White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/234130/

Monograph:
Vu, Thu and STAFYLAS, DIMITRIOS (2025) A New Measure for Investors' Private Information. Working Paper.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

A New Measure for Investors' Private Information

Thu Vu [a], Dimitrios Stafylas [b *]

[a] *Aston Business School, Aston University*
[b] *School for Business and Society, University of York*

ABSTRACT

We develop a new measure for investors' private information (SDPIN) embedded in a stock price. Unlike existing measures such as PIN and DPIN, it considers both trade order size and trade frequency; we argue that private information can also be revealed through trade order size, thereby applying to a broader range of stock trading scenarios. We test the accuracy of our new private information measure against PIN and DPIN and find it to be more precise. Consequenlty, it reduces information asymmetries among investors and helps managers better assess the private information reflected in their firms' stock prices, enabling them to make more informed (optimal) decisions.

Keywords: Informed Trading, Price Discovery, Private Information, Stock Price Non-synchronicity.

JEL Classification: C49; G14; G19

* Corresponding author: dimitrios.stafylas@york.ac.uk. University of York, Church Lane Building, YO10 5ZF, York, UK.

1. Introduction

Investors who possess private information about a public firm (informed investors) use it to exploit investment opportunities unavailable to other investors. The current literature indicates that the actions of informed investors influence not only the firm's stock price but also managerial decisions (Bakke and Whited, 2010; Zuo, 2016)[1], and consequenlty impact economic development (Bond et al., 2012). Some studies also suggest that private information of investors may assist managers in making more optimal decisions, such as regarding mergers and acquisitions (Luo, 2005; Kau et al., 2008), investment projects (Chen et al., 2007; Bakke and Whited, 2010), and in more accurately forecasting the firm's earnings (Loureiro and Taboada, 2015; Zuo, 2016). However, the level of private information varies across stocks due to the asymmetric costs involved in acquiring new information (Grossman and Stiglitz, 1980; Keiber, 2007; Nezafat and Schroder, 2022).

The stock price is the ultimate measure of the shareholders' wealth. Therefore, managers should carefully identify the key drivers of the stock price (Rappaport, 1987). Numerous reasons can cause a stock proce to change, one of which is the presence of private information hold by a few investors. Therefore, measuring the amount of investors' private information embedded in the stock price is important. However, it is very challenging task because private information is known by very few investors who intentionaly try to conceal it from the market. Informed investors employ various strategies to hide their trading activities: they break their trades into smaller orders, using the so-called "smokescreen" trading (Keim and Madhavan, 1995; Chordia and Subrahmanyam, 2004), they set up multiple accounts and and use different drading platforms to anonymise trades, through "dark pools" (Yeoh, 2010; Bayona et al., 2023)[2], and they can strategically trade on unrelated assets (Hsu, 2018) to mislead other investors. Besides these strategies that increase the complexity of measuring private information in stock prices, market sentiment, irrational behaviors, and unpredictable macroeconomic and geopolitical factors also influence stock prices events.

The first attempt to measure the level of private 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. Roll (1998) advocates that stock price non-synchronicity is a good proxy for private information and attributes the actual price shifts of stocks primarily to the presence of firm-specific information. This measure of private information has been used in several studies, such as those by Morck et al. (2000), Jin and Myers (2006), Hutton et al. (2009), and Zuo (2016).

However, Roll's (1998) measure is not effective when firm-specific information does not fully represent all the private information held by investors. Therefore, two new measures were introduced:

[1] In the Appendix A.1, we illustrate the overlap between the information possessd by informed investors and that owned by the firm's managers.

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

the probability of informed trading (PIN) and the dynamic measure for the probability of informed trading (DPIN). The former was developed by Easley et al. (2002), estimating the likelihood of informed trading based on a sequential trade model, while the latter, developed by Chang et al. (2014), is a dynamic private information measure particularly useful for high-speed trading scenarios.

The PIN and DPIN measures proved to be more accurate than Roll's (1998) price non-synchronicity measure. Unlike Roll's (1998) measure, these measures not only rely on price movements but also consider the characteristics of trading, which often convey information not yet reflected in the stock price. Both 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 address those limitations.

We note that estimating the PIN measure using high-frequency trade data can be notably time-consuming due to the several intricate steps involved in the process. Specifically, high-frequency trade data require a high volume of trades to be processed and classified accurately, which necesitates the use of sophisticated algorithms, such as the quote rule, tick rule, or the Lee-Ready, for trade classification across stock exchanges. Moreover, estimating 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 handling floating-point exceptions[3] and boundary solutions[4]. Additionally, parameter estimation over multiple days demands significant computational power and time due to the need to manage a large volume of high-frequency data points to ensure accuracy and statistical significance. The meticulous nature of these computations, combined with the volume of data to be processed and analysed, makes calculating the PIN measure a complex and relatively time-consuming process task.

