to-book ratio and earnings management are justifiable by the previous literature. For instance, Beatty and Weber (2003) and Sweeney (1994) find evidence of earnings management to avoid reducing dividends. Growth firms have strong stock market incentives not to miss earnings thresholds (Franz, 2014) because the market penalizes these firms more severely for missing projected earnings targets.

Table 3.7: This table presents results for the relationship between investors' private information and upward and downward earnings management. PI is the investors' private information, measured by the PIN, DPIN, and SDPIN. PIN, DPIN, and SDPIN are the measures for investors' private information, calculated according to Eq.(13), Eq.(6), and Eq.(8), respectively. DAC is earnings management (also called discretionary accruals) and is computed through the modified Jones model (Raman and Shahrur, 2008); CONTROL variables include the firm size (SIZE) given by market capitalization divided by 10^6; RETURN represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarterly); INST denotes institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; DUAL is CEO duality and is equal 1 if the CEO of the company also serves as the chairman of the board of directors and equal 0 otherwise; GROWTH is firm growth opportunity; ROE presents profitability. Standard errors are in parentheses.

	Upward DAC			**Downward DAC**		
	PIN Model	**DPIN Model**	**SDPIN Model**	**PIN Model**	**DPIN Model**	**SDPIN Model**
R-Square	0.183	0.191	0.194	0.093	0.107	0.133
F Test	701.21***	58.09***	132.53***	209.72***	1,294.2***	507.84***
	Coef. (SE)	**Coef. (SE)**	**Coef. (SE)**	**Coef. (SE)**	**Coef. (SE)**	**Coef. (SE)**
_Cons	1.522***	1.149***	1.343***	0.537**	0.669*	0.134**
PI	-1.300*** (0.217)	-0.739** (0.344)	-0.212*** (0.005)	-0.681 (0.705)	-0.184* (0.107)	-0.133* (0.074)
RETURN	0.043*** (0.005)	0.026*** (0.001)	0.016*** (0.001)	0.436* (0.240)	0.061 (0.092)	-0.075* (0.044)
INST	-0.003*** (0.001)	-0.008*** (0.003)	-0.011*** (0.003)	-0.021 (0.046)	-0.547* (0.316)	-0.118* (0.065)
RETURN x PI	1.994* (1.010)	3.572* (2.151)	2.049** (1.018)	-1.957 (2.005)	-2.882 (2.194)	1.627 (2.047)
PI x INST	-1.673* (0.996)	-3.380** (1.602)	-1.522** (0.724)	-1.681* (0.974)	-1.445* (0.900)	-0.635* (0.384)
SIZE	-10.1E-06*** (2.7E-07)	-2.3E-05*** (4.4E-06)	-3.4E-06*** (7.2E-07)	-6.8E-05** (3.3E-05)	-4.2E-05** (2.09E-05)	-7.5E-05** (3.9E-05)
DUAL	0.029** (0.015)	0.055** (0.027)	0.108** (0.049)	-0.970 (1.26)	0.688 (0.492)	0.516 (0.737)
GROWTH	0.114 (0.260)	0.023 (0.029)	0.172* (0.095)	-0.018 (0.034)	-0.331* (0.192)	-0.146 (0.220)
ROE	-0.047 (0.037)	0.229 (0.185)	0.104* (0.059)	0.362 (0.384)	-0.175** (0.082)	-0.068* (0.40)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes

*Note: *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

3.4.3.2 Covid pandemic period

Table 3.8 presents the results of the second robustness check for the period excluding the COVID-19 time. The period of 2018-2021 includes two years of the Covid pandemic 2020 and 2021. During this time the world economy in general and the stock markets in particular were heavily affected. The U.S. stock market has witnessed big drops in the 2nd and 3rd quarters of the year 2020 and still suffers in the following period. Therefore, I provide a robust check for the non-Covid period from 2018 to 2019 to verify the results. Regarding the connection between earnings management and private information, only DPIN and SDPIN measures have significant and negative coefficients. PIN also shows a negative effect on earnings management, but this result is statistically insignificant. The outcomes for variables stock return (RETURN), the interaction term RETURN x PI, and institutional holdings (INST) are still valid as in the main model. However, in the period from 2018 to 2019, institutional investors (INST) presents no significant role in the relationship between earnings management and investors' private information on stock prices. Similarly, firm profitability (ROE) is found to have an insignificant effect on the managers' decisions to manage earnings. The market-to-book ratio (GROWTH), however, has positive coefficients that are statistically significant at the levels of 5% (PIN and SDPIN models) and 10% (DPIN model), indicating that earnings manipulation is more common in firms with higher growth rates.

Table 3.8: This table presents results for the relationship between investors' private information and earnings management in the non-covid time from 2018 to 2019. PI is the investors' private information, measured by the PIN, DPIN, and SDPIN. PIN, DPIN, and SDPIN are the measures for investors' private information, calculated according to Eq.(13), Eq.(6), and Eq.(8), respectively. DAC is earnings management (also called discretionary accruals) and is computed through the modified Jones model (Raman and Shahrur, 2008); CONTROL variables include the firm size (SIZE) given by market capitalization divided by 10^6; RETURN represents stock return which is the buy-and-hold return of a firm's stock over the period studied (quarterly); INST denotes institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; DUAL is CEO duality and is equal 1 if the CEO of the company also serves as the chairman of the board of directors and equal 0 otherwise; GROWTH is firm growth opportunity; ROE presents profitability. Standard errors are in parentheses.

	PIN Model	**DPIN Model**	**SDPIN Model**
R-Square	0.104	0.100	0.132
F Test	189.52***	37.40***	76.09***
	Coef. (SE)	Coef. (SE)	Coef. (SE)
_Cons	1.228	0.099	0.563
PI	-2.008 (1.404)	-0.186** (0.088)	-1.210** (0.545)
RETURN	0.075** (0.038)	0.013** (0.061)	0.096*** (0.005)
INST	-0.046** (0.022)	-0.109** (0.051)	-0.038** (0.016)
RETURN x PI	2.953* (1.772)	1.704* (1.022)	0.947** (0.464)
PI x INST	-0.486 (0.630)	0.089 (3.126)	-0.055 (0.081)
SIZE	-8.9E-06*** (6.7E-07)	-1.5E-06*** (4.4E-07)	-2.3E-06*** (2.2E-07)
DUAL	0.020* (0.011)	0.049** (0.025)	0.040* (0.024)
GROWTH	0.009** (0.004)	0.037* (0.061)	0.095** (0.048)
ROE	0.075** (0.033)	0.588* (0.311)	0.161** (0.075)
Year Dummies	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes

*Note: *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

3.5. Conclusion

This study contributes to the long-standing and important debate on whether financial markets affect the economy through the managerial learning channel. I investigate the effect of investors' private information on the managers' decisions regarding earnings management. By using both the traditional measures and an alternative measure for private information (see Chapter 2), I provide the literature with a more diverse set of measures. I perform multiple tests to estimate the effect of investors' private information on managers' decisions. The results show that over factors related to firms' characteristics, investors' private information is a relevant determinant of managers' decisions on earnings management, regardless of whether the private information is known or unknown to the managers. Therefore, this work provides a new perspective on the managerial learning channel and incentives channel mechanism through which the stock market affects the economy. Outside investors, although not involved in the day-to-day operations of the business, can discover information that is hidden behind the balance sheets and income statements of the firm and/or generate information that is valuable and not yet known by the managers.

The methodology diverges from previous literature by employing a multi-level approach to analyze earnings management. Unlike prior studies that rely on dummy variables for sector analysis, the methodology addresses specific industry factors, including private investor information. This approach allows for a more comprehensive examination, revealing that intrinsic firm characteristics are the primary drivers of earnings management changes. While industry-level variables play a significant role, I found that they primarily stem from factors other than investors' private information.

The findings have relevant practical implications. Specifically, the reciprocal influence resulting from informed trading enhances firms' financial management efficiency by reducing discretionary accruals in accounting profits and therefore, promoting more transparent corporate information disclosures. This aids finance managers in integrating more precise private information into their decision-making processes and prompts regulators to ensure disclosure of information pertaining to trading orders, potentially impacting market efficiency overall. The research has some limitations. First, the analysis is confined to S&P 500 companies. Although the S&P 500 serves as a robust sample for analyzing U.S. market dynamics, representing a significant portion of the U.S. market, the findings may not be generalizable to companies outside the U.S. or in other geographic locations, where market behaviors and dynamics could differ significantly. Second, while quantitative methods are valuable for analyzing the relationship between private information and earnings management, they have inherent limitations in capturing the nuanced qualitative aspects of managerial decision-making. Third, the study does not incorporate an autoregressive component for the dependent variable, earnings management. Including an autoregressive term could help capture the temporal persistence of earnings management decisions, offering insights into how past behavior influences current decisions. The absence of this component may limit the model's ability to fully account for these dynamics, and this limitation is acknowledged in the study.

Additionally, there is a frequency mismatch between the informed trading variables, derived from high-frequency data, and discretionary accrual (DAC) models, which are typically constructed on a quarterly or annual basis. Despite efforts to aggregate high-frequency variables to match the DAC model’s timeframe, this mismatch may still introduce challenges, potentially leading to unexpected or misleading results.

Therefore, future research endeavors could broaden the sample to include companies from other indices and countries. This expansion would facilitate a more comprehensive understanding of the relationship between private information and earnings management across different segments of the market, thereby enhancing the validity and applicability of the findings. In addition, future research could greatly benefit from integrating qualitative methods, such as interviews or case studies, alongside quantitative analyses. This combined approach is crucial as quantitative methods can achieve research objectives that qualitative methods may not, and vice versa (Quinlan et al., 2015; Saunders et al., 2016). Qualitative methods, for instance, allow for a deeper exploration of concepts within the research domain. They offer a unique opportunity to gain insights into managers’ thoughts and experiences, shedding light on how they utilize both stock price behavior and investors’ private information to inform their decisions regarding earnings management. By employing a combination of qualitative and quantitative methodologies, researchers can attain a more comprehensive understanding of how managers respond to private information. This holistic approach provides deeper insights into their decision-making processes and unveils the implications for earnings management.

Chapter 4. Recent Advances in the Relationship between Stock Liquidity and Informed Trading

4.1. Introduction

The relationship between information asymmetry, market efficiency, and liquidity is a cornerstone of financial economics research. In *"The Cost of Transacting"*, Demsetz (1968) highlights the role of transaction costs, such as the bid-ask spread, in shaping market functioning, liquidity, and efficiency. He identifies these costs as compensation for market makers who facilitate trades, manage inventories, and mitigate risks. The study reveals that transaction costs are influenced by factors such as trading volume, price volatility, and market structure. Higher trading volumes lower transaction costs through economies of scale, while increased price volatility raises costs due to heightened risk exposure for market makers. Moreover, competition among market participants tends to reduce transaction costs, fostering a more efficient market environment. This work laid the foundation for subsequent research into the interplay between transaction costs, liquidity, and market dynamics, forming a crucial part of market microstructure theory.

Building on these insights, other classical microstructure models, such as those of Kyle (1985), Glosten and Milgrom (1985), and Easley and O'Hara (1992, 1997a,b), provide critical frameworks for understanding how information asymmetry, especially informed trading, influences liquidity and market dynamics. Informed traders are equipped with superior knowledge about a stock's value and prospects for the firm's business which gives them a competitive advantage in some trading decisions. Hence, the existence of informed trading may enhance market efficiency, leading stock prices to be closer to their "true" values. But informed trading also creates an environment where liquidity providers become more wary due to adverse selection and increase transaction costs to protect themselves. Consequently, other traders may trade less frequently (Glosten and Milgrom, 1985), so there is a reduction in trading activity and an increase in transaction costs that, collectively, surpass the liquidity enhancement created by the existence of informed trading. Therefore, overall, the market liquidity decreases, and I conclude that informed trading penalizes liquidity.

The argument presented above has been made considering traditional financial markets, where liquidity is maintained by market makers or specialists[24] who engage in continuous buying and selling activities to ensure smooth and orderly trading, which, in turn, aligns with their own interests as market

[24] Both market makers and specialists provide liquidity and facilitate trading, market makers operate in a competitive environment across a range of markets and securities, focusing on profiting from the bid-ask spread. Specialists, on the other hand, are assigned specific securities on an exchange and have a more central role in ensuring orderly trading and market stability for those securities. I note that, in modern exchanges such as the NYSE, specialists are often referred to as Designated Market Makers (DMMs) and these entities have similar responsibilities as the traditional specialists.

participants. These intermediaries play a vital role in guaranteeing the presence of both buyers and sellers. However, a different scenario emerges in the current modern stock market, which has evolved significantly due to technological advancements and changes in market structure.

In modern markets[25], where financial markets rely on computerized systems for matching a buy with a sell order on an electronic trading platform, the order book is continually updated in real-time as traders submit, adjust, or cancel their trade orders. This continuous order book updating is instrumental in ensuring that there are available orders to facilitate trade execution, thus contributing to overall market liquidity. Some studies in the literature provide compelling evidence that high-frequency trading[26], one characteristic of automated market, significantly enhances market liquidity (Ammar et al., 2020; Heng et al, 2020). In this context, the role of maintaining liquidity differs significantly from that in traditional markets. In modern markets, there might not be a single specialist or a group of dealers tasked with the responsibility of providing liquidity. Instead, liquidity results from the collective actions of various traders who place limit orders[27] at different prices. Importantly, these collective liquidity providers include both informed and uninformed traders and those who are informed have an informational advantage and may adopt an aggressive trading strategy to maximize profits and compensate for the resources and efforts they have invested in acquiring private information (Agudelo et al., 2015, Wong et al., 2009). This peculiar characteristic of modern markets has the potential to fundamentally alter the relationship between informed trading and liquidity. To show whether that is the case is the goal of this paper.

Particularly noteworthy is the introduction of anonymity in certain financial markets, which adds a layer of complexity to the connection between informed trading and liquidity. For instance, Dufour and Engle (2000) have uncovered evidence suggesting an increased presence of informed traders during the most active periods in the New York Stock Exchange (NYSE). In contrast, no such evidence was found in the case of infrequently traded stocks, as indicated by Manganelli (2005). The

---

[25] The modern stock market is often described by few different terms, depending on the specific aspect that I want to emphasize, such as Electronic Stock Market, Digital Stock Market, and Automated Stock Market. The Electronic Stock Market refers to the use of electronic trading platforms, where transactions are conducted digitally rather than on traditional trading floors. The Digital Stock Market emphasizes the integration of digital technologies, including online trading platforms and digital brokerage services, which have revolutionized stock trading. The Automated Stock Market highlights the use of automated systems and sophisticated algorithms, encompassing practices like high-frequency trading and algorithmic trading, to execute trades with minimal human intervention.

[26] Also known as algo or algorithmic trading and refers to computerized trading using proprietary algorithms. There are two types of high frequency trading. Execution trading is when an order (often a large order) is executed via a computerized algorithm. The program is designed to get the best possible price. It may split the order into smaller pieces and execute at different times. The second type of high frequency trading is not executing a set order but looking for small trading opportunities in the market. It is estimated that more than 50 percent of stock trading volume in the U.S. is currently being driven by computer-backed high frequency trading.

[27] When an investor places an order to buy or sell a stock, they can choose between "market order" and "limit order". Market orders are transactions meant to execute as quickly as possible at the current market price. Limit orders set the maximum or minimum price at which the investor is willing to buy or sell. However, there is no assurance of execution. Limit orders are designed to give investors more control over the buying and selling prices of their trades.

shift towards modern markets represents a fundamental change in the way financial markets operate. It does not only alter the dynamics of liquidity provision but also presents a new context for examining the relationship between informed trading and liquidity. Although there is extensive literature focusing on stock liquidity (Armanious and Zhao, 2024; Elshandidy and Elsayed, 2024; Marks and Shang, 2024; Yao and Qiu, 2024) and informed trading (Le et al., 2019; Pedraza, 2020), few studies have examined whether the relationship between stock liquidity and informed trading. Most of the existing literature relies on data collected from periods when algorithm trading was not yet popular. Considering the prevalence of modern markets these days, it is important to study the aforementioned relationship. For which I use different data timespans, including daily, weekly, monthly, quarterly, and yearly data to observe patterns in the relationship.

In the existing literature, there are various measures for informed trading, such as the stock price non-synchronicity (Roll, 1988), probability of informed trading (PIN) (Easley et al., 1997b, 2002), probability of informed trading (DPIN) (Chang et al., 2014), and probability of informed trading with size effects (SDPIN) (see Chapter 2).[28] Among these measures, the latter two are dynamic private information indicators and are particularly valuable in the realm of high-speed trading. DPIN and SDPIN allow us to estimate the probability of informed trading at much finer frequencies, specifically within the trading day. Since such frequencies are more in line with the speed at which traders react to and digest information in modern financial markets. Therefore, DPIN and SDPIN may be better suited to more accurately capture information-based trading activity at higher frequencies, compared to other existing models. Therefore, I use those dynamic measures (DPIN and SDPIN) to estimate the proportion of informed trading within the new (automated) stock trade market environment.

