A New Measure for Investors' Private Information Embedded in the Stock Price

Thu Vu[1]
Aston Business School, Aston University; Research Areas: 140, 150, 180, 350; Email: 190208860@aston.ac.uk

Alcino Azevedo
Aston Business School, Aston University; Research Areas: 130, 310, 430, 450; Email: a.azevedo@aston.ac.uk

Dimitrios Stafylas
School for Business and Society, York University; Research Areas: 310, 350, 370, 380; Email: dimitrios.stafylas@york.ac.uk

Abstract

We develop a new measure for the private information that is embedded in stock prices which considers both the trade order size and the trade frequency (volume). The currently available private information measures neglect the possibility that private information may also be embedded in the trade volume. We test the validity of our new private information measure and study how it relates to the stock price synchronicity for which we used various private information proxy measures. We show that our new private information measure is more accurate and potentially more useful for both investors and financial market regulators than those that are currently available.

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

EFM Classification: 350

JEL Classification: C49; G14; G19

[1] Corresponding author: 190208860@aston.ac.uk.

1. Introduction

There is the belief that stock prices encompass investors' private information: that is, investors often buy and sell stocks after having acknowledged information that is not yet available to the firm. Glosten and Milgrom (1985) and Kyle (1985) also believe that prices move upon new information that is generated by outside investors for their own speculative trading and capitalized into price via the trading process. Having access to private information is potentially very valuable because those investors who own that privileged information (the so-called "informed investors") can use it to buy/sell stocks and make a profit which otherwise would not be possible. On the other hand, the behavior of informed investors also affects firms' behavior. This is because firms' managers, although not aware of the exact private information the informed investors had access to, believe it may exist, and embed this belief in their decisions (see, Bakke and Whited, 2010; Zuo, 2016).

Consequently, there is an obvious interest in measuring the level of private information that is embedded in a stock price. Roll (1988) provided the first measure for investors' private information, the so-called stock price non-synchronicity model.[2] In Roll's (1988) justification for the model, he divides new information that results in the stock price's movement into three hierarchical levels: market-wide information, industry information, and firm-specific information. Accordingly, stock prices co-move with the market due to the arrival of new information related to the market and/or industry, and move in a non-synchronous way with the market due to the release of firm-specific information. We note, however, that firm-specific information does not comprise information confidentially owned by the traders only, but also information about the firm that is public. As highlighted by Roll (1988): "firm-specific return variation might be partly caused by private information". This means that, the stock price non-synchronicity measure might not fully reflect the information that is privately owned by the informed investors.

We also note that the level of confidential information may vary across stocks, for instance, as a consequence of different costs of personal information acquisition (see, Grossman and Stiglitz, 1980 among others). Such fees are difficult to measure directly. Hence, it is not easy to assess the level of private information in a stock price. More recently, two new measures for investors' private information were developed: the "probability of informed trading" (PIN) and the "dynamic measure for the probability of informed trading" (DPIN). The PIN measure was proposed by Easley et al. (2002) and estimates the probability of informed trading based on a sequential trade model and the numerical optimization for estimation. It is one of the most widely used methods to estimate the probability of informed trading and can be applied to observe the role of informed trading in many finance areas such as the informativeness of share price, corporate governance, firm decisions regarding investment, and

[2] Also known as "firm-specific return variation".

mergers and acquisitions, among others. The DPIN measure was developed by Chang et al. (2014) with the aim of providing a dynamic measure of investors' private information for high-speed trading. It is quick to compute and can be aggregated over several short intervals of time (e.g., a day, or shorter intervals of time). However, due to the complex nature of the information that is produced, transmitted, and aggregated between market participants, both PIN and DPIN also have limitations and can be improved. This paper proposes a new measure arguing that the above-described measures of Easley et al. (2002) and Chang et al. (2014) neglect the information that is potentially revealed in the stock trade volume or the size of the stock trade orders. Thus, we construct a new measure for investors' private information which also considers the aforementioned factors, a unique contribution to the literature. We follow Chang et al. (2014) model to identify informed trades in a day, and then come up with a new formula to estimate a new dynamic DPIN taking into account all large, medium, and small-sized orders. Specifically, we raise the following question: "Is there a dynamic measurement for the probability of informed stock trading that better reflects the size order effect?" We examine the relationship between stock return synchronicity and private information using high-frequency data to test the validity of the measures.

Our findings suggest that the accuracy of our new private information measure is higher than those of the aforementioned PIN and DPIN. We note that, in our new private information measure, the measurement of investors' private information in stock price is distinguished from the estimation of the probability of informed trading orders, which enables us to determine the content of the private information embedded in the daily stock price. With this new measure, the feedback effect of the stock price on decision-makers in the real side of the economy can be better gauged and estimated. As highlighted by Bond et al. (2012), "There is substantial scope for further research to advance our understanding of the implications of the feedback loop between financial markets and the real economy, whereby financial markets affect and reflect the events in the real economy".

Our findings can have relevant implications for public firms and investors in general, as well as financial market regulators. We note that, financial market regulations aim to reduce unnecessary asymmetric information between firms and thus, avoid duplicate efforts to look for information underlying the stock prices. Our new measures enable investors to rely less on private information, therefore, abnormal returns will be less likely and the market behaviour more in line with the efficient market hypothesis. In the "imperfect" world investors live, stock price movements can be a source of information not only for finding over or undervalued securities, but also to acknowledge what are the market's expectations for the firm's value and performance. Our new private information measure also helps reconcile both management values and market values. We corroborate the opinion of Rappaport (1987) who said that: "Managements who ignore the important signals from stock price, particularly in today's environment of corporate takeovers and restructurings, do so at their peril".