Previous studies have highlighted that investors' private information can also be identified through the size of trade orders. For instance, Easley and O'Hara (1987) show that informed traders typically initiate larger orders which contain greater informational content compared to smaller orders whereas Barclay and Warner (1993), Chakravarty (2001), and Alexander & Peterson (2007) show that most of the cumulative stock-price change is due to medium-size trades. This evidence aligns with the hypothesis that informed trades are concentrated in the medium-size category and that price movements are mainly driven by informed traders' private information. However, order size is an aspect neglected by both PIN and DPIN private information measures.

[3] Floating-point exceptions occur during computations involving floating-point numbers when operations produce results that are mathematically undefined or unrepresentable within the constraints of the floating-point system. These exceptions include dividing by zero, excedding numerical limits, performing invalid operations like square roots of negative numbers, and handling infinite values.

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

This paper aims to develop a new investors' private information measure that overcomes the limitations of the PIN and DPIN measures mentioned earlier. Our new measure for private information (SDPIN) builds on the DPIN but considers the trade order size (volume). The DPIN measure is effective when investors, including informed investors, trade stocks using same-size trade orders. However, empirical evidence shows that that this is rarely the case, so we argue that our SDPIN measure is more effective than the DPIN because it also captures private information revealed by choosing different trade order sizes. Our methodology for estimating the SDPIN follows that of Chang et al. (2014) in identifying informed trades but incorporates large, medium, and small trade orders sizes.

We study the relationship between stock return synchronicity and investors' private information using high-frequency data. We also examine the accuracy of our SDPIN measure compared to DPIN measures, and find that it is more precise, especially during algorithmic trading.[5] According to NYSE's rule 127.10, a large trade typically involves over 10,000 shares, with the minimum trade size set at 100 shares. However, what qualifies as a large or small order varies across stocks because of factors like the average daily trading volume and its influence on the stock market. Our data shows that, for the NYSE and NASDAQ, trade orders range from 1 to about 15,000 shares, lthough some investors place larger trade orders, ranging from hundreds of thousands to millions of shares. This empirical evidence highlights the wide variation in trade order sizes chosen by investors, motivating us to propose a new private information measure that incorporates trade order sizes, as well.

This research contributes to the financial markets literature as follows. Firstly, our SDPIN private information measure is more flexible than the existing (PIN and DPIN) measures, because it accommodates a broader range of trade order sizes, making it applicable to more diverse stock trade scenarios. This feature of our SDPIN measure enhances its reliability over existing private information measures, especially in high-frequency trading environments or when algorithmic trading is involved. Secondly, our SDPIN measure is easier to compute than the PIN measure when working with high-frequency data, providing an advantage over the more complex and time-consuming PIN estimation process. Thirdly, our SDPIN measure can also help us better understand the feedback loop between stock prices and the economy since a more precise estimation of the private information embedded in a stock price allows for a clearer understanding of how stock prices and and their movements influence managers' behaviours. Therefore, the improved accuracy of our SDPIN measure in detecting private information embedded in stock prices sheds light on how private information-driven stock prices impact managerial decisions and, ultimately, economic outcomes development.

Our new SDPIN private information measure can impact the behaviour of various financial market participants. It helps investors identify the level of private information embedded in a stock price, enabling them to can make more informed investment and risk management decisions. It benefits

[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.

all market players, including market regulators, by further information asymmetry among them. It is also valuable for firms' managers because it makes them more aware of the private information in their firm's stock price, allowing for more informed (optimal) managerial decisions. Additionally, it is useful for market regulators, as they can more accurately measure the presence of investors' private information and implement new regulaltions to reduce asymmetric information among investors and enehance transparency in the financial market.

The rest of the paper is structured as follows. Section 2 reviews the relevant literature and states the research hypotheses. Section 3 describes the data sample and research methodology. Section 4 presents our empirical findings. Section 5 concludes the paper.

2. Literature Review

2.1. Probability of Informed Trading

The idea of measuring the private information embedded in a stock price is credited to Easley et al. (1997), 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. (1997) 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 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. (1997) methodology – see Ellis et al. (2000), Lin and Ke (2011), and Yan and Zhang (2012), for instance.

