INVESTOR TYPES, LIQUIDITY AND PRICE FORMATION: EVIDENCE FROM THE STOCK EXCHANGE OF THAILAND

STEFANO ALDERIGHI, PEDRO GURROLA-PEREZ

SEPTEMBER 2021

WFE WORLD FEDERATION OF EXCHANGES

Investor types, liquidity and price formation: evidence from the Stock Exchange of Thailand

Stefano Alderighi[(1)], Pedro Gurrola-Perez[(2)]

September 2021

Abstract Using timestamped orders and trades data from the Stock Exchange of Thailand, we study how different market participants - retail investors, domestic and foreign institutions - influence price formation for different liquidity levels. We find that trading participants contribute heterogeneously to price formation, and that liquidity affects the size, but not the relative importance, of their contribution to prices. Retail investors' trades are associated with the highest information content for both high- and low-liquidity stocks. In particular, retail investors' trade imbalances are associated with the highest five-minutes returns' predictability for almost all liquidity levels, while trades between individual traders are associated with sizeable price impacts both over a five-minute and a daily horizon. Foreign investors appear to be the second most important contributors to price formation. Interestingly, we show that trades between foreign and retail traders are the ones that convey most information, hinting that the interaction between two informed categories is beneficial for price formation. Finally, we find that domestic institutions contribute only marginally to price formation, consistently with them being largely buy-and-hold participants.

Keywords: Investor types, liquidity, price efficiency, price discovery, emerging markets.
JEL classification: G10, G12, G14.

[(1)] UK Endorsement Board. Email: s.alderighi@endorsement-board.uk. Stefano Alderighi worked on this paper while he was at the World Federation of Exchanges.
[(2)] World Federation of Exchanges. Email: pgurrola@world-exchanges.org

1 Introduction

One of the two main functions of stock exchanges is to ensure the efficient and orderly functioning of secondary markets.[1] By setting fair and transparent trading rules, and by offering trading in a wide set of companies, stock markets aim to attract secondary market participants. These provide liquidity to the market, and, by engaging in trading, allow prices to incorporate new, relevant information (the process of price discovery), bringing them closer to their fundamental value (the concept of price efficiency). While the concepts of liquidity, price discovery and price efficiency are distinct, they are closely related: not only has liquidity been found to lead to both better price discovery (Riordan and Storkenmaier, 2012; Frijns et al., 2018) and enhanced price efficiency (Chordia et al., 2008; Chung and Hrazdil, 2010), but the literature suggests that prices that more readily incorporate new information are also more efficient (Chordia et al., 2008; Chung and Hrazdil, 2010). In other words, greater informational efficiency leads prices to more closely follow a random-walk process, i.e. to be more efficient in an asset pricing meaning of the term (O'Hara, 2011, 2003).

The positive relation between liquidity and price formation is predicted by the theory and is well documented empirically, yet the literature is somewhat silent on how this very same relation interacts with market participation. It is well-established however that different types of market participants (i.e. retail investors, domestic institutions, foreign institutions) contribute differently to trading activity, depending on a jurisdiction's historical, cultural and institutional factors (WFE and Oliver Wyman, 2016; Alderighi, 2017); that is, trading participants contribute heterogeneously to liquidity (Grinblatt, 2000).[2] Even more importantly, the literature agrees that different types of trading participants possess different levels of information (Albuquerque et al., 2009; Chan et al., 2007; Froot and Seasholes, 2001; Goncalves and Eid, 2017; Grinblatt, 2000; Lang et al., 2020; Xu and Wan, 2015) and abilities to interpret market movements (Dvorak, 2005), and therefore contribute heterogeneously to price formation. Moreover, information asymmetries across market participants are found to have an impact on stock prices (Chan et al., 2008).

Following the considerations above, a natural question is whether the contribution of different types of trading participants to price formation also varies depending on the liquidity of the stock. This question is important for many reasons. To start with, exchanges often struggle to achieve satisfactory levels of liquidity and price efficiency in smaller capitalised stocks (Menkveld and Wang, 2013). Understanding the extent to which small-stock market participants are characterised by information asymmetries would provide exchanges with useful information on where to intervene to enhance their liquidity. Who are the uninformed liquidity suppliers? What intervention(s) can be proposed to bring more of them to the market and help counterbalance the activity of informed traders, thus reducing price impacts? Would enhanced liquidity attract even more informed traders, thus positively contributing to price formation? Understanding these dynamics is no less important for large-cap stocks. Is information on large cap stocks impounded into prices thanks to one dominating group of traders? Or do blue-chip stocks benefit from a healthy mix of informed trading participants? If not, what can an exchange do to reduce their dependence on a individual category of informed traders? To the best of our knowledge, however, no paper has attempted to answer the questions above by investigating the interaction between liquidity, market participation and price formation. This paper aims to fill this gap.[3]

We investigate this question using proprietary timestamped limit order book and trades data from the Stock Exchange of Thailand (SET). SET's data is ideal for this purpose for several reasons. Firstly,

---

[1]The other one being allocating risk capital to companies through primary markets.

[2]Asian markets for example are typically characterised by high levels of retail participation, while European and Latin American markets dominated by institutional investors (WFE and Oliver Wyman, 2016).

[3]As noted in Boehmer and Wu (2008), "understanding how trader-type specific order flow affects prices has important implications for modeling the dynamics of liquidity provision, trader behaviour, and market design". Adding the liquidity dimension seems to the authors a natural step towards enhancing this understanding.

their data allows to attribute orders and trades to different types of trading participants (i.e. retail, domestic institutions, foreign institutions).[4] This setting is an improvement as compared to previous studies on this topic, which largely relied on either ownership or shareholding data (Ferreira et al., 2017; Grinblatt, 2000; Lang et al., 2020), used lower-frequency transactional data from stock exchanges (Goncalves and Eid, 2017; Xu and Wan, 2015), or had information on trades but not order submissions (Dvorak, 2005). Secondly, liquidity in SET is spread across stocks of all sizes, thus allowing us to use a representative sample of the whole market, and not just a selection of highly liquid stocks. Thirdly, SET presents several features that make the estimation of liquidity and price formation particularly clean: it is characterised by high levels of transparency, as the vast majority of trades pass through the order book; it has no fragmentation (public Thai companies are only traded on SET, and not on dark pools, multilateral trading facilities, or other markets); and it has a relatively contained amount of algorithmic/high frequency trading (AT/HFT) (Likitapiwat, 2016), mostly concentrated on large cap stocks.

Our focus on an emerging market is also important as most of the studies on the topic are set on developed markets: the United States (Brogaard et al., 2014; Chordia et al., 2008; Hasbrouck and Saar, 2013; Hendershott et al., 2011), Canada (Riordan et al., 2013), the United Kingdom (Benos and Sagade, 2016), Germany (Riordan and Storkenmaier, 2012) or are comparisons between advanced markets (Frijns et al., 2018, 2015; Wallace et al., 2019).[5] With over US\$500 billion market capitalisation as of end of 2019, nearly US\$30 billion of monthly trading activity on average in 2019, high levels of liquidity, a sophisticated trading platform, low cost of trading and a burgeoning underlying economy, the Stock Exchange of Thailand has established itself as an important emerging market exchange, suitable to expand the external validity of the evidence on the topic.