The remaining of the paper is as follows. Section 2 provides the related literature and the hypothesis development. Section 3 provides our empirical specifications. Section 4 provides the empirical analysis and the discussion. Section 5 concludes our study.

2. Literature Review

2.1. The concept of private information

In equilibrium, a specific stock might have different types of information in its price. The conventional wisdom says that managers own complete information about the firm and that the stock price is passive and merely reflects the investors' expectations about the present value of the firm's future cash flows. Accordingly, the secondary markets' trading either has no impact on the real economy or influences the real economy only to the extent to which "ex-post liquidity affects the firm's cost of capital in the primary markets" (Bond et al., 2012). However, it is questionable to treat secondary financial markets as a sideshow, since market prices can also be a valuable source of information and influence managers' decisions (Bond et al., 2012).

The market price is efficient and comprises information aggregated from various sources, hence, 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 mean of communication with the companies' managers apart from the trading process (Chen et al., 2007). For instance, Subrahmanyam and Titman (1999) elucidated that through day-to-day operations, traders might, incidentally, discover valuable information about the companies. This information, called private, personal, or confidential, can be reflected into prices via outside investors' trading activities. Therefore, stock prices can reveal confidential news held by speculators that are otherwise not available to the firm's managers. If so, managers can learn from this information about their own firm's prospects to make decisions. As a result, it is argued that stock markets can affect the real economy because of this information transmission (see, eg., Bond et al., 2012). The current literature describes cases in which private information is potentially a good guidance to decision-makers, such as in the evaluation of mergers and consolidation opportunities (see Luo, 2005), making decisions on firm investments (see Bakke and Whited, 2010; Chen et al., 2007), and earnings forecast disclosure (see Zuo, 2016; Loureiro and Taboada, 2015).

Figure 1[3] illustrates the separation among pieces of information owned by investors and managers. Information possessed by informed investors includes both public information and private information. We can define private information as the information that has not been published yet and is

[3] We do not assume the correlation between the size of each area with the amount of the information in this Figure.

confidentially held by informed investors and this information content is either entirely new for the manager (area A) or is the one also known by the managers (area C).

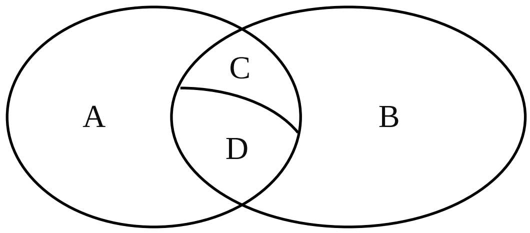

Figure 1: Relation between information owned by different participants

Source: Developed by the authors from 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

2.2 Measure for investors' private information in stock prices

2.2.1. Stock price non-synchronicity

Some studies employed stock price non-synchronicity as an alternative estimation for private information (see Chen et al., 2007; Durnev et al., 2004). This indicator was first proposed by Roll (1988) and is also called firm-specific return variation. According to Roll (1988), prices move upon new information coming from three sources. The first is a general revaluation of stock through information about the macroeconomy and market, the second is from the industry firm belongs to, and the third is from the firm's specific which includes both information publicly released and private information gathered and possessed by the trading activity of speculators. Information related to the first two sources causes the price to co-move with the market while the firm's stock price capitalized with more specific information will be asynchronous with the market. As a result, Roll (1988) also suggested that price non-synchronicity (or firm-specific return variation) is correlated with private information.

One limitation of the stock price synchronicity is that it measures the degree to which the stock price co-moves with the market, but this co-movement may be due to private and public information, so it may not accurately reflect the amount of private information in the stock price. We support the view there is a close relationship between price non-synchronicity and personal information within the stock price. However, there is a difference between them because each component represents a different amount of information embedded in stock prices. Firm-specific return variation includes all the public and private information about a firm whereas the type of new information studied in this paper can be firm-specific or market-wide (even if the amount of market-wide is small) confidentially found and owned by market participants. Nevertheless, the two concepts overlap each other instead of being the same.

2.2.2 Probability of informed trading (PIN)

The PIN measure was developed by Easley et al. (2002) and named EHO PIN. However, the idea of PIN was initially revealed in Easly et al. (1997). When trying to analyze the information content in different trade-sized orders, Easly et al. (1997) find that the trading behaviour of uninformed traders is highly history dependent. Uninformed traders are more likely to be active and copy the trades that have recently occurred. This behaviour implies that sequences of trades are less informative but that reversals in order flow are very informative. From this discovery, Easly et al. (1997) formulate a structural market microstructure model in which they demonstrated how the numbers of buys, sells, and no-trading intervals could be used to estimate the proportion of trades that are likely to be motivated by private information - which is called by Easley et al. (2002) the probabilities of information-based trading (PIN). The EHO PIN measure has also been empirically examined widely in the literature since the time it was suggested. There are several studies on the EHO's method used to estimate the PIN, pointing out its strengths and weaknesses, and suggesting improvements for it (e.g. Ellis et al, 2000); Lin and Ke, 2011; Yan and Zhang, 2012). The PIN is estimated by using some trading algorithms to classify buy orders and sell orders and by maximizing the likelihood function specified by the model. Limitations and improvements for EHO PIN are discussed below.