We note that estimating the PIN measure depends on algorithms that classify trades into *buy* and *sell* orders and on a likelihood function for a stock single trading day to estimate the daily rates of noise and informed trades (Easley et al., 1997; Easley et al., 2002). Specifically, Easley et al. (1997) classify trades following Lee and Ready (1991), who develop a procedure that combines both the quote rule and the tick rule[7]; a trade above or below the midpoint is classified using the quote rule, while a trade at the midpoint is classified using the tick rule. They also compare trade prices to the trade quotes from the five seconds prior to the transaction. The algorithm used achieves accuracy between 73% and 91%, depending on the stock exchange (Finucane, 2000; Lee and Radhakrishna, 2000; Odders-White,

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

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

2000). Ellis et al. (2000) propose a new classification algorithm for NASDAQ, called the EMO algorithm. It is pointed out 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 occure away from the quotes.

Later on, Easley et al. (2012) introduced a high-frequency estimate of the Probability of Informed Trading (PIN), called the Volume-synchronized Probability of Informed Trading (VPIN) or the VPIN flow-toxicity metric. They devide time bars, which are short trading intervals within a day, into volume buckets where each contains an equal amount of traded volume, denoted as V. VPIN is then calculated as the average imbalance between buy and sell volumes across these buckets, serving as an important measure for detecting informed trading activities. The process requires a method to split volume into buys and sells, and a new volume classification method called Bulk Volume Classification (BVC) is proposed. This departs from standard trade classification schemes in two ways: first, volume is classified in bulk, and second, this method classifies part of a bar's volume as buy, and the rest as sell. Additionally, the BVC algorithm effectively estimates order flow toxity[8]; for example, it outperforms the tick rule and Lee-Ready's algorithm in some equities, gold, and oil futures markets.

However, 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. Pöppe et al. (2016) add that the BVC is not robust to the choice of classification algorithm, while traditional trade-by-trade classification algorithms, such as those 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 used is biased because large *buy/sell* transactions may trigger the power function embedded in the likelihood, producing a numerical value that exceeds the range of what computer software can handle. This problem is called a floating-point exception and, according to Lin and Ke (2011), it can significantly affect the estimation of the PIN measure. They suggest reformulating the likelihood function to address its shortcomings. Yan and Zhang (2012) raise another concern related the boundary values (0 and 1) of the variable representing the probability that traders acquire private information about the firm's fundamentals. If the probability is 0 or 1, it indicates that no private information event or uninformed trade ever occurred during a given period, which is unlikely. Therefore they propose using a modified factorized likelihood function (the LK factorization) to resolve the problem.

2.2. Probability of Informed Trading Dynamic Measure

Another key limitation of the PIN measure is that it is very difficult to apply with high-frequency data. This issue is worsened by the increasing popularity of high-frequency trading, and we observe that estimating informed trading using daily or weekly data does not fully capture all the information potentailly linked to stock trades. Regarding this perspective, Chang et al. (2014) introduce

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

a dynamic intraday measure for the probability of informed trading (DPIN). They emply a new methodology, based on Schwert (1990) and Avramov et al. (2006), to calculate the proportion of informed trades; this new measure also offers the benefit of being significantly easier to compute bacause it does not require, as the PIN measure does, estimating a function for the numerical optimization.

However, the DPIN measure relies solely on the number of trades, so it ovelooks information that could be embedded in investors' choice of trading stocks with different order sizes. Hence, we argue that it should be improved. Chang et al. (2014) were the first to highlight this issue and proposed two new measures, the DPIN_SIZE and DPIN_SMALL, which estimate private information based on either large or small trade orders. However, this remains insufficient because informed and uninformed investors can strategically use different order sizes to conceal informed trades. To address this problem, we propose a new dynamic intraday measure with size order effects (SDPIN) that considers both the number and size of informed trades.

2.3. Developing Research Hypotheses

The current microstructure literature recognizes the connection between stock price asynchrony and investors' private information. Stock return asynchrony results from firm-specific information, which occurs when a stock trade is driven by firm-specific information only. According to Roll (1998), private information accounts for a certain proportion of stock price asynchrony and indicates to what extent changes in stock return are influenced 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) argue that investors actively seek out not-yet-public firm-specific information to earn abnormal returns – this happens when some investors access private information before it becomes public and leverage this advantage. The privileged position of some (informed) investors impacts stock prices and increases stock return non-synchronicity.

The above literature assumes a positive relationship between investors' private information and the stock price non-synchronicity.[9] Therefore, we set the following research hypotheses:

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

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

3. Data Sample and Methodology

[9] In Appendix B.1, we summarize the results from previous studies on the relationship between stock price non-synchronicity and various factors, including investors' private information, idiosyncratic volatility, firm size, stock trading volume, Amihud's (2002) illiquidity measure, and stock return.