In the regression analyses, I use various timeframes - daily, weekly, monthly, quarterly, and yearly. The goal is to examine whether the results hold across these different data frequencies. Specifically, I find that informed trading enhances stock liquidity when I use daily and weekly, but that such an effect does not hold when I use monthly, quarterly, and annual). The finding holds relevance within the context of the current market dynamics. In today's rapidly evolving financial landscape, characterized by technological advancements, algorithmic trading, and regulatory uncertainty, the impact of informed trading on liquidity may vary across different time horizons. The results for the daily and weekly timeframes align with the notion that short-term trading activities, driven by informed investors, contribute to market liquidity by enhancing trading volumes and market depth. The absence of a statistically significant relationship between informed trading and stock liquidity, when I use

[28] SDPIN follows method of DPIN by Chang et al. (2014) regarding the identification of informed trades, but it considers the effect of (large, medium, and small) the trade order sizes.

monthly, quarterly, and yearly data, suggests that other factors may play a role in such a relationship in the medium and long term.

I also find that time constraints and market sentiment significantly moderate the relationship between informed trading and liquidity. During periods of heightened market sentiment, the impact of informed trading on liquidity increases, leading to more pronounced changes in bid-ask spreads and trading volumes. Time constraints further amplify the effect of informed trading on stock liquidity during specific trading windows such as during economic announcements.

Furthermore, the results show that the new SDPIN liquidity measure performs better than the DPIN liquidity measure in some cases. The SDPIN measure is a dynamic measure for the probability of informed trading considering the trade size effect; it builds upon the foundations laid by Chang et al. (2014) with the probabilities of informed trading (DPIN) measure but also takes into account the trade order size and volume.

Regarding liquidity measures, the outcomes also highlight that high-frequency measures are superior in tracking liquidity in shorter timeframes (daily and weekly data), while low-frequency measures like Amih are more effective in analyzing longer-term liquidity trends over monthly, quarterly, and yearly periods. This distinction underscores the importance of selecting the appropriate liquidity measure depending on the specific time horizon being analyzed.

This study has several contributions. Firstly, it addresses a significant gap in the literature by examining the dynamics of informed trading and its impact on stock liquidity considering modern stock trading conditions. I adopt the DPIN and SDPIN as proxies for high-speed trading among informed participants which has a particular relevance in the contemporary financial markets. This approach helps mitigate measurement errors, caused by other traditional measures such as PIN by Easley et al., (1997b, 2012) that are systematically linked to market conditions,[29] thereby yielding more accurate results (Putnins and Michayluk, 2018). Secondly, the findings reveal that there is a positive impact of informed trading on stock liquidity, recognizing the fundamental shifts in this relationship from traditional market setups. Understanding how this relationship evolves over time and across different data timeframes is also important for all market participants and policymakers. To specify, the findings, may help to increase investors' confidence as informed trading is recognized as beneficial for liquidity, at least on a daily and weekly trading basis. As a result, they might adjust their trading strategies to incorporate more informed trading practices. Regulators might reconsider some of the stringent regulations aimed

[29] The traditional PIN measure has some well-known limiting features. Most notably, in order to estimate PIN, one must aggregate very fine intraday data, which occur at approximately five-minute intervals within the trading day, across multiple days (Easley et al., 1997b). The resulting estimate measures informed trading over a very long macro horizon — typically from one month to one year. Arguably, the variation and information content of intraday trades is diluted, or possibly even lost, when combining over such long time periods, especially in modern financial markets where information is short-lived and traders act with increasing alacrity (Chang et al., 2014).

at curbing informed trading, potentially leading to more nuanced and balanced regulations that protect market integrity while promoting liquidity. Also, the findings could stimulate further academic investigation into the intricacies of this relationship, prompting more studies to re-evaluate the relationship. This could lead to a paradigm shift in how scholars understand the dynamics between informed trading and liquidity.

The remaining of the paper proceeds as follows. Section 2 provides the relevant literature review and states the research hypotheses. Section 3 presents the data sample and research methodology. Section 4 presents the empirical findings. Section 5 concludes the study.

4.2. Literature Review

This section reviews the literature on liquidity and its relationship with informed trading. It begins by discussing the concept of liquidity and examining its key characteristics, including depth, tightness, immediacy, and resiliency. The review then explores various liquidity measures, such as transaction cost-based, volume-based, price impact-based, and multidimensional indicators. Next, the literature on the relationship between liquidity and informed trading is examined, highlighting the complex interactions between informed trading and liquidity in contemporary markets. Based on these insights, hypotheses are proposed regarding the dynamic impact of informed trading on liquidity.

4.2.1. Liquidity

Liquidity and its associated issues have long been central topics in financial literature, drawing significant attention from researchers (e.g., Acharya & Pedersen, 2005; Amihud, 2002; Amihud et al., 2005; Chordia et al., 2000; Gregoriou, 2013). Despite the extensive body of work on liquidity and related concepts, a universally accepted definition and a standardized liquidity measure for all markets remain elusive. This is largely due to the multifaceted nature of liquidity, which includes factors such as trade volume, trading speed, and price impact. One commonly accepted definition of liquidity by Liu (2006) describes liquidity as the ability to trade large volumes of stocks quickly, at a low cost, and with minimal price impact.

Amihud and Mendelson (2012) and Sarr and Lybek (2002) identify four key characteristics of liquid markets: (1) Depth, which refers to the ability to trade a large volume of assets without significantly impacting their quoted prices; (2) Tightness, associated with low transaction costs, which enables the simultaneous buying and selling of assets at similar prices; (3) Immediacy, indicating the efficiency of trading, or how quickly orders can be executed; and (4) Resiliency, the market's ability to absorb new orders swiftly and correct imbalances. This definition shows that liquidity has four different aspects (trade costs, trading volume, speed of trading, and price impact). However, it is crucial to acknowledge that no single liquidity measure fully captures all these dimensions.

Based on the different aspects they capture, liquidity measures can be classified into several types: Transaction cost-based measures, Volume-based measures, Price impact-based measures, and Multidimension-based measures.

Transaction cost-based measures focus on the costs incurred when executing trades. These can be further divided into explicit costs (e.g., fees) and implicit costs, which are less visible but often more significant. Implicit costs include factors such as bid-ask spreads, transaction size, and trade execution timing. Bid-ask spreads, for instance, reflect the price discrepancy between buying and selling a security, and they are influenced by market makers' order processing, information asymmetry, and inventory costs. Measures like Roll's (1984) implicit effective spread attempt to capture these hidden costs, while other measures, such as the High-Low spread by Corwin and Schultz (2012), utilize price volatility to estimate liquidity.

Volume-based measures, on the other hand, use trading volume as a proxy for liquidity. Higher trading volumes typically indicate more liquid markets, as large numbers of trades help reduce transaction costs. One such measure is the turnover ratio, which compares the number of shares traded to the total shares outstanding, offering insights into how frequently a stock is traded (Gabrielsen et al., 2011). A higher turnover ratio usually correlates with better liquidity. However, volume-based measures can have limitations, such as not accounting for the frequency of trades and potential impacts of large, irregular orders, which may distort liquidity assessments (Gabrielsen et al., 2011).

Price impact-based measures focus on how price changes relate to trading volume. Amihud's (2002) illiquidity ratio is one of the most commonly used measures in this category, showing how sensitive a stock's price is to changes in volume. This method assumes that higher trading volume leads to more stable prices and reduced liquidity costs. However, the illiquidity ratio has some drawbacks, including a size bias, as larger companies tend to have lower illiquidity measures due to their market size. Other measures, such as the Florackis et al. (2011) price impact ratio, called Return to turnover ratio (RtoTR), aim to address these biases by incorporating factors like the turnover ratio, thus improving the accuracy of liquidity assessments.

Multidimension-based measures, like Liu's (2006) turnover-adjusted number of zero daily trading volume, combine multiple liquidity dimensions. These measures consider factors such as the quantity of trades, the speed at which trades occur, and transaction costs. By focusing on zero-volume days and adjusting for turnover, these measures aim to capture a broader picture of liquidity, reflecting not only the volume and price impacts but also the timing and frequency of trading activity. This approach offers a more comprehensive view of liquidity, incorporating both the costs and the availability of assets for trade, thus providing a deeper insight into market behavior.

Based on data frequency, liquidity proxies can be categorized into two main strands: high-frequency (intraday) measures and low-frequency (daily) measures (Le and Gregoriou, 2020).

High-frequency liquidity measures are derived from intraday data, typically recorded at very short intervals (often at the millisecond or second level). These measures capture real-time liquidity dynamics, making them particularly useful for analyzing market microstructure and short-term trading behavior. They require advanced computational resources and programming due to the large volume of transactions involved, making them more applicable to well-established markets, such as those in the U.S. (e.g., Huang and Stoll, 1997; Hasbrouck, 2009). Common examples of high-frequency measures include the Bid-Ask Spread, which measures the difference between the buying and selling price at any given time. Short-term fluctuations in this spread are essential for understanding immediate liquidity. Market Depth, which examines the number of orders at different price levels near the current market price, provides insight into how quickly liquidity can be absorbed at various levels. Price Impact measures how much the price changes in response to a trade, with a large price impact indicating lower liquidity, as trades significantly move prices. Trade Volume (Tick-by-Tick) analyzes the frequency and size of trades over very short periods, assessing liquidity in terms of trade activity.

High-frequency measures are characterized by granular data that can reveal short-term price movements. They are sensitive to market microstructure, allowing for a detailed understanding of liquidity at each tick or transaction level. These measures are often employed in algorithmic trading and for studying intraday market dynamics. However, their use requires significant computational power, limiting their widespread application to well-established markets with ample data infrastructure.

On the other hand, low-frequency liquidity proxies are based on data collected over longer periods, such as daily, weekly, or monthly intervals. These measures are more widely accessible and applicable across various market types, including emerging markets, since they do not require the high computational power needed for high-frequency analysis. They provide a broader view of liquidity, reflecting longer-term liquidity conditions and trends, which is useful for assessing the overall health and stability of markets. Examples of low-frequency measures include the Amihud Illiquidity Measure (Daily), which measures the daily price impact per unit of volume and is commonly used to assess illiquidity over longer time horizons. The Turnover Ratio is the ratio of trading volume to the number of shares outstanding or market capitalization, capturing asset trading frequency over time. Volatility-Based Measures, such as Roll's (1984) Spread Estimate, derive from the relationship between price volatility and trading volume over extended periods. The Effective Spread is calculated from daily prices and reflects the difference between the actual transaction price and the midpoint of the bid-ask spread over a longer time frame.

Low-frequency measures are characterized by their ability to provide an aggregate view of liquidity that is less sensitive to microstructural factors. They are useful for analyzing long-term trends and overall market stability rather than real-time fluctuations. These measures are commonly employed in long-term investment and portfolio management analysis, offering valuable insights into the liquidity environment of various markets.

4.2.2. Informed Trading and Liquidity

Informed traders possess an informational advantage and use this advantage to enhance their trade profits. Thus, informational advantage should in principle contribute to a greater market efficiency since price movements toward the fundamental value of the asset are more likely. However, the presence of information asymmetry adds a few challenges to the liquidity providers.

Market microstructure theory posits that informed trading reduces liquidity. The market maker widens the bid-ask spread and/or increases the cost of large trades, anticipating the adverse selection problem faced for trading with informed traders (Hasbrouck, 1991; Kyle, 1985). It means that they must transact with a mix of informed and uninformed traders, with no easy way to distinguish them. A consequence of the adverse selection is that the liquidity providers are more likely to trade with the informed traders, who by definition are better informed about the true value of the asset that is being traded (Kyle, 1985). In this regard, Easley et al. (2010) derive a theoretical relationship between the probability of informed trading (PIN) and the bid-ask spreads, and Easley et al. (1997b, 1996) provide empirical evidence that this relationship is negative, relying on U.S. stocks. They assert that when there is a higher likelihood of informed trading, the liquidity providers are more cautious, so they require greater compensation for the risk of adverse selection, which leads to a rise in transaction costs and a decrease in stock liquidity. That is, the higher the number of informed traders, the less willing are the liquidity providers to participate in the market.

Furthermore, information asymmetry allows informed traders to earn extra returns at the expense of uninformed traders (Glosten and Milgrom, 1985; Easley and O'Hara, 1992), since uninformed traders may trade at prices that are less favorable due to the presence of informed traders. This phenomenon is typically referred to as the "winner's curse" because uninformed traders may unknowingly be on the losing side of a trade when they transact with informed traders.[30] Some studies find empirical evidence for a negative relationship between informed trading and stock liquidity. For instance, Agudelo et al. (2015) test this implication using the PIN model (Easley et al., 2008) as a time-

[30] The "winner's curse" refers to a scenario in which uninformed traders, due to information asymmetry, inadvertently find themselves on the losing end of a transaction when they interact with informed traders. Essentially, uninformed traders may unknowingly agree to transactions at less favorable prices or terms, not realizing that better-informed traders have more accurate information about the asset's value or market conditions (Amyx and Luehlfing, 2006; Foreman and Murnighan, 1996).

varying measure of informed trading in the six largest Latin American stock markets. Under alternative specifications and robustness tests, the results suggest that signed dynamic PIN is related to returns, as a proxy for information asymmetry rather than just liquidity effects. These results contribute to the ongoing discussion on whether PIN is a valid informed trading measure, and to a better understanding of price formation in emerging markets. Wong et al. (2009) investigate the issue of informed trading and its relation to liquidity in the Shanghai Stock Exchange. Consistent with the hypothesis that information-based trade exists for all stocks, their findings suggest an increased presence of informed trading in both liquid and illiquid stocks when markets are active. Moreover, for the actively traded stocks, the results support the price formation model of Foster and Viswanathan (1990) that activities of informed traders deter uninformed investors from trading, thereby reducing market liquidity. Chung et al. (2005), using the PIN measure, provide evidence that larger information asymmetry increases the price impact of trades. Finally, both Lei and Wu (2005) and Easley et al. (2008) proceed further, by using the time-varying PIN estimations to predict bid-ask spreads and provide evidence on the positive correlation between PIN and bid-ask spread or negative link between PIN and stock liquidity.

However, more recently, Collin-Dufresne and Fos (2015) present a contrasting perspective using hedge funds data. It is argued that activist hedge funds privately increase their ownership in target firms before they are required to publicly disclose their ownership positions and strategically time their trades to avoid illiquidity. Thus, they find a positive relationship between liquidity and informed trading. Hedge funds and proprietary traders can make investments that allow them to analyze the fundamental valuation of the firm. Financial intermediaries can invest in information and trading networks that help them acquire supply and demand information for a security’s valuation. All traders recognize large financial intermediaries by name, so that when traders’ identity is displayed in such a market, uninformed traders get information at no cost, thus becoming more informed themselves. Consequently, they make more aggressive limit orders, increasing liquidity. Hence, uninformed market makers can be considered as liquidity suppliers thanks to the effect of pre-trade transparency. Moreover, Collin-Dufresne and Fos (2015) claim that if the private information is long-lived, informed investors will choose to trade when liquidity is high to limit the price impact of their trades which leads to a positive relationship between informed trading and liquidity.

Prior to Collin-Dufresne and Fos (2015) Rindi (2008) also poses a similar view while studying the impact of pre-trade transparency on liquidity in a market where risk-averse traders accommodate the liquidity demand of noise traders. According to the study, when some risk-averse investors become informed, an adverse selection problem ensues for others, making them reluctant to supply liquidity. Hence the disclosure of traders’ identities improves liquidity by mitigating adverse selection. However, informed investors are effective liquidity suppliers, as their adverse selection and inventory costs are minimized. With endogenous information acquisition, transparency reduces the number of informed investors, thus decreasing liquidity as Mattias Levin wrote in the CEPS Task Force report (2003) “Overall, transparency is no panacea and there is “disquieting evidence” that too much transparency may harm the market quality, as it effectively disables some liquidity provision.”

The recent literature examining the relationship between informed trading and stock liquidity is notably sparse. Putnins and Michayluk (2018) and Rosu (2020) are among the few contributions on this topic. Both studies emphasize the essential role of informed traders in providing limit orders, a key feature of modern financial exchanges worldwide. Rosu (2020) finds that informed traders extensively use limit orders. He devises a theoretical model to explore the connection between the activity of informed traders and stock liquidity, concluding that a larger fraction of informed traders generally improves liquidity—using the bid-ask spread and market resiliency as proxies for liquidity—but has no effect on the price impact of orders. This result is driven by two key model features: competition among informed traders, where each trader considers the future arrivals of other informed traders, and the long-lived nature of private information, as information about fundamental value is revealed to the public only through order flow. Consequently, a larger share of informed traders can enhance market efficiency to a level that it offsets the effect of an increase in adverse selection.