We first study the relation between liquidity, investor participation and price formation by assessing whether trade imbalances predict returns, a relation associated with the presence of informed traders (Huang and Stoll, 1994; Hanke and Weigerding, 2015), but also potentially caused by inventory effects by uninformed liquidity suppliers (Boehmer and Wu, 2008). Consistently with the literature, we find that the imbalance of aggressive orders (Lee et al., 2004) is a positive predictor of future five-minute returns, and that returns' predictability almost monotonically increases with illiquidity (Chordia et al., 2008). We contribute to this strand of literature by showing that this relation depends on market participation, with order imbalances by foreign investors being the strongest predictor of five-minute returns (economically and statistically) for the highest liquidity segment, and order imbalances by retail investors being the strongest predictor of five-minute returns (economically and statistically) for all other segments. We attribute this result to retail investors being informed traders, i.e. better at analysing the market and interpreting market trends (Engelberg et al., 2012; Vega, 2006). These finding suggest that retail investors have a preponderant role in conveying information to the market.

We then move on to study the relation between liquidity, price discovery and investors' type. We begin by estimating the price impact of trades over five-minute horizons, following Riordan and Storkenmaier (2012). We test, and find support for, the notion that different investors' categories are characterised by heterogeneous information contents, in line with the literature on the topic (Dvorak, 2005; Goncalves and Eid, 2017; Lang et al., 2020; Xu and Wan, 2015). To do so, we estimate the price impact of trades by different investor categories, focusing on trades within the same category (e.g. retail to retail) and between different categories (e.g. retail to foreign). We find that the most informative trades are between retail investors, between foreign institutions, and, notably, between retail investors and foreign institutions. Trades involving domestic institutions are instead the least informative. This evidence suggests that

---

[4]A fourth category of investors documented by previous literature on SET (Pavabutr and Sirodom, 2010; Phansatan et al., 2012), proprietary traders, can't be distinguished from institutional investors in the data.

[5]Though the literature on the contribution of domestic and foreign investors to price discovery has also focused on emerging markets, for example: China (Chan et al., 2007; Xu and Wan, 2015), Indonesia (Dvorak, 2005), Brazil (Goncalves and Eid, 2017).

retail and foreign investors contribute proportionally more to price discovery than domestic institutions, the latter ones likely to be largely buy-and-hold institutions.

We provide additional evidence by estimating the price impact function using the Hasbrouck's SVAR model (Hasbrouck, 1991a,b). We contribute to the literature by estimating the Hasbrouck's price impact of trades by different investor categories, performing an exercise similar in spirit to (Benos and Sagade, 2016) but, to the best of our knowledge, never performed before. We find that trades between retail and foreign investors are associated with the highest price impacts, followed by trades between retail investors and trades between foreign investors. Domestic institutions play instead a marginal role in the price discovery process. We find that this pattern is consistent across all liquidity levels, suggesting that the relative contribution of different trading participants to price formation does not depend on liquidity. These findings are consistent with the ones obtained from previous analyses, reinforcing the notion that retail and foreign investors are the most informed categories; that is, the ones likely possessing the highest dexterity at interpreting and profiting from market trends and available information (Engelberg et al., 2012; Vega, 2006).

The rest of the paper is as follows. Section 2 summarises the relevant literature. Section 3 describes the database. Section 4 discusses the empirical models and presents the results. Section 6 concludes and provides hints for future research.

2   Related literature

2.1   Liquidity and price formation

The prevailing view in the literature is that more liquidity leads to better price formation.[6] Illiquid stocks are indeed typically characterised by wider spreads, which reflect adverse selection costs arising from the presence of informed traders, among other things (such as inventory and order processing costs) (Brennan and Subrahmanyam, 1996; Glosten and Milgrom, 1985; Huang and Stoll, 1997). For traders, the bid/ask spread is essentially a transaction cost (O'Hara, 2003; Tetlock, 2008), henceforth higher liquidity (i.e. smaller spreads) represents an opportunity for better informed investors to transact more because they can make higher profits using their trading skills (Frijns et al., 2018; Menkhoff and Schmeling, 2010). In addition, it must be noted that uninformed traders typically protect themselves against better informed traders by supplying liquidity to the market, thus reinforcing this mechanism (Admati and Pfleiderer, 1988). These considerations lead us to conclude that liquidity should in principle be associated with better price formation.[7]

Several empirical studies confirm these predictions. Chordia et al. (2008) focuses on the relation between liquidity and price efficiency. Using ten years of data from the New York Stock Exchange (NYSE), the authors find that subsequent reductions in tick sizes (from the eighth to the sixteenth regime in July 1997, and then from the sixteenth to the decimal regime in January 2001), notably associated with increases in liquidity (Aitken and Comerton-Forde, 2005), led to an increase in price efficiency. Trade imbalances are also more likely to predict five-minute returns during moments of high illiquidity. Aktas et al. (2008) and Chung and Hrazdil (2010) confirm the evidence provided in Chordia et al. (2008).

---

[6]The behavioural finance literature postulates a negative relation between liquidity and price efficiency, ground on dealers' limited cognitive ability (Barberis et al., 1998) or liquid markets attracting a higher number of noise traders (de Long et al., 1990). Based on these conjectures, some experimental studies investigate the relation between liquidity and price efficiency using data on bets (such as bets on horse races or football matches), because this settings allow the researcher to observe the price fundamental (i.e. the money one would win/lose from the bet on a particular horse/team) in a controlled environment that works very similarly to an exchange. However, to date the experimental evidence on the relation between liquidity and price efficiency is controversial. Tetlock (2008) uses data on sports bets to find that higher liquidity is associated with lower price efficiency. Using a similar framework, Brown and Yang (2016) find the opposite result.

[7]A positive relation between liquidity and price efficiency is found for different asset classes. For example, Brandt and Kavajecz (2004) find that liquidity is positively associated with price discovery in the U.S. treasury market.

Riordan and Storkenmaier (2012) study whether the introduction of the Xetra trading system at Deutsche Börse, aimed at reducing latency on the market, led to an increase in liquidity and an improvement of price discovery.[8] They find that the introduction of Xetra greatly reduced quoted and effective spreads on Deutsche Börse. At the same time, the introduction of Xetra is negatively related to both the five-minute and the permanent price impact of trades, which the authors interpret as an improvement in price discovery. While the authors do not directly link the increase in liquidity to the enhanced price discovery, their findings suggest that the two are contingent. Interestingly, the authors perform these analyses on subsamples identified by market capitalisation quartiles, finding that the introduction of Xetra leads to a proportionally higher improvement in price discovery for small cap (i.e. relatively illiquid) stocks.