In the first place, some main trade classification algorithms have been extensively used in microstructure studies: (*i*) the quote rule, (*ii*) the tick rule, (*iii*) Lee and Ready's, (*iv*) Ellis et al (2000), and (*v*) the bulk volume classification. The quote rule classifies a transaction as a buy if the associated trade price is above the midpoint of the bid and the ask or as a sell if the trade price is below the midpoint quote. The tick rule classification is based on price movements relative to previous trades. The tick rule classifies each trade into an uptick or a downtick. If the price of a transaction is above (below) the previous price, then it is a buy (sell). Quote rule and tick rule each have their own pros and cons. While the quote rule cannot classify trades if they are at the midpoint or when the bid and ask prices are not available, the tick rule can do the job. On the contrary, the quote rule can work with opening trades that the tick rule can not classify. Lee and Ready's (1991) procedure is essentially a combination of these above rules: first, classify a trade according to the quote rule (above or below the midpoint), and then classify the midpoint transaction using the tick rule. Also, they suggested comparing transaction prices to quotes reported at least five seconds before the transaction was reported. The Lee and Ready (1991) algorithm performs fairly well, with a 73–91% accuracy depending on the study and the market (Finucane, 2000). Research by Ellis et al. (2000) also contended that the Lee and Ready (1991) algorithm produced high-performance accuracy in classifying trades and they provided an improvement over Lee and Ready's algorithm for the Nasdaq market and it is later called in the literature EMO algorithm.[4]

After Lee-Ready's algorithm and specific EMO algorithm for the Nasdaq exchange, an alternative algorithm to classify trades was proposed by Easley et al. (2012). This method of classification is characterized by high speeds: the bulk volume classification (BVC), which focuses on a fixed period of time, volume, or trading interval called bars. Easley et al. (2012) find that BVC leads to more useful results for the purposes of estimating flow toxicity[5] and works better than the tick rule and LR's algorithm in the Future market (equities, gold, and oil). However, Chakrabarty et al. (2015) test the BVC accuracy in the equities market and show that both tick rule and Lee and Ready's (1991) algorithm have significantly higher precision than the BVC method with regard to both volume and order imbalance. They conclude that the tick rule and Lee and Ready's (1991) algorithm surpass the BVC in identifying periods of high and persistent order flow toxicity in the equities market. Similarly, a study by Pöppe et al. (2016) argued that BVC is not robust to the choice of classification algorithm while traditional trade-by-trade classification algorithms (Lee and Ready's (1991) and EMO algorithm) have been evaluated on most financial markets of interest and literature reviews and a new evaluation of proprietary signed trading data shows they perform reasonably well, with accuracy rates of up to 90%.

The second type of improvement for PIN value was proposed by prior researchers regarding the second and third steps related to the likelihood function for a single trading day of stock. Lin and Ke (2011) argued that the likelihood function in estimating EHO PIN contains bias. They explained that "during the PIN estimation process, large buys or 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 problem is called the floating-point exception (FPE). Yan and Zhang (2012), were concerned with the boundary solutions of the probability that some traders acquire new (private) information about the firm fundamental (0 and 1), which might influence the estimate of PIN substantially. The value 0 or 1 of the parameter means that no private information event or no uninformed trade ever occurs during a period of time, which is impossible. Therefore, in reality, the value of this parameter should not be on the boundary. Yan and Zhang (2012) also acknowledged the problem indicated by Lin and Ke (2011) and modified the original likelihood function of EHO PIN to resolve all of these above issues.

2.2.3. Dynamic measurement for the probability of informed trading (DPIN)

The traditional PIN measure has its own limitations. This method is well-known and commonly used

---

[4] Ellis et al. (2000) focused more on the Nasdaq market in the US as they noticed "there has been, to date, no study of the accuracy of trade algorithms using Nasdaq data". They pointed out that outside-the-spread problem/trade appears specific to Nasdaq data and that the tick rule performs more reliably than the quote rule for trades away from the quotes.

[5] Order flow toxicity measures a trader's exposure to the risk that counterparties possess private information or other informational advantages.

to estimate the amount of private information for a relatively long period of time, usually a month, a quarter, or a year. We note that it is very challenging and time-consuming (if not impossible) to use these methods relying on intraday data. High-frequency trading and machine learning speed up the unprecedented number of trading transactions executed per second. Therefore, there is a larger amount of information including new information aggregated and incorporated in stock price within a very short time, and this type of information might be missed or uncaptured by traditional measures for private information which are generally calculated for longer periods of time. However, stock prices can have high fluctuations during a week or a month despite the fact that the closing price of that week or that month is close to the opening price. For this reason, the estimation of informed trading over a longer time horizon reduces the information content within the intraweek or intramonth trades. We advocate that it is essential to capture that content of information within higher frequencies as well. Chang et al. (2014) developed a new measure for private information called "The dynamic intraday measure of the probability of informed trading (DPIN)". They propose a new way to calculate the proportion of informed trades over a given time interval in a day based on the Avramov et al. (2006) and Schwert (1990) trading regression model.

According to Chang et al. (2014), DPIN can overcome the main disadvantage of the traditional measure PIN as it can be aggregated over a high number of short time intervals, and it is dynamic, flexible, and easy to compute. DPIN is dynamic and flexible because it is aggregated from a high number of time intervals to make a comparison with the previous macro-horizon models, which usually estimate investors' confidential information in the stock price in long timeframes such as a month or a year. DPIN is easier and simpler to compute as it requires no function of numerical optimization to estimate. It is therefore more straightforward and less time-consuming to deal with large datasets involving intraday trading. Moreover, DPIN is an effective model to estimate investors' personal information in share price for similarly long horizons as PIN, "we find numerical estimates that are generally consistent with existing measures for the probability of informed trading. DPIN measures are remarkably close in terms of location, spread, and skewness when combining across firms and years" (Chang et al., 2014).