According to Putnins and Michayluk (2018), with the demise of traditional market makers and the proliferation of trade execution algorithms that mix market and limit orders, it is no longer clear who provides liquidity in limit order book markets and what determines their liquidity provision decisions. They develop two measures for informed traders’ order choice and test them using data collected from the Australian Stock Exchange (ASX), covering a time period between 2008 and 2011. They conclude that informed traders use limit orders and market orders, and informed traders provide liquidity with limit orders during periods of high stock price uncertainty. During such times, mispricing persists for longer, therefore, informed traders are more patiently trying to obtain better execution prices, higher selling prices, and lower buying prices, with informed traders’ order choice acting as an uncertainty multiplier.

Ahern (2020) investigates the impact of liquidity on informed trading. He tests whether the existing stock liquidity measures enable us to detect informed trading, using insider trading as a measure for informed trading. The study shows that trading volume is positively related to the intensity of informed trading, whereas there is no relationship between absolute returns and market factors. This positive relationship, between trading volume and insider trading, is consistent with the idea that insider trading increases trading volume above its historical firm-average and cross-sectional event-day average. It is noted, however, that in the realm of real-world dynamics, firms targeted by informed traders are not randomly selected. Initially, informed trading hinges on possessing exclusive, valuable insights, a trait that varies among firms in a non-random manner. For example, information signaling a firm's vulnerability to acquisition holds significant value, yet firms are not chosen randomly as potential targets. Additionally, the dissemination of information follows a pattern rather than occurring haphazardly. Some firms may heavily rely on external contractors who are inclined to distribute privileged information. Given the non-random allocation of informed trading across firms, the exclusion of a variable could precipitate both liquidity and informed trading. For instance, high-tech firms may exhibit a heightened propensity for informed trading due to their increased likelihood of attracting acquisition interest. Concurrently, these same firms may also face heightened liquidity due to unrelated factors like a scarcity of institutional investor participation. Consequently, omitted variables have the potential to fabricate a misleading association between illiquidity and informed trading.

4.2.3. Hypothesis Development

It is essential to consider that the impact of informed trading on stock liquidity can vary depending on the nature of the informed information, the trading strategies used, stock market conditions, and the actions of other market participants. Moreover, modern markets, with their rapid execution and real-time data processing, can either magnify or mitigate these effects compared to traditional markets. Additionally, market regulators play a critical role in monitoring and addressing informed trading to maintain the integrity and fairness of the financial markets. The effect of informed trading on liquidity in modern markets is a complex interplay that involves several dynamics; informed trading, which is driven by traders possessing non-public information, can have short-term and long-term impacts on stock liquidity. Given the current landscape dominated by high-speed trading, the focus lies predominantly on the short-term relationship between these two factors. That is, in the short term, informed trading can lead to a temporary increase in liquidity in the context of a modern market with high-speed trading (Putnins and Michayluk, 2018). The execution of stock trade orders by informed traders in automated markets leads to a temporary increase in liquidity because of the efficiency and speed at which information is processed and acted upon and because these informed traders have access to sophisticated IT tools that allow them to execute trades more quickly and with more precision. On the other hand, by leveraging real-time data feeds and using advanced analytical tools and techniques, they are able to make more accurately anticipate market movements and adjustments in their trading strategies, which enhances the market depth and reduces bid-ask spreads, and as a result, increases the stock liquidity.

Furthermore, the presence of informed traders can attract liquidity providers who seek to profit from market inefficiencies or arbitrage opportunities. In automated markets, these liquidity providers may include market makers and algorithmic trading firms, which play a vital role in maintaining orderly trading conditions by continuously quoting bid and ask prices. Informed trading activity can stimulate the participation of these liquidity providers, thereby enhancing overall market liquidity and reducing trading costs for other market participants. Also, other market participants, who may not have access to the same information, might be attracted to the trading activity, too. These participants might be drawn to the heightened trading activity, either looking to trade with or against the informed traders. This sudden influx of interest can lead to a transient increase in trading volumes and liquidity.

Informed traders often want to execute their orders quickly and may be willing to trade at the existing market prices. This can lead to narrower bid-ask spreads as they provide liquidity by placing market orders (Chung et al., 1999). In an automated market, this can reduce the cost of trading for other participants. The presence of informed traders can lead to price movements, and the magnitude of these price changes can influence short-term liquidity. Rapid price changes can lead to increased trading activity as traders respond to the new information, either amplifying or mitigating the short-term liquidity effects. Overall, this initial burst of activity generated by informed traders can create a temporary boost in liquidity as other market participants respond to the new information or increased trading activity (Hayashi and Nishide, 2024; Hirschey, 2021).

Informed trading can be effectively captured by examining its influence on price impact and its subsequent effect on liquidity. As informed traders enter the market with private or superior information, they contribute to more efficient price discovery, which enhances overall liquidity. By making trades based on superior knowledge, informed traders help correct any mispricing, leading to a more accurate reflection of a security's true value. This process often results in narrower bid-ask spreads, a key indicator of improved liquidity, as market makers adjust their prices to better align with the asset's fundamental value. Furthermore, informed trading can increase market depth by providing liquidity to other participants through larger, well-informed trades, which facilitates quicker execution and reduces transaction costs. Although larger trades typically exert a greater price impact, informed traders are able to execute these transactions without significantly disrupting market prices, as their trades are consistent with the underlying market fundamentals. In addition, informed traders may also utilize limit orders to manage their trades strategically, which further mitigates price impact by providing liquidity at specified price levels. These limit orders can help reduce market volatility by absorbing the supply and demand imbalances, ultimately contributing to more stable and efficient markets. This enhanced price discovery process, supported by the use of limit orders, contributes to a more stable and liquid market environment.

The Amihud illiquidity measure, which quantifies price impact relative to trading volume, further reinforces this positive relationship. In markets where informed trading is prevalent, price changes become more predictable and less volatile, leading to a more liquid market in which participants can execute transactions at favorable prices. As informed trading increases, the volume of trades that align with the true value of the asset tends to rise, which can reduce the Amihud illiquidity measure, signaling an improvement in liquidity. However, this relationship is context-dependent: while informed trading can reduce bid-ask spreads and improve price efficiency over time, it may also induce short-term volatility, temporarily increasing the Amihud measure due to rapid price changes.

Consequently, informed trading may lead to both short-term fluctuations in liquidity as well as long-term improvements, depending on factors such as the frequency and size of trades and the market's capacity to absorb them without significant price distortions.

Therefore, I posit that within a short timeframe, the entry of informed traders into the market might lead to a temporary surge in liquidity. Building upon this premise, we, first, propose the first hypothesis **H1** on the positive impact of informed trading on stock liquidity for daily data frequency as we, first, measure both of these variables on a daily basis. Daily measurements offer readily available data, facilitating statistical analysis and a comprehensive understanding of market dynamics, including the impact of news events and the effectiveness of price discovery mechanisms.

Hypothesis 1 (H1): *Informed trading is positively related to stock liquidity, relying on daily data.*

Expanding on this overarching hypothesis, I continue to examine the Hypotheses H1A and H1B for other short timeframes, including weekly and monthly frequencies, as the following:

Hypothesis 1A (H1A): *Informed trading has a positive effect on stock liquidity, relying on weekly data.*

Hypothesis 1B (H1B): *Informed trading has a positive effect on stock liquidity, relying on monthly data.*

4.3. Methodology

This section outlines the data used and methodology designed to test the hypothesis regarding the impact of informed trading on liquidity, specifically examining how private information—measured by the DPIN and SDPIN indices—affects various liquidity measures in financial markets. The research posits that within a short timeframe (daily, weekly, and monthly), the entry of informed traders into the market may lead to a temporary surge in liquidity. To empirically test these hypotheses, a series of regression models will be employed, with liquidity measures serving as the dependent variables and DPIN, SDPIN, along with other control variables, included as key independent predictors. The following subsections will detail the model specifications, define the variables, and outline the data sources used in the analysis.

4.3.1. Data Sample

The paper uses a data sample that comprises the S&P 500 stocks. With the increasing adoption of automated trading systems and algorithms, the dynamics of the U.S. stock market have undergone significant transformation. Therefore, the S&P 500, as a representative index of the U.S. equity market, offers us valuable insights into how automated trading strategies impact market performance and liquidity across various sectors. In the realm of automated markets, where trades are executed by computers guided by algorithms based on predefined rules, the S&P 500 is a good benchmark for evaluating the efficacy of automated trade systems. The firm data sample is collected from Bloomberg and “Wharton Research Data Services” (WRDS), whereas the intraday trading data was collected from the “Trade and Quote” (TAQ). Notice that, for a stock to be included in the sample, the firm has to be in the S&P 500 for at least a year. Overall, the data sample comprises 475 firms covering the time period between 2018 and 2021.

4.3.2. Liquidity Measure

The study uses four measures for stock liquidity: Bid-Ask spread (Spread), Effective spread (ESpread), Amihud liquidity measure (Amih), and Volume-based Amihud liquidity (VAmih).[31] The selection of liquidity measures in this study—Amihud illiquidity measure (Amih), Volume-based Amihud illiquidity (VAmih), Bid-Ask spread (Spread), and Effective spread (ESpread)—is driven by their ability to capture different aspects of market liquidity, particularly in the context of informed trading and high-frequency trading (HFT).

Bid-Ask spread (Spread)

The percent quoted Bid-Ask spread is calculated as follows:

$$Spread = \frac{A_t - B_t}{M_t} \tag{35}$$

where *Spread* is Bid-Ask spread, A_t is the best Ask quoted at time t, B_t is the best Bid quoted at time t, and M_t is the midpoint of A_t and B_t.

I follow Holden and Jacobsen (2014), aggregating the spread at the daily level by computing the time-weighted average of the intraday quoted spreads. Each observed intraday spread is weighted by the duration it remains active.

Effective spread (ESpread)

The bid-ask spread, as quoted, assumes trades exclusively at the ask or bid prices. However, the effective spread offers a more nuanced view, accounting for trades occurring within this quoted spread range. By subtracting the midpoint of the National Best Bid and Offer (NBBO) quotes from the trading price, the effective spread estimates the discrepancy between the actual transaction price and a proxy for the true price (i.e., the midpoint). This discrepancy, indicative of market friction, is then

[31] The Bid-Ask spread (Spread), Effective spread (ESpread), Amihud illiquidity measure (Amih), and the Volume-based Amihud illiquidity (VAmih) are direct measures of illiquidity or reverse measures of liquidity. A stock that has large Spreads, ESpreads or higher values of Amih and VAmih means that is less liquid.

doubled to derive the complete spread. Consequently, the percentage effective spread per trade is computed by dividing this doubled difference by the trading price.

$$ESpread = 2\frac{|P_k - M_k|}{M_k} \quad (36)$$

Where P_k is the executed price of order k, and M_k is the midpoint of the quotes prevailing when order k occurs. The effective spread is aggregated to the daily level by taking the dollar-volume-weighted average of the effective spread across all trades per day.

The Bid-Ask spread (Spread) and Effective spread (ESpread) are chosen for their high-frequency nature and relevance in capturing liquidity dynamics in environments characterized by high-frequency trading. These measures are more sensitive to short-term price movements, which are common in high-frequency trading strategies. The Bid-Ask spread reflects the immediate liquidity available to traders, with narrower spreads indicating higher liquidity, while the Effective spread provides a more accurate picture of trading costs by accounting for the actual execution price relative to the best quotes. These measures are directly aligned with high-frequency trading contexts, where informed traders can quickly adjust prices and liquidity in response to market conditions.

Amihud liquidity measure (Amih)

I construct the Amihud illiquidity measure (*Amih*) following Amihud (2002), defined as the absolute value of stock return divided by the dollar trading volume on a given trading day.

$$Amih = \frac{|R_t|}{DV_t} \quad (37)$$

Where *Amih* is the Amihud illiquidity measure, R_t is the stock return on the trading day *t*, DV_t is the dollar volume on the trading day *t*.

The Amihud illiquidity measure (Amih) has both advantages and disadvantages. According to Le and Gregoriou (2020), this low-frequency liquidity measure is effective at capturing liquidity benchmarks based on intraday data. However, it cannot compare stocks with different market capitalizations, resulting in a size bias. This means that small-cap stocks are automatically characterized as "illiquid" due to their size (Azevedo et al., 2014; Florackis et al., 2011; Le and Gregoriou, 2022).

The volume-based Amihud liquidity (VAmih)

The volume-based Amihud illiquidity (*VAmih*) is calculated as the absolute value of stock return divided by the number of shares traded on a given trading day.

$$VAmih = \frac{|R_t|}{V_t} \quad (38)$$

Where $VAmih$ the volume-based Amihud illiquidity, R_t is the stock return on the trading day t, V_t is the trading volume on day t.

Although the Amihud illiquidity measure (Amih) and Volume-based Amihud illiquidity (VAmih) are not high-frequency measures per se, they are still highly relevant in this study for capturing the broader price-impact aspects of liquidity. These measures are particularly effective in assessing how large trades, often associated with informed trading, influence price changes in the market. The Amihud illiquidity measure (Amih), which calculates the absolute stock return relative to dollar trading volume, reflects the price impact of trades, helping to gauge liquidity at a daily level. Similarly, the Volume-based Amihud illiquidity (VAmih), which uses the number of shares traded, adjusts the liquidity measure to be more volume-centric. Both of these measures, though lower-frequency, remain useful for understanding the influence of informed trading, as informed traders typically engage in larger trades that can significantly affect stock prices, even on a daily or weekly basis. By combining both high-frequency measures (Spread and ESpread) with the price-impact measures (Amihud and VAmih), the study captures a comprehensive view of liquidity dynamics influenced by informed trading across different time scales.

Table 4.1 presents the descriptive statistics on the variables used in the main regression models. Across the measures of informed trading, both DPIN and SDPIN exhibit considerable variability, with DPIN ranging from 0.000 to 0.943 and SDPIN from 0.000 to 0.995. This suggests a diverse range of levels of informed trading activity in the dataset. Notably, the measures of stock liquidity, including Amih and VAmih, demonstrate substantial differences in their distributions, with Amih ranging from 0.000 to 14.592 and VAmih ranging from 0.000 to 78.447. This indicates varying degrees of liquidity across different stocks, possibly reflecting differing levels of market activity or trading volume. Similarly, bid-ask spread (*Spread*) and effective spread (*ESpread*) exhibit differences in their distributions, with effective spread showing narrower variability compared to bid-ask spread. Investor sentiment, as captured by SENT, also varies notably, with values ranging from 0.294 to 0.674, indicating fluctuations in market sentiment over the observed period.

Table 4.2 presents the correlation matrix which shows that the correlation coefficients are in general below 0.5.

Table 4.1: This table presents the statistical descriptions of regression variables. IT is informed trading, measured by two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015), and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). TimeUrge is a measure of the time remaining (number of days, weeks, or months) before the private information is publicly announced in the nearest quarterly reports. RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 106; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by the institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. The number of observations is 407,168.

Variable	Min	Max	Mean	Std Dev	Kurtosis	Skewness
DPIN	0.000	0.947	0.205	0.166	0.375	-0.851
SDPIN	0.000	0.995	0.235	0.219	0.453	0.276
Amih	0.000	14.592	0.018	0.170	2.152	8.970
VAmih	0.000	78.447	1.674	4.968	2.120	6.443
Spread	0.000	0.521	0.024	0.020	4.779	24.505
ESpead	0.000	0.526	0.015	0.114	5.871	73.150
TimeUrge	0.007	0.111	0.018	0.011	3.241	4.247
SENT	0.294	0.674	0.491	0.099	1.136	17.556
RET	-0.539	0.742	0.001	0.023	0.399	-0.797
VOL	0.000	428.000	4.385	9.287	51.424	1005.680
INST	0.461	0.932	0.826	0.215	0.866	0.301
SIZE	2,356	109,856	67,311	1,559	14.780	32.513

Table 4.2: This table presents the matrix of correlations between variables. IT is informed trading, measured by two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015), and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019. RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by the institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6.

	DPIN	SDPIN	SENT	RET	INST	VOL	SIZE
DPIN	1						
SDPIN	0.894	1					
SENT	0.367	0.348	1				
RET	-0.034	-0.028	0.002	1			
INST	0.022	0.010	-0.167	-0.001	1		
VOL	0.108	0.060	0.066	0.044	0.036	1	
SIZE	0.339	0.321	0.369	-0.002	-0.046	0.087	1

4.3.3. The Regression Model

To investigate the research hypotheses, I construct a main model (Model 1) that serves as an analytical framework to explore the multifaceted nature of liquidity within financial markets. It examines the aforementioned relationship using daily, weekly, and monthly data. Several control variables are included in the models, including informed trading (IT), investor sentiment (SENT), institutional ownership (INST),[32] stock size (SIZE), and market activity metrics for instance returns (RET) and trade volume (VOL). The rationale behind the models lies in their ability to capture the impact of these variables on stock liquidity, offering insights into how informed trading, investor sentiment, and other factors influence liquidity dynamics over different time horizons.

When the market is stable without much fluctuation, a sudden increase in stock liquidity in a short time can be a signal of informed trading. In recessionary market conditions, very liquid stocks can also be a sign of informed trading. However, it would be questionable to infer that there is informed trading merely based on the increase of stock liquidity where there is a positive market sentiment. Hence, I include in the regression model an interaction term (*SENT x IT*) between the informed trading variable and the market sentiment variable to capture the moderating role of market sentiment on the relationship between informed trading and market liquidity.