Evidence on the relation between liquidity and price efficiency is also found in a related strand of literature that exploits the quasi-experimental setting provided by cross-listings (of stocks, or otherwise) to study price formation (Hasbrouck, 1995).[9] In theory, the prices of instruments cross-listed on different markets should react to the same information, and therefore converge to the same fundamental value. In practice, this is not the case because of information (and other) frictions, leading investors to forego arbitrage opportunities between different venues. These frictions allow the researcher to study how different market characteristics (such as liquidity) affect price discovery.

Frijns et al. (2018), for example, use a sample of Canadian stocks cross-listed in the United States to show that price discovery takes place in the more liquid market (the United States). Pascual et al. (2006), using a sample of Spanish stocks cross-listed on BME Spanish Exchanges and the New York Stock Exchange, find that price discovery happens on the leading market in terms of trading activity for those stocks, the Spanish one. This suggests that this dimension of liquidity is important to explain price discovery (Hasbrouck, 1995). Along similar lines, Wallace et al. (2019) compare and contrast derivatives and equity markets by studying a derivatives contract (the E-mini future) and an ETF (the SPY ETF) linked to the same underlying (the S&P 500). The authors find that the E-mini futures contract, characterised by higher liquidity (a lower spread, higher trading turnover and higher trading volume), dominates the SPY ETF in terms of price discovery. These findings support the proposition that liquidity and price discovery are positively correlated.

Based on theoretical predictions and empirical findings we expect to find better price discovery and more efficient prices when liquidity is higher.

2.2 Market participation and price formation

The literature has investigated whether different market participant types possess different levels of information largely focusing on the dichotomy foreign/domestic investors. The reason is that information asymmetries between domestic and foreign investors are a determinant of the home bias (Alderighi et al., 2019; Dvorak, 2005); as such, understanding how to level these information asymmetries is important to enhance cross-border equity flows towards emerging economies (Alderighi et al., 2019; Lang et al., 2020). From a theoretical perspective, it is unclear who should possess better information between domestic and foreign investors, as on one hand international investors have to face cultural and language barriers that put domestic investors at an informational advantage, though on the other hand foreign investors are typically more experienced traders than domestic ones (Albuquerque et al., 2009; Bae et al., 2008; Froot and Seasholes, 2001). Therefore, determining whether information is largely held by domestic or foreign participants is an empirical question.

Empirically, the literature has typically approached this problem by studying whether international

---

[8]For a description of Xetra, see https://www.xetra.com/xetra-en/trading and Reck (2020)

[9]Strictly speaking, the focus is on the same underlying, not cross-listing as such. Chakravarty et al. (2004) for example use the (Hasbrouck, 1995) framework to study price discovery in stock and option markets, finding that a relevant share of price formation happens on the derivatives market.

investors are better performers than domestic ones, thus attributing the ability to make better returns to superior information. The results are contrasting. Dvorak (2005) for example, using data from the Jakarta Stock Exchange (JSX), finds that domestic investors have an overall information advantage, though some foreign institutions are better performers thanks to their superior experience. He also finds that brokerage firms, who also sell research and advisory services, have a role in providing better information, as clients of global brokers have higher long-run profits than clients of either local or Asian brokerages. Lang et al. (2020) study the effect of Multilateral Memorandum of Under- standing (MMoU), a form of cross-border cooperation, on global mutual fund portfolio allocation. They find that foreign mutual fund companies facing higher informational asymmetries are the ones benefitting the most from MMoUs, thus investing proportionally more in US cross-listed companies than other mutual funds. These findings are consistent with international investors being at an informational disadvantage. Other contributions (Ferreira et al., 2017; Kang and Stulz, 1997) find a consistent result.

A number of papers find however that foreign investors are the ones possessing superior information. Grinblatt (2000) uses a comprehensive database of Finnish stock holdings to study the trading behaviour and performance of different investor categories, including retail investors (households). He finds that international investors are the most sophisticated players on the Finnish market, while domestic investors (and in particular retail investors) pursue contrarian strategies that hint for lower levels of sophistication. Froot and Seasholes (2001), using a database of 25 destination countries, show that institutional portfolio inflows have a positive effect on domestic stock prices attributable to fundamentals, suggesting that international investors possess better information than domestic ones.

Chan et al. (2007) follow a different approach from the papers cited so far, as they evaluate the information content of foreign and domestic trades by comparing A-shares (reserved to domestic investors) and B- shares (reserved to foreign investors) on the Chinese stock market using the Hasbrouck (1995) model, noting that in 2001 some domestic investors were allowed to trade in B-shares too. They find that A-shares dominate price discovery, suggesting that domestic investors possess better information on Chinese stocks.

A different literature (excellently summarised in Hanke and Weigerding (2015)) looks at the relation between order imbalances and stock prices, with some contributions focusing on the contribution of different trading participants to price discovery. Boehmer and Wu (2008), using data from the New York Stock Exchange, find that domestic institutions' trade imbalance positively affect returns, a finding they attribute to information impounding. They find that retail investors, on the contrary, are uninformed liquidity providers. Lee et al. (2004), using data from the Taiwan Stock Exchange, provide similar evidence.

Based on the theoretical predictions and empirical findings herein summarised we expect that different trading participant types would contribute heterogeneously to price discovery, though in absence of theoretical arguments and comparable empirical findings we are agnostic as to whether their heterogeneous contributions to price discovery would depend on liquidity.

3 Data

In this section we start by providing institutional details about the SET market. We will then describe the database we will use as well as the main manipulations we performed in order to use the data for our estimations. The section concludes providing descriptive statistics of the data.

3.1 The Stock Exchange of Thailand: Institutional details

The Stock Exchange of Thailand (SET) is an advanced emerging market in the South East Asian region, according to the FTSE Russell classification.[10] As of December 2019, it had over 726 listed companies for a market capitalisation of nearly US\$570 billion, which placed SET in the tenth position in terms of capitalisation in the Asian region. SET is a very liquid market: over the course of 2019, the monthly average Electronic Order Book (EOB) stock trading on SET was over nearly US\$30 billion, for a share turnover velocity equal to 63% (Source: World Federation of Exchanges). Much of the liquidity on SET is due to actively trading retail investors, who dominate the market (Pavabutr and Sirodom, 2010; Rhee and Wang, 2009; WFE and Oliver Wyman, 2016; Richards, 2005): individual trading activity accounted for 54% of EOB value traded and 78% of EOB number of trades as of December 2017 (source: SET). Trading activity on SET shows no fragmentation and is characterised by a relatively contained amount of algorithmic trading (Likitapiwat, 2016).[11] According to SET, algo-trading was roughly 22% of their trading activity over the course of 2018.

