Research on Information-Driven Trades in China

Juan Tao, Xi’an Jiaotong-Liverpool University, China*
Dongqi Sun, Xi’an Jiaotong-Liverpool University, China
Yingying Wu, Xi’an Jiaotong-Liverpool University, China

ABSTRACT

The authors examine the information-driven trades and informed traders’ order size strategies in China’s stock market. They find the aggregate U-shaped informed trading is not only explained by the time-of-day effect but is also related to the order size strategy, which is shown by intraday variations in the composition of small, medium, and large trades. The evidence of information predictability from early morning to market close and from late afternoon to the next day provides additional insights into the intraday informed trading pattern. They identify the non-negligible price impact (PI) of large trades and propose a modified model, VDPIN-PI, which better captures the trades with information advantage compared to the baseline model.

KEYWORDS

Camouflage, Contrarian Trades, Information Advantage, Information-Driven Trades, Intraday Pattern, Order Size, Predictability, Price Impact

1. INTRODUCTION

By examining information driven trades and trading strategies, this study unveils the information transmission and dissemination in China’s stock market. At the time when China gradually opens up the capital market (Ge et al., 2020) by introducing the Qualified Foreign Institutional Investor (QFII), RMB Qualified Foreign Institutional Investor (RQFII) and Stock Connect Programs, this research helps prospective international investors to better understand the China’s market, where the issue of information asymmetry is perceived to be more severe compared to more developed markets. The main cause of asymmetric information may include information leakage, superior forecasting of public information of some traders or their proprietary information collection.

One well-documented phenomenon about informed trades is that they concentrate at the opening, and sometimes the closing, period of the exchange (Madhavan et al., 1997; Barclay and Hendershott, 2003). Admati and Pfleiderer (1988) argue that, since the market opening and close are distinguished by the fact that they fall just after and before the exchange is closed, respectively, they may cause increased trading volume because of a rush to trade by informed and liquidity traders. Further, Gao et al. (2018) suggest that the intraday return predictability can be partially caused by late-informed...
trading near the market close to avoid overnight risk. Another strand of research focuses on the order submission strategies of informed traders. Kyle (1985) suggests that profit-maximising informed investors may attempt to camouflage their information and reduce the price impact by spreading trades over time. Barclay and Warner (1993) find that informed traders will concentrate on medium-sized trades, given their concerns such as costs related to regulatory requirements, delays and brokerage commission. In contrast, Barardehi et al. (2018) argue that aggressive market orders in large sizes may be used during high-liquidity times when the price impacts are small. The present study attempts to examine the trading strategies of informed traders in China’s stock market, which is dominated by retail investors who are usually perceived to be irrational and impatient. To maximise their profits, do informed traders in China tend to behave cautiously and camouflage their trades or to trade aggressively, given the atmosphere of urgency and the likely consequence of high delay costs?

In this study, we examine the presence of informed trading and informed traders’ strategies concerning price impact and order size and offer three contributions to the related literature. First, we propose an intuitive method for examining the price impact of influential trades. Then we develop a modified informed trading measure by considering the price impact. Second, we investigate the intraday informed trades in China’s stock market and detect that these have an aggregated U-shape. We discover that in addition to the time-of-day effect, the overall U-shape stems mainly from variations in the trade size composition. Third, to gain more insights into the right-side peak of the U-shape, we analyse the unexpected return predictability. Thus we discover that the late afternoon informed trading is motivated not only by information retained from the early morning but also by privileged access to private information supposed to arrive the next day, which provides additional explanation for the high informed trading at the opening and near the market close. The information advantage of informed traders suggests their power to forecast future stock returns (Agrawal and Mittal, 2019), which explains the intraday return predictability documented in different markets and assets classes (Gao et al. (2018), Gao et al. (2019) and Zhang et al. (2020)). Our findings of intraday pattern of information driven trades also shed lights on information dissemination and price discovery in stock markets of developing countries.

A large body of literature is available on informed trading measures and the relevant trading strategies of informed traders. One such measure that Easley et al. (1996) proposed relies on trade imbalance, driven by information shock, to infer the probability of informed trading (PIN). Another measure, first proposed by Avramov et al. (2006) and then developed by Chang et al. (2014) and Chang and Wang (2019), is proxied by contrarian trades. The rationale for this proxy measure is that those who trade against the market, rather than engaging in herding, are akin to informed traders. It defines buy trades in the presence of negative unexpected returns and sell trades in the presence of positive unexpected returns as contrarian trades. Piotroski and Roulstone (2005) also establish insiders possessing superior information as contrarian traders who reverse the trajectory of past returns, sending signals that prices are under or overvalued. We conjecture that in China’s market, where retail investors dominate, contrarian trades are likely to be executed by informed traders.

In this study, we first use contrarian trades to represent informed trades, noted as the baseline volume based dynamic probability of informed trading (baseline VDPIN). The overall probability of informed trading at each interval reflects the time-of-day effect of such trading. However, the estimated results cannot illustrate the strategies of informed traders on order size. This concern stems from their conflicting interests between the cost of large trade price impact and of information delay. The analysis of the trade size composition offers alternative explanations and implies that the time-of-day effect alone cannot explain the U-shape. We are motivated to further explore the causes of the overall higher percentage of informed trades at the two special periods. Thus, we compare the composition of the three trade size categories throughout the day and establish that the aggregate U-shape of informed trades is also explained by changes in the composition of small, medium, and large trades within the day. In addition, at daily opening and close, when informed trading is intensive, large trades are observed with a decrease in the average trade size but an increase in the
proportion. Meanwhile, midsized trades show an increase in the trade size and the proportion. These results imply that large orders are split into medium-sized ones, perhaps because of informed traders’ camouflage motive. We conjecture that, without considering the substantial price impact of large trades, the baseline VDPIN may fail to capture some large informed trades and mistakenly include some uninformed contrarian trades instead. To examine this argument, for each trade size category we assess the trade price impact in two ways. Thus we find that large trades have a significantly higher price impact compared with small and medium trades. Hasbrouck (1991) measures the information content of stock trades as the ultimate price impact of the trade innovation and also finds that large trades have a larger price impact.

These findings lead to another aim of our research - to develop an informed trading measure that considers the price impact (PI). Acknowledging the non-negligible price impact of large trades and its influence on unexpected returns, we first divide the time intervals with large trades into impact intervals and non-impact intervals. At impact intervals, we argue that contrarian trades are unable to capture the large informed trades because high information content is embodied in the unexpected returns. Specifically, a positive (negative) information shock manifests itself as positive (negative) unexpected returns. Therefore, non-contrarian trades actually represent the informed trades in impact intervals. We propose a modified model in which we adjust the baseline VDPIN for large trade price impact within impact intervals and make no adjustment within non-impact intervals. The overall results of the modified model VDPIN-PI show a more pronounced U-shape of informed trading than that from the baseline VDPIN. Information advantage test suggests that a higher proportion of informed trades is captured by VDPIN-PI than the baseline model. Also, autocorrelation test shows a clear return reversal for those trades executed by uninformed traders measured by VDPIN-PI. This result is supported by the argument that price changes caused by (uninformed) liquidity trades should be reversed later (Campbell et al., 1993). These results support the effectiveness of the modification. To shed lights on the information flow between the market open and close, we further conduct the predictability test and find that the late afternoon information (represented by unexpected returns) predicts that of the next day. Intraday predictability from morning to market close is also present but only significant in intervals with large trades.