The DPIN measure is not absent from problems particularly those related to the method of estimation. Specifically, the DPIN can be defined as the rate of informed orders over total orders, so the estimation is only based on the number of trades without considering the size of the trade orders. However, we know that trade volume varies across the orders. Therefore, to overcome this limitation, Chang et al. (2014) developed two other separate measures, named DPIN_SIZE and DPIN_SMALL, relying on two opposite views of informed trading size. One, the DPIN_SIZE measure follows the view that informed traders often place only large-sized orders, whereas the DPIN_SMALL measure follows the view that informed investors tend to break up their large orders into a series of small trades. However, with the presence of the two opposing views on the size of trading orders in literature, the complete support for any of the above is controversial. An investor can utilize large or small orders at different times in one or more days. On the other hand, at a specific time interval, both large and small-sized informed orders can be placed. Therefore, the inclusion of either large-sized trades or just small-sized trades in the calculation of DPIN is not a sufficient method to capture the relevant amount of private information in stock price. There is a need to develop a measure that can overcome this limitation; and this is something we deal with in this study.

2.3 Developing Research Hypotheses

Roll (1998) suggests that confidential information might play a crucial role in explaining stock price movement. That is, firm-specific stock returns variation might be partly caused by the existence of investors' private information. There are numerous studies, both within and across countries, showing that there is a negative (positive) relationship between the stock price synchronicity (non-synchronicity) and the investors' private information (see, Roll, 1988; Morck et al., 2000; Durnev et al., 2004; and Zuo, 2016, among others). These articles' respective analyses rest upon the validity of the idea that stock price non-synchronicity is indeed partially triggered by private information, and eventually provides indirect and circumstantial evidences for their possible relationship. In an efficient market, investors are motivated to seek for newer information to gain abnormal returns. Thus, it leads to more firm-specific variation in stock price and, therefore, higher SYNCH and a lower degree of return synchronicity (and vice versa).

Below are our research hypotheses. We provide an in-depth validity of the improved measures for investors' private information by investigating the linkage between them and firm-specific return variation or stock price synchronicity. Specifically, our research relies on two main research hypotheses, stated below, that study the relationship between stock return synchronicity, measured by the market model of Roll (1988), and private information, measured by the DPIN and SDPIN models.

H1: *Dynamic probability of informed trading (DPIN) is positively related to stock price non-synchronicity.*

H2: *Dynamic probability of informed trading with size effect (SDPIN) is positively related to stock price non-synchronicity.*

3. Data Sample and Methodology

3.1. Data Sample

All data used in this paper refers to equities from the energy sector traded in the "New York Stock Exchange" (NYSE) and the "National Association of Securities Dealers Automated Quotations" (NASDAQ) stock market. The energy sector is one of the most important industries both in the U.S. and globally. Hence, fluctuations in oil and gas prices tend to affect significantly other industries, as the ongoing energy crisis in Europe shows (see, Elder and Serletis, 2010; Yazdi et al., 2022). The energy sector is also a very specialized sector due to the complex nature of oil and gas extraction techniques, energy distribution, and associated logistics. Hence, the current stock price of energy firms can be explained in great part by the existence of new information related to the economy, the industry, and the firm itself, which might be publicly available or private – for instance, information that is known, somehow, by the investors only.

The data regarding the firms was collected from Bloomberg and "Wharton Research Data Services" (WRDS), whereas the intraday trading data was collected from the "Trade and Quote" (TAQ), used to estimate private information in stock price (PI). Our data sample comprises information on 236 U.S. energy stocks listed on the NYSE and NASDAQ covering the time period between January 2018 and December 2020 - our sample time period covers the Covid-19 pandemic.

In order to include a stock into our sample, there is the requirement that the stock is listed in the aforementioned exchanges for at least a year, which led to a drop in the number of stocks in our sample, from 261 to 236 stocks. Moreover, some stocks were also removed from our initial sample due to absence of data. We use daily data and our final sample comprises information on 236 stocks and has 154,797 observations. Table 1 describes the details of our data sample. NYSE contains the majority of energy firms (179) compared to the NASDAQ (57). The oil & gas exploration & production subcategory contains the largest number of firms for NYSE (89) and NASDAQ (28).

[Table 1 here]

3.2. Regression Models

3.2.1. Main Model

In this section, we use the following regression model (1) to estimate the relationship between the stock return synchronicity, measured by Roll (1988) model, and the investor's private information, measured by the DPIN and the SDPIN models, where *SYNCH* is the proxy for the stock price non-synchronicity using (dependent variable); so a higher *SYNCH* means that the stock price is less synchronous with the market.

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

where $i = 1, \ldots, n$ and $t = 1, \ldots, m$, where *i* denotes the stock and *t* denotes the day; *PI* is the measure for the investors' private information, measured by the DPIN and the SDPIN for k = 1 and 2, respectively; the CONTROL variable is a vector of control variables which includes the idiosyncratic volatility (IVOL), measured by the three-factor model Fama-French (1993), the firm size (SIZE) given by market capitalization divided by 10^6, the volume (VOL) accounts for the stock trade volume divided by 10^6, the bid-ask spread (SPREAD) that accounts for the difference between the highest ask price and the lowest bid price; illiquidity risk (ILLIQ) is measured by Amihud (2002), and the stock return at day *t* (RETURN) is calculated by taking the difference between the closing price at day *t* and the day before divided by the closing price at day *t-1*. To account for lag effects, the level-one lag of each control variable is also added to the vector of controls ($CONTROL_{i,t-1}$).