To further control the strategic timing of informed investors, I follow Ahern (2020) and create a variable to measure the urgency of trading. Hence, the variable *TimeUrge* measures the urgency of the trading. When there is high urgency, informed investors have less freedom to strategically time their trades, so the timing of trades is closer to the idealized experiment of random assignment of trading days. The *TimeUrge* variable increases as the public announcement date gets closer to the current date, therefore, both illiquidity and informed trading tends to increase as the announcement date approaches.

Model 1: using daily, weekly, and monthly data:

$$Liquidity_{i,t} = \beta_0 + \beta_1 IT_{i,t} + \beta_2 TimeUrge_{i,t} + \beta_3 IT_{i,t}\ x\ TimeUrge_{i,t} + \beta_4 SENT_{i,t} + \beta_5 SENT_{i,t}\ x\ IT + \beta_6 INST_{i,t} + \beta_7 SIZE_{i,t} + \beta_8 RET_{i,t} + \beta_9 VOL_{i,t} + \varepsilon_t \quad (39)$$

Where *IT* is informed trading, measured by two metrics: the dynamic probability of informed trading (*DPIN*) based on Chang et al. (2014) and Chang and Wang (2015), and the dynamic probability of informed trading with the size order effects (*SDPIN*) developed in Chapter 2. *Liquidity* is stock liquidity. The paper uses 4 different measures for this variable, including Bid-Ask spread (*Spread*), effective spread (*ESpread*), Amihud illiquidity measure (*Amih*), and the volume-based Amihud illiquidity (*VAmih*).

SENT is the investor sentiment index which is calculated based on Bouteska (2019) which is an AAII Sentiment Survey to compute investor sentiment variables.[33] It is the ratio of the percentage of bullish

[32] Institutional Investors (INST) refers to large entities such as mutual funds, pension funds, hedge funds, insurance companies, and other organizations that manage significant amounts of capital on behalf of individuals or other institutions. The variable INST represents the proportion of a firm's shares owned by institutional investors, which serves as a measure of institutional ownership. This is calculated by dividing the number of shares held by institutional investors by the total number of shares outstanding.

[33] The AAII Sentiment Survey offers insight into the opinions of individual investors by asking them their thoughts on where the market is heading in the next six months and has been doing so since 1987. The sentiment survey measures the percentage

investors divided by the sum of the percentage of bullish and bearish investors each week and use this proxy for individual investor sentiment. In financial circles, this measure is popularly known as the bull-bear spread. This methodology is also employed by Fisher and Statman (2000, 2006) and Kurov (2010).

$$SENT = \frac{BULL}{BULL+BEAR} \tag{40}$$

where *BULL* is the strength of the bullish market and *BEAR* is the strength of the bearish market.

TimeUrge is a measure of the urgency of the trading *and* is calculated as follows

$$TimeUrge = \frac{1}{N} \tag{413}$$

where *TimeUrge* is the urgency of trading. *N* is a measure of the time remaining (number of days, weeks, or months) before the private information is publicly announced in the nearest quarterly reports. *RET* is the absolute value of daily stock return; (*VOL*) accounts for the daily average of stock trade volume divided by 10^6; *INST* denotes the institutional investors variable, defined by the proportion of firm shares owned by the institutional investors. Firm size (*SIZE*) is given by market capitalization divided by 10^6.

4.4. Empirical Findings

This section presents the regression results and discussion on the relationship between informed trading and stock liquidity. The analysis is structured by frequency, providing insights for daily, weekly, and monthly data. Additionally, results for quarterly and yearly frequencies are included, along with a robust test focusing on the COVID-19 period to ensure the findings' reliability across varying market conditions. Each subsection will detail the outcomes of the regression models, highlighting key patterns, statistical significance, and practical implications of the relationship between informed trading and liquidity. Discussions integrate theoretical perspectives and prior empirical evidence to contextualize the findings and offer interpretations for observed trends.

4.4.1 Findings: daily frequency

Table 4.3 presents the results for the daily frequency data.

Main Results

We can see that all stock liquidity measures, including *Amih*, *VAmih*, *Spread*, and *ESpread*, have negative and statistically significant coefficients for *DPIN* and *SDPIN*, suggesting that informed trading is positively related to stock liquidity, as I would expect. We can confirm the hypothesis H1.

of individual investors who are bullish, bearish, and neutral on the stock market short term; individuals are polled from the AAII Web site on a weekly basis.

This finding contradicts those of previous studies that rely on classical microstructure models and data from when automated trading was less predominant or not yet available (Glosten and Milgrom, 1985; Easley and O'Hara, 1992). These results shed light on the effect of automated trading on stock liquidity. As previously stated, automated markets enable traders to rapidly incorporate newly available information into stock prices, and informed traders who can leverage sophisticated tools and real-time data feeds can more accurately anticipate market movements and act in a timely manner, leading to higher stock liquidity. And their activity might also attract liquidity providers like market makers and other market participants (uninformed investors), enhancing overall market liquidity and reducing trading costs.

The coefficients for the variable *TimeUrge* and its interactions with informed trading indicators (*DPIN* and *SDPIN*) also have a negative and statistically significant relation to both *Spread* and *ESpread* but are statistically insignificant for the *Amih* and *VAmih* liquidity measures. Furthermore, the effect of informed trading on liquidity remains negative and significant under the moderating effect of the TimeUrge. This result suggests that informed trading could stem from investors possessing privileged information directly linked to forthcoming financial reports. In such instances, informed investors may encounter constraints in strategically timing their trades and feel compelled to execute more transactions as the public announcement date approaches. This scenario could lead to heightened levels of informed trading and subsequently, increased liquidity in the market.

The variable volume (*VOL*) is statistically significant and inversely related to all four liquidity measures. This finding is relevant as increased trading volume typically corresponds with higher market activity and liquidity. When trading volumes surge, it often indicates greater participation from market participants, including both informed traders and liquidity providers. This increased activity can lead to improved market depth, narrower bid-ask spreads, and enhanced price discovery mechanisms, all of which contribute to reduced levels of stock illiquidity.

The negative coefficients for market sentiment (*SENT*) across all models suggest a positive relationship with stock liquidity, as expected. In periods of positive sentiment, investors tend to be more optimistic about the prospects of the market, leading to increased participation and trading activity. This increased activity can contribute to improved market liquidity as more buyers and sellers are willing to transact at various price levels (Kumari, 2019; Liu, 2015). The coefficients of the interaction terms between informed trading (*DPIN* and *SDPIN*) and market sentiment (*SENT*) maintain significance levels at 1%, 5%, and 10%, implying a consistent association between informed trading and stock liquidity across various levels of market sentiment. The continued significantly positive impact of DPIN and SDPIN on liquidity, even under the moderating influence of market sentiment, underscores the robustness of this finding. When investors' sentiment is high, informed trading tends to have an even greater positive impact on stock liquidity and vice versa. This heightened sentiment amplifies the effect of informed trading, resulting in more efficient market functioning and improved liquidity conditions. Consequently, stocks experience higher trading volumes and narrower bid-ask spreads, reflecting the enhanced market confidence and activity.

Performance of liquidity measure

In models with daily data, when comparing the performance of high-frequency measures (Spread and Effective Spread (Espread)) to low-frequency measures (Amihud illiquidity (Amih) and Volume-Amihud (Vamih)), the high-frequency measures consistently demonstrate superior sensitivity and statistical significance, particularly in capturing liquidity dynamics driven by behavioral and informational variables.

One key area where high-frequency measures outperform is their responsiveness to private information variables, such as DPIN and SDPIN. For both Spread and Espread, DPIN and SDPIN are highly statistically significant, often at the 1% or 5% level, underscoring their capacity to reflect the immediate effects of private information on market liquidity. In contrast, Amih and Vamih, while significant for DPIN, tend to exhibit weaker statistical significance (5% or 10% level), especially for SDPIN, where Vamih fails to reach significance. This finding suggests that informed trading, as captured by *DPIN* and *SDPIN*, has a particularly pronounced effect on bid-ask spreads and effective spreads for daily data, indicating that these metrics are sensitive to the influence of informed trading behavior.

Behavioral variables such as sentiment (SENT) also exhibit stronger predictive power for high-frequency measures. SENT is consistently significant at the 1% level for Spread and Espread, emphasizing their ability to capture the short-term emotional and psychological factors influencing liquidity. Conversely, SENT’s significance for low-frequency measures like Amih and Vamih is weaker, typically at the 5% or 10% level, indicating that these measures may not fully account for rapid shifts in market sentiment.

High-frequency measures are particularly adept at capturing the effects of interaction terms, such as DailyUrge x DPIN and SENT x DPIN, which highlight complex relationships between behavioral and informational drivers. Spread and Espread consistently show significant results for these terms, often at the 1% or 5% level, whereas Amih and Vamih fail to detect meaningful relationships in many cases. This further reinforces the dynamic adaptability of high-frequency measures.

In conclusion, in the daily model, high-frequency measures such as Spread and Espread exhibit superior performance compared to low-frequency measures like Amih and Vamih in terms of both sensitivity to key liquidity drivers and statistical significance.

Table 4.3: This table presents results for the relationship between stock liquidity and informed trading. Result for daily data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (*DPIN*) based on the model by Chang et al. (2014) and Chang and Wang (2015). And the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (Vamih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). DailyUrge is the number of days remaining before the private information is publicly announced in the nearest quarterly reports. RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses. Total number of observations is 407,168 days.

	Amih		**Vamih**		**Spread**		**Espread**	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.024*** (0.001)	0.294*** (0.050)	2.105*** (0.0076)	3.680*** (0.015)	0.025*** (0.003)	0.010*** (0.000)	0.014*** (0.003)	0.006*** (0.002)
DPIN	-0.016** (0.007)		-0.129* (0.074)		-0.0488** (0.039)		-0.007** (0.003)	
SDPIN		-0.049** (0.026)		-1.237* (0.643)		-1.004*** (0.021)		-0.061** (0.029)
DailyUrge	-0.025* (0.016)	-0.028* (0.017)	-5.058 (4.937)	-6.210 (6.490)	-0.004** (0.002)	-0.003** (0.002)	-0.019** (0.009)	-0.053** (0.027)
DailyUrge x DPIN	-0.097* (0.059)		11.002 (20.406)		-0.014* (0.076)		-0.016*** (0.000)	
DailyUrge x SDPIN		-0.172* (0.088)		8.358 (6.091)		-0.336* (0.207)		-0.425*** (0.002)
RET	-0.019* (0.011)	-0.157** (0.079)	-2.418* (1.209)	-3.532* (1.860)	-0.064** (0.031)	-0.054** (0.028)	-0.114** (0.052)	-0.451** (0.199)
VOL	-1.59E-05*** (1.18E-06)	-4.26E-05*** (9.81E-06)	-6.31E-04*** (4.08E-05)	-2.09E-04*** (3.27E-05)	-1.05E-06*** (4.39E-08)	-3.39-06*** (3.65E-08)	-7.19E-05*** (2.01E-06)	-5.09E-05*** (1.22E-06)
SENT	-0.025** (0.013)	-0.025** (0.012)	-0.722** (0.363)	-0.264*** (0.008)	-0.043*** (0.007)	-0.060*** (0.139)	-0.074*** (0.016)	-0.072*** (0.014)
SENT x DPIN	-0.347 (0.396)		-2.049* (1.147)		-0.175*** (0.009)		-0.996* (0.561)	
SENT x SDPIN		-0.904* (0.433)		-1.885* (103)		-1.240** (0.041)		-0.713* (0.377)
INST	-0.009** (0.004)	-0.008** (0.005)	-0.521** (0.148)	-0.472* (0.279)	-0.006*** (0.002)	-0.007*** (0.002)	-0.009** (0.004)	-0.010** (0.005)
SIZE	-3.24E-06** (2.02E-07)	-3.03E-06** (2.02E-07)	-5.29E-05* (2.84E-05)	-1.61E-05* (9.04E-06)	-3.43E-06*** (6.57E-07)	-3.49E-06*** (6.62E-07)	-2.40E-06*** (6.63E-07)	-3.18E-06*** (3.95E-07)

***Note:** *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

4.4.2 Findings: weekly frequency

Table 4.4 presents regression results for weekly frequency, offering further insights into the relationship between informed trading and stock liquidity measures. Two liquidity measures, Spread and Espread, exhibit consistent outcomes comparable to the daily frequency, maintaining statistically significant levels at 5% and 10%, respectively. This consistency suggests that the impact of informed trading on these liquidity metrics persists across different time frames, indicating a robust relationship between informed trading behavior and bid-ask spread as well as effective spread on a weekly basis. However, for the other two stock liquidity measures (Amih and Vamih), only SDPIN has a significant impact on Amih, albeit at a significance level of 10% only. This finding suggests that while SDPIN influences Amih, the relationship is less pronounced compared to that for Spread and Espread. This outcome underscores the complexity of liquidity dynamics and highlights the varying effects of informed trading on different liquidity metrics.

Table 4.4: This table presents results for the relationship between stock liquidity and informed trading. Result for Weeky data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). And the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (Vamih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). WeeklyUrge is the number of weeks remaining before the private information is publicly announced in the nearest quarterly reports. RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses. Total number of observations is 92,331 weeks.

	Amih		**Vamih**		**Spread**		**Espread**	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.353*** (0.012)	0.496*** (0.003)	1.438*** (0.130)	1.079*** (0.335)	0.065*** (0.019)	0.188*** (0.067)	0.993*** (0.116)	0.527*** (0.025)
DPIN	-0.042 (0.070)		-3.562 (4.911)		-0.262** (0.125)		-0.183* (0.109)	
SDPIN		-0.011* (0.007)		-0.188 (0.315)		-1.506** (0.691)		-0.006* (0.003)
WeeklyUrge	-0.003* (0.002)	-0.002* (0.001)	11.715 (15.009)	3.039 (2.497)	-0.001* (0.000)	-0.003* (0.002)	0.002 (0.002)	-0.044 (0.057)
WeeklyUrge x DPIN	-0.006* (0.004)		12.884 (19.612)		-0.048 (0.165)		-0.107 (0.100)	
WeeklyUrge x SDPIN		-0.094* (0.052)		-10.475 (8.961)		-0.158 (0.303)		-0.009 (0.035)
RET	-0.283* (0.173)	-0.267* (1.545)	-1.024** (0.445)	-3.359** (1.600)	0.057 (0.053)	0.054 (0.072)	1.358 (2.094)	0.464 (1.007)
VOL	-4.21E-05*** (1.86E-06)	-1.38E-05*** (5.33E-06)	-4.75E-04* (2.83E-04)	-3.88E-04** (1.97E-04)	-9.11E-06*** (1.64E-06)	-8.04E-06*** (1.62E-06)	-4.39E-06** (2.14E-06)	-1.25E-06** (6.28E-06)
SENT	-0.055* (0.032)	-0.084* (0.046)	-1.747* (0.924)	-0.516** (0.228)	-0.010** (0.005)	-0.026** (0.014)	-0.049** (0.025)	-0.523** (0.263)
SENT x DPIN	-1.057* (0.625)		9.022 (11.748)		-5.468* (3.117)		-2.395* (1.267)	
SENT x SDPIN		-0.896** (0.427)		6.384 (6.076)		-4.433** (2.198)		-2.722* (1.448)
INST	-0.008*** (0.000)	-0.001*** (0.000)	-0.652** (0.295)	-1.002* (0.527)	-0.002*** (0.000)	-0.007*** (0.001)	-0.064** (0.033)	-0.002** (0.000)
SIZE	-6.41E-05** (2.98-05)	-6.20E-05** (3.13E-05)	-4.73E-04* (2.74E-04)	-4.416E-04* (2.374E-04)	-1.92E-06* (1.00E-06)	-1.16E-06* (5.01E-07)	-1.57E-06** (7.32E-07)	-2.26E-06*** (3.19E-07)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

4.4.3 Findings: monthly frequency

Based on the regression results presented in Table 4.5, it is evident that DPIN and SDPIN have a statistically significant impact on the Amihud illiquidity measure (Amih) in the monthly data sample. Specifically, both DPIN and SDPIN show a negative relationship with stock liquidity, as indicated by the positive coefficients for Amih, meaning that higher values of DPIN or SDPIN are associated with lower liquidity. This supports the expectation that informed trading—whether measured by the probability of informed trading (DPIN) or by the size of trading orders (SDPIN)—increases market illiquidity, as informed traders can create price movements that reduce liquidity for other participants.

While this effect is observed for Amih, it is not consistently significant across all liquidity measures, such as the Bid-Ask Spread (Spread) and Effective Spread (ESpread), where the relationship with DPIN and SDPIN is weaker or statistically insignificant. This discrepancy suggests that the Amihud illiquidity measure is more sensitive to informed trading in the context of monthly data, making it a better fit for this frequency of data and the research questions being investigated. The Amih measure captures the price impact of liquidity, and the results indicate that informed trading is more readily reflected in price changes, making it more suitable for the monthly timeframe compared to measures like Spread or ESpread, which may capture more short-term fluctuations in liquidity.