The remainder of this article proceeds as follows. Section 2 reviews the literature and Section 3 describes the data. Section 4 introduces the baseline model to capture informed trades. Section 5 discusses trade size compositions, examines the trade price impact, and develops the informed trading measure (VDPIN-PI) modified for price impact. Section 6 extends the analysis to the autocorrelation and information predictability tests, and Section 7 concludes.

2. LITERATURE REVIEW

Informed traders play an important role in stock price discovery by incorporating information into stock prices and subsequently improving the market efficiency (Piotroski and Roulstone, 2004). One strand of research aims to discern the disparate characters and influences of informed trades from liquidity trades. For example, Campbell et al. (1993) argue that informed trades and liquidity trades should differ in that liquidity trades will cause a temporary price change that will subsequently be reversed. In contrast, price reversals are not expected to occur following the price changes generated by informed trades. Conversely, Admati and Pfleiderer (1988) examine the interacting strategic decisions of informed traders and discretionary liquidity traders (who can time their trading) and develop a theory that states discretionary liquidity trading is typically concentrated and informed traders trade more actively in periods when liquidity trading is concentrated. Meanwhile, liquidity traders also benefit from the increased entry of informed traders who compete with each other, which typically improves the welfare of liquidity traders. Admati and Pfleiderer (1988) argue that their theory of the strategic behaviour of liquidity traders and informed traders provides a partial explanation for the intraday U-shaped patterns of volume and price variability.
Other scholars are interested in informed investors’ strategic choices of order size. Some models indicate that informed traders prefer to trade large amounts at any given price because they face competition from other informed traders and the privacy of their information could be short-lived (Easley and O’Hara, 1987; Grundy and McNichols, 1989; Holthausen and Verrecchia, 1990; Kim and Verrecchia, 1991). Alternative theories suggest that the expected price impact of large trades (i.e., price concessions) increases with trade size, and profit-maximising informed investors may attempt to camouflage their information by spreading trades over time (Kyle, 1985). To address more specifically informed investors’ trade-size choices, Barclay and Warner (1993) analyse these informed investors’ trade-off between the cost of price impact and of information delay, the possibility of detection and prosecution of illegal trading on private information and their concern about the brokerage commission cost. They hypothesise that informed traders will concentrate their trades in medium sizes (i.e., stealth trading hypothesis). The price impact of large trades is also documented in other studies. For example, Chan and Fong (2000) find a significant return impact of order imbalance, and that the order imbalance in the large trade size category affects the return more than that in smaller trade size categories for stocks traded on the New York Stock Exchange (NYSE). As regards the London Stock Exchange (LSE), Sun and Ibikunle (2017) also find a positive (negative) relationship between informed trading and the permanent price impact of block purchases (sales), thereby suggesting that private information is impounded via block trading on this exchange.

Several models have been proposed in the literature to measure informed trades in the securities market. For example, Roll (1988) is the first to suggest that price non-synchronicity (or firm-specific return variation) can be a suitable proxy for private information. Easley et al. (1996, 1997) develop a structural market microstructure model to measure the probability of informed trading (PIN), based on observable data on the number of buys and sells from the trading process. Chen et al. (2007) apply both approaches in examining price informativeness and investment sensitivity to stock price and suggest that the PIN measure of Easley et al. (1996, 1997) may capture the source of information (trading activities of informed traders) reflected in price, whereas price non-synchronicity may capture the result of this information on the price. Duarte and Young (2009) indicate that the PIN of Easley et al. (1996, 1997) can actually be decomposed into two components and that only one component (adjusted PIN) measures asymmetric information whereas the other (probability of symmetric order flow shock) measures illiquidity effects unrelated to information asymmetry. However, a problem associated with the PIN of Easley et al. (1996, 1997) is that it may be sensible to estimate certain parameters in the PIN model over a long macro horizon. Over long horizons, it is likely that the actual effects of short-lived information may be diluted or masked by other factors (Chang et al., 2014; Yaokumah et al., 2019). Other models that do not require intermediate numerical estimation of non-observable parameters include those by Easley et al. (2012) and Chang and Wang (2019).

Some studies have examined the rationality of investors by studying herding and contrarian behaviour in financial markets (Omigie et al., 2020; Peng et al., 2020). For example, Avery and Zemsky (1998) indicate that information cascades - when it becomes rational to ignore one’s own private information and instead follow one’s predecessors’ decisions (herding) - cannot occur in a simple sequential asset market because a flexible market price incorporates all publicly available information. To test this theory, Drehmann et al. (2005) designed an internet experiment based on a sequential asset market with privately informed traders. They do not find evidence of herding, which supports Avery and Zemsky’s (1998) prediction. In contrast, they find that informed subjects frequently act as contrarians. Following previous research on rational and irrational investors (Friedman, 1953), positive feedback investment strategies (e.g. Cutler et al., 1990; DeLong et al., 1990) and herding investors (Froot et al., 1992), Avramov et al. (2006) conjecture that herding or positive feedback behaviour represents uninformed trades, whereas contrarian trades are akin to informed trades.
3. DATA AND DESCRIPTIVE STATISTICS

This study is based on quote and trade data from China’s stock market. Compared with developed markets, this market is believed to have less effective implementation of regulations on corporate disclosure and insider trading. Further, its market participant composition differs in that it has more retail investors and fewer institutional investors. Given the higher expected level of information asymmetry, it is worth examining China’s market to study informed trading behaviour, which may differ from that of other developed markets. The sample data, including tick-by-tick quote, trade price and transaction volume, is obtained from the Wind database. The dataset consists of 300 constituent stocks of the CSI300 index. The CSI300 stocks are from the Shanghai and Shenzhen Stock Exchanges and account for about 60% of China’s overall stock market capitalisation. Our sample includes data from January 2012 to December 2014. The Shanghai and Shenzhen Stock Exchanges both employ the call auction trading mechanism from 9:15 to 9:30 am to open the market and Shenzhen Stock Exchange also employs this mechanism from 14:57 to 15:00 pm to close the market. We exclude all call auction periods of the two exchanges from our sample because quote and trade data are not available. Finally, our dataset includes all transactions of the 300 stocks during the continuous auction periods of 9:30 to 11:30 in the morning and 13:00 to 15:00 (13:00 to 14:57 for Shenzhen Stock Exchange) in the afternoon during the three years.