SET is a continuous auction limit order market. Trading can happen through the order book ("Automatic Matching") or via negotiated deals ("Put Through"), however trading through the limit order book represented roughly 96% of the trading activity in 2018 (source: SET), a sign that equity trading on SET is characterised by high levels of transparency. In addition, SET does not have designated dealers in any listed companies. Most orders happen through the matching of limit orders, although market orders are also commonly used on the market. Any unmatched quantity arising from an executed market order is cancelled from the trading system, however SET caters for order types that allow unmatched quantities to "walk the order book" (Hasbrouck, 2007). Hidden orders are not allowed, which is another desirable feature as hidden liquidity has confounding effects on price formation (Hasbrouck, 2010).

As per pricing of stocks, SET applies a sliding scale tick size, with the tick size increasing in the size of the stock. This system is aimed at encouraging retail participation, as retail investors prefer trading in smaller stocks, while institutional investors' trading is more concentrated in bigger companies. SET has circuit breakers (Pavabutr and Sirodom, 2010; Alderighi et al., 2021), although since they were implemented, they were triggered only 3 times before the sample period (on December 19th, 2006; on October 10th and 27th, 2008) (source: SET).[12] Therefore, volatility interruptions do not represent a source of noise or disruption in the trading data (Engle and Patton, 2004; Riordan and Storkenmaier, 2012).

3.2 The database

We collected intraday limit order book and deals information for a sample of SET stocks for April 2018. To select our stocks, we used a stratified sampling procedure (Brogaard et al., 2014; Engle and Patton, 2004) that allows us to retain information on the whole distribution of the companies listed on the exchange (Chung and Hrazdil, 2010). To identify our strata, we segmented the market into large, medium, small

---

[10]For additional institutional information on the Stock Exchange of Thailand see Pavabutr and Sirodom (2010); Phansatan et al. (2012); Likitapiwat (2016)

[11]Both fragmentation and the presence of high frequency traders are potentially confounding factors. Fragmentation can be a confounding factor as when liquidity is fragmented price discovery might take place somewhere else (Frijns et al., 2010, 2015, 2018; Hasbrouck, 1995; Otsubo, 2014; Pascual et al., 2006). Based on this argument, Benos and Sagade (2016) study stocks listed in the UK because they have lower levels of fragmentation than, say, US stocks. The presence of high frequency trading is a confounding factor as HFT is shown to have its own influence on both liquidity (Hendershott et al., 2011) and price discovery (Brogaard et al., 2014; Hasbrouck and Saar, 2013; Riordan and Storkenmaier, 2012). In a market where HFT is a contained proportion of the total, it is something that can be controlled for in a regression specification. However, when HFT activity is preponderant, high frequency traders would necessarily become the main focus of the study.

[12]Circuit breakers were triggered three additional times in March 2020 during the period of heightened stock market volatility due to the COVID-19 pandemic. See Alderighi et al. (2021)

and micro companies using a WFE categorisation.[13] Figure 1 shows the number of companies listed on SET broken down by the four identified categories, while Figure 2 shows median bid/ask spreads for the very same categories. From Figure 2 it is immediately evident that trading activity on SET is spread across the entirety of the market, with relatively contained spreads for small and micro companies (around 80 bps for small companies, and around 100 bps for micro companies). This is a result of the high levels of retail participation in the market, as well as of the application of the tick-size sliding scale (Pavabutr and Sirodom, 2010).

Figure 1: Number of companies listed on SET
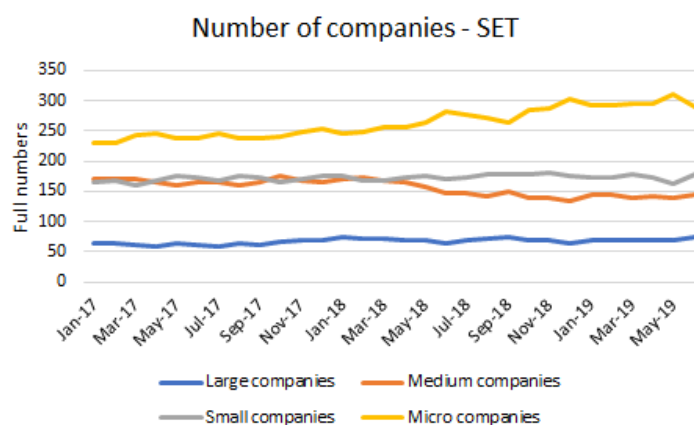
Source: WFE Database

Figure 2: Median Simple Spread (MSS) - SET
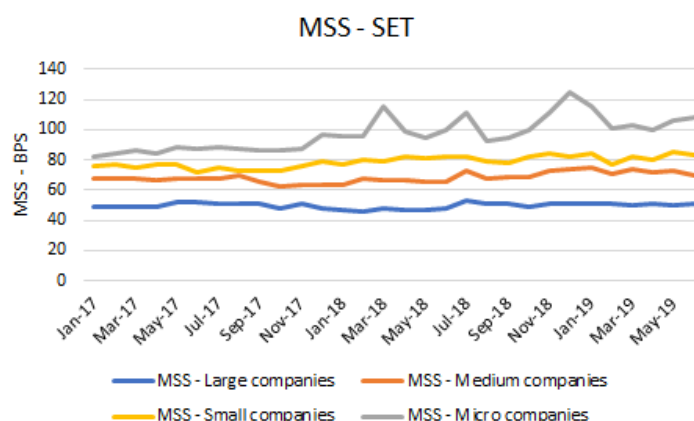
Source: WFE. Median Simple Spread (MSS) is a spread-measure of market liquidity reporting the median quoted spread broken down by company size. For more details on MSS see the WFE definition manual.

SET provided the authors with 25 randomly drawn stocks from each of the above categories, for a total of 100 stocks. Of those, 17 were companies listed on mai, their SME board. Our stock selection

---

[13]This segmentation is used for WFE Median Simple Spread data, a market-level measure of liquidity. The categorisation is as follows: Large market cap segment: market cap $\geq$ US\$1.3 bn: Mid market cap segment: US\$1.3 bn $\geq$ market cap $\geq$ US\$200 m; Small market cap segment: US\$ 200 m $\geq$ market cap $\geq$ US\$ 65 m; Micro market cap segment: market cap $\leq$ US\$ 65 m. For more details on Median Simple Spread data, see the WFE definition manual.

procedure is equivalent in spirit to that of Engle and Patton (2004) , who study the impact of trades on quote prices for frequently and unfrequently traded stocks, thus performing a comparable exercise. No stocks were affected by stock splits over the sample period. This is important, as stock splits are relatively common in Thailand (Pavabutr and Sirodom, 2010) and are a potentially confounding factor.

Intraday orders and trades data were provided at a hundred-of-a-second frequency. The database contains only orders submitted to and matched through the limit order book ("automatching"); no bilateral deals were included in the database. SET provided information on all orders and trades that took place during their trading hours, including opening and closing auctions.