3.1. Data Sample

Our data sample inlcudes daily and intraday information on energy sector firms listed on the New York Stock Exchange (NYSE) or the National Association of Securities Dealers Automated Quotations (NASDAQ), covering the period from January 2018 to December 2020. Due to limited computational power, we could not use a data sample that includes all U.S. public firms. Instead, we selected the energy sector to apply our new measures for investors' private information. Although other sectors could have been considered, the energy sector is particularly suitable for testing our new private information measure because it is significantly influenced by new regulations and regulation changes, faces technological uncertainty, and is characterized as capital-intensive with specialized market segments. Additionally, it is more sensitive to geopolitical tensions among countries and plays an essential role in the performance of other sectors and the global economy, as demonstrated by the ongoing Russian-Ukraine war (Elder and Serletis, 2010; Yazdi et al., 2022).

The firms' data sample was collected from "Bloomberg" and "Wharton Research Data Services" (WRDS), while the intraday trading data was obtained from the "Trade and Quote" (TAQ). Our initial sample includes information on 236 stocks; we note that for a stock to be included in our sample, the firm must be listed on the exchanges for at least a year. Overall, the sample consists of 236 firms with 154,797 observations. Table 1, provides additional details on the data sample.

*****Insert Table 1 around here*****

3.2. Regression Models

3.2.1. Main Model

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

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

where $i = 1, \dots, n$ and $t = 1, \dots, m$, with *i* denoting the stock and *t* the day; $PI_{i,t}^{DPIN}$ represents investors' private information measured by the DPIN, and $PI_{i,t}^{SDPIN}$ represents investors' private information measured by the SDPIN; the CONTROL variable accounts for control variables, a vector that includes 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 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's (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 control vector ($CONTROL_{i,t-1}$); and ε_{it} is a zero-mean residual.[10]

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

3.2.2 The Estimation of Stock Price Non-Synchronicity

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

$$IR_{i,j,t} = \alpha_{i,0} + \alpha_{i,m}MR_{m,t} + \alpha_{i,j}SR_{j,t} + \varepsilon_{i,t} \quad (2)$$

where $IR_{i,j,t}$ is the return of stock of a firm *i* operating in sector *j*; $MR_{m,t}$ is the market return, and $SR_{j,t}$ is the return of industry *j*; the stock price synchronicity is represented 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 industry changes from those driven by overall economic changes. Additionally, industry returns are often disproportionally influenced by a small group of firms, so stock price movements may not acurately reflect the entire industry. Therefore, adding industry return to model (2) can produce misleading results. Moreover, since R^2 values are bounded between 0 and 1, they may not be ideal as a dependent variable. Consequently, instead of using the original non-synchronicity model proposed by Roll (1988) (Model 2 above), we adopt an adjusted model proposed by Morck et al. (2000), where we first calculate the coefficient of determination (R^2) from the following regression:

$$IR_{jt} = \alpha_{i0} + \beta_{im}MR_{mt} + \varepsilon_{it} \quad (3)$$

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

$$SYNCH_{it} = Log\frac{1-R^2}{R^2} \quad (4)$$

where R^2 is the stock price synchronicity.

3.2.3. DPIN Estimation

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

[10] Appendix C.1 outlines the types of data that must be collected for each variable.

(unexpected returns) based on the residual of the autoregressive model developed by Schwert (1990), which was subsequenlty modified by Jones et al. (1994) and Avramov et al. (2006) as follows:[11]

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

where $R_{i,j}$ is the dependent variable and represents the return on stock *i* at intraday interval *j,* with $j = 1, \dots, 26$; 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 during 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, while a *sell* order with a positive (negative) unexpected return is classified as an informed (uninformed) trade. The DPIN measure is, thus, constructed, as follows:

$$DPIN_{BASE_{i,j}} = \frac{NB_{i,j}}{NT_{i,j}}(\varepsilon_{i,j} < 0) + \frac{NS_{i,j}}{NT_{i,j}}(\varepsilon_{i,j} > 0) \quad (6)$$

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

3.2.4. SDPIN Estimation

In this subsection, we introduce our new measure of investors' private information, which is based on the private information measure by Chang et al. (2014). We proceed as follows: first, we use the Lee and Ready (1991) algorithm to classify orders; second, we isolate the unexpected component of returns (the residuals) from the following regression:[12]

$$R_{i,j} = \gamma_0 + \sum_{k=1}^{5}\gamma_{1i,k}D_k^{day} + \sum_{k=1}^{22}\gamma_{3i,k}R_{i,j-k} + \varepsilon_{i,j} \quad (7)$$

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

To identify informed trades, we 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 of the total volume of informed transactions (both buy-informed and sell-informed trades) to the total trading volume:

$$SDPIN_{i,j} = \frac{\sum V_{i,j}^{B}}{TV_{i,j}}\left(\varepsilon_{i,j} < 0\right) + \frac{\sum V_{i,j}^{S}}{TV_{i,j}}\left(\varepsilon_{i,k} > 0\right) \quad (8)$$

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.