The sample period of 729 business days has 1,222.19 million trades in total, including 643.48 million (52.65%) buyer-initiated (buy) trades and 578.71 million (47.35%) seller-initiated (sell) trades. We stratify all trades into three trade size groups according to the classification by:

**Straight Flush:** A local trading platform most widely used by market participants in China.

All trades are labelled as small, medium or large trades, depending on whether the trading volume is below 20,000 shares, between 20,000 and 100,000 shares, or above (including) 100,000 shares, respectively. The individual stock returns are calculated over five-minute intervals as changes in the midpoints of the bid-ask spread to avoid any bid-ask bounce. Each trading day has 48 five-minute intervals throughout the continuous auction periods.

Table 1. Summary statistics on trades and interval returns

<table>
<thead>
<tr>
<th>Trades of small, medium, and large sizes</th>
<th>No. of trades (million)</th>
<th>No. of trades (%)</th>
<th>Vol. of trades (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>926.84</td>
<td>75.88</td>
<td>52.85</td>
</tr>
<tr>
<td>Medium</td>
<td>279.33</td>
<td>22.87</td>
<td>35.96</td>
</tr>
<tr>
<td>Large</td>
<td>15.34</td>
<td>1.25</td>
<td>11.19</td>
</tr>
<tr>
<td>Total</td>
<td>1,221.51</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5-minute interval returns</th>
<th>Mean</th>
<th>S.D.</th>
<th>Median</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2.41*10^{-6}</td>
<td>0.0031</td>
<td>8.31*10^{-7}</td>
<td>−0.28</td>
<td>5.10</td>
</tr>
<tr>
<td>Positive</td>
<td>4.29*10^{-5}</td>
<td>0.0029</td>
<td>4.61*10^{-5}</td>
<td>0.37</td>
<td>4.97</td>
</tr>
<tr>
<td>Negative</td>
<td>−4.27*10^{-5}</td>
<td>0.0028</td>
<td>−4.48*10^{-5}</td>
<td>−0.41</td>
<td>5.21</td>
</tr>
</tbody>
</table>

Note: The upper panel presents the descriptive statistics of small, medium and large trades from January 2012 to December 2014. The ‘No. of trades (million)’ column reports the total number of trades in each trade size category. The ‘No. of trades (%)’ and ‘Vol. of trades (%)’ columns report the percentage of the number of trades and trading volume relative to the total trades for each category. The lower panel presents the summary statistics on the five-minute interval returns for all 300 stocks over the sample period. The statistics were first computed for individual stocks over time and then averaged for each return group: all returns, positive returns, and negative returns.
The upper panel of Table 1 summarises the statistics of each trade classification. The trades are dominated by small trades. The lower panel of Table 1 provides the summary statistics of the five-minute returns. The maximum and minimum returns are not reported because they are capped by $\pm 10\%$ of the daily price limits applied in China’s stock market. The average interval return is $2.41 \times 10^{-6}$, which is positive over the whole sample. The positive and negative returns have similar standard deviations and kurtosis, but skewness of opposite signs.$^3$

4. BASELINE VDPIN MODEL

In this section, we construct a baseline VDPIN model and employ it to estimate the probability of informed trading in the whole sample. Avramov et al. (2006) suggest that contrarian trades (according to the direction of unexpected returns) are akin to informed trades, which is supported by the empirical findings of their own research as well as that of Chang et al. (2014) and Chang and Wang (2019). The intuition and explanations are that the unexpected return over an interval (after excluding the influence of past information and some routine day and interval effects) reflects the price movement direction within the current interval. For example, a positive unexpected return over an interval is a result of upward price movement. Then sells in this interval are defined as contrarian informed trades because they do not follow the broad market and are more likely driven by private information not known to the market yet. Following Avramov et al. (2006), we first isolate the unexpected component of returns as the residuals from the following regression in Equation (1), and then designate buy (sell) trades in the presence of negative (positive) unexpected returns as informed (contrarian) buys (sells):

$$R_{i,j} = \sum_{k=1}^{5} \gamma_{i,k} D_{k}^{\text{Day}} + \sum_{k=1}^{48} \gamma_{2i,k} D_{k}^{\text{Int}} + \sum_{k=1}^{12} \gamma_{3i,k} R_{i,j-k} + \varepsilon_{i,j}$$

where $R_{i,j}$ is the five-minute interval return of stock $i$ at intraday interval $j$ ($j = 1, 2, ..., 48$), $D_{k}^{\text{Day}}$ is a dummy variable measuring the day-of-week effect, and $D_{k}^{\text{Int}}$ is a dummy variable measuring the effect of a particular five-minute interval. Thus, the residual $\varepsilon_{i,j}$ captures the unexpected return of stock $i$ at interval $j$, after accounting for average day-of-week effects, average interval-of-day effects and return autocorrelation effects.$^{12}$

In a given interval, those trades that buy/sell in the same direction as the sign of the residual demonstrate herding behaviour and are classified as uninformed trades. By contrast, those trades against the sign of the residual are termed contrarian trades and are classified as informed trades.

Our baseline measure $VDPIN_{\text{Base}_{i,j}}$ for stock $i$ in each five-minute interval $j$ is specified in Equation (2). Chang et al. (2014) note that their dynamic probability of informed trading ($DPIN$) measure fails to capture the widely known U-shaped intraday pattern of informed trading, which is successfully documented by their $DPIN_{\text{SIZE}}$ measure with size effects incorporated. This finding suggests the importance of considering the size effect in the informed trading measure; thus, we use trading volumes (or shares traded), rather than the number of trades, to develop our baseline $VDPIN$ measure:

$$VDPIN_{\text{Base}_{i,j}} = \frac{VB_{i,j}}{VT_{i,j}} (\varepsilon_{i,j} < 0) + \frac{VS_{i,j}}{VT_{i,j}} (\varepsilon_{i,j} \geq 0)$$

where $VB_{i,j}$ is the total volume of buys for stock $i$ at interval $j$, $VS_{i,j}$ is the total volume of sells for stock $i$ at interval $j$ and $VT_{i,j}$ is the total volume of trades (buys as well as sells) for stock $i$ at
interval \( j \). The percentage of informed (contrarian) buys (or sells) is denoted as \( VB_{i,j} / VT_{i,j} \), \( \epsilon_{i,j} < 0 \) (or \( VS_{i,j} / VT_{i,j} \), \( \epsilon_{i,j} \geq 0 \)). \( \epsilon_{i,j} < 0 \) is an indicator variable that equals 1 when the unexpected return is negative, and 0 otherwise (defining contrarian buys), and \( \epsilon_{i,j} \geq 0 \) is an indicator variable that equals 1 when the unexpected return is positive, and 0 otherwise (defining contrarian sells).