3.3   Data preparation

For the purposes of this research, it is important to determine whether trades are buyer or seller initiated. Absent this information, the literature typically applies the Lee-Ready algorithm (Lee and Ready, 1991). Despite its widespread use, studies have found that the Lee-Ready algorithm erroneously allocates trades in a non-negligible percentage of cases (Aitken and Frino, 1996; Ellis et al., 2000; Finucane, 2000; Odders-White, 2000; Theissen, 2001), especially during quote revision (Finucane, 2000). Therefore, as we have information on each trade and quote event for each stock/day in our sample, we decided to apply a chronological criterion to determine whether a trade is a buy or a sell, following previous literature (Aitken and Frino, 1996; Odders-White, 2000). To allocate trades, we first identified the pair of orders associated with each trade. Then, we identified the more recent order between the two. We considered a trade buyer (seller) initiated if the buy (sell) order related to the trade is inserted later than its matching sell (buy) order. As customary in the literature, trade direction D_{it} is equal to +1 if the order is buyer-initiated, and equal to -1 if the order is seller-initiated.

One limitation of the SET database is that it does not indicate the best bid and ask prices for each timestamp (i.e. the "top of the book"), and henceforth this information needs to be inferred. To reconstruct the top of the book, we implemented the following procedure. To start with, for each stock/day we looped over every single order insertion, and calculated the best bid (ask) price as the maximum (minimum) price for an unmatched limit order till that particular event (step one). This simple procedure leads however to negative spreads, as all inserted buy (sell) limit orders with a price higher (lower) than the best ask (bid) would feature as best bid (ask) prices. We therefore considered all inserted buy (sell) limit orders with a price higher (lower) than the best ask (bid) as immediately executed orders (Hasbrouck and Saar, 2009; Khanna et al., 2009; Menkhoff and Schmeling, 2010), thus removing their posted prices from the best bid/ask prices (step two). We iterated steps one and two till we obtained no negative spreads for each stock/day.

We used best bid/ask quotes to calculate several spread-indicators routinely used in the literature, namely:

- the quoted spread
- the quote midpoint
- the effective spread
- the realised spread
- the price impact of trades over a five-minute horizon

For the calculation of these indicators we used the formulas reported in Riordan and Storkenmaier (2012).

As customary in the literature, we restricted our econometric analyses to orders and trades during regular trading hours, and we excluded the first and the last five minutes of trading activity.[14]

---

[14]SET has a lunch break, which we ignore for simplicity. We perform our main analyses considering introducing pre- and

Finally, and as indicated in Hasbrouck (1991b) and remarked in Benos and Sagade (2016), it is important to choose whether to analyse data in 'clock time' or 'event time'. Our units of analysis vary depending on the estimations we perform, though we never analyse the data in clock-time. In each of the empirical results sections we clearly describe how our data is arranged.

3.4 Description of the data: market participation, liquidity and price formation

We herein provide summary statistics. In particular, we show how liquidity is distributed in the cross-section of stocks in the sample.[15]

As mentioned above (see Section 3.1), SET is a retail-dominated market. Participation of individuals, however, is not uniform across stock sizes. While trading in small and smaller mid-cap stocks is overwhelmingly dominated by retail investors, trading in larger mid-cap stocks is more balanced between the three investors' categories, while trading in large stocks is mostly performed by foreign and domestic institutions. Figure 3 below describes market participation (herein understood as the investor category that initiated the trade) broken down per stock size over the sample period.

Figure 3: Market participation per quartiles of market capitalisation
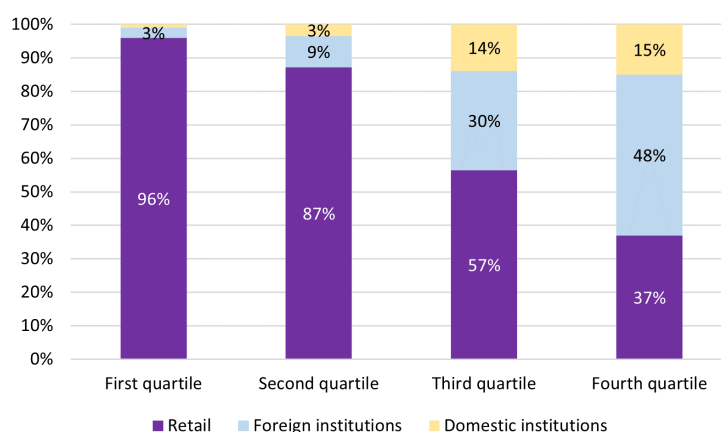
Source: SET

Similar to other Asian exchanges (Barber et al., 2009), individual traders on SET are active, leading to a reasonably high levels of trading activity in small and medium cap stocks. Table 1 shows average daily trading activity over the sample period, broken down by quartiles of market capitalisation, together with additional descriptive statistics.

As evident from the Table 1, large cap stocks see disproportionately more daily Baht turnover than small cap stocks.[16] The average daily trading value for stocks in the fourth quartile of market capitalisation is 1.25 billion Baht, more than 30 times the trading value for stocks in the first quartile, amounting to 32.24 million Baht. This result is however driven by the higher prices of large cap stocks, as small

---

post-lunch break dummies as a robustness check. The results (available upon request) are not affected by the introduction of these variables.

[15]In this section we segment by market capitalisation, however the analyses in the rest of the paper are segmented by liquidity. The rationale is to segment using a different variable from the one we aim to describe.

[16]To allow the reader to contextualise the Thai Baht numbers, note that 100 Baht are roughly US$3.22. For example, 50,000 Baht would be roughly US$1,500.

Table 1: Liquidity indicators per quartiles of market capitalization (averages)

	(1) Whole sample	(2) Quartile 1	(3) Quartile 2	(4) Quartile 3	(5) Quartile 4
Market capitalisation (Baht millions)	199.29	1.54	4.98	28.08	302.44
Trade value (Baht thousands)	195.09	27.27	43.19	80.24	269.8
Daily trading volume (millions)	44.57	18.83	12.66	31.26	55.43
Daily trading value (Baht millions)	841.17	32.24	36.08	164.36	1,255.57
Price of a trade (Baht units)	45.45	2.75	7.99	15.68	64.52
Average value of a trade (Baht thousands)	195.11	27.25	43.2	80.23	269.83
Turnover velocity (Units)	0.15	0.43	0.17	0.13	0.13
Effective spread (bps)	29.75	60.06	37.35	33.11	25.03
Realised spread (bps)	6.66	23.46	6.91	8.18	4.62
Price impact (bps)	23.1	36.97	30.44	24.93	20.39
Observations	2,047,774	110,930	151,631	483,843	1,301,370

and mid-cap stocks see overall comparable trading volumes to large-cap stocks. For example, the average daily volume traded for stocks in the fourth quartile of market capitalisation is 55.43 million units, only 2.9 times larger than the average daily volume for stocks in the first quartile, amounting to 18.83 million units. In other words, investors seem to engage in a conspicuous amount of trades in smaller-cap securities. In addition, it must be noted that the (total) Baht value traded in small stocks is large as compared to the size of the companies: turnover velocity (the ratio between value traded in monetary terms, and market capitalisation, a commonly used measure of liquidity) is indeed higher for small and mid-cap stocks than it is for large stocks. For example, the average turnover velocity for stocks in the first quartile of market capitalisation is 43%, as opposed to 13% for stocks in the fourth quartile.