In Table 2, we summarise the results from previous studies.

[Table 2 here]

The main model uses three measures for stock liquidity: the idiosyncratic volatility (IVOL), the bid-ask spread (SPREAD), and the Amihud (2002) illiquidity measure (ILLIQ). The bid-ask spread is the factor measuring market liquidity. Typically, the stocks listed in the Exchanges of developed markets are more liquid than those listed in the Exchanges of developing or emerging markets and this is usually reflected in their lower bid-ask spreads. Information disclosure transparency in developed markets also helps to improve stock liquidity, thus, we conjecture that stock liquidity is positively related to SYNCH, and therefore to a lower co-move with the market. We follow Chang et al. (2014) to exploit daily volume (VOL) and daily return (RETURN) in the model. All the Lags of control variables are expected to have the same direction of effect on SYNCH with its variable.

3.2.2 Stock Price Non-Synchronicity Measure

As stated before, Roll (1998) decomposes the variation of a stock return into three different sources: market-related, industry-related, and specific to the firm; the first two sources relate to systematic variations, whereas the third is firm-specific. Like in Roll (1998), synchronicity is measured by the coefficient of determination (R^2) underlying the following regression:

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

where $IR_{i,j,t}$ is the return of stock *i* in industry *j*; $MR_{m,t}$ is the market return, and $SR_{j,t}$ is the return of the industry (sector) *j*; the stock price synchronicity is given by the R^2 being 1-R^2 the stock price non-synchronicity.

Morck et al. (2000) argue it is difficult to separate the impact on the stock price of changes in the sector and changes in the market. Furthermore, the industry returns are often derived from a few firms and so it reflects more the reality of those firms and related news than the industry. Therefore, adding the return of the sector to model (2) above can yield spurious results. Secondly, as R^2 values are bounded within the unit interval [0, 1], this might not serve as an appropriate dependent variable. Thus, instead of using the original non-synchronicity model of Roll (1988) as stated in (1), we use the adjusted measure proposed by Morck et al. (2000).

Thus, we define first the coefficient of determination (R-square) underlying 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.

Then, we use the above-defined *SYNCH variable as* the measure of 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. Dynamic Probability of Informed Trading (DPIN)

We note that the DPIN is calculated first using the Lee-Ready algorithm, to delineate informed (contrarian) and uninformed (herding) trades, after which, Chang et al. (2005) isolate the unexpected components of returns from the residual of the autoregressive model by Schwert (1990), modified by Jones et al. (1994) and Avramov et al. (2006):

$$R_{i,j} = \gamma_0 + \sum_1^4 \gamma_{1i,k} D_k^{day} + \sum_{k=1}^{26} \gamma_{2i,k} D_k^{Int} + \sum_1^{12} \gamma_{3i,k} R_{i,j-k} + \varepsilon_{i,j} \quad (5)$$

where $R_{i,j}$ is the return on stock *i* at the intraday interval j, with $j = 1, \dots ,26$; D_k^{Day} represents the day-of-week dummy variables for Tuesday through Friday, and D^{Int} represents dummy variables corresponding to each 15-minute interval over the day *t*; $\varepsilon_{i,j}$ serving as a measure for the unexpected return; Chang et al. (2014) follow Avramov et al. (2006) as the baseline of DPIN. A buy order which generates a negative (positive) unexpected return is classified as an informed (uninformed) order, whereas a sell order with a positive (negative) unexpected return is classified as an informed (uninformed) trade. Then the DPIN measure is 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 $NB_{i,j}$, $NS_{i,j}$, $NT_{i,j}$ are the number of buy, sell, and total trades, respectively, for stock *i* at day *j*.

There are two main modifications made in the model compared to the original autoregressive model by Schwert (1990). First, 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 (equivalent to 2 weeks) instead of 22 lags (equivalent to one month period) in the model by Schwert (1990). However, it seems no clear justification for this change in their methodology. Another aspect to consider is the interval used to estimate the DPIN. Notice that, the DPIN is calculated for each 15-minute interval of the day. Thus, the daily DPIN is the average value of the DPIN values of 26-time intervals. According to Chang et al. (2014), this is an advantage compared to the traditional PIN measure because over a longer time period (typically one year is used to estimate the PIN), it is possible that the actual impact of short-lived confidential information will be offset or masked by other factors. However, the use of a 15-minute time interval to estimate DPIN is also questionable. We argue that a too-long or a too-short time period used to estimate for the probability of informed trading is both ineffective.