We aggregate \( VDPIN_{Baseline} \) across the 300 stocks over three years for each five-minute interval. The resulting intraday dynamics of informed trades are presented as the dashed line in Figure 3 (placed in Section 5.3 to ease comparisons with the results of the modified model). Consistent with the widely documented findings in the literature, informed trading shows an overall U-shaped intraday pattern. However, we note a spike at the opening periods, with the second five-minute interval exhibiting higher probability of informed trading than the first five-minute interval. This spike occurs primarily because China’s stock exchanges apply a call auction (9:15 to 9:30 am) to open the market prior to the formal opening of the continuous market at 9:30 am. Part of the information accumulated overnight is already disseminated into prices in the call auction, which provides feedback that requires time for other participants to reflect. Evidence of the call auction’s contribution to price discovery and market quality has also been found for the London Stock Exchange (Ellul et al., 2005) and the NASDAQ (Pagano et al., 2013).

5. LARGE TRADE PRICE IMPACT AND \( VDPIN-PI \)

Section 5.1 first analyses the proportions of different trade sizes. Section 5.2 follows to examine the price impact of the large trades. These analyses provide insights regarding why the baseline \( VDPIN \) may fail to capture informed trades of large sizes and instead may mistakenly include some non-informed contrarian trades because of the price impact. Section 5.3 explains the modified model, \( VDPIN-PI \). Then in Section 5.4, we perform information advantage test, which examines whether informed trades captured by the two models have an information advantage, compared with later prices. The results indicate that the \( VDPIN-PI \) model with the modification for price impact can identify informed trades more effectively than the baseline model.

5.1 Trade Size and Composition

To gain further insights into informed trading strategies, we divide the whole sample into three different trade size categories, i.e. small, medium and large. We present the proportions of small, medium, and large trades, in terms of both number of trades and trading volume, for each interval on the left side of Figure 1. The large- and medium-sized trades show a U-shape, whereas the small trades exhibit a reversed one, which means that a lower proportion of small-sized trades occurs at the beginning and end of the day. This implies that during these periods, (informed) investors use more medium- or even large-sized orders to acquire positions and minimise information delay costs.

The right side of Figure 1 displays the average number of trades and trade sizes at each five-minute interval for the three size categories. Small-sized trades, similar to medium- and large-sized trades, peak in terms of the number of trades at the beginning and end of the day. Different from the similar patterns observed for the number of trades, the average trade sizes show distinct features among these categories. The key difference is as regards large trades, because they present a reversed U-shape, opposite to the ones from the small and medium trades. This finding suggests that investors who trade in large sizes choose to submit orders with smaller sizes at the market open and close periods, which may be attributed to the camouflage motive and stealth trading of informed traders (Kyle, 1985; Barclay and Warner, 1993). The average sizes in the first and last 10 minutes are higher compared with those in the other intervals at the beginning and end of the day. This result could be explained by that the delay cost outweighs the price impact cost faced by informed traders or that large discretionary liquidity traders concentrate their trades at these particular periods (Barclay and Warner,
1993; Madhavan et al., 1997). Figure 1 reinforces that the trades of all three subgroups increase at the beginning and end of the day, but the medium and large trades do so more than the small trades.

### 5.2 Large Trade Price Impact

The baseline VDPIN builds on the implicit assumption that informed trades do not have a significant price impact. However, large trades are more likely to contain information and have a price impact, which will be transmitted into the contemporaneous unexpected returns. Therefore, unexpected returns in intervals with large trades can proxy for the information shock carried by large trades. This transmission mechanism is also documented by Hasbrouck (1991).

The left-side panels of Figure 1 show the composition for small, medium, and large trades within a day. They represent the average proportions of each trade size group to all trades in terms of both number of trades and trade volume at each interval. The time displayed on the horizontal axis denotes the ending time of each five-minute interval within the day. The right-side panels show the average number of trades and trade sizes for small, medium and large trades. Small, medium and large trade categories are displayed at the upper, middle, and lower panels separately. The scale for average numbers of trade is displayed on the left vertical axis (S-Number, M-Number and L-Number). The scale for average trade sizes is displayed on the right vertical axis (S-Size, M-Size and L-Size) in the unit of ‘round block’, which represents 100 shares. The time displayed on the horizontal axis denotes the ending time of each five-minute interval within the day.

He develops a framework that relates the trade (via the inferred private information) and the quote revision and finds a higher price impact for large trades. Price impact is also considered in informed trading modelling by Kitamura (2016). Hung and Lien (2019) suggest that higher trading aggressiveness, based on trader’s own private information, will more quickly push stock price to its

---

**Figure 1. Intraday average numbers of trades, trade sizes and intraday proportions for small, medium, and large trades**
new equilibrium level. They find trading aggressiveness is positively related to price impact, which implies the importance of considering price impact when measuring informed trading. Through the following examinations, we aim to illustrate the non-trivial price impact, the influence on unexpected returns, and the implications for the model modification.

Next, we examine our conjecture regarding the large trade price impact and its effect on the informed trades proxy by contrarian trades, using two methods. For both methods, intervals are separated into two groups, depending on the presence or absence of large trades in the interval. In total, there are 581,057 intervals with, and 9,287,598 intervals without, large trades. In the first method, we investigate whether the price impact of trades differs in two interval groups. Following the literature (see, e.g. Sun and Ibikunle (2017)), the price impact is measured at a five-minute frequency with Equations (3) and (4):

$$\text{Percentage price impact}\,(PPI)_s = \frac{2D_s(M_{s+5} - M_s)}{M_s}$$

(3)

$$\text{Average } PPI_{i,j} = \frac{1}{N} \sum_{s=1}^{N} PPI_s$$

(4)

where $PPI_s$ denotes the percentage price impact for trade $s$, $D_s$ is a dummy variable that indicates the direction of trade $s$ and equals 1 when trade $s$ is a buyer-initiated trade and -1 when trade $s$ is a seller-initiated trade, $M_s$ is the bid-ask midpoint at the time that trade $s$ occurs, $M_{s+5}$ is the bid-ask midpoint five minutes after trade $s$. $\text{Average } PPI_{i,j}$ refers to the average percentage price impact for trades of stock $i$ at each interval $j$ and $N$ denotes the total number of trades for stock $i$ at each interval $j$. The upper panel of Table 2 demonstrates the significant price impact caused by large trades. The average percentage price impact of trades at intervals with large trades (0.0927%) is more than three times of that without large trades (0.0254%), with the difference being significant at 1% level using the Z-test.6

<table>
<thead>
<tr>
<th>$\text{Avg } PPI_{i,j}$</th>
<th>Intervals with large trades</th>
<th>Intervals without large trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0927%</td>
<td>$TI_{i,j} \geq 0$</td>
<td>$TI_{i,j} &lt; 0$</td>
</tr>
<tr>
<td>$\varepsilon_{i,j} \geq 0$</td>
<td>74.13%</td>
<td>25.87%</td>
</tr>
<tr>
<td>$\varepsilon_{i,j} &lt; 0$</td>
<td>20.51%</td>
<td>79.49%</td>
</tr>
</tbody>
</table>

Note: In the upper panel of the table, the row of $\text{Avg } PPI_{i,j}$ presents the average percentage price impact of trades for stock $i$ at each interval $j$, according to Equation (4) for intervals with and without large trades. The lower panel demonstrates the proportions of positive and negative unexpected returns and the interactions between the unexpected return and trade imbalances for both interval categories (with and without large trades). $\varepsilon_{i,j} \geq 0$ and $\varepsilon_{i,j} < 0$ refer to the direction of unexpected returns in each interval. The columns of $\varepsilon TI_{i,j} \geq 0$ and $\varepsilon TI_{i,j} < 0$ represent the percentages of the direction of trade imbalance given the sign of unexpected return. Thus, the numbers on the diagonal and off the diagonal represent the percentage of intervals that have the same or opposite signs to $TI$ and $\varepsilon$. 