Turnover velocity alone would give a misleading picture of how liquidity is distributed. For example, looking at the effective spread (a commonly used indicator of the adverse selection cost paid by liquidity demanders), it is immediately evident that it is monotonically decreasing in stock size. This implies that potentially informed traders demanding for immediate execution face much higher hidden costs when trading in smaller cap stocks than when trading in larger cap stocks. Consistently, trades in small cap stocks have higher price impacts than trades in large cap stocks, hinting that smaller cap stocks are characterised by lower price efficiency: the average price impact of trades in the first quartile of market capitalisation is almost 37 bps, and the corresponding figure for stocks in the fourth quartile of market capitalisation is roughly 20 bps. As effective spreads are higher than price impacts for all company sizes, liquidity suppliers enjoy positive profits (the realised spread) across all the sample, with the average realised spread being monotonically decreasing in market size, another hint that smaller cap stocks are characterised by higher frictions.

While small cap stocks appear to be relatively more illiquid when looking at spread measures, it must be noted that spreads are overall comparable across different stock sizes, thanks to the already emphasised high levels of trading activity in small cap stocks and to a tick size sliding scale proportional to stock prices, a market structure feature that keeps spreads uniform across different stock price levels. These statistics suggest that active retail traders are important to keep liquidity relatively high in Thai small cap stocks.

4 Empirical models and results

4.1 Returns' predictability of order imbalances

We will now study the relation between order imbalances and price formation for different trading participants, and how this relation is affected by the underlying levels of stock liquidity. Order imbalances are theoretically and empirically associated with higher stock returns, as documented by copious literature on the topic (Hanke and Weigerding, 2015). In the theoretical framework proposed by Huang and Stoll (1994), a positive relation between imbalances and stock returns is due to information being impounded into stock prices. However, and as noted in Boehmer and Wu (2008), in principle this relation could be also caused by inventory adjustments performed by uninformed liquidity suppliers. As such, whether the relation between imbalances and returns can be attributed to information impounding is an empirical question. Research has found contrasting results. Chordia et al. (2008) (see also Aktas et al. (2008); Chung and Hrazdil (2010)) show that trade imbalances are a stronger predictor of five-minute returns when liquidity is low, and finding they associate with a positive relation between liquidity and price efficiency. On the other hand, Lee et al. (2004), focusing their attention on the imbalance of aggressive immediately executable orders orders, attribute a positive relation between order imbalances and stock returns to information impounding. In this paper, we follow Lee et al. (2004), and consider the imbalance of aggressive orders as informative, based on the underlying assumption that informed traders would demand for immediate execution to cash-in their information advantage.[17]

More in detail, to provide evidence on the relation between liquidity and price efficiency, we firstly estimate it for different levels of stock liquidity, identified through percentiles of the effective spread. We then estimate whose investors' category order imbalances (i.e. retail, domestic institutions, foreign institutions) contribute the most to returns' predictability for different levels of liquidity.

We follow Chordia et al. (2008) and Cont et al. (2014) and for each stock/trade in our sample calculate our dependent variable, returns, as the log difference of quote midpoints over a five minute interval. We calculate our main regressor, imbalance of aggressive orders, as below:

$$\text{Imbalance}_{it} = \frac{\sum_t^{t+5} \text{Volume of buy marketable orders}_{it} - \sum_t^{t+5} \text{Volume of sell marketable orders}_{it}}{\sum_t^{t+5} \text{Volume of buy marketable orders}_{it} + \sum_t^{t+5} \text{Volume of sell marketable orders}_{it}}, \tag{1}$$

where marketable orders are either market orders, or limit orders the price of which is equal or better than the current best bid/ask price. Volumes are calculated over 5-minutes intervals. We estimate the following linear specification:

$$E\left[\text{Returns}_{it} \mid X\right] = \beta_0 + \beta_1 \text{Imbalance}_{it-1} + \text{Day fixed effects} + \text{Stock fixed effects} + \gamma' \text{Controls} \tag{2}$$

Where i refers to the stock, and t refers to subsequent five-minute intervals. We include a broad set of controls to limit omitted variable bias, namely: the lagged number of trades over five-minutes intervals, as past literature found that trade intensity has a positive impact on prices (Dufour and Engle, 2000); the lagged number of order cancellations on order insertions over five minutes intervals, to control for the presence of high frequency traders[18]; lagged five-minutes-interval volatility of stock returns; company

---

[17]See also the theoretical framework proposed in Easley et al. (2002), and empirically tested by Vega (2006), showing that better information is reflected into order imbalances, and that abnormal imbalances have an effect on returns.

[18]It is commonly believed that high frequency traders cancel orders at a way higher rate than non-HFT trading participants, see for example SEC (2010). Subrahmanyam and Zheng (2016), using Nasdaq data, show however that HFT traders are not more likely to cancel orders than other trading participants. In our sample we find that the cancellation rate of orders is the lowest for retail investors (19%) while it is just above 50% for foreign institutions and nearly 30% for domestic institutions. To the extent that HFTs are more likely to be large, foreign participants (given that Thailand is an emerging market), we believe our proxy is valid in capturing HFT activity, though more research is needed on this topic.

size, as measured by end of previous day market capitalisation; the number of research reports and the number of corporate events for each stock/day, to control for company-level public information arrival (Riordan et al., 2013). We saturate the model by introducing stock fixed effects (to capture time-invariant individual heterogeneity) and day fixed effects (to capture market-wide time series movements, Hasbrouck and Seppi (2001)).[19] Inference is performed using standard errors clustered by stock.

We estimate Equation (2) over the whole sample and for different levels of liquidity, identified through percentiles of the effective spread. We consider six segments:

ES(< 10)= effective spreads smaller than the 10th percentile (the most liquid);

ES($10, 25$)= effective spreads between the 10th and the 25th percentiles;

ES($25, 50$)= effective spreads between the 25th and the 50th percentiles;

ES($50, 75$)= effective spreads between the 50th and the 75th percentiles;

ES($75, 90$)= effective spreads between the 75th and the 90th percentiles;

ES(> 90) = effective spread higher than 90th percentile (the least liquid).