In conclusion, the Amihud measure offers a more reliable indicator of the relationship between informed trading and stock liquidity in the context of monthly data, as it consistently shows the expected patterns of behavior.

Table 4.5: This table presents results for the relationship between stock liquidity and informed trading. Result for Monthly data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023 Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). MonthlyUrge is the number of months remaining before the private information is publicly announced in the nearest quarterly reports. RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses. There are 21,523 observations.

	Amih		**VAmih**		**Spread**		**ESpread**	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.076*** (0.025)	0.353*** (0.009)	0.742** (0.346)	0.284** (0.143)	0.191* (0.130)	0.287* (0.152)	0.359** (0.164)	1.677* (0.958)
DPIN	-0.003* (0.002)		0.807 (1.391)		-1.621 (2.079)		1.602 (2.080)	
SDPIN		-0.005* (0.003)		-1.004 (1.066)		-1.840 (2.016)		0.911 (1.165)
MonthlyUrge	-0.003* (0.002)	-0.003* (0.002)	3.597 (10.614)	1.1437 (2.592)	-0.079 (0.122)	-0.077 (0.605)	-0.084* (0.45)	-0.045* (0.024)
MonthlyUrge x DPIN	0.108 (0.255)		-22.010 (28.046)		-0.038 (0.059)		-0.584 (1.049)	
MonthlyUrge x SDPIN		-0.094 (0.073)		-19.816 (17.522)		0.264* (0.211)		-0.709 (1.646)
RET	0.157* (0.084)	0.139* (0.073)	2.514 (4.923)	0.899 (0.911)	0.057* (0.033)	0.055* (0.030)	0.066* (0.035)	0.437** (0.192)
VOL	-9.04E-05** (4.18E-05)	-9.11E-04 (8.25E-04)	-4.17E-03 (7.02E-03)	-1.03E-03 (8.98E-03)	6.26E-05 (9.00E-05)	6.79E-05 (9.83E-05)	-7.19E-05** (3.51E-05)	-5.84E-05*** (3.09E-06)
SENT	-0.0728* (0.039)	-0.084** (0.43)	-0.508** (0.254)	-2.116 (3.270)	-0.024* (0.012)	-0.0217* (0.011)	-0.440** (0.190)	-1.072* (0.564)
SENT x DPIN	0.167 (0.663)		-2.902* (1.517)		2.072 (2.685)		1.091 (1.037)	
SENT x SDPIN		0.228 (0.231)		4.915 (5.048)		0.043* (0.023)		0.095* (0.052)
INST	-0.146 (0.352)	-0.149* (0.079)	-0.314 (0.401)	-5.529 (4.833)	-0.124* (0.014)	-0.675* (1.114)	-0.006*** (0.000)	-0.003*** (0.000)
SIZE	-8.12E-06** (3.63E-06)	-8.07E-06*** (1.59E-06)	-7.19E-05* (3.80E-05)	-4.73E-05** (2.07E-05)	-4.11E-06*** (3.38E-07)	-4.43E-06*** (3.51E-07)	-7.82E-06** (3.93E-07)	-6.35E-06* (3.57E-07)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

4.4.4. Findings: quarterly and yearly frequencies

In addition to the daily, weekly, and monthly data analyses, the paper also uses another model (Model 2) for quarterly and yearly data. These tests enable us to gain a deeper understanding of the relationship between informed trading and stock liquidity.

Model 2: using quarterly and yearly data

$$Liquidity_{i,t} = \beta_0 + \beta_1 IT_{i,t} + \beta_2 SENT_{i,t} + \beta_3 SENT_{i,t} * IT + \beta_4 INST_{i,t} + \beta_5 SIZE_{i,t} + \\ + \beta_6 RET_{i,t} + \beta_7 VOL_{i,t} + \varepsilon_t \quad (42)$$

where IT is informed trading, measured by the two metrics: dynamic probability of informed trading (*DPIN*) and dynamic probability of informed trading with the size order effects (*SDPIN*). *Liquidity* is stock liquidity. *SENT* is the investor sentiment index which is calculated according to Bouteska (2019). *RET* is the absolute value of daily stock return; (*VOL*) accounts for the daily average of stock trade volume divided by 10^6; *INST* denotes the institutional investors variable, defined by the proportion of firm shares owned by the institutional investors. Firm size (*SIZE*) is given by market capitalization divided by 10^6.

Tables 4.6 and 4.7 present the regression findings for quarterly and yearly frequencies, respectively. Notably, I can see that there are no statistically significant coefficients for both the DPIN and SDPIN. Therefore, we can conclude that in the medium and long term, there is no empirical evidence supporting the idea that there is a relationship between informed trading and stock liquidity.

Overall, the regression findings across different frequencies provide us with interesting insights regarding the impact of informed trading on stock liquidity. When relying on daily and weekly data, the results show that informed trading affects stock liquidity. The negative signs of all coefficients suggest that higher levels of informed trading are associated with lower liquidity, which is consistent with existing literature. However, this is not the case when I use in the analysis monthly, quarterly, and yearly data, for which most of the regression coefficients are insignificant, showing that there is no empirical evidence that a relationship between informed trading and stock liquidity do exist. Hence, we can conclude that the effect of informed trading on stock liquidity exists in the short term only. This might be because daily and weekly data capture the more immediate reactions to new information or changes in trading strategies, whereas monthly, quarterly, and yearly data capture broader market trends and macroeconomic trends which overshadow the individual impact of stock trading activities.

In both the quarterly and yearly datasets, the Amih measure consistently outperforms VAmih, as well as the high-frequency Spread and ESpread measures. Amih provides more stable and significant relationships with key financial variables, such as stock returns, trading volume, and institutional ownership, reflecting long-term liquidity dynamics. On the other hand, VAmih and the high-frequency measures show greater sensitivity to short-term fluctuations, making them less reliable in capturing broader market trends. As such, Amih is the superior measure in terms of its ability to consistently reflect the underlying liquidity conditions over time.

In Table 4.6 (quarterly data), the Amih measure, a low-frequency liquidity metric, outperforms both VAmih and the high-frequency measures such as Spread and ESpread in several key aspects. Amih consistently demonstrates statistically significant and stronger relationships with important variables like stock returns (RET), trading volume (VOL), and institutional ownership (INST). For example, Amih shows a significant positive correlation with RET, indicating a clear link between illiquidity and stock return behavior. On the other hand, while VAmih and the high-frequency measures (Spread and ESpread) show some correlations with RET, they are less stable and often statistically insignificant. Amih also exhibits a stronger relationship with institutional ownership (INST) and firm size (SIZE), both of which reflect long-term liquidity patterns. In contrast, VAmih struggles to show consistent results, especially in terms of its interaction with SENT (investor sentiment) and VOL. Overall, the Amih measure provides a more reliable reflection of liquidity, capturing broader market trends, while VAmih and the high-frequency measures are more prone to short-term fluctuations.

In Table 4.7 (yearly data), the pattern observed in the quarterly data is mirrored in the yearly data (Table 4.7). Amih again outperforms VAmih, as well as Spread and ESpread, in terms of its ability to establish strong, statistically significant relationships with stock returns (RET), trading volume (VOL), and institutional ownership (INST). For instance, Amih shows a significant relationship with RET, where higher illiquidity is associated with negative stock returns. While VAmih demonstrates some correlations with RET and VOL, its results are less stable and sometimes statistically insignificant. Furthermore, Amih exhibits more consistent correlations with institutional ownership (INST) and firm size (SIZE) than VAmih, suggesting that the low-frequency measure captures liquidity dynamics more effectively over time. In contrast, VAmih and the high-frequency measures (Spread and ESpread) exhibit weaker and more volatile relationships, often reflecting short-term liquidity shifts rather than longer-term trends.

Table 4.6: This table presents results for the relationship between stock liquidity and informed trading. Result for quarterly data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses.

	Amih		**VAmih**		**Spread**		**ESpread**	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.935**	0.180**	1.304*	0.257**	0.544**	1.097*	0.218*	0.672*
	(0.434)	(0.091)	(0.776)	(0.120)	(0.236)	(0.590)	(0.121)	(0.356)
DPIN	-0.014		-11.266		0.055		0.995	
	(0.084)		(10.420)		(0.081)		(1.035)	
SDPIN		0.162		-4.008		0.706		0.740
		(0.197)		(5.056)		(0.790)		(0.619)
RET	0.157*	0.004	1.599*	0.974**	0.160*	0.074**	1.332	0.358*
	(0.088)	(0.169)	(0.946)	(0.444)	(0.085)	(0.033)	(1.086)	(0.211)
VOL	-2.16E-0***	-1.31E-04**	8.42E-04*	5.15E-04	-2.04E-05**	-2.99E-05**	-6.32E-05**	-5.25E-05**
	(3.68E-05)	(5.90E-05)	(4.98E-04)	(5.37E-04)	(9.14E-06)	(1.37E-05)	(2.98E-05)	(2.41E-05)
SENT	-0.094*	-0.086*	1.285*	0.573	-0.693*	-0.374*	-0.008	-0.004
	(0.049)	(0.048)	(0.681)	(0.656)	(0.410)	(0.1820)	(0.009)	(0.006)
SENT x DPIN	3.133		10.837		0.597		2.807	
	(2.498)		(12.014)		(0.622)		(3.525)	
SENT x SDPIN		4.457		5.365		0.528		2.049
		(4.680)		(9.280)		(0.533)		(1.161)
INST	-0.0035**	-0.0059**	-0.971*	-1.027**	0.105	0.114*	-0.050	-0.748
	(0.002)	(0.003)	(0.490)	(0.478)	(0.056)	(0.066)	(0.063)	(1.336)
SIZE	-7.62E-06**	-7.57E-06**	-3.25E-05*	-3.21-05	-6.00E-07*	-6.32E-07**	-1.61E-07*	-1.63E-07**
	(3.82E-06)	(3.46E-06)	(1.92E-05)	(7.03E-05)	(3.09E-07)	(2.82E-07)	(8.43E-08)	(7.69E-08)
Observations	7,039	7,039	7,039	7,039	7,039	7,039	7,039	7,039

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

Table 4.7: This table presents results for the relationship between stock liquidity and informed trading. Result for yearly data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses.

	Amih		**VAmih**		**Spread**		**ESpread**	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.442*	0.778*	0.920	1.185**	0.036*	0.905*	1.307*	1.669**
	(0.4261)	(0.407)	(1.097)	(0.566)	(0.023)	(0.487)	(0.773)	(0.76=76)
DPIN	0.759		2.641		1.009		0.846	
	(0.502)		(2.105)		(0.816)		(1.077)	
SDPIN		1.201		-0.362		1.129		1.237
		(1.405)		(0.297)		(1.187)		(1.490)
RET	0.834*	0.062**	2.677**	2.626**	0.090*	0.083*	0.244*	0.241*
	(0.439)	(0.003)	(1.345)	(1.257)	(0.050)	(0.061)	(0.141)	(0.137)
VOL	-1.47E-04**	-2.09E-04**	-4.37E-03*	-3.95E-03*	-6.16E-05**	-6.19E-05**	-4.87E-05**	-4.46E-05**
	(5.51E-05)	(9.68E-05)	(2.40E-03)	(2.24E-03)	(2.76E-05)	(2.85E-05)	(2.25E-05)	(2.25E-05)
SENT	-0.062*	-0.063*	-0.249	-0.246*	-0.009**	-0.010**	-1.427*	-1.301*
	(0.037)	(0.038)	(0.165)	(0.149)	(0.042)	(0.054)	(0.755)	(0.680)
SENT x DPIN	1.007		10.837		-0.597		-1.877	
	(1.473)		(10.014)		(0.422)		(1.525)	
SENT x SDPIN		2.126		11.365		0.528*		2.049
		(1.188)		(10.280)		(0.336)		(3.161)
INST	-0.009**	-0.007**	-0.015**	-0.016**	-0.024*	-0.018*	-0.359*	-0.451*
	(0.005)	(0.004)	(0.007)	(0.008)	(0.013)	(0.011)	(0.203)	(0.239)
SIZE	-5.05E-06***	-5.01E-06***	-9.97E-05	-9.13E-05	-1.13E-06**	-1.12E-06**	-7.55E-06*	-7.29E-06**
	(2.92E-07)	(2.91E-07)	(9.28E-05)	(8.07E-05)	(5.08E-07)	(5.08E-07)	(3.42E-06)	(3.45E-06)
Observations	1,847	1,847	1,847	1,847	1,847	1,847	1,847	1,847

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

4.4.5. Effect of Covid-19 Pandemic

During the Covid-19 pandemic, global economies, including stock markets, experienced significant disruptions and high uncertainty. Specifically, the U.S. stock market witnessed substantial declines, particularly during the second and third quarters of 2020, with ongoing repercussions in subsequent periods. Given the extraordinary nature of the COVID-19 pandemic and its potential effects on the empirical findings, I run some robustness checks to account for this. To control for the COVID-19 effect on the results I add the following regression variable, *COV*. It takes into account the COVID-19 shock and takes a value of 1 during the period affected by the COVID-19 pandemic (from January 2020 to December 2021) and 0 otherwise. The new regression models are below.

Model 3: using daily, weekly, and monthly data

$$LIQUIDITY_{i,t} = \beta_0 + \beta_1 IT_{i,t} + \beta_2 TimeUrge_{i,t} + \beta_3 IT\ x\ TimeUrge_{i,t} + \beta_4 SENT_{i,t} + \beta_5 SENT\ x\ IT_{i,t} + \beta_6 COV_{i,t} + \\ + \beta_7 COV_{i,t} * IT_{i,t} + \beta_8 RET_{i,t} + \beta_9 VOL_{i,t} + \beta_{10} INST_{i,t} + \beta_{11} SIZE_{i,t} + \varepsilon_t \quad (43)$$

Model 4: using quarterly and yearly data

$$Liquidity_{i,t} = \beta_0 + \beta_1 IT_{i,t} + \beta_2 SENT_{i,t} + \beta_3 SENT\ x\ IT_{i,t} + \beta_4 COV_{i,t} + \beta_5 COV_{i,t}\ x\ IT_{i,t} + \beta_6 RET_{i,t} + \beta_7 VOL_{i,t} + \\ + \beta_8 INST_{i,t} + \beta_9 SIZE_{i,t} + \varepsilon_t \quad (44)$$

where *IT* is informed trading, measured by the two metrics: dynamic probability of informed trading (*DPIN*) and dynamic probability of informed trading with the size order effects (*SDPIN*); *Liquidity* is stock liquidity. *SENT* is the investor sentiment index which is calculated according to Bouteska (2019); *TimeUrge* is a measure of the time remaining (number of days, weeks, or months) before the private information is publicly announced in the nearest quarterly reports; *COV* is a variable that accounts for the COVID-19 shock, which takes a value of 1 during the period affected by the COVID-19 pandemic and 0 otherwise*;* *RET* is the absolute value of daily stock return; (*VOL*) accounts for the daily average of stock trade volume divided by 10^6; *INST* denotes the institutional investors variable, defined by the proportion of firm shares owned by the institutional investors; Firm size (*SIZE*) is given by market capitalization divided by 10^6.

The outcomes of the robustness checks shown in Tables 4.8, 4.9, and 4.10 confirm the findings of the main model across different frequencies. Specifically, for daily and weekly data, DPIN and SDPIN continue to exhibit a significant and negative impact on all illiquidity proxies, of Amih, VAmih, Spread, and ESpread. These results reaffirm the robustness of the main model's findings in capturing the relationship between informed trading and stock liquidity at shorter intervals. For monthly data, I do not find any evidence supporting the significant relationship between informed trading and liquidity.

For longer timeframes, including quarterly and yearly panel data, the relationship between informed trading and liquidity metrics is insignificant. This result consistency across robustness checks strengthens the reliability of the main outcomes regarding the influence of informed trading on stock liquidity across varying timeframes. Across all models and frequencies, the COVID-19 shock (COV) consistently demonstrates a strong impact on stock liquidity. The majority of coefficients associated with COV are positive, suggesting a significant decrease in stock liquidity during the COVID-19 period. This finding underscores the profound effect of the COVID-19 pandemic on financial markets, leading to reduced liquidity levels.

Table 4.8: This table presents results for the Robust Test. Result for Daily and Weekly data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by the Bid-Ask spread, effective spread, Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses.