[11] Chang et al. (2014) follow Jones et al. (1994) and Avramov et al. (2006) by regressing the daily return of individual stocks on its own 12 lags (covering about two weeks) instead of on the 22 lags (covering about one month) as in Schwert (1990) model. The DPIN is calculated for each 15-minute interval during the day, meaning that its average is based on 26 DPIN values per day.

[12] We estimate SDPIN directly for each trading day, replacing dummy variables for each 15-minute interval with dummy variables for each trading day of the week and including 22 lags of the daily return.

4. Empirical Results

4.1. Descriptive Analysis

In Table 2, we present the descriptive statistics of the variables used in our regression models. Panel A shows the statistics per year, while Panel B summarises the data for the entire sample period. Panel C shows the number of observations each year. Specifically, Panel A reveals that the mean value of the SDPIN is lower than that of the DPIN across all years. The mean value of the SYNCH consistently decreased from 2018 to 2020, while the mean values of both the DPIN and SDPIN increased. For example, between 2019 and 2020, the DPIN and SDPIN increased by 51.81% and 52.19%, respectively, whereas SYNCH decreased by 13%. This 13% drop indicates that stock prices had a higher level of co-movement with the market in 2020 compared to 2019. This change could be attributed to the impact of the COVID-19 crisis, which adversely affected the global market, with the energy industry being one of the most severely affected.[13] Panel B provides further evidence of the COVID-19 crisis's effects on the U.S. market. There was a significant decline in market liquidity in 2020, compared to 2019. Specifically, the mean values of the illiquidity measures IVOL and ILLIQ increased by 85.49 % and 75.44% respectively. The trade volume (VOL) rose by 56.48%, while market capitalization (SIZE) decreased by 31.14%. All these results align with those repoted by Chung and Chuwonganant (2023).

*****Insert Table 2 around here*****

Table 3 presents the correlation coefficient matrix; overall, it indicates that most correlations are below 0.5, except from certain pairs of regression-independent variables: IVOL vs. LagIVOL, SIZE vs. LagSIZE, SPREAD vs. LagSPREAD, and VOL vs. LagVOL. These pairs clearly suggest a multicollinearity issue. To assess this, we use the "variance inflation factors" (VIF) measure to evaluate the collinearity between the regressors in the model(s), with results shown in Table 4.[14] Additionally, within the correlation matrix, the two measures for private information, DPIN and SDPIN, are highly correlated. Notably, our SDPIN measure has a slightly stronger correlation with SYNCH than with the DPIN measure, supporting the use of our SDPIN measure.

*****Insert Table 3 around here*****

*****Insert Table 4 around here*****

4.2. Main Findings

In Tables 5 and 6, we present the results for the main model (Model 1); these show the relationship between stock return variations (measured by SYNCH) and investors' private information (measured by the DPIN and SDPIN measures). Specifically, while Table 5 shows the results for the

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

[14] The VIF measures the size of the variance of the coefficient estimates for regressors that have been inflated due to collinearity with other regressors. In Table 4, we present the VIF 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). As a result, the LagSIZE and LagIVOL independent variables were removed from the regression model.

entire sample, Table 6 shows results by year (2018-20). From Table 5, we observe that our SDPIN measure fits the market slightly better than the DPIN measure does, for all the years – its R^2 is 0.144 whereas that of the DPIN is 0.139. All the regression coefficients are statistically significant at the 1% level, for both the DPIN and SDPIN measures, except for RETURN and LagRETURN. This indicates that investors' private information reduces the synchronicity of stock price movements with market movements. This finding supports our research hypotheses *H1* and *H2* and aligns with previous literature (Roll, 1988; Durnev et al., 2004). We can confidently state that our SDPIN measure is a reliable measure for investors' private information and, when applied to our data sample, it slightly exceeds the reliability of the DPIN measure.