The results are summarised in Figure 4 and briefly commented on below.[20]

Figure 4: Regression coefficients for imbalance of aggressive orders (Equation 2). The coefficient represents the basis points increase in five-minute returns associated with a 100% increase in the imbalance of aggressive orders.
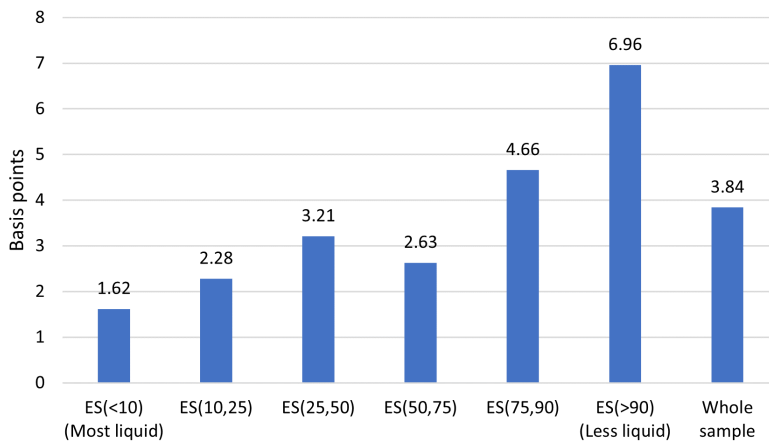
We firstly look at the results over the whole sample. Our main regressor, (lagged) imbalance of aggressive orders, has a strong positive correlation with five-minute returns (p-value: 0.00), hinting that trade imbalances are associated with return predictability in the short-run. To interpret the coefficient, imagine observing no trade imbalances for a particular stock (i.e. buys exactly equal sells) in a given five-minute interval. Then suppose that in the following five-minute interval all trades in that particular stock are buyer-initiated, making the regressor jump from zero to 100%. For the average stock, such unit

---

[19]We estimated the models by explicitly introducing a set of common factors (past-day returns on the SET broad market index; past-day returns on the SET SME Index; past-day returns on the MSCI EM index; past-day ten-year-yields on government bonds) instead of day fixed effects, and the results are quantitatively and qualitatively comparable.

[20]As a robustness check, we estimated these regression models using order imbalances as more commonly defined (difference between buy and sell trades).

change (100 percentage points) would be associated with an increase in five-minute returns equal to 3.8 basis points. A trade imbalance increase equal to a standard deviation (0.69) would imply higher returns for 2.6 bps. This magnitude is sizeable, as in our sample five-minute returns are in the range of basis points (between -10 bps and 10 bps) in over 43% of the cases, have a zero median, and a mean equal to -1.8 bps.

The finding that order imbalances are associated with short-run returns' predictability is not new to the literature. In the seminal Chordia et al. (2008), focusing on the NYSE, the authors report that a unit change in trade imbalances would be associated with a 4.2 bps in five-minutes return, a magnitude comparable to the one reported in this paper. Focusing on Asian markets, Narayan et al. (2015) show that order imbalances are associated with returns' predictability in the Chinese stock markets. Evidence that order imbalances were a significant predictor of stock returns in China during the 2007-2008 crisis is provided in Lao et al. (2018). Yamamoto (2012) reports significant order imbalance returns' predictability for Nikkei stocks. Interestingly, Visaltanachoti and Luo (2009) perform an analysis similar to ours using SET data, though at a lower frequency, and show that order imbalances are not a significant predictor of returns over a 30-minute horizon.

Importantly, in the graph we also show the results segmented by percentiles of the effective spread. Comparing the estimation results in the most liquid segment (i.e. when effective spreads are in the bottom ten percent of the distribution) with those in the least liquid segment (i.e. when effective spreads are in the top ten percent of the distribution) it is evident that order imbalances have a much stronger impact on stock prices when liquidity is lower, consistently with the literature (Chordia et al., 2008; Riordan and Storkenmaier, 2012). A unit change in the imbalance of aggressive orders is associated with a 6.96 bps increase in returns in the low-liquidity segment; the corresponding figure for the high-liquidity segment is more than four times smaller, 1.62 bps.

We move on to investigate which market participants (retail, domestic institution or foreign institution) are associated with higher returns predictability, and whether this is dependent on liquidity. We estimate the following equation:

$$
\begin{aligned}
E\left[\text{Returns}_{it} \mid X\right] \quad = \quad & \beta_0 + \beta_1 \text{Imbalance Retail}_{it-1} + \beta_2 \text{ Imbalance Domestic Inst}_{it-1} + \\
& + \quad \beta_3 \text{Imbalance Foreign Inst}_{it-1} + \text{Day fixed effects} + \\
& + \quad \text{Stock fixed effects} + \gamma' \text{Controls}
\end{aligned}
\tag{3}
$$

Where the set of controls is the same as the one included in Equation 2. Results are reported in Table 2. We first look at the results estimated using the whole sample (Column 1). To start with, we note that imbalances of aggressive orders are a predictor of returns for all trading participants. Retail investors, however, are characterised by the most sizeable coefficient: a 100% increase in retail investors' imbalances is associated with a 4.4 bps increase in five-minute returns, while a 100% increase in foreign institutional (domestic institutional) investors' trade imbalances is associated with a 1.59 (1.05) bps increase in five-minute returns. Over the entire sample, individuals are the investor category whose order imbalances affect returns the most, suggesting that retail investors possess a superior ability to interpret market trends and publicly available information (Engelberg et al., 2012; Vega, 2006).[21] Order imbalances by foreign and domestic institutions are characterised by a smaller, although not-negligible, degree of returns' predictability.

When segmenting by liquidity, we find that retail investors' trade imbalances are the most sizeable and (overall) significant positive predictors of returns among the three investor categories for all liquidity

---

[21]This is plausible as retail traders on SET are, as documented, very active, and therefore also likely to be experienced. Trading experience is associated with a number of positive outcomes, including the tendency of showing less remarked behavioural biases (Korniotis and Kumar, 2011). In addition, it must be noted that SET has historically supported their individual investors' base with a wide range of financial literacy programs and other initiatives, thus investing considerably in building up the trading skills of their investors (WFE, 2017).

Table 2: Returns predictability of trade imbalances broken down by liquidity thresholds. Fixed effects linear regression models. Dependent variable: stock returns over five minutes intervals. Clustered Standard errors in parentheses.$*p \leq 0.1, **p \leq 0.05, ***p \leq 0.01$.