	Daily Frequency							
	Amih		VAmih		Spread		ESpread	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
DPIN	-0.117** (0.053)		-1.028* (0.608)		-0.743*** (0.063)		-0.107** (0.054)	
SDPIN		-0.321** (0.142)		-2.237* (1.165)		-1.002*** (0.024)		-0.536*** (0.125)
COV	0.029*** (0.009)	0.027*** (0.009)	2.042*** (0.619)	2.041*** (0.620)	0.105*** (0.007)	0.106*** (0.007)	0.062*** (0.004)	0.059*** (0.004)
DPIN x COV	-0.732* (0.398)		6.091 (6.148)		-3.382** (1.700)		-1.167* (0.698)	
SDPIN x COV		-0.634* (0.330)		-6.709 (5.334)		-1.816** (1.849)		-0.016* (0.010)
	Weekly Frequency							
	Amih		VAmih		Spread		ESpread	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
DPIN	-0.072* (0.040)		-1.746* (0.914)		-0.520** (0.229)		-0.260** (0.139)	
SDPIN		-0.018** (0.010)		-1.824* (0.965)		-1.228** (0.620)		-0.510** (0.250)
COV	0.182** (0.089)	1.164** (0.539)	2.005*** (0.607)	2.029*** (0.441)	0.292** (0.178)	0.078*** (0.0090)	0.107** (0.049)	0.083*** (0.021)
DPIN x COV	-1.353 (2.048)		11.759 (10.285)		-1.264 (1.098)		-0.744* (0.619)	
SDPIN x COV		-0.997 (1.004)		12.938 (13.746)		-1.177* (0.701)		-0.953* (0.784)

***Note:** *, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

Table 4.9: This table presents results for the Robust Test. Result for Monthly and Quarterly data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses.

	Monthly Frequency							
	Amih		VAmih		Spread		ESpread	
DPIN	-0.310* (0.278)		0.801 (0.109)		-1.138 (0.966)		0.606 (1.042)	
SDPIN		-0.848* (0.508)		-1.003* (0.622)		-1.852 (1.673)		0.944 (0.705)
COV	0.307** (0.145)	0.146** (0.078)	2.907** (1.312)	3.028*** (0.772)	0.336* (0.183)	0.029** (0.015)	0.294** (0.016)	0.076*** (0.013)
DPIN x COV	2.902 (3.077)		16.167 (19.841)		2.072* (1.108)		1.091 (1.037)	
SDPIN x COV		-0.954 (0.884)		12.016 (14.255)		-0.964 (1.011)		0.709 (0.714)
	Quarterly Frequency							
	Amih		VAmih		Spread		ESpread	
DPIN	-0.062 (0.071)		-1.248 (1.413)		0.487 (0.513)		0.100 (0.138)	
SDPIN		-0.162 (0.197)		-2.932 (3.170)		0.724 (0.619)		1.050 (0.998)
COV	0.618* (0.367)	0.104** (0.053)	2.010** (0.918)	2.029** (0.939)	0.138** (0.062)	0.447** (0.195)	0.513*** (0.049)	0.579*** (0.054)
DPIN x COV	2.115 (1.937)		10.837* (6.193)		-1.597 (1.122)		2.807 (3.416)	
SDPIN x COV		3.036* (1.607)		5.206 (6.948)		0.745 (0.728)		-3.266 (2.975)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

Table 4.10: This table presents results for the Robust Test. Result for Yearly data. IT is informed trading, measured by the two metrics: the dynamic probability of informed trading (DPIN) based on the model by Chang et al. (2014) and Chang and Wang (2015). and the dynamic probability of informed trading with the size order effects (SDPIN) by Thu et al. (2023). Stock liquidity is measured by Bid-Ask spread (Spread), effective spread (ESpread), Amihud illiquidity measure (Amih), and the volume-based Amihud illiquidity (VAmih). SENT is the investor sentiment index which is calculated according to Bouteska (2019). RET is the absolute value of daily stock return; (VOL) accounts for the daily average of stock trade volume divided by 10^6; INST denotes the institutional investors variable, defined by the proportion of firm shares owned by institutional investors. Firm size (SIZE) is given by market capitalization divided by 10^6. Standard Errors are presented in parentheses.

Liquidity Measures & Interaction Terms	**Yearly Frequency**							
	Amih		VAmih		Spread		ESpread	
DPIN	0.632 (0.479)		1.758 (1.020)		2.010 (1.733)		0.839 (0.936)	
SDPIN		0.228 (0.424)		-1.307 (1.0216)		1.558 (1.274)		1.233 (1.000)
COV	0.139** (0.064)	0.127** (0.056)	1.739 (1.696)	1.105 (1.342)	0.137** (0.069)	0.538* (0.276)	0.013* (0.006)	1.008** (0.454)
DPIN x COV	-1.170* (0.692)		-8.167 (10.054)		2.072 (1.685)		1.802 (1.969)	
SDPIN x COV		-2.262* (1.184)		-8.016 (9.762)		1.264 (1.951)		-2.834 (2.774)

Note: **, **, and *** indicate p <0.1, 0.05, and 0.01, respectively*

4.5. Conclusion

There is a multifaceted relationship between information asymmetry, market efficiency, and stock liquidity, especially in the context of financial markets that are characterized by a prevalence of high-speed trading and the use of new sophisticated IT tools. Classical microstructure models have long underscored the impact of information imbalances on liquidity, particularly in traditional market setups where intermediaries, market makers, and specialists, play a crucial role in maintaining liquidity. However, with the rise of automated trading characterized by continuous order book updates, the dynamics between informed trading and liquidity have undergone significant changes. In contrast to traditional markets where liquidity is primarily provided by market makers and specialists, modern markets derive liquidity from the collective actions of a diverse range of traders, including both informed and uninformed participants. This shift introduces new complexities, such as the potential for more aggressive trading strategies used by informed traders seeking to capitalize on their informational advantages. The study aims to explore this association considering the existence of automated markets with high-frequency trading, contributing to a deeper understanding of liquidity dynamics in modern financial markets.

The research utilizes the DPIN and SDPIN informed trading measures and examines their impact on stock liquidity across different timeframes. The findings suggest that informed trading tends to promote liquidity in short timeframes (daily and weekly frequencies), aligning with the idea that trading activities enhance trading volumes and market depth in the short term. However, over longer timeframes, the impact of informed trading becomes insignificant, suggesting a diminishing effect on stock liquidity.

The positive impact of informed trading on stock liquidity when I rely on daily and weekly data is relevant to the literature in the context of modern markets. By facilitating faster information dissemination, attracting liquidity providers, and promoting more informed trading decisions, they create favorable conditions for the alignment of informed trading activities with liquidity enhancement efforts.

The research has a significant contribution. It studies the impact of informed trading on stock liquidity relying on a wide set of time-frequency data samples, which enables us to distinguish the effect of informed trading in the short frequencies and the long frequencies. I specifically use the DPIN and SDPIN as high-speed trading proxies for informed trading, which is particularly relevant in contemporary financial markets. These measures address the limitations of traditional metrics like PIN, developed by Easley et al. (1996, 2012) and based on the assumptions that one must aggregate very fine intraday data, which occur at approximately five-minute intervals within the trading day, across multiple days (Easley et al., 1997b). These assumptions make PIN a suitable estimation of informed trading for long-horizon trading (a month or a year) rather than for high-speed trading. The finding on the positive impact of informed trading on stock liquidity provides valuable guidance for both market participants and policymakers, informing them for better adaptation to changing market conditions. For instance, institutional investors, traders, and market makers can adjust their strategies more effectively to manage risks. Additionally, scholars can use these insights to develop enhanced market analysis models, while regulators can leverage them to craft fairer and more transparent policies. This, in turn, can strengthen surveillance measures aimed at preventing market manipulation in automated trading.

The analysis of liquidity measures across daily, weekly, monthly, quarterly, and yearly data reveals distinct patterns in performance based on the frequency of the data. High-frequency measures such as Spread and ESpread perform better in capturing liquidity dynamics for daily and weekly data. These measures show stronger and more statistically significant relationships with key financial variables like stock returns (RET) and trading volume (VOL), reflecting their ability to capture short-term fluctuations in market liquidity. However, as the data frequency increases to monthly, quarterly, and yearly intervals, the performance of the high-frequency measures diminishes, and low-frequency measures like Amih demonstrate a clear advantage. Amih, in particular, performs better over these longer time periods, exhibiting more consistent and significant relationships with market variables such as institutional ownership (INST), firm size (SIZE), and investor sentiment (SENT). This suggests that Amih is more suited for capturing long-term liquidity trends and the underlying stability of the market, while high-frequency measures are more sensitive to short-term market noise.

One limitation of this research is that it might not fully account for the unique characteristics of different modern markets, such as market structure, technology advancement, regulation, and liquidity levels, which can significantly influence the dynamics of informed trading and liquidity. Additionally, the liquidity measures used in this study, including Bid-Ask spread, Effective spread, Amihud illiquidity measure (Amihud, 2002), and Volume-based Amihud illiquidity, are well-suited for capturing short-term liquidity, often on a daily basis. Therefore, these metrics might introduce biases when used to measure liquidity over longer periods, potentially leading to misinterpretations of the long-term impact of informed trading on liquidity.

Future research should incorporate market-specific characteristics to better understand how different factors influence the relationship between informed trading and liquidity. Developing or utilizing alternative liquidity measures that are more suitable for longer timeframes could help address potential biases. Comparative studies across different markets and regions can provide insights into how market-specific factors influence these dynamics while investigating the impact of particular technological advancements and regulatory frameworks that could offer new perspectives. Incorporating behavioral finance perspectives and extending the temporal scope to include periods of market stress or financial crises could provide a deeper understanding of market behaviors.

Chapter 5. Conclusion

This thesis encompasses three studies that focus on the effects of the investors' private information embedded in stock prices. Some of the findings challenge the conventional wisdom that says that secondary markets reflect the investors' current expectations about the future cash that can be generated by firms. It shows that financial markets (i.e., stock prices and their evolution over time) can also affect managerial decisions, thereby affecting economic activities.

In essence, a financial market serves as a dynamic arena where a multitude of investors, each armed with unique pieces of information, converge to trade securities with the aim of capitalizing on their insights. Within this ecosystem, certain investors may possess superior information about the quality of the firms thereby achieving higher stock investment returns. This privileged information is named in the microstructure literature as "private information". Crucially, stock prices serve as the aggregation point for this diverse array of private information. As a result, they reflect a comprehensive evaluation of a firm's value, incorporating both publicly available data and the insights gleaned from investors' private information.

The first empirical study aims at estimating the amount of investors' private information incorporated in the stock prices. I review the existing private information measures in the literature, highlighting their strengths and weaknesses, and develop a new investors' private information measure that I named the dynamic measure for the probability of informed trading with size order effects (SDPIN) and built on the dynamic measure for the probability of informed trading (DPIN) of Chang et al. (2014) but considers the size of the stock trade orders (trade volume). The result of the first study shows that SDPIN performs better in the research models than the DPIN of Chang et al. (2014) within the research models.

In the second paper, the application of the SDPIN measure is employed to explore the impact of investors' private information on managerial decisions regarding earnings management. The findings of this study yield intriguing insights, indicating that managers do take into account investors' private information while making decisions on earnings management. Specifically, I conclude that managers exhibit a reduced inclination to engage in earnings management when stock prices are more heavily influenced by investors' private information, and vice versa. Furthermore, managers who have a propensity to inflate earnings demonstrate a stronger reaction to private information in stock prices compared to managers who engage in earnings deflation. This distinction underscores the complex interplay between market dynamics, managerial incentives, and the influence of private information on corporate financial reporting practices.

In the third study, I examine the dynamics of the impact of informed trading on stock liquidity, considering the intricacies of the modern financial markets where high-frequency and algorithm trading is common. I use various data frequencies, from daily to yearly. The findings show that informed trading has a positive effect on stock liquidity when I rely on daily and weekly data frequencies, but such a relationship disappears when rely on monthly, quarterly, and yearly data frequencies. This finding highlights the importance of accounting for the temporal dimension and frequency of data analysis when studying the relationship between informed trading and stock liquidity. Moreover, it suggests that the impact of informed trading on liquidity may vary depending on the speed and frequency of trading activities. This finding is important since modern financial markets are characterized by high-frequency trading.

Overall, the thesis and its findings make some significant contributions, both theoretically and empirically. Firstly, it provides a methodological improvement in estimating investors' personal information in stock prices. The newly private information measure serves as a valuable tool for all participants in financial markets, especially investors and managers. It enhances the ability to estimate the level of private information embedded in stock prices, providing managers and investors with a powerful tool to evaluate market dynamics more effectively. By understanding the extent of private information in stock prices, investors can make better-informed decisions, such as identifying opportunities to capitalize on information asymmetries or assessing the risks associated with trading in certain securities. This is particularly relevant in markets where informed trading is prevalent, as it underscores the importance of timing and strategic decision-making in achieving superior returns. Secondly, the SDPIN measure facilitates a deeper comprehension of the feedback loop between stock prices and the economy. By offering a more precise gauge of private information, it illuminates the influence of stock prices and movements on managerial behavior. Future research utilizing this measure promises to advance the understanding of how financial markets interconnect with the real economy, unraveling the relationship where financial markets both influence and reflect real-world events. The increased accuracy of the SDPIN measure in assessing the level of private information inherent in stock prices provides insights into how such prices are shaped and, consequently, impact managerial decisions and economic progress.

Additionally, the findings of the second study regarding the impact of investors' private information on earnings management are empirical evidence supporting the transmission of information from secondary markets to the broader economy. This underscores the pivotal role that financial markets play, extending beyond their conventional role and highlighting their potential for significant feedback effects on the economy. The study fills a crucial gap in the existing literature by examining a novel external factor that influences managerial decisions on earnings management and corporate disclosure. This expansion of knowledge contributes to a more comprehensive understanding of these critical facets of corporate behavior. Furthermore, the second study can offer critical insights for investors who rely on corporate disclosures to inform their strategies. The finding that managers are less inclined to manipulate earnings when stock prices are heavily influenced by private information suggests that markets with higher levels of informed trading might exhibit greater transparency and reliability in corporate reporting. For investors, this emphasizes the value of participating in markets where informed trading is robust, as such environments are likely to feature a higher quality of financial information. Furthermore, investors can use the degree of private information embedded in stock prices as a proxy to assess the credibility of corporate disclosures, enabling them to differentiate between firms with potentially inflated earnings and those with more reliable reporting practices.

In the third study, the findings, which underscore the positive impact of informed trading on stock liquidity, recognize the evolving nature of this relationship in modern market setups. The findings have the potential to catalyze further academic inquiry into the nuances of this relationship, prompting a re-evaluation of existing studies. Understanding the evolution of the relationship between the two factors across different timeframes is crucial for all market participants and policymakers. The finding that informed trading enhances stock liquidity in shorter timeframes, such as daily and weekly frequencies, but not in longer periods, such as monthly or yearly data, underscores the need for investors to account for the temporal dimension of their strategies. Active investors, particularly those employing high-frequency trading or algorithmic trading strategies, can leverage this relationship to improve trade execution and reduce transaction costs. On the other hand, long-term investors might need to adjust their strategies by focusing on securities where liquidity trends remain stable over extended periods, minimizing exposure to the temporary effects of informed trading.

Moreover, the evolving nature of financial markets, with the rise of high-frequency trading and algorithmic strategies, requires investors to stay informed about the implications of these changes on liquidity and price formation. The findings highlight the importance of adopting adaptive strategies that account for the speed and frequency of trading activities in modern markets. For instance, retail investors might benefit from understanding how informed trading influences liquidity and adjust their timing to avoid periods of heightened informed activity, where the risk of adverse selection is higher.

A common limitation across the three studies lies in their focus on specific sectors, markets, or indices, which restricts the generalizability of their findings. Each market or sector has unique dynamics, regulatory environments, and competitive landscapes that may not align with other regions or industries, limiting the applicability of the results to broader contexts.

The first study, while providing valuable insights, does not account for trading costs such as brokerage fees and spreads. These factors are critical as they directly influence investor behavior and the acquisition of private information, potentially affecting the accuracy of findings related to trading dynamics. Additionally, its focus on a single sector with unique characteristics limits the extent to which its findings can be extrapolated to other sectors with different market structures or competitive environments.

In the second study, its reliance on purely quantitative approaches limits its ability to capture qualitative aspects, such as managerial discretion in earnings management or nuanced decision-making processes, which could deepen the understanding of the observed relationships.

The third study faces challenges in addressing the complexities of modern financial markets. It does not fully incorporate the influence of evolving market characteristics such as technological advancements, varying regulatory frameworks, or liquidity changes. These factors significantly impact the relationship between informed trading and market liquidity. Additionally, its use of liquidity measures optimized for short-term analysis may not capture long-term trends effectively. This limitation could introduce biases when interpreting the sustained impact of informed trading on liquidity over extended periods.

References

Abedifar, P., Bouslah, K., and Zheng, Y., (2021). Stock price synchronicity and price informativeness: Evidence from a regulatory change in the U.S. banking industry. *Finance Research Letters*, 40, 101678

Agudelo, D. A., Giraldo, S., and Villarraga, E. (2015). Does pin measure information? Informed trading effects on returns and liquidity in six emerging markets. *International Review of Economics and Finance*, *39*, 149–161.

Ahern, K. R. (2020). Do Proxies for Informed Trading Measure Informed Trading? Evidence from Illegal Insider Trades. *The Review of Asset Pricing Studies*, 10(3), 397–440.