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

In Table 6, we present the annual results for the DPIN and SDPIN regression models. The R^2 and the coefficient of DPIN and SDPIN was lower in 2020 compared to 2018 and 2019. This could be due to the negative impact of the Covid-19 crisis on the U.S.. The coefficient of SDPIN is statistically significant and positively related to SYNCH, across all years. The relationship between DPIN and SYNCH in 2018 is negative but not significant. Furthermore, the three liquidity measures (ILLIQ, SPREAD, and IVOL) and their Lags (LagILLIQ, LagSPREAD, and Lag IVOL) are negative and statistically significant at the 1% level, as expected. These findings align with those of Morck et al. (2000) and Zuo (2016), who show that stocks with low R^2 exhibit the least trading and face the greatestobstacles to informed trading. According to Morck et al. (2000), markets with low R^2 stocks indicate a more efficient information environment for investors. For instance, stronger investor protection can enhance market liquidity, leading to a higher level of firm-specific information reflected 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 market movements. Consequently, if R^2 estimation is based on the value-weighted index, a positive relationship between the market capitalization of assets and stock return synchronicity (or a negative relationship with SYNCH) is expected. Previous literature shows that larger companies with high trading volumes of shares tend to attract more analysts (see Alford and Berger, 1999). Furthermore, Bhushan (1989) argues that the supply of analyst services is also influenced by the correlation between stock returns and market returns. For a given level of information acquisition

cost related to macro variables, a higher correlation (greater stock return synchronicity) reduces the marginal cost of acquiring infrmation,thereby increasing the supply of analyst services.

Notice that the level of trading volume affects stock return synchronicity because it influences the speed of price adjustments. We find that the coefficients for VOL and LagVOL are negative and statistically significant at a 1% level. This is because stocks traded very frequently react to market information promptly, making their individual price movements more synchronous with the market, whiles infrequently traded stocks experience greater delays in their price reactions, leading to lower stock returns synchronicity.

The signs of the coefficients for RETURN and LagRETURN differ from those for SYNCH, but they are not statistically significant. Finally, Table 6 also indicastes that in 2019, the coefficients for RETURN and Lag RETURN are positive and statistically significant. These results align with those of Chan and Chan's (2014) results.

*****Insert Table 5 around here*****

*****Insert Table 6 around here*****

4.3. Robustness Tests

For the robustness tests, like Chang et al. (2014), we follow the Fama and MacBeth (1973) framework[15] and use cross-sectional regressions to to estimate the regression parameters/coefficients. Then, we calculate the average across all days in the tiem series to get the parameter estimates. Specifically, we perform two robustness tests using the same data samplefrom Model (1) and the first-order differentiated data (the difference in value between data for the current day and the previous day) for the sensitivity test. The two Fama–MacBeth models are specified as follows:

Fama–MacBeth Robust Test 1:

$$SYNCH_{i,t} = \beta_0 + \beta_{1,t}PI_{i,t} + \beta_{2,t}PI_{i,t-1} + \beta_3 CONTROL_{i,t} + \beta_4 CONTROL_{i,t-1} + \varepsilon_{i,t} \quad (9)$$

Fama–MacBeth Robust Test 2:

$$\Delta SYNCH_{i,t} = \beta_0 + \beta_{1,t}\Delta PI_{i,t} + \beta_{2,t}\Delta PI_{i,t-1} + \beta_3 \Delta CONTROL_{i,t} + \beta_4 \Delta CONTROL_{i,t-1} + \varepsilon_{i,t} \quad (10)$$

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

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

In Tables 7 and 8, we present the results of the two Fama–MacBeth robustness tests. In the first robustness test, the parameters are estimated using the time-series average of cross-sectional

[15] The Fama and MacBeth's (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-step process that helps analyze how different variables or factors influence asset returns. The first step involves estimating *N* cross-sectional regressions, and the second step involves calculating *T* time-series averages of the coefficients from the *N*-cross-sectional regressions.

regressions. We find that the coefficients of the DPIN and SDPIN and the LagDPIN and LagSDPIN are positive and statistically significant at the 10% level; this supports the idea that the SPDIN is a reliable measure of investors' private information. Concerning the control variables, IVOL, SIZE, and SPREAD are significant at the 5%, 10%, and 10% levels, respectively, for both the DPIN and SDPIN robustness tests models.

For the second robustness check, the measures of the daily change of DPIN (ΔDPIN) and SDPIN (ΔSDPIN) indicate a negative relationship with the daily change of SYNCH (ΔSYNCH), but these coefficients are not statistically significant. The coefficient of ΔIVOL is negative and significant at the 10% level.

*****Insert Table 7 around here*****

*****Insert Table 8 around here*****

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 paper introduces a new measure of investors' private information, called SDPIN. It builds on Chang et al.'s (2014) DPIN measure but also considers the size of stock trade orders (trade volume), a parameter that has so been ovelloked by the aforementioned measures. We argue that private information can be revealed through trade size orders.