The conventional wisdom is that informed investors define their own price range to trade in a particular period (weeks or months) based on the private information they have. When the market price falls below the target price, the investor may wish to place a buy order to take advantage of the foreseen profit and vice-versa. Furthermore, in order to make a buy or a sell decision, traders consider various factors that are previously defined in the trading system and or by the brokers, such as the trading fee, and time allowed to settle their last orders. Therefore, the frequency of 15 minutes might be too short and yield biased results for DPIN. We estimate 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, including 22 lags of daily return. The modified process is as follows: first, we use the Lee and Ready (1991) algorithm for the order classification, second we isolate the unexpected component of returns, the residuals from the following regression:

$$R_{i,j} = \gamma_0 + \sum_1^5 \gamma_{1i,k} D_k^{day} + \sum_1^{22} \gamma_{3i,k} R_{i,j-k} + \varepsilon_{i,j} \tag{7}$$

where $R_{i,j}$ is the return on stock *i* at the day *k* (*j*=1,…,4), D_k^{Day} represents day-of-week dummy variables for Monday through Friday. In order to identify the informed trades, $\gamma_{1i,k}$ it is used the $\varepsilon_{i,j}$ as a proxy for unexpected returns.

Buying trades in the presence of negative (positive) unexpected returns are classified as informed (uninformed) trades. Sell trades in the presence of positive (negative) unexpected returns are classified as informed (uninformed) trades. The DPIN measure is constructed according to the Eq. (6).

3.2.4. The DPIN with the Size Order Effects

Although the PIN and DPIN measures are estimated in different ways, both represent the rate of orders proposed by informed traders over the total trades. However, the trade volume can vary across trade orders. If the monopolists or informed traders want to camouflage their trading activity by making several small-sized orders rather than one large order, the rate of orders proposed by informed traders over total trades may be close to the rate of given by the number of informed orders over the total number of trading. However, if only one or a few large-sized tradings are made by an informed monopolist trader and the other small-sized tradings are uninformed or noise, there is a more peculiar case. Hence, using the PIN and DPIN measures might not be enough to accurately reflect the effect of investors' private information in stock price. We argue therefore that a new measure should be developed to capture the above-described possible characteristics of informed trading. Thus the new measure should include both the number and the volume of the informed trade.

We contend that, if all the informed traders wish to hide their tradings under small or medium-sized trades, then the DPIN, which represents the proportion of informed trades in the total number of trades executed in a day will reflect exactly what it is meant to be, but if the participants in the market behave differently and large informed orders can exist, our new measure is more effective. Our new measure is an extension of that of Chang et al. (2014). We propose a new way of computing the DPIN which also considers the order size, so it is more precise in measuring the amount of private information embedded in the stock price.

We use Lee and Ready (1991) to classify the number of buy orders (B) and the number of sell orders (S) in a single trading day. Then, we isolate the unexpected component of returns as the residuals from the regression, after which we use the residuals $\varepsilon_{i,j}$ as a proxy for unexpected returns, and identify informed trades. The amount of investors' private information is given by the ratio total volume of all informed transactions (including both buy and sell-initiated trades) over the total trading volume, as follows:

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

where $V_{i,J}^{B}$ and $V_{i,j}^{S}$ are the total volume of all buy-informed and sell-informed orders, respectively. $TV_{i,j}$ is the total trading volume (of both informed and uninformed orders).

4. Empirical Results

4.1. Descriptive Analysis

In table 3, we present the descriptive statistics on the variables used in our main regression model; in panel A, the descriptive statistics are provided per year, whereas in panel B are provided considering the whole sample time period. In panel C, we provide the number of sample observations per year.

Specifically, panel A shows the mean, minimum, and maximum of the variables used in our main regression model per year (from 2018 to 2020). It reveals that the mean value of the SDPIN is slightly lower than that of the DPIN, in all years. Interestingly, it appears to be an opposite moving direction between the mean of SYNCH and PI (DPIN and SDPIN) over three years. While the mean value of the

SYNCH consistently decreased from 2018 to 2020, the mean values of both the DPIN and the SDPIN increased. Noticeably, in 2020 the DPIN and the SDPIN increased, by 51.81% and 52.19% respectively, when compared to 2019, and the SYNCH decreased by nearly by 13%. The oil and gas industry was among the most severely affected by the Covid-19 crisis.[6] The lower value of the SYNCH in 2020 compared to 2019 (dropped by 13%) means that the energy sector prices co-move more with the market than in 2019. However, the increase in the degree of this co-movement is much lower than the increase of DPIN and SDPIN (by 51.81% for DPIN and 52.19% for SDPIN) and can be explained as it is offset by the bigger drop in energy price (especially oil and gas sector) compared to the whole market. Panel B provides more empirical evidence of the effects on the U.S. market of the Covid-19 pandemic. There was a significant decrease in market liquidity in the year 2020 when compared to 2019 (presented by a significant increase in the mean value of the liquidity variables). For instance, IVOL and ILLIQ increased by 85.49 % and 75.44% respectively, the trade volume (VOL) increased by 56.48%, and the market capitalization (SIZE) decreased by 31.14%.

[Table 3 here]

Table 4 presents the correlation matrix which shows that the correlation coefficients are in general below 0.5. The exceptions to this rule relate to 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. Thus, we use the “variance inflation factors” (VIF) measure to determine the level of collinearity between the regressors of the model(s).[7] Within the correlation matrix, the two measures for private information DPIN and SDPIN are strongly correlated at the level of 0.896. SDPIN shows a stronger correlation with SYNCH than DPIN. This appears to be the first sign of a stronger association between SDPIN and SYNCH than DPIN and SYNCH.

[Table 4 here]

[Table 5 here]

4.2. Main Findings

Tables 6 and 7 present our results of the main model. They show the relation between the stock return variations (measured by SYNCH) and investors’ private information (measured by DPIN and SDPIN).

[6] We note that the price of oil contracts delivered in May 2020 dropped to negative values -$37.63/barrel on the April 20, 2020 expiration date (Corbet et al., 2021).