Alexander, G., and Peterson, M. (2007). An analysis of trade-size clustering and its relation to stealth trading. *Journal of Finance and Economics*, 84(2), 435–471.

Alford, A., and Berger, P., (1999). A simultaneous equation analysis of forecast accuracy, analyst following, and trading volume. *Journal of Accounting, Auditing and Finance*, 14, 219–246.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56.

Ammar, I. B., Hellara, S., and Ghadhab, I. (2020). High-frequency trading and stock liquidity: An intraday analysis. *Research in International Business and Finance*, *53*.

Amyx, D. A., and Luehlfing, M. S. (2006). Winner's curse and parallel sales channels—online auctions linked within e-tail websites. Information and Management, 43(8), 919–927.

Armanious, A., and Zhao, R. (2024). Stock liquidity effect on leverage: The role of debt security, financial constraint, and risk around the global financial crisis and Covid-19 pandemic. *International Review of Financial Analysis*, *92*.

Azevedo, A., Karim, M., Gregoriou, A. & Rhodes, M. (2014). Stock price and volume effects associated with changes in the composition of the FTSE Bursa Malaysian KLCI. *Journal of International Financial Markets, Institutions and Money*, 28, 20-35.

Avramov, D., Chordia, T., and Goyal, A. (2006). The impact of trades on daily volatility. *Review of Financial Studies,* 19(4), 1241–1277.

Bakke, T. E., and Whited, T. M. (2010). Which firms follow the market? An analysis of corporate investment decisions. *Review of Financial Studies,* 23(5), 1941-1980.

Barclay, M., and Warner, J. (1993). Stealth trading and volatility: which trades move prices? *Journal of Finance and Economics,* 34(3), 281–306.

Baumol, W. J. (1965). The Stock Market and Economic Efficiency. New York: Fordham University Press.

Bayona, A., Dumitrescu, A., and Manzano, C. (2023). Information and optimal trading strategies with dark pools. *Economic Modelling*, *126*.

Beatty, A., Ke, B., and Petroni, K. (2002). Earnings management to avoid earnings declines across publicly and privately held banks. *Accounting Review,* 77, 547-570.

Beneish, M. D. (2001). Earnings management: a perspective. *Managerial Finance,* 27(12), 3-17.

Bharath, S. T., Jayaraman, S., and Nagar, V. (2013). Exit as Governance: An Empirical Analysis. *Journal of Finance*, 68(6), 2515-2547.

Bhushan, R., 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics*, 22, 255–274.

Bond, P., Edmans, A., and Goldstein, I. (2012). The real effects of financial markets. *Annual Review of Financial Economics,* 4, 339–360.

Bond, P., Goldstein, I., and Prescott, E. S. (2010). Market-Based Corrective Actions. *The Review of Financial Studies*, *23*(2), 781.

Boot, A. W. A., and Thakor, A. V. (1997). Financial system architecture. *Review of Financial Studies,* 10(3), 693

Bouteska, A. (2019). The effect of investor sentiment on market reactions to financial earnings restatements: lessons from the United States. *Journal of Behavioral and Experimental Finance*, *24*.

Brown, S., and Hillegeist, S.A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, 12(2/3), 443–477.

Chakrabarty, B., Pascual, R., and Shkilko, A. (2015). Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the Lee-Ready algorithm. *Journal of Financial Markets*, 25, 52-79.

Chakravarty, S. (2001). Stealth trading: which trader's trades move prices? *Journal of Finance and Economics,* 61(2), 289–307.

Chan, K., and Hameed, A. (2006). Stock price synchronicity and analyst coverage in emerging markets. Journal of Financial Economics 80(1): 115-147.

Chan, K., Chan, L. K. C., Jegadeesh, N., and Lakonishok, J. (2006). Earnings quality and stock returns. *The Journal of Business*, 79(3), 1041–1082. https://doi.org/10.1086/500669

Chan, K., Chan, Y. C. (2014). Price informativeness and stock return synchronicity: Evidence from the pricing of seasoned equity offerings. *Journal of Financial Economics*, 114(1), 36–53.

Chang, S. S., and Wang, F. A. (2015). Adverse selection and the presence of informed trading. Journal of Empirical Finance, 33, 19–33.

Chang, S. S., Chang, L. V., and Wang, F. A. (2014). A dynamic intraday measure of the probability of informed trading and firm-specific return variation. *Journal of Empirical Finance*, *29*, 80–94.

Charfeddine, L., Riahi, R., and Omri, A. (2013). The Determinants of Earnings Management in Developing Countries: A Study in the Tunisian Context. *IUP Journal of Corporate Governance*, 12(1), 35–19.

Chen, Q., Goldstein, I., and Jiang, W. (2007). Price informativeness and investment sensitivity to stock price. *Review of Financial Studies,* 20(3), 619–650.

Cheng, Q., and Warfield, T. D. (2005). Equity incentives and earnings management. *Accounting Review*, 80(2), 441-476.

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

Chung, K. H., and Zhang, H. (2011). Corporate Governance and Institutional Ownership. *Journal of Financial and Quantitative Analysis* 46(1): 247-273.

Chung, K. H., Van Ness, B. F., and Van Ness, R. A. (1999). Limit orders and the bid-ask spread. *Journal of Financial Economics*, *53*(2), 255–287.

Chung, K., Li, M., and McInish, T. (2005). Information-based trading, price impact of trades, and trade autocorrelation. Journal of Banking and Finance, 29,1645–1669.

Chung, K.H., Chuwonganant, C. (2023). COVID-19 pandemic and the stock market: Liquidity, price efficiency, and Trading. *Journal of Financial Markets*, 64, 100803

Collin-Dufresne, P., and Fos, V. (2015). Do prices reveal the presence of informed trading? Journal of Finance 70, 1555–1582.

Corbet, S., Hou, Y. (Greg), Hu, Y., Oxley, L. (2021). Volatility spillovers during market supply shocks: The case of negative oil prices. *Resources Policy*, 74.

Dai, Y., Kong, D., and Wang, L. (2013). Information asymmetry, mutual funds and earnings management: Evidence from China. *China Journal of Accounting Research*, 3, 187-209.

Dasgupta, S., Gan, J., Gao, N., 2010. Transparency, price informativeness, and stock return synchronicity: theory and evidence. *Journal of Financial and Quantitative Analysis*, 45(5): 1189–1220.

Dechow, P. M., and Dichev, I. D. (2002). The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors. *The Accounting Review*, 77, 35–59.

Degeorge, F., Patel, J., and Zeckhauser, R. (1999). Earnings Management to Exceed Thresholds *. The Journal of Business, 72(1), 1–33.

Demirtas, K. O., and Zirek., D. (2011). Aggregate earnings and expected stock returns in emerging markets. *Emerging Markets Finance and Trade*, 47(3).

Dufour, A., and Engle, R. F. (2000). Time and the price impact of a trade. The Journal of Finance, 55, 2467−2498.

Durnev, A., Morck, R., and Yeung, B. (2004). Value-enhancing capital budgeting and firm-specific return variation. *Journal of Finance,* 59(1), 65-105.

Easley, D., & O' Hara, M. (1987). Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics*, *19*(1), 69.

Easley, D., and O'Hara, M. (1992). Time and the process of security price adjustment. Journal of Finance, 47,577–605.

Easley, D., Engle, R., O'Hara, M., and Wu, L. (2008). Time-varying arrival rates of informed and uninformed trades. Journal of Financial Econometrics, 6,171–207.

Easley, D., Hvidkjaer, S., and O'Hara, M. (2002). Is information risk a determinant of asset returns? *Journal of Finance*, 57, 2185–2221.

Easley, D., Hvidkjaer, S., and O'Hara, M. (2010). Factoring information into returns. Journal of Financial and Quantitative Analysis, 45,293–309.

Easley, D., Kiefer, N. H, O'Hara, M., and Paperman, J. (1996). Liquidity, information, and infrequently traded stocks. Journal of Finance, 51,1405–1436.

Easley, D., Kiefer, N. H., and O'Hara, M. (1997a). One day in the life of a very common stock. Review of Financial Studies, 10,805–835.

Easley, D., Kiefer, N. H., and O'Hara M. (1997b). The information content of the trading process. *Journal of Empirical Finance*, 4, 159-186.

Easley, D., Prado, L. M. M., and O'Hara, M. (2012). Flow Toxicity and Liquidity in a High-frequency World. *Review of Financial Studies*, 25(5), 1457–1493.

Edmans, A., Jayaraman, S., and Schneemeier, J. (2017). The source of information in prices and investment-price sensitivity. *Journal of Financial Economics*, 126(1), 74-96.

Elder, J., and Serletis, A., (2010). Oil Price Uncertainty. *Journal of Money, Credit and Banking*, 42(6), 1137–1159.

Ellis, K., Michaely, R., and O'Hara, M. (2000). The accuracy of trade classification rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis*, 35, 529-551.

Elshandidy, T., and Elsayed, M. (2024). The UK financial firms' textual risk disclosure and market liquidity. *Finance Research Letters*, *63*.

Fakhfakh H. and Nasfi F. (2012). The determinants of Earnings Management by Acquiring Firms. *Journal of Business Studies Quarterly*, 3(4), 43-57.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3.

Fama, E., and MacBeth, J., (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, 81(3), 607–636.

Fan, Q., and Jahan-Parvar, M. R. (2012). U.S. industry-level returns and oil prices. *International Review of Economics and Finance*, *22*(1), 112–128. https://doi.org/10.1016/j.iref.2011.09.004

Farooq, O., and Ahmed, S. (2014). Stock Price Synchronicity and Corporate Governance Mechanisms: Evidence from an Emerging Market. *International Journal of Accounting, Auditing and Performance Evaluation* 10(4): 395-409.

Fatemeh, D., and Narjes, M. (2013). A study on the effect of information asymmetry on earning management: Evidence from Tehran Stock Exchange. *Management Science Letters*, 3(7), 2161–2166.

Fathi, J. (2013). The determinants of quality of financial information disclosed by French listed companies. *Mediterranean Journal of Social Sciences*, 4(2).

Fernandes, N., and Ferreira, M. A. (20090. Insider Trading Laws and Stock Price Informativeness. *The Review of Financial Studies,* 22: 1845–887.

Finucane, T. J. 2000. A Direct Test of Methods for Inferring Trade Direction from Intra-Day Data. *Journal of Financial and Quantitative Analysis*, 35(4), 553–576.

Fisher, K., and Statman, M. (2000). Investor sentiment and stock returns. *Financial Analysts Journal*, 56 (2), 16–23.

Fisher, K.L., and Statman, M. (2006). Market timing in regressions and reality. *Journal of Financial Research*, 29 (3), 293–304.

Fishman, M. J., and Hagerty, K. M. (1989). Disclosure decisions by firms and the competition for price efficiency. *Journal of Finance,* 44(3), 633-646.

Florackis, C., Gregoriou, A. and Kostakis, A. (2011). Trading frequency and asset pricing on the London Stock Exchange: evidence firm a new price impact ratio. *Journal of Banking and Finance*, 35(12), 3335-3350.

Foreman, P., and Murnighan, J. K. (1996). Learning to avoid the winner's curse. Organizational Behavior and Human Decision Processes, 67(2), 170–180.

Foster, F. D., and Viswanathan, S. (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. Review of Financial Studies, 3, 593−624.

Foucault, T., and Fresard, L. (2012). Cross-listing, investment sensitivity to stock price and the learning hypothesis. *The Review of Financial Studies*, 25(11), 3305–3350.

Franz, D. R., Hassab Elnaby, H. R., and Lobo, G. J. (2014). Impact of proximity to debt covenant violation on earnings management. *Review of Accounting Studies*, 19(1), 473–505. https://doi.org/10.1007/s11142-013-9252-9

Gao, P., and Liang, P. J. (2013). Informational feedback, adverse selection, and optimal disclosure policy. *Journal of Accounting Research,* 51(5), 1133–1158.

Giannetti, M., and Simonov, A. (2006). Which Investors Fear Expropriation? Evidence from Investors' Portfolio Choices. *Journal of Finance* 61(3): 1507–1547

Glosten, L. R., and Milgrom, P. R. (1985). Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14, 71–100.

Goh, J., Lee, H., and Lee, J. (2013). Majority Shareholder Ownership and Real Earnings Management: Evidence from Korea. *Journal of International Financial Management and Accounting*, 24(1), 26–61.

Graham, J. R., Harvey, C. R., and Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, 40(1), 3–73.

Grossman, S. J., and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review,* 70(3), 393-408.

Halioui, K., Neifar, S. and Abdelaziz. F. (2016). "Corporate governance, CEO compensation and tax aggressiveness: evidence from American firms listed on the NASDAQ 100". *Review of Accounting and Finance*, Vol. 15 No. 4, pp. 445-462.

Hasbrouck, J. (1991). Measuring the information content of stock trades. Journal of Finance, 46,179–207.

Hayashi, T., and Nishide, K. (2024). Strategic liquidity provision in high-frequency trading. *International Review of Financial Analysis*, *93*.

Hayek, F. (1945). The use of knowledge in society. *American Economic Review*, 35, 519–530.

He, W., Li, D., Shen, J., and Zhang, B. 2013. Large Foreign Ownership and Stock Price Informativeness Around the World. *Journal of International Money and Finance* 36: 211-230.

Healy, P. M., and Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, 13(4), 365–384.

Heck, R. H., Thomas, S. L., and Tabata, L. N. (2014). Multi-level and longitudinal modeling with IBM SPSS (2nd ed). Routledge: New York.

Heng, P., Niblock, S. J., Harrison, J. L., and Hu, H. (2020). The Impact of High-Frequency Trading on Australian Futures Market Liquidity and Efficiency. *Journal of Derivatives*, *27*(4), 51–76.

Hirschey, N. (2021). Do High-Frequency Traders Anticipate Buying and Selling Pressure? *Management Science*, *67*(6), 3321–3345. https://doi.org/10.1287/mnsc.2020.3608

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

Holthausen, R. W., Larcker, D. F., and Sloan, R. G. (1995). Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics*, 19(2), 29–74.

Hribar, P., and Collins, D. W. (2002). Errors in estimating accruals: Implications for empirical research. Journal of Accounting Research, 40(1), 105–134.

Hsu, C.H. 2016. Strategic noise trading of later-informed traders in a multi-market framework. *Economic Modelling*, *54*, 235–243.

Hu, C., Liu, S., (2013). The Implications of Low R2: Evidence from China. *Emerging Markets Finance and Trade*, 49(1), 17–32.

Hurwitz, H. (2018). Investor sentiment and management earnings forecast bias. *Journal of Business Finance and Accounting*, 45(1/2), 166–183.

Hutton, A. P., Marcus, A. J., and Tehranian, H. (2009). Opaque financial reports, R^2, and crash risk. *Journal of Financial Economics*, 94(1), 67–86.

Ibikunle, G., Gregoriou, A., Hoepner, A. G. F., Rhodes, M., 2016. Liquidity and market efficiency in the world’s largest carbon market. *The British Accounting Review*, 48(4), 431–447.

Inekwe, J. N. 2020. Liquidity connectedness and output synchronization. *Journal of International Financial Markets, Institutions and Money*, 66.

Jin, L., Myers, S. C., (2006). R2 Around the World: New Theory and New Tests. *Journal of Financial Economics*, 79: 257–292.

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, 29(2), 193-228.

Jones, C., G. Kaul, and M. Lipson. (1994). Transactions, Volume, and Volatility. *Review of Financial Studies*, 7, 631–651.

Kang, Q., and Liu, Q. (2008). Stock trading, information production, and executive incentives. *Journal of Corporate Finance,* 14(4), 484-98

Kau, J. B., Linck, J. S., Rubin, P. H. (2008). Do managers listen to the market?. *Corporate Finance*, 14(4):34

Keiber, K. L., 2007. Reconsidering the Impossibility of Informationally Efficient Markets. *Applied Financial Economics*, 17(13–15), 1113–1122.

Keim, D., and Madhavan, A. 1995. Anatomy of the trading process: empirical evidence on the behavior of institutional traders. *Journal of Finance and Economics,* 37(3), 371–399.

Kelly, P. J., 2014. Information Efficiency and Firm-Specific Return Variation. *Quarterly Journal of Finance*, 4(4), 1–44.

Kumari, J. (2019). Investor sentiment and stock market liquidity: Evidence from an emerging economy. *Journal of Behavioral and Experimental Finance*, *23*, 166–180.

Kurov, A. (2010). Investor sentiment and the stock market's reaction to monetary policy. *Journal of Banking and Finance*, 34 (1), 139–149.

Kyle, A, S. (1985). Continuous auctions and insider trading. *Econometrica,* 53(6),1315-36

La Porta, R., Lopez, S. F., Shleifer, A., & Vishny, R. W. (2000). Agency problems and dividend policies around the worl. *The Journal of Finance*, 55(1), 1-33.

Le H, Gregoriou A. (2020). How Do You Capture Liquidity? A Review of The Literature on Low-Frequency Stock Liquidity. *Journal of Economic Surveys*, 34, 1170-1186.