We test the accuracy of our SPDIN measure and compare it to those of the PIN and DPIN measures using intraday data from the U.S. energy sector which inlcudes information on 236 firms listed on the NYSE and NASDAQ from 2018 to 2020. We find a positive and significant correlation between the SDPIN measure and stock price non-synchronicity, supporting the reliability of this new private information measure. It proves to be more effective than the PIN and DPIN measures, especially during algorithm trading. As algorithm trading becomes more popular and is expected to grow, develpoing our SDPIN measure can be seen as a valuable contribution to the literature.

The use of the SDPIN measure can have positive implications for investors and the firms' managers. Specifically, it enables market participants to more accurately assess the level of private information embedded in a stock price, thereby reducing information asymmetries among investors and increasing financial market transparency. The firms' managers can more easily determine the amount of private information inlcuded in the stock price, and using this information allows them to make more informed (optimal) decisions.

For future research, we suggest testing our SDPIN measure in other sectors and financial markets, including those in developing countries. It would be valuable to see how our SDPIN measure performs compared to the PIN and DPIN measures in these new markets and microstructure enviroments. Additionally, studying the relationship between our SDPIN measure and stock liquidity as well as firms' earnings management, would be interesting.

References

Abedifar, P., Bouslah, K., 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

Alexander, G., 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., Berger, P. 1999. A simultaneous equation analysis of forecast accuracy, analyst following, and trading volume. *Journal of Accounting, Auditing & 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. https://doi.org/10.1016/S1386-4181(01)00024-6

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

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

Bayona, A., Dumitrescu, A., Manzano, C. 2023. Information and optimal trading strategies with dark pools. *Economic Modelling*, *126.* https://doi.org/10.1016/j.econmod.2023.106376

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

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

Chakrabarty, B., Pascual, R., 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., 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.

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

Chang, S. S., Chang, L. V., 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.

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

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

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

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.

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.

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

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

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

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

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

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

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

Fernandes, N., and Ferreira, M. A. 2009. 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 & Quantitative Analysis*, 35(4), 553–576.

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

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

Hsu, C.H. 2016. Strategic noise trading of later-informed traders in a multi-market framework. *Economic Modelling*, *54*, 235–243. https://doi.org/10.1016/j.econmod.2015.12.026

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

Huang, S., O'Hara, M., & Zhong, Z. 2021. Innovation and Informed Trading: Evidence from Industry ETFs. *Review of Financial Studies*, 34(3), 1280–1316.

Hutton, A. P., Marcus, A. J., Tehranian, H. 2009. Opaque financial reports, R2, 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 & Money*, 66.

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

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., 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.

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

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

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

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

Loureiro, G, R., Taboada, A, G. 2015. Do Improvements in the Information Environment Enhance Insiders' Ability to Learn from Outsiders? *Journal of Accounting Research*, 53(4), 863-905. http://dx.doi.org/10.2139/ssrn.2587836

Mirman, L. J., Santugini, M. 2019. The Informational Role of Prices. *Scandinavian Journal of Economics*, 121(2), 606–629.

Morck, R., Yeung, B., 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.

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

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

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

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

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.

Roll, R., 1988. Presidential address: R2. *Journal of Finance*, 43, 540–566.

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

Yan, Y., 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.

Yazdi, A. K., Wanke, P. F., Hanne, T., Abdi, F., Sarfaraz, A. H. 2022. Supplier selection in the oil & 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), 391-41.

Table 1: This table presents the number of stocks of our 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

Table 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 time period (2018-20), and in Panel C are the number of our 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	0.187	0.195	0.297	0.000	0.000	0.000	1.000	1.000	0.910
SDPIN	0.180	0.184	0.281	0.000	0.000	0.000	1.000	1.000	0.996
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*
SYNCH	2.032	-0.807	13.820	1.355
DPIN	0.227	0.000	1.000	0.171
SDPIN	0.216	0.000	1.000	0.183
IVOL	0.033	0.002	0.156	0.021
SIZE	1,989	2.770	35,522	4,330
SPREAD	0.737	0.000	27.270	0.970
VOL	0.914	0.000	146.500	2.000
ILLIQ	0.080	0.000	63.420	0.630
RETURN	0.000	-0.568	1.604	0.049

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

Table 3: This table presents the correlation coefficients between the different pairs of our 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 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