In the second method, the previously detected higher price impact from large trades is further examined to determine the potential influence on price changes and the unexpected returns. Intuitively, the net effect of the trade imbalance between buys and sells will be realised on the price change.\(^7\) For instance, the trade imbalance of net buys is associated with a price increase and positive unexpected returns. Thus we study the directional relationships between the trade imbalance and unexpected return and compare the results at intervals with and without large trades. Such results may indicate the existence of an influential price impact of large trades if the directional relationship varies on the inclusion of large trades. Following Sun and Ibikunle (2017), we introduce trade imbalance in Equation (5) as follows:

\[
TI_{i,j} = VB_{i,j} - VS_{i,j}
\]

where \(TI_{i,j}\) denotes the trade imbalance factor for stock \(i\) at interval \(j\), \(VB_{i,j}\) is the total volume of all buyer-initiated trades and \(VS_{i,j}\) is the total volume of all seller-initiated trades in the same interval.

The lower panel of Table 2 reports the relationship between \(TI\) (trade imbalance) and \(\varepsilon\) (unexpected return) for intervals with and without large trades. It shows that the values on the diagonal are higher than those off the diagonal for all intervals. This can be inferred from the effect of net buy (sell) pressure on price change and hence the sign of the relevant unexpected return. However, for intervals with large trades, \(TI\) and \(\varepsilon\) of the same sign (74.13% and 79.49%, respectively) occur more frequently than these do for intervals without large trades (60.54% and 63.97%, respectively). This result suggests that large trades have a much higher impact on prices; hence, the signs of unexpected returns in those intervals are likely to be driven by (and to be consistent with) the trade imbalance. Therefore, if informed traders submit large orders, those large trades with price impact would present as non-contrarian trades. For example, upon the arrival of positive news in a given interval, small- and medium-sized informed buys may have no substantial impact on prices. These buys would be identified as informed contrarian buys using the baseline model. However, with the same positive information, if instead a very large sized order is submitted, this informed buy may exert a dramatic positive impact on the price, and the unexpected return estimated for this interval would appear to be positive. In this case, this large informed buy would not appear to be contrarian and would be missed in the baseline \(VDPIN\) measure. Worse, the baseline \(VDPIN\) measure would mistakenly capture sells as informed trades since sells show as contrarian trades. Thus, we modify the informed trading measure by incorporating the price impact factor. In addition, even for the intervals with large trades, not all trade imbalances align with the directions of unexpected returns. This observation reflects informed traders’ order size adjustments to available liquidity; that is, not all large trades will cause an influential price impact.

### 5.3 VDPIN Modification With Price Impact

Given the substantial price impact of large trades, we develop the modified \(VDPIN\) model by contrasting the trade imbalance of large trades and unexpected returns. The model development is demonstrated with a tree diagram of five interval scenarios in Figure 2. All intervals are first divided into ones with and without large trades. For intervals with large trades, when unexpected returns \(\varepsilon\) and the trade imbalance of large trades \(TI(L)\) have the same sign, we propose this interval to be the impact interval. Therefore, the intervals under scenarios 1 and 4 represent the intervals with a large trade price impact (impact intervals). In contrast, when \(\varepsilon\) and \(TI(L)\) have different signs, we designate these intervals to be non-impact intervals. Scenarios 2 and 3 represent non-impact intervals.
Considering the lower impact of small and medium trades, we perceive intervals without large trades as non-impact intervals under scenario 5.

Through distinguishing the price impact of large trades, our modified measure of informed trades uses contrarian trades only for non-impact intervals 2, 3, and 5 indicated in Figure 2. For impact intervals 1 and 4, non-contrarian trades are used to represent informed trades. We admit that not all the non-contrarian trades in the impact intervals are informed trades. Nevertheless, the large trades with a price impact in these intervals are more likely to be informed trades, which should not be left out. In summary, the modified measure that incorporates the large trade price impact (VDPIN-PI) is described in Equation (6):

\[
VDPIN_{i,j} - PI = \left[ \frac{VB_{i,j}}{VT_{i,j}}(\varepsilon_{i,j} < 0) + \frac{VS_{i,j}}{VT_{i,j}}(\varepsilon_{i,j} \geq 0) \right] (2, 3, 5) \\
+ \left[ \frac{VB_{i,j}}{VT_{i,j}}(\varepsilon_{i,j} \geq 0) + \frac{VS_{i,j}}{VT_{i,j}}(\varepsilon_{i,j} < 0) \right] (1, 4)
\]

(6)

By aggregating VDPIN-PI over all stocks and the whole sample period, the intraday informed trades over the 48 five-minute intervals is shown (as the solid line) in Figure 3. The informed trades exhibit a more prominent U-shape than that of the baseline VDPIN, thereby implying that the modified VDPIN captures more informed trades concentrating at the opening and closing periods of the day, which is in line with the likely fact that information is more intensive during these periods.

Figure 3 shows the percentages of informed trades for each interval aggregated across all stocks and over all trading days from the baseline VDPIN and modified VDPIN (VDPIN-PI) models. The time displayed on the X-axis denotes the ending time of each five-minute interval within a day.

Beyond the difference at the overall level, it is of particular interest to note the influence of the modification on each trade size category, especially for the large trades. An anatomy of informed trades of each size category is illustrated in Table 3, compared with their counterparts from the baseline measure. Comparing the results of the two models, we observe an overall decrease in the

---

Figure 2. Tree diagram of impact and non-impact intervals. Note: \(TI(L)_{i,j}\) is the trade imbalance of large trades for stock \(i\) at interval \(j\); \(TI(L)_{i,j} = LVB_{i,j} - LVS_{i,j}\), where \(LVB_{i,j}\) is the sum volume of all large buy trades and \(LVS_{i,j}\) is the sum volume of all large sell trades.
percentage of informed trades identified by the VDPIN-PI in terms of number of trades (from 44.28% to 43.86%). However, the volume of informed trades reported by the VDPIN-PI increases (from 41.83% to 43.99%). Through scrutinising the changes in results from the baseline to the modified model, we find that the small size category experiences a drop in the number of informed trades (from 45.06% to 43.86%), but an increase in the volume of trades (from 40.09% to 41.74%). This result implies a higher average size of informed trades in the small-sized subgroup. In contrast, for
the large-sized category, we identify a significant increase in the number of informed trades (from 40.12% to 57.54%), yet less so in the volume (from 47.81% to 52.59%). This result suggests that the average size of large informed trades decreases. The least change occurs in the medium-sized category. Combining these observations, we conclude that in the VDPIN-PI model trades with information tend to converge to the medium trade size, which lends further support to the camouflage argument of Kyle (1985) and the stealth trading hypothesis of Barclay and Warner (1993). Meanwhile, the size changes discovered in the informed trades provide further support for constructing the measures using the volume rather than the number of trades. If we use the method of Chang and Wang (2019) to define the probability of informed trades as the number of medium sized contrarian trades against the total trade numbers (PCM), a corresponding PCM of 9.86%, is obtained over the sample period of 2012 to 2014. The higher level of informed trading in China is in line with our expectation given its more severe information asymmetry compared to developed markets.