[7] 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 5, we provide our 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, we removed the LagSIZE and LagIVOL independent variables from our regression model.

Table 6 shows our results for the entire sample and Table 7 shows our results per year (2028-20). From Table 7, we conclude that in all the years, the SDPIN measure fits better with the market than the DPIN measure, since its R-squared (R^2) is 0.144, whereas the R^2 for the DPIN measure is 0.139. For both the DPIN and the SDPIN measures, the coefficients of all the independent variables are statistically significant at the 1% level, apart from RETURN and LagRETURN, showing that they affect SYNCH and return synchronicity. This means that more private information embedded in the stock price reduces its level of co-movement with the market. This finding corroborates our research hypotheses and is in line with prior studies such as those of Roll (1988) and Durnev et al. (2004), who show that a higher level of investors' private information leads to a higher proportion of the stock price fluctuations being caused by firm-specific (idiosyncratic) information and, therefore, a decrease in the stock price synchronicity. It also shows that both DPIN and SDPIN are valid measures for the private information of investors in share prices. These pieces of private information are confidentially discovered and integrated into prices by investors in order to acquire abnormal returns.

Countries with well-developed financial markets, like the U.S, often exhibit a low R^2. This means that the movement of the stock prices is due to firm-specific information, including private information, and not because of influences from the market. Although we argue that stock price non-synchronicity and investors' private information are two distinct measures, our findings show that they are very closely linked. Using the same control variables in the regression models for DPIN and SDPIN, we find that the coefficient for the SDPIN is 0.429, with a standard error of 0.033, and the coefficient for the SYNCH is 0.254, with a standard error of 0.036. This suggests that SDPIN shows a stronger (positive) relationship to SYNCH than DPIN.

Table 7, presents our results for the DPIN and the SDPIN regression models per year. The value of R^2 and the coefficient of DPIN and SDPIN is lower in 2020 than in 2018 and 2019. This might be because of the negative impact on the U.S. of the Covid-19 pandemic. The SDPIN coefficient is statistically significant and positively associated with SYNCH in all the years. Conversely, the effect of the DPIN on SYNCH in 2018 is negative and insignificant. As expected, the three measures for liquidity (ILLIQ, SPREAD, and IVOL) along with their Lags (LagILLIQ, LagSPREAD, and Lag IVOL) present a statistically significant negative impact on SYNCH, at the 1% level. These results are in line with previous findings of Morck et al. (2000) and Zuo (2016) who show that low R^2 stocks are those with the least trade and with the greatest impediments to informed trades. A stock market with low R^2 stocks can be interpreted to have a good or efficient information environment (such as the US market). The information environment comprises of two levels: the market level (e.g., government indexes, corporate law) and the firm level (e.g., corporate governance mechanism, reporting, and disclosure). Strong investor protection in an efficient market can help to improve market liquidity as it encourages investors to make informed trading, leading to a higher amount of firm-specific information reflected in stock price and therefore, lower synchronicity (see, e.g., Morck et al., 2000).

The coefficient of SIZE is statistically significant at the 1% level and negatively associated with SYNCH. According to Chan and Hameed (2006), when the number of stocks within an index is small, a few large companies dominate market movements. Consequently, when 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 negative relationship with SYNCH); and they also confirm this result in the tests. Previous literature shows that larger companies with high trading volumes of shares tend to attract a larger number of analysts as they have an incentive of more brokerage commissions (Alford and Berger, 1999). We also note that Bhushan (1989) claims that the supply of analyst services is also affected by the correlation between return and market return. That is, for a given level of information acquisition cost relating to macro variables, a higher correlation between firm return and the market return (also means higher stock return synchronicity) facilitates a lower marginal cost to acquire information, leading to an increase in the supply of analyst services.

The level of a stock's trading volume affects stock return synchronicity because it influences the speed of price adjustments. We include VOL and LagVOL in the regression model and get similar results. Specifically, the VOL and the LagVOL are statistically significant at a 1% level and negatively associated with SYNCH. 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, resulting in lower stock return synchronicity.

Although the signs of the coefficients of RETURN and LagRETURN are different from those found for SYNCH, they are statistically insignificant. Only the coefficients of RETURN and LagRETURN are not statistically significant – therefore, we did not find evidence of their effect on return synchronicity. In Table 7 we can also see that in 2019 both RETURN and its Lag (Lag RETURN) have positive and statistically significant impacts on SYNCH. Our findings are in line with those of Chan and Chan (2014). Chang et al. (2014) found mixed results on the relationship between SYNCH and the two measures of stock return.

[Table 6 here]

[Table 7 here]

4.3. Robustness Tests

For the robustness tests, we use the framework of Fama and MacBeth (1973), following Chang et al. (2014). We run cross-sectional regressions to obtain estimates for regression parameters after which we use the time-series average across all days to arrive at parameter estimates. Specifically, we perform two robustness tests based on: (*i*) the same dataset with the initial model, and (*ii*) the use of first-order differentiated data.