Table 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 Model		SDPIN Model	
Variables	*Coef.*	*SE*	*Coef.*	*SE*
Cons.	2.022***	0.013	1.987***	0.013
DPIN	0.254***	0.036	-	-
SDPIN	-	-	0.429***	0.033
VOL	-4.33E-08***	3.62E-09	-4.35E-08***	3.61E-09
ILLIQ	-0.235***	0.009	-0.228***	0.008
IVOL	-4.880***	0.293	-3.968***	0.288
SIZE	-5.00E-05***	1.40E-06	-4.00E-05***	1.39E-06
SPREAD	-0.095***	0.009	-0.097***	0.009
RETURN	-0.004	0.115	-0.014	0.114
LagDPIN	0.251***	0.036	-	-
LagSDPIN	-	-	0.4268***	0.0331
LagSPREAD	-0.096***	0.009	-0.100***	0.009
LagVOL	-4.15E-08***	3.58E-09	-4.20E-08***	3.57E-09
LagILLIQ	-0.239***	0.09	-0.231***	0.009
LagRETURN	0.253	0.109	0.260	0.109

Panel B	DPIN	SDPIN
R-square	0.139	0.144
F-test	823.96***	856.8***

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

Table 6: This table presents results for the main regression model across the years of our 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 Model			SDPIN Model		
	2018	2019	2020	2018	2019	2020
R-Square	0.446	0.348	0.206	0.446	0.356	0.209
T-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*

Table 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		SDPIN	
	Coef.	SE	Coef.	SE
DPIN	0.794*	0.305	-	-
SDPIN	-	-	0.739*	0.367
IVOL	-152.110**	94.019	-126.050**	93.3570
SIZE	-3.01E-07*	2.39E-07	-2.87E-07*	2.38E-07
VOL	-1.05E-07	2.37E-07	-1.05E-07	2.36E-07
ILLIQ	1.933	1.639	1.911	1.644
SPREAD	-0.070*	0.029	-0.069*	0.028
RETURN	0.612	4.856	0.649	4.820
LagDPIN	0.249*	0.133	-	-
LagSDPIN	-	-	0.186*	0.106
LagVOL	-9.26E-08	2.33E-07	-9.05E-08	2.32E-07
LagILLIQ	1.847	1.614	1.822	1.621
LagSPREAD	-0.112	0.286	-0.114	0.283
LagRETURN	0.733	4.726	0.776	4.695

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

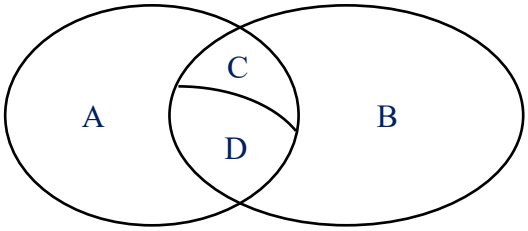
Table 8: This table presents the results of the robustness test 2. SYNCH is the proxy for stock price non-synchronicity and is a dependent variable; it is the proxy for stock price non-synchronicity, measured by Eq.(4). DPIN and SDPIN are the measures for investors' private information, calculated according 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	0.794	-	-
ΔSDPIN	-	-	-0.005	0.039
ΔIVOL	-62.080*	40.190	-62.460*	44.276
ΔVOL	1.79E-09	2.38E-09	1.72E-09	1.17E-09
ΔILLIQ	-0.008	0.039	-0.009	0.094
ΔSPREAD	2.00E-04	2.895E-04	5.00E-04	1.40E-04
RETURN	-0.012	0.056	-0.015	0.059
ΔLagDPIN	-0.002	0.094	-	-
ΔLagSDPIN	-	-	-0.006	0.039
ΔLagVOL	0.14E-09	2.33E-09	1.40E-09	1.196E-09
ΔLagILLIQ	-0.015	0.014	-0.015	0.009
ΔLagSPREAD	5.00E-04	0.286	6.00E-04	0.013
LagRETURN	-0.030	0.026	-0.037	0.063

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

Appendix A

Figure A.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 B

Table B.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 our study by the DPIN and SDPIN measures; the control variables used in our 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) Ibikunle et al. (2016) Inekwe (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) Inekwe (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.

Appendix C

Table C.1: This table presents 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)	The closing price of the stock	Daily
			S&P 500 index	Daily
2	DPIN	Investors' private information measured by DPIN	The closing price of a stock	Daily
			Executed price for each trade	Intraday
			Executed volume for each trade	Intraday
			The 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	The daily closing price of a stock	Daily
			Executed price for each trade	Intraday
			Executed volume for each trade	Intraday
			The 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 the Fama-French model	The closing price of the stock	Daily
			S&P 500 index	Daily
4	SIZE	Firm size	The 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)	The closing price of the stock	Daily
			Trading volume	Daily