5.4 Information Advantage Test

An informed trading measure can either overestimate or underestimate the probability of informed trading. For instance, liquidity trades, given their portfolio or inventory motives for trade, can be mistakenly captured by our measure. The robustness of an informed trading measure could be tested by examining the extent to which the trades captured in this measure have an information advantage over followers. We examine whether the proportion of the trades captured in the baseline VDPIN (and VDPIN-PI) can indeed trade at a better price than the subsequent price that has reflected most of the relevant information. The results of the autocorrelation tests imply that, in China’s stock market, it takes about 30 minutes for stock prices to fully impound new information (can be found in Section 6.1). Hence, we postulate that a buy (sell) trade has information advantage if traders can buy (sell) at a price that is lower (higher) than the price of 30 minutes later.

As a benchmark, we first examine the proportion of all trades that have information advantage using Equation (7). The informed trades captured by the baseline VDPIN and VDPIN-PI can be tested for information advantage through Equations (8) and (9) as follows:

\[
PT_{i,j}^{IA} = \frac{VB_{i,j}}{VT_{i,j}}(P_t < P_{t+6}) + \frac{VS_{i,j}}{VT_{i,j}}(P_t \geq P_{t+6})
\]

\[
VDPIN_{Bais_{i,j}}^{IA} = \frac{VB_{i,j}(\varepsilon_{i,j} < 0, P_t < P_{t+6})}{VB_{i,j}(\varepsilon_{i,j} < 0)} + \frac{VS_{i,j}(\varepsilon_{i,j} \geq 0, P_t \geq P_{t+6})}{VS_{i,j}(\varepsilon_{i,j} \geq 0)}
\]

\[
VDPIN - PT_{i,j}^{IA} = \left[\frac{VB_{i,j}(\varepsilon_{i,j} < 0, P_t < P_{t+6})}{VB_{i,j}(\varepsilon_{i,j} < 0)} + \frac{VS_{i,j}(\varepsilon_{i,j} \geq 0, P_t \geq P_{t+6})}{VS_{i,j}(\varepsilon_{i,j} \geq 0)}\right]_{(2,3,5)}
\]

\[
\left[\frac{VB_{i,j}(\varepsilon_{i,j} \geq 0, P_t < P_{t+6})}{VB_{i,j}(\varepsilon_{i,j} \geq 0)} + \frac{VS_{i,j}(\varepsilon_{i,j} < 0, P_t \geq P_{t+6})}{VS_{i,j}(\varepsilon_{i,j} < 0)}\right]_{(1,4)}
\]

where \(P_t\) and \(P_{t+6}\) represent the transaction prices at time \(t\) and time \(t+6\) (30 minutes later), respectively. \(PT_{i,j}^{IA}\) represents the proportion of all trades that have information advantage for each stock and at
each interval. \( VDPIN_{\text{Base}_{i,j}}^{IA} \left( VDPIN - PI_{i,j}^{IA} \right) \) represents the proportion of trades captured by \( VDPIN_{\text{Base}_{i,j}} \left( VDPIN - PI_{i,j} \right) \) that have information advantage, because the traders can buy (sell) at a price that is lower (higher) than the price of 30 minutes later. The intraday proportion of trades that have information advantage is presented in Figure 4 for all trades as well as informed trades captured by the baseline and modified models.

Figure 4 clearly shows that, compared with all trades and trades captured by the baseline \( VDPIN \), the trades captured by the \( VDPIN-PI \) indeed have information advantage throughout the trading day. This result strongly suggests that \( VDPIN-PI \) is a more efficient proxy of informed trades than the baseline \( VDPIN \).

In Figure 4, the line for ‘All Trades’ represents the proportion of all trades that have information advantage (\( PT_{i,j}^{IA} \) in Equation (7)). The line for ‘Baseline \( VDPIN \) Trades’ represents the proportion of trades captured by the baseline \( VDPIN \) that have information advantage (\( VDPIN_{\text{Base}_{i,j}}^{IA} \) in Equation (8)). The line for ‘\( VDPIN-PI \) Trades’ represents the proportion of trades captured by \( VDPIN-PI \) that have information advantage (\( VDPIN - PI_{i,j}^{IA} \) in Equation (9)). All results are obtained first from computation of individual stocks, and then averaged across stocks.

6. FURTHER ANALYSES

In this section, we perform two additional analyses. First, we conduct an autocorrelation test to determine whether there are reversals in price changes caused by informed and uninformed trades captured by our models. Second, to gain insights into information transmission mechanism over time, we test the predictability of information shock.

Figure 4. Intraday pattern of proportion of trades with information advantage
6.1 Autocorrelation Test

We first employ the autocorrelation regression test proposed by Avramov et al. (2006), which is based on Campbell et al.’s (1993) model (CGW). The CGW model suggests that price decreases should result from either negative news or liquidity-driven selling. Uninformed liquidity-driven selling can lead to temporary price decreases, which should subsequently be reversed after the release of selling pressure, whereas informed selling should cause permanent price changes. Avramov et al. (2006) perform an autocorrelation test for the next day for their daily data. To accommodate our five-minute high-frequency data, we consider up to 30 minutes to document the process of return reversals for uninformed trades. The regression is presented in Equation (10) as follows:

\[
\varepsilon_{i,j+n} = \phi + \left[ \delta_{i,0} \text{Vol}_{i,j} + \delta_{i,1} \frac{V_{i,j}}{V_{i,j}}(\text{informed}) + \delta_{i,2} \frac{V_{i,j}}{V_{i,j}}(\text{uninformed}) \right] \varepsilon_{i,j} + u_{i,j+n} \quad (10)
\]

where \( \varepsilon_{i,j+n} \) is the unexpected return for stock \( i \) at the interval from \( j \) to \( j + n \). Given that we use five-minute returns, \( \varepsilon_{i,j+6} \), for example, is the unexpected return for the 30-minute interval from \( j \) to \( j + 6 \). \( \text{Vol}_{i,j} \) is the volume of stock \( i \) in interval \( j \) to control for the volume effect on return reversal. \( V_{i,j} / V_{i,j} \) (\text{informed}) represents the informed sell and \( V_{i,j} / V_{i,j} \) (\text{uninformed}) represents the uninformed sell of stock \( i \) at interval \( j \). If our model is able to identify informed trades, we would expect to find that \( \delta_1 \) insignificant and \( \delta_2 \) significantly negative, as suggested by CGW and Avramov et al. (2006). Given that we found the price impact from large trades and proposed the subsequent modification to intervals with large trades, it is more meaningful to report and analyse the findings between the baseline and modified measures for these intervals in particular:

Table 4 presents the results of the autocorrelation test for intervals with large trades for the baseline and modified models. \( \delta_2 \) is significantly negative for the modified model, which

<table>
<thead>
<tr>
<th>Baseline</th>
<th>VDPIN</th>
<th>VDPIN-PI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \delta_0 \times 10^4 )</td>
<td>( \delta_1 )</td>
</tr>
<tr>
<td>( j + 1 )</td>
<td>-1.71*</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>(-1.73)</td>
<td>(-1.29)</td>
</tr>
<tr>
<td>( j + 2 )</td>
<td>-1.98*</td>
<td>-0.34</td>
</tr>
<tr>
<td></td>
<td>(-1.86)</td>
<td>(-1.23)</td>
</tr>
<tr>
<td>( j + 3 )</td>
<td>-2.87*</td>
<td>-0.40</td>
</tr>
<tr>
<td></td>
<td>(-1.85)</td>
<td>(-1.10)</td>
</tr>
<tr>
<td>( j + 4 )</td>
<td>-2.62*</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>(-1.79)</td>
<td>(-0.98)</td>
</tr>
<tr>
<td>( j + 5 )</td>
<td>-2.30</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>(-1.61)</td>
<td>(-1.06)</td>
</tr>
<tr>
<td>( j + 6 )</td>
<td>-2.11</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

Note: This table summarises the results of the autocorrelation test as in Equation (10). The test was undertaken for five-, 10-, 15-, 20-, 25- and 30-minute intervals. All the results were achieved from regressions of individual stocks first, and then averaged across stocks. The averages of the Newey and West (1987) robust t-statistics are shown in the parentheses, and significance at the 1%, 5% and 10% levels is denoted by ***, ** and *, respectively.
indicates that the return reversal is indeed observed for the uninformed trades measured by the modified model, consistent with the CGW theory. Meanwhile, the large informed trades captured by the modified measure since they do not show subsequent price reversals ($\delta_1$ are all positive) are unlikely to be executed by liquidity traders. However, a positive $\delta_2$ is observed for the baseline model, thereby, implying that uninformed trading is not properly measured. The results suggest that the modified measure is able to distinguish large informed trades from large liquidity trades.10

6.2 Predictive Regression Analysis on the Information at Market Close

The literature generally suggests that the opening session of a trading day embeds information accumulated overnight, which explains the formation of the left-side peak of the intraday U-shaped pattern of informed trading. Yet, the formation of the right-side peak is less understood. Gao et al. (2018) document the intraday predictability of the last half-hour returns by using data on the first half-hour returns, which they find is supported not only by the infrequent portfolio rebalancing theory but also by the model of late-informed trading. Investors who receive information late or are slow in processing information will act before the market close when liquidity is high. We consider this argument merits further investigation. Do informed traders in the last half-hour trade only on the information retained from the early morning? Kurov et al. (2019) find that prices begin to move in the ‘correct’ direction before a macroeconomic news announcement. They suggest that some traders have private information about macroeconomic fundamentals. The source of their private information may be information leakage, comparative advantage some traders possess in collecting and processing information. Kaniel et al. (2012) and Campbell et al. (2009) document pre-announcement informed trading not only around the macroeconomic news, but also on corporate news. Given the aforementioned discussion, we examine the attributes of the information near the market close by testing the unexpected return predictability from market open to market close as well as from market close to the next day. Considering the high percentage of informed trading within the half-hour before and after market close in Figure 3, we analyse the predictive regressions at the half-hour frequency.

First, to examine the intraday predictability we focus on how the last half-hour unexpected return is predicted by that of the first half-hour using Equation (11):

$$\varepsilon_{i,14:30–15:00} = \alpha_1 + \beta_1 \varepsilon_{i,9:30–10:00} + \varepsilon_{i,14:30–15:00}$$  (11)

where $\varepsilon_{i,14:30–15:00}$ denotes the unexpected return of stock $i$ spanning from 14:30 to 15:00 (from 9:30 to 10:00). It is calculated by Equation (1) adjusted for half-hour frequency.

To assess the predictability of the unexpected return from the last half-hour to the next day, we estimate Equation (12):

$$\varepsilon_{i,15:00–10:00[+1]} = \alpha_2 + \beta_2 \varepsilon_{i,14:30–15:00} + \varepsilon_{i,15:00–10:00[+1]}$$  (12)

where $\varepsilon_{i,15:00–10:00[+1]}$ denotes the unexpected return of stock $i$ spanning from the previous day’s close at 15:00 to 10:00 of the day (from 14:30 to 15:00). In this analysis of inter-day predictability, we include the overnight price changes in estimating the unexpected return for the next day. This is because part of the overnight information is already reflected in the opening price of the next day through the opening auction. Beyond the half-hour interval, we also extend the next day return to both the whole morning session and the whole trading day, to investigate the predictive ability for a longer horizon.11
The left part of Table 5 reports the unexpected return predictability for the whole sample. The upper panel shows that the intraday predictability of the first half-hour to the last half-hour is not statistically significant at conventional levels, regardless of the positive slope of 0.5 and $R^2$ of 1.33%. However, the lower panel shows that the unexpected return predictability from the last half-hour interval to the next day is significant, implying the predictability of informed trading given the way informed contrarian trades are constructed.

As discussed earlier, in intervals with large trades, high information content is embodied in the unexpected returns given the significant price impact of large trades. A positive (negative) information shock causes positive (negative) unexpected returns. Therefore, we are particularly interested in the predictive ability of the unexpected returns of the intervals with large trades. The results, reported in the right part of Table 5, demonstrates both economically and statistically significant predictability, despite the small sample size. It is noteworthy that the intraday predictability, shown in the upper panel, turns highly significant when focusing on the intervals with large trades. Both $R^2$ and the $t$-statistic increase remarkably compared with that for the results from all intervals. This indicates that the opening sessions with large trades are more likely to have information retained until market close. Specifically, high predictability of market open information to market close information is suggested by the statistically significant slope of 0.76. The lower panel indicates significant predictability of the last half-hour to the next day information, although with lower slope values. As the predicted variable expands from the first half-hour to the next morning session and to the whole