The two Fama–MacBeth models are specified as follows:

Fama–MacBeth Robust Test 1:

$$SYNCH_{i,t} = \beta_0 + \beta_{1,t}PI_{i,t} + \beta_{2,t}PI_{i,t-1} + \beta_3CONTROL_{i,t} + \beta_4CONTROL_{i,t-1} + \varepsilon_{i,t} \quad (10)$$

Fama–MacBeth Robust Test 2:

$$\Delta SYNCH_{i,t} = \beta_0 + \beta_{1,t}\Delta PI_{i,t} + \beta_{2,t}\Delta PI_{i,t-1} + \beta_3\Delta CONTROL_{i,t} + \beta_4\Delta CONTROL_{i,t-1} + \varepsilon_{i,t} \quad (11)$$

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

Variables in the first robust test are the same as the original model (Equation 1) regarding both independent and control variables. Similar to Chang et al. (2014), we include contemporaneous and lagged returns (RET) in the model, but not its daily difference. The difference in firm size (SIZE) and change for its lag is not included in Test 2 as these are will be highly correlated with stock return.

In Tables 8 and 9, we present the results for the two Fama–MacBeth robust tests. Within the first robustness test, the parameters are estimated from the time-series average of cross-sectional regressions. The DPIN and SDPIN and the LagDPIN and LagSDPIN are statistically significant at the 10% level and positively associated with SYNCH, which supports the idea that DPIN and SPDIN are valid measures for private information. Of all the control variables, IVOL, SIZE, and SPREAD are statistically significant at 5%, 10%, and 10% levels respectively, in both DPIN and SDPIN robust models.

For the second robustness check, the daily change of DPIN (ΔDPIN) and SDPIN (ΔSDPIN) shows the negative but insignificant relationship to the daily change of SYNCH (ΔSYNCH). The changes in DPIN and SDPIN insignificantly explain the changes in SYNCH or price co-movement. ΔIVOL is the only variable that shows a significant and negative impact on SYNCH at the level of 10%.

[Table 8 here]

[Table 9 here]

5. Conclusion

This study provides an alternative way of estimating investors' private information in stock prices. We extend the high-frequency measure for the probability of informed trading (DPIN) by Chang et al. (2014). Particularly, we incorporate the effect of the volume of all different-sized orders to get the newly developed dynamic measure for private information (SDPIN). We then test the econometric models against the market data to check the validity and reliability of SPDIN. Our findings show a positive and significant correlation between SDPIN and stock price non-synchronicity, signaling that SDPIN can be a valid method to determine the amount of confidential news that spectaculars transmit into the price through their trading activities. In some cases, SDPIN exhibits an even better fit with the models than DPIN.

SDPIN inherits various advantages from its original version DPIN. This dynamic and flexible measure allows calculating the content of private information for one day or even for a shorter interval by using intraday data of trading. Moreover, when it is used to estimate for similarly long horizons, there are consistent numerical estimates like PIN. Our methodology makes SDPIN the first private information measurement which can capture quite accurately the effect of the trading volume. All different types of order sizes, are taken into account. Our results assist finance managers in incorporating more accurate private information in their decisions, and challenges regulators to ensure information related to trading orders is disclosed which might have an impact on the market efficiency, overall. The application of SPDIN in other sectors and the whole market could be a topic for future research.

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.

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

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.

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.

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.

Chan, K., and 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., and 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., 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.

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

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

Dasgupta, S., Gan, J., and 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., and Yeung, B. (2004). Value-enhancing capital budgeting and firm-specific return variation. *Journal of Finance*, 59(1), 65-105.

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., Kiefer, N. H., and O'Hara M. 1997. 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.

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.

Fama, E., and 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. *Review of Financial Studies*, 22(5), 1845–1887.

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.

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

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

Hutton, A. P., Marcus, A. J., and 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., and 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., and Myers, S. C. (2006). R2 Around the World: New Theory and New Tests. *Journal of Financial Economics,* 79: 257–292.

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

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.

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.

Patton, A. J., and 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., and 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.

Rao, L., and 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., 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.

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

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 shows 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 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) Ibikunlea 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.

Table 3: 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 4: This table shows 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 5: This table shows the variance inflation factors (VIF) of all the regression variables. Panel A shows the VIF and the 1/VIF for before the removal of the LagSIZE and LagIVOL variables (in bolt below) and Panel B shows the VIF and the 1/VIF for 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 6: 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*
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***

Table 7: This table presents results for the main regression model regarding 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			SDPIN		
	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)

Table 8: This table shows our robust test 1. 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.

Variables	DPIN		SDPIN	
	Coef.	SE	Coef.	SE
DPIN	0.305*	0.794	-	-
SDPIN	-	-	0.367*	0.739
IVOL	-152.110**	940.190	-126.050**	933.570
SIZE	-3.01E-07*	2.39E-06	-2.87E-07*	2.38E-06
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*	0.290	-0.069*	0.287
RETURN	0.612	4.856	0.649	4.820
LagDPIN	0.305*	0.794	-	-
LagSDPIN	-	-	0.3674*	0.7388
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

Table 9: This table shows results of Robust test 2. 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 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	90.039
ΔIVOL	-62.080*	940.190	-62.460*	44.276
ΔVOL	1.79E-09	2.38E-07	1.72E-09	1.17E-08
ΔILLIQ	-0.008	1.639	-0.009	0.094
ΔSPREAD	2.000E-04	0.2895	5.00E-04	0.0140
RETURN	-0.012	4.856	-0.015	0.259
ΔLagDPIN	-0.002	0.794	-	-
ΔLagSDPIN	-	-	-0.006	0.039
ΔLagVOL	0.14E-09	2.33E-07	1.40E-09	1.196E-08
ΔLagILLIQ	-0.015	1.614	-0.015	0.092
ΔLagSPREAD	5.00E-04	0.286	6.00E-04	0.013
LagRETURN	-0.030	4.726	-0.037	0.263