Le H, Gregoriou A. (2022). Liquidity and asset pricing: evidence from a new free-float-adjusted price impact ratio. *Journal of Economic Studies*, 49, 751-771.

Le, A., Yin, X., and Zhao, J. (2019). Informed trading around earnings announcements in Australia. *Pacific-Basin Finance Journal*, *58*.

Lee, C. M. C., and Radhakrishna, B. (2000). Inferring Investor Behavior: Evidence from TORQ Data. *Journal of Financial Markets*, 3(2), 83–111

Lee, C. M. C., and Ready, M. L. (1991). Inferring Trade Direction from Intraday Data. *The Journal of Finance*, 46(2), 733–746.

Lei, Q., and Wu, G. (2005). Time-varying informed and uninformed trading activities. Journal of Financial Markets, 8,153–181.

Lin, H. W., and Ke, W. C. (2011). A computing bias in estimating the probability of informed trading. *Journal of Financial Markets*, 14, 625–640.

Liu, S. (2015). Investor Sentiment and Stock Market Liquidity. *Journal of Behavioral Finance*, *16*(1), 51–67.

Loureiro, G, R., and Taboada, A, G. (2015). Do Improvements in the Information Environment Enhance Insiders' Ability to Learn from Outsiders? *Journal of Accounting Research*, 53(4).

Luo, Y. (2005). Do insiders learn from outsiders? Evidence from mergers and acquisitions. *Journal of Finance,* 60(4), 19(5), 1-82.

Mackay, P., and Phillips, G. (2005). How does industry affect firm financial structure? *The Review of Financial Studies*, 18, 1433–1466.

Manganelli, S. (2005). Duration, volume and volatility impact on trades. Journal of Financial markets, 8,377−399.

Marks, J. M., and Shang, C. (2024). Business seasonality and stock liquidity. *Journal of Financial Markets*, *67*.

Moardi, M., Salehi, M., Poursasan, S., and Molavi, H. (2020). Relationship between earnings management, ceo compensation, and stock return on Tehran stock exchange*. International Journal of Organization Theory and Behavior*, 23(1), 1–22. https://doi.org/10.1108/IJOTB-12-2018-0133

Morck, R., Yeung, B., and Yu, W. (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58(1): 215-260.

Nagar, V., Nanda, D., and Wysocki, P. (2003). Discretionary disclosure and stock-based incentives. *Journal of Accounting and Economics*, 34(1), 283–309.

Nezafat, M., and Schroder, M., (2022). Private Information, Securities Lending, and Asset Prices. *Review of Financial Studies*, 35(2), 1009–1063.

Nuanpradit, S. (2019). Real earnings management in Thailand: CEO duality and serviced early years. *Asia - Pacific Journal of Business Administration*, *11*(1), 88–108.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets*, 3, 259.

Park, K. (2017). Pay disparities within top management teams and earning management. *Journal of Accounting and Public Policy*, *36*(1), 59–81.

Patton, A. J., Verardo, M., 2012. Does beta move with news? Firm-specific information flows and learning about profitability. *Review Financial Studies*, 25 (9), 2789–2839

Pedraza, A. (2020). Informed Trading in Business Groups. *The World Bank Economic Review*, *34*(2), 351–370.

Peress, J. (2014). Learning from Stock Prices and Economic Growth. *The Review of Financial Studies*, *27*(10), 2998–3059.

Piotroski, J. D., Roulstone, B. T., (2004). The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *Accounting Review,* 79: 1119–1151.

Poppe, T., Moos, S., and Schiereck, D. (2016). The sensitivity of VPIN to the choice of trade classification algorithm. *Journal of Banking and Finance*, 73, 165-181.

Putnins, T.J., and Michayluk, D. (2018). When do informed traders provide liquidity? Available at SSRN: http://dx.doi.org/10.2139/ssrn.2312330

Quinlan, C., Babin, B., Carr, J., Griffin, M., and Zikmund, W. (2015). *Business research methods.* Andover, Hampshire: Cengage Learning.

Raman, K., and Shahrur, H. (2008). Relationship-specific investments and earnings management: Evidence on corporate suppliers and customers. *The Accounting Review*, 83(7), 1041–1081.

Rao, L., Zhou, L., (2019). Crash risk, institutional investors and stock returns. *North American Journal of Economics and Finance*, 50.

Rappaport, A. (1987). Stock market signals to managers. *Harvard Business Review*, 65(6), 57–62.

Richardson, V. J. (2000). Information Asymmetry and Earnings Management: Some Evidence. *Review of Quantitative Finance and Accounting*, 15(4), 325.

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

Roll, R. (1988). Presidential address: R^2. *Journal of Finance*. 43, 540–566.

Ronen, J., and Yaari, V. (2007). Demand for the truth in principal–agent relationships. *Review of Accounting Studies*, 12(1), 125.

Roşu, I. (2020). Liquidity and Information in Limit Order Markets. *Journal of Financial and Quantitative Analysis*, *55*(6), 1792–1839.

Santana, S. C. V., Guimarães dos Santos, L. P., Valentim de Oliveira Carvalho Júnior, C., and Lopo Martinez, A. (2020). Investor sentiment and earnings management in Brazil. *Revista Contabilidade and Finanças - USP*, 31(83), 283–301.

Saunders, M., Lewis, P., and Thornhill, A. (2016). *Research methods for business students (7th ed.).* Harlow: Pearson.

Schipper, K. (1989). Commentary on earnings management. *Accounting Horizons*, 3(4), 91-102.

Schwert, G. W. (1990). Stock Volatility and the Crash of '87. *Review of Financial Studies*, 3(1), 77–102.

Simpson, A. (2013). Does investor sentiment affect earnings management? *Journal of Business Finance and Accounting*, 40(7), 869-900.

Sletten, E. (2012.). The effect of stock price on discretionary disclosure. *Review of Accounting Studies*, *17*(1), 96–133. .

Subrahmanyam, A., and Titman, S. (1999). The going public decision and the development of financial markets. *Journal of Finance,* 54(3), 1045–1082.

Sweeney, A. P. 1994. Debt-covenant violations and managers' accounting response. Journal of Accounting and Economics17 (3): 281–30

Verrecchia, R. E. (1982). Information acquisition in a noisy rational expectations economy. *Econometrica,* 50(5), 1415–1430.

Warfield, T., Wild, J., and Wild, K. L. (1995). Managerial Ownership, Accounting Choices and Informativeness of Earnings. Journal of Accounting and Economics, 20 (1), 61–91. https://doi.org/10.1016/0165-4101(94) 00393-J

Watts, R. L., and Zimmerman, J. L. (1978). Towards a positive theory of the determination of accounting standards. *Accounting Review*, 53(1), 112-134.

Wong, W. K., Tan, D., and Tian, Y. (2009). Informed trading and liquidity in the Shanghai stock exchange. *International Review of Financial Analysis*, *18*(1-2), 66–73.

Yan, Y., and Zhang, S. (2012). An improved estimation method and empirical properties of the probability of informed trading. *Journal of Banking and Finance*, 36(2), 454–467.

Yao, L., and Qiu, Y. (2024). Market liquidity, credit maturity structure and asset mismatch in manufacturing firms. *Finance Research Letters*, 105400.

Yazdi, A. K., Wanke, P. F., Hanne, T., Abdi, F., Sarfaraz, A. H., 2022. Supplier selection in the oil and gas industry: A comprehensive approach for Multi-Criteria Decision Analysis. *Socio-Economic Planning Sciences*, 79.

Yeoh, P. 2010. Dark pools. *Business Law Review*, *31*(1).

Zuo, L. (2016). The informational feedback effect of stock prices on management forecasts. *Journal of Accounting and Economics*, 61(2-3), 391-413.

Beneish, M. D. (1997): Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance, Journal of Accounting and Public Policy, 16(3):271- 309.

McNichols, M., & Wilson, G. P. (1988): Evidence of Earnings Management from the Provision for Bad Debts, Journal of Accounting Research, 26:1-31.

Cecchini, M., Jackson, S. B. & Liu, X. (2012): Do initial public offering firms manage accruals? Evidence from individual accounts, Review of Accounting Studies, 17(1):22-40.

Burgstahler, D. & Dichev, I. (1997): Earnings management to avoid earnings decreases and losses, Journal of Accounting and Economics, 24(1):99-126.

Degeorge, F., Patel, J. & Zeckhauser, R. (1999): Earnings Management to Exceed Thresholds, The Journal of Business, 72(1):1-33.

Gore, P., Pope, P. F. & Singh, A. K. (2007): Earnings management and the distribution of earnings relative to targets: UK evidence, Accounting & Business Research (Wolters Kluwer UK), 37(2):123-149.

Dechow, P.M., Sloan, R.G., Sweeney, A.P., 1995. Detecting earnings management. The Accounting Review 70 (2), 193-225.

Dechow, P. M., Hutton, A. P., Kim, J. H. & Sloan, R. G. (2012) Detecting Earnings Management: A New Approach, Journal of Accounting Research, 50(2):275-334

Kothari, S. P., Leone A. J. & Wasley, C. E. (2005): Performance matched discretionary accrual measures, Journal of Accounting and Economics, 39(1):163-197. 21.

Peasnell, K. V., Pope, P. F. & Young, S. (2000): Detecting earnings management using cross-sectional abnormal accruals models. Accounting and Business Research, 30(4):313-326

Marai, A. & Pavlović, V. (2014). An Overview Of Earnings Management Measurement Approaches: Development And Evaluation. 11. 21-36.

McNichols, M. F. (2000): Research design issues in earnings management studies. Journal of Accounting and Public Policy, 19:313-345.

Francis, J. R., Khurana, I. K., & Pereira, R. (2005). Disclosure incentives and effects on cost of capital around the world. *Accounting Review*, 80(4), 1125-1162.

Kothari, S. P. (2001). Capital markets research in Accounting. *Journal of Accounting and Economics*, 31,105–231.

Healy, P.M., Wahlen, J.M., 1999. A review of the earnings management literature and its implications for standard setting. Accounting Horizons 13 (14), 365-383.

DeAngelo, L., 1986. Accounting numbers as market valuation substitutes: a study of management buyouts of public stockholders. The Accounting Review 61 (3), 400-420

Liu, W. (2006) A liquidity-augmented capital asset pricing model. Journal of Financial Economics 82: 631–671

Amihud, Y. and Mendelson, H. (2012) Liquidity, the value of the firm, and corporate finance. Journal of AppliedCorporate Finance 24: 17–32.

Sarr, A. and Lybek, T. (2002) Measuring liquidity in financial markets. IMF Working Paper, 1–64.

Gregoriou, A. (2013) Earnings announcements and the components of the bid-ask spread: evidence from theLondon stock exchange. Journal of Economic Studies 40: 112–126

Chordia, T., Roll, R. and Subrahmanyam, A. (2000) Commonality in liquidity. Journal of Financial Economics56: 3–28.

Roll, R. (1984). A simple implicit measure of the effective bid–ask spread in an efficient market. Journal of Finance, 39(4): 1127–1139.

Acharya, V. and Pedersen, L. (2005) Asset pricing with liquidity risk. Journal of Financial Economics 77:375–410.

Amihud, Y., Mendelson, H. and Pedersen, L. (2005) Liquidity and asset prices. Foundations and Trends inFinance 1: 269–364.

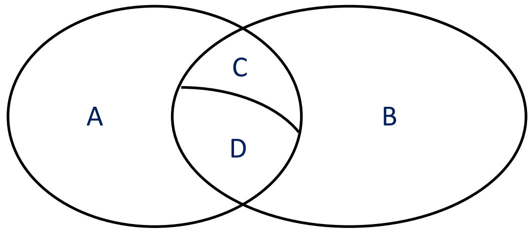
Corwin, S., & Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. Journal of Finance 67, 719–760.

Gabrielsen, A., Marzo, M., & Zagaglia, P. (2011). Measuring market liquidity: an introductory survey. SSRN Electronic Journal

Demsetz, H. (1968). The Cost of Transacting. The Quarterly Journal of Economics, Vol. 82 (1), 33-53.

Appendix 1

Figure 1 - Relation between Information Owned by Different Participants



Source: Developed by the authors following Zuo (2016)

A: Information owned only by informed investors
B: Information owned only by managers
C: Information owned by managers and informed investors (uninformed investors do not know this information)
D: Public information known by all parties including managers, informed and uninformed traders

Manager's total information = B + C + D
Informed investor's total information = A + C + D

Informed investors' private information = A + C

Appendix 2

Table 1: This table reports the regression coefficient signs of the existing literature for the relationship between the stock price non-synchronicity (SYNCH) and the variables used in those studies. The PI accounts for the investors' private information and it is proxied in the study by the DPIN and SDPIN measures; the control variables used in the study include: the idiosyncratic volatility (IVOL) measured by the three-factor model of Fama-French (1993), firm size (SIZE) measured by the market capitalization divided by 10^6, volume (VOL) measured by the stock trade volume divided by 10^6, bid-ask spread (SPREAD) which represents the difference between the highest ask price and the lowest bid price, illiquidity risk (ILLIQ) proxied by the Amihud (2002) illiquidity measure, and stock return (RETURN) which is the difference between the closing price at day *t* and the day before divided by the closing price at day *t*-1; the LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, LagRETURN are the lag level 1 of the following variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

Variables	*Correlation Coefficient*		
	Positive (+)	Negative (-)	Expected sign
PI (DPIN, SDPIN)	Roll (1988) Morck et al. (2000) Durnev et al. (2004) Jin and Myers (2006) Chen et al. (2007) Fernandes and Ferreira (2009) Hu and Liu (2013) Chang et al. (2014) Zuo (2016)	Chan and Hammed (2006) Dasgupta et al. (2010) Kelly (2014)	+
IVOL	Kelly (2014) Chan and Chan (2014) Rao and Zhou (2019)	Morck et al. (2000) Durnev et al. (2004) Piotroski and Roulstone (2004)	+
SIZE		Chan and Hameed (2006) Hutton et al. (2009) Chan and Chan (2014) Chang et al. (2014) Abedifar et al. (2021)	-
SPREAD	Kelly (2014) Chan et al. (2013)	Patton and Verardo (2012) Ibikunlea et al. (2016) InekI (2019)	+
VOL		Chan and Hameed (2006) Chang et al. (2014)	-
ILLIQ	Chan and Chan (2014) Rao and Zhou (2019) Abedifar, et al. (2021)	Chang et al. (2014) InekI (2019)	+
RETURN	Chan and Chan (2014)	Chang et al. (2014)	+
LagSDPIN	Chang et al. (2014)		+
LagSPREAD	Kelly (2014)		+
LagVOL		Chang et al. (2014)	-
LagILLIQ		Chang et al. (2014)	+
LagRETURN	Chang et al. (2014)	Chang et al. (2014)	+

Note: Chang et al. (2014) find both positive and negative signs for the relationship between LagRETURN and SYNCH.

Table 2: This table present the type of data that need to be collected for each variable. SYNCH is the proxy for stock price non-synchronicity and is a dependent variable; it is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according to Eq.(6) and Eq.(8), respectively. Control variables include Idiosyncratic Volatility (IVOL) measured by Fama-French (1993) in their three-factor model; Firm Size (SIZE) is the firm's daily market capitalization divided by 10^6; Volume (VOL) is the Stock daily volume divided by 10^6; Bid-Ask Spread (SPREAD) is calculated as the difference between the highest Ask Price and the lowest Bid Price in the day; illiquidity risk (ILLIQ) is Amihud (2002) illiquidity measure; and Daily Stock Return (RETURN) is estimated as the return of day *t* compared to day *t*-1. LagSDPIN, lag SPREAD, LagVOL, LagILLIQ, LagVOL, and LagRETURN are the lag level 1 of each variable SDPIN, VOL, ILLIQ, VOL, RETURN, respectively.

No.	Variable	Description	Data Type	Timeframe
1	SYNCH	Stock price non-synchronicity (Dependent variable)	Closing price of the stock	Daily
			S&P 500 index	Daily
2	DPIN	Investors' private information measured by DPIN	Closing price of a stock	Daily
			Executed price for each trade	Intraday
			Executed volume for each trade	Intraday
			Lowest bid price for each executed trade	Intraday
			Highest ask price for each executed trade	Intraday
			Quote for each trade	Intraday
3	SDPIN	Investors' private information measured by SDPIN	Daily closing price of a stock	Daily
			Executed price for each trade	Intraday
			Executed volume for each trade	Intraday
			Lowest bid price for each executed trade	Intraday
			Highest ask price for each executed trade	Intraday
			Quote for each trade	Intraday
3	IVOL	Idiosyncratic volatility measured by Fama-French model	Closing price of the stock	Daily
			S& P 500 index	Daily
4	SIZE	Firm size	Closing price of the stock	Daily
			Number of outstanding shares	Daily
5	VOL	Stock trade volume	Stock Trading Volume	Daily
6	SPREAD	Bid-ask spread	Bid price	Daily
			Ask price	Daily
7	ILLIQ	Stock illiquidity measured by Amihud (2002)	Closing price of the stock	Daily
			Trading volume	Daily