<table>
<thead>
<tr>
<th>Predictor: $\epsilon_{i,9:30-10:00}$</th>
<th>Predictor: $\epsilon_{i,9:30-10:00}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>$R^2$ (%)</td>
</tr>
<tr>
<td>$\epsilon_{i,14:30-15:00}$</td>
<td>0.50</td>
</tr>
<tr>
<td>(1.13)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor: $\epsilon_{i,14:30-15:00}$</th>
<th>Predictor: $\epsilon_{i,14:30-15:00}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2$</td>
<td>$R^2$ (%)</td>
</tr>
<tr>
<td>$\epsilon_{i,15:00-10:00}^{(+1)}$</td>
<td>0.067***</td>
</tr>
<tr>
<td>(2.61)</td>
<td></td>
</tr>
<tr>
<td>$\epsilon_{i,15:00-11:30}^{(+1)}$</td>
<td>0.05**</td>
</tr>
<tr>
<td>(2.26)</td>
<td></td>
</tr>
<tr>
<td>$\epsilon_{i,15:00-15:00}^{(+1)}$</td>
<td>0.031*</td>
</tr>
<tr>
<td>(1.86)</td>
<td></td>
</tr>
</tbody>
</table>

Note: This table reports the predictive regression results of the unexpected return from market open to market close using Equation (11) as well as from market close to the next day using Equation (12). The evidence for all intervals in the whole sample and for the predictor intervals with large trades are displayed in the left part and right part respectively. All the results were obtained from the regressions of individual stocks first, and then averaged across stocks. The averages of the Newey and West (1987) robust $t$-statistics are shown in the parentheses, and significance at the 1%, 5% and 10% levels is denoted by ***, ** and *, respectively.
next day, both $R^2$ and the $t$-statistic monotonically decrease. This may be partly due to the diluting effects from new information inflows during the day.

In summary, there is strong evidence that the late afternoon unexpected returns predict that of the next day. Intraday predictability from morning to market close is also present but only significant in intervals with large trades. This suggests strong predictability of informed trading in large size from early morning to market close and from late afternoon to the next day. Together, this evidence provides insightful explanations for the U-shaped informed trading pattern, especially the right-side peak, which is associated with both information retained from the morning and private information that is supposed to arrive the next day.

7. CONCLUSION

Contrarian trades have been proposed and applied by many studies to represent informed trades. The notion is that, considering the rationality of investors - especially those with private information - informed subjects frequently act as contrarians (Drehmann et al. (2005)). However, large trade price impact poses a challenge to the proxy of informed trades by contrarian trades. This issue is linked to another strand of research on the camouflage behaviour of informed traders, which means they may split large trades into medium sizes to mitigate the price impact cost. Hence, an examination of trade size variations is essential to understand informed trades. This article conducts, at the transaction level, an investigation of informed trades on the CSI300 component stocks in China’s market. We begin by applying the baseline measure ($VDPIN$), which uses contrarian trades to proxy for informed trades. Further examination of price impact leads us to incorporate price impact into the modified model, that is, $VDPIN-PI$. We also conduct a predictability test to seek explanations of the detected intraday pattern of informed trading.

Our main findings are as follows: (i) We identify the aggregate intraday U-shaped informed trades for CSI300 constituent stocks, similar to the findings from other markets. (ii) As revealed by a trade composition analysis, the aggregate U-shape of informed trades is essentially driven by the change of trade size composition within the day. (iii) Large-sized trades exert a substantial price impact that affects the detection of informed trades using the contrarian trades. By accounting for price impact, the modified model ($VDPIN-PI$) can capture informed trades more effectively, as confirmed by the additional analyses. (iv) The autocorrelation test results suggest that the market takes about 30 minutes to digest information. (v) The results from the predictability analysis, especially of the intervals with large-sized trades, imply that the informed trading at market close is driven not only by information retained from early morning (late-informed trading) but also by private information supposed to arrive the next day. This finding provides new insights for the high informed trading found at market close.

In summary, the results are generally consistent with those of Admati and Pfleiderer (1988) regarding intraday patterns of volume and price variability, which suggests that informed (and liquidity) trading should concentrate at the open and close at the aggregate level. Further, this study incorporates the unexplored dimensions of trade size composition and predictability into the empirical investigation of the aggregate U-shape pattern of informed trading. The findings stated in (ii) and (v), in addition to the time-of-day effect, explain the intraday U-shaped informed trading. Moreover, given the non-negligible nature of price impact, this study reveals the importance of incorporating price impact in informed trading modelling.

FUNDING AGENCY

Publisher has waived the Open Access publishing fee.
REFERENCES


ENDNOTES

1 In the report of ‘Shanghai Stock Exchange market and infrastructure’ published by Shanghai Stock Exchange, it is reported that ‘Individual investors make up an absolutely majority of investors. In terms of existing investor accounts, individual investors make up 99% of the total accounts. By the end of 2016, there were 193 million investor accounts in the Shanghai Stock market, including 192 million individual investor accounts and 681,000 institutional investor accounts. In 2016, among active accounts, retail investor accounted for the most (59.7%), followed by small investors (27.8%). Medium investors, large and extremely large investors all had a much lower number of active accounts.’ Accessed from http://english.sse.com.cn/news/publications/sseinfrastructure/ at 12th January 2022.

2 Buyer-and seller-initiated trades are indicated for each trade price. More information about the Wind database can be found from the website: www.wind.com.cn

3 We also divide each trading day into sixteen 15-minute intervals and analyse the intraday pattern of informed trading. Similar but less prominent results are obtained compared with those for the five-minute interval. We report the five-minute interval results, which allow us to have more insights and understanding of information occurrence, informed trading behaviour and the market in general. The analysis results for the 15-minute intervals are not reported but are available upon request.

4 The serial dependence on returns is caused by the lagged adjustment to information, exchange-mandated price smoothing and other market microstructure imperfections (Hasbrouck, 1991).

5 The trade size classification method is explained in Section 3 ‘Data and Descriptive Statistics’.

6 Given the prevalence of medium-sized trades, it is not sensible to further separate the intervals between with and without midsized trades.

7 To clarify, the directional relationship in this study refers to the consistency of the signs of two variables, $TI_{i,j}$ and $\epsilon_{i,j}$.

8 Chang and Wang (2019) reported about 5% informed trades (PCM) in the U.S. market over the period between 1993 and 2012.

9 The results for the whole sample are available upon request.

10 We find large trades have the highest percentage of informed trades among all trade size categories. However, one concern is that some of the large trades are probably executed by liquidity traders, such as some financial institutions (Admati and Pfleiderer, 1988), and these large liquidity trades could be mistakenly measured as large informed trades. Unlike developed markets, China’s stock market is dominated by individual traders. Over 70% of investors are actually retail investors. Financial institutions, such as pension funds, were not allowed to invest in the stock market until April 2017. Therefore, large trades in China’s stock market are less likely to be executed by institutions for liquidity reasons.

11 For longer horizons of half day and whole day, the predictive regression of Equation (12) similarly can be rewritten as $\epsilon_{15,00-1:15,00} = \alpha_k + \beta_1 \epsilon_{15,00-1:15,00} + \epsilon_{15,00-1:15,00}$ and $\epsilon_{15,00-1:15,00} = \alpha_k + \beta_1 \epsilon_{15,00-1:15,00} + \epsilon_{15,00-1:15,00}$ respectively.

12 The serial dependence on returns is caused by the lagged adjustment to information, exchange-mandated price smoothing and other market microstructure imperfections (Hasbrouck, 1991).