	(1) Full sample	(2) $ES(<10)$	(3) $ES(10,25)$	(4) $ES(25,50)$	(5) $ES(50,75)$	(6) $ES(75,90)$	(7) $ES(>90)$
Trade imbalance:							
Retail	4.44***	0.53	2.05*	2.82***	3.76***	7.25***	8.54***
	(0.55)	(0.66)	(1.06)	(0.71)	(0.71)	(1.29)	(1.32)
Foreign	1.59***	1.87**	2.46	1.96**	0.86	2.70	1.10
	(0.53)	(0.79)	(2.01)	(0.83)	(0.83)	(2.17)	(1.28)
Domestic	1.05*	0.24	1.80	1.73*	1.09	3.54**	-2.052
	(0.55)	(0.66)	(1.09)	(0.87)	(0.99)	(1.54)	(1.87)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared (within)	0.007	0.030	0.017	0.013	0.009	0.015	0.010
R-squared (between)	0.003	0.300	0.000	0.006	0.030	0.003	0.053
Adjusted R-squared	0.007	0.023	0.012	0.011	0.008	0.012	0.008
Number of stocks	97	9	16	41	54	31	50
Average	617.351	399.667	341.750	314.732	340.778	311.903	196.860
Observations	59883	3597	5468	12904	18402	9669	9843

segments apart from the most liquid one. In other words, the market tends to move in the direction of retail investors' aggressive buy and sell orders (note that both trade imbalances and returns can take both positive and negative values), hinting that their order submissions/trading activity convey information. When liquidity is the highest (effective spreads lower than the 10th percentile), however, foreign investors are the category whose imbalances are associated with the highest influence on prices: a 100% increase in retail investors' trade imbalances is associated with an increase in five-minute returns equal to 1.9 bps. This coefficient is significant at the 5% level. Retail and domestic institutions' coefficients are less sizeable and statistically insignificant.

As liquidity goes down, we find that retail investors' imbalances are associated with an increasingly sizeable and statistically significant influence on five-minute returns. We don't find the same result for foreign and domestic institutions (excluding the segment $ES(25, 50)$): returns' predictability is largely driven by retail investors' imbalances. We attribute this finding to retail investors' imbalances conveying relevant information to the market, which is incorporated in stock prices, though we caveat that in lower liquidity segments this result might be expected given that the vast majority of trades happens indeed between retail investors. Domestic institutions' show a sizeable and statistically significant coefficient in segment $ES(75, 90)$, where a 100% increase in domestic institutions trade imbalances is associated with an increase in five-minute returns equal to 3.5 bps. In section 4.2 we show that trades between domestic and foreign institutions are the most informative in the segment $ES(75, 90)$, suggesting that some informed trading between domestic and foreign institutions might happen on stocks that fell in that liquidity segment.

As a robustness check we test whether the coefficients in the high liquidity portfolio ($ES(<10)$) are statistically different from the coefficients in the low liquidity portfolios ($ES(>90)$). We find that only the retail trade imbalance coefficients are statistically different from each other at the canonical 5% level (Chi-squared test value: 8.45. p-value: 0.0037). For the foreign trade imbalance coefficients the null cannot be rejected (Chi-squared test value: 2.35. p-value: 0.125), and for the domestic institution trade imbalance coefficients neither (Chi-squared test value: 2.67. p-value: 0.102).

4.2   Properties of the price impact function

In this section we provide evidence on the relation between liquidity and price discovery by studying the properties of the price impact function. We follow the academic literature (and in particular, Riordan and Storkenmaier (2012)), as well as industry practice, and define the price impact function as the difference between effective and realised spreads:

$$
\begin{aligned}
\text{Price Impact}_{it} &= \text{Effective Spread}_{it} - \text{Realised Spread}_{it} \\
&= D_{it}\left(\frac{\text{Trade Price}_{it} - \text{Midpoint}_{it}}{\text{Midpoint}_{it}}\right) - D_{it}\left(\frac{\text{Trade Price}_{it} - \text{Midpoint}_{it+5}}{\text{Midpoint}_{it}}\right) \\
&= D_{it}\left(\frac{\text{Midpoint}_{it+5} - \text{Midpoint}_{it}}{\text{Midpoint}_{it}}\right)
\end{aligned}
\tag{4}
$$

Here D_{it} is trade direction (equal to 1 for a buy, and to -1 for a sell). The price impact function measures the information content of a trade, and can be seen as the adverse selection cost paid by liquidity demanders (or as the revenue made by liquidity suppliers). Wider effective spreads imply that liquidity demanders need to pay a higher price to obtain immediate execution. As a consequence, they will demand liquidity when (they believe) the information they possess is truly valuable, i.e. when they expect the price impact of their trade to be high enough to cover the cost represented by the effective spread. Therefore, we expect price impact to be increasing in the effective spread measure. Table 3 below confirms our expectations:

Table 3: Effective spread, realised spread and price impact function per liquidity segments. All indicators are expressed in basis points

	(1) $ES(< 10)$	(2) $ES(10, 25)$	(3) $ES(25, 50)$	(4) $ES(50, 75)$	(5) $ES(75, 90)$	(6) $ES(> 90)$
Effective spread	13.78	18.79	24.35	30.96	37.27	61.57
Realised spread	2.69	2.56	4.25	8.33	5.10	20.90
Price impact	11.09	16.22	20.10	22.63	32.11	40.84
Observations	205,154	307,081	510,861	515,008	306,303	203,367

It is worth noting that realised spreads are on average positive in all segments, meaning that liquidity providers (and not liquidity demanders) are the ones profiting (on average) from trades. Across the whole sample, realised spreads are negative in roughly 38% of the cases, suggesting that liquidity demanders profit from trades in a sensibly lower number of cases. This evidence is consistent with Hasbrouck (1991a).

This simple analysis, although illustrative, has the drawback of being very endogenous, as we segmented the sample using one of the measures analysed in the table. To show more compelling evidence on the relation between liquidity and price discovery, we exploit a well-documented relation in the literature: the price impact function is a monotonically increasing concave function of trade size, as found in Hasbrouck (1991a) and noted in more recent literature (Cont et al., 2014; Dufour and Engle, 2000; Kempf and Korn, 1999; Potters and Bouchaud, 2003; Lillo et al., 2003; Weber and Rosenow, 2005).[22]

We estimate the model:

$$
\begin{aligned}
E\left[\text{Price Impact}_{it} \mid X\right] = \;&\beta_0 + \beta_1 \ln(\text{Trade size}_{it}) + \text{Day fixed effects} + \\
&+ \text{Stock fixed effects} + \gamma'\text{Controls}
\end{aligned}
\tag{5}
$$

---

[22]For additional analyses and comments on the concavity of the price impact function we refer to the Supplementary Materials, Section 3.

Where i refers to the stock, and t identifies a trade event (we estimate the model in tick-time).[23,24] As an indicator of trade size we use the Thai Baht value of a trade. The logarithmic function caters for the concave relation between trade size and price impact. Similarly to the results displayed in Table 2, we segment the sample for different levels of liquidity. We introduce the same set of controls as in Equation 2, as well as day and stock fixed effects. Inference is performed on a clustered variance-covariance matrix (by stock). Our expectation is that trades would have a more sizeable price impact when liquidity is low (i.e. spreads are high) as in Lim and Coggins (2005).

It should be noted that in a linear-log model the regression coefficient (β_1 in our case) represents the absolute change in the dependent variable triggered by a percentage change in the regressors. This interpretation is however not very insightful in our case, as we are interested in understanding the different impacts of small and large trades on prices. To obtain estimates that are consistent with this interpretation, we calculate marginal effects. Unlike in a linear model, in a logarithmic specification the marginal impact of the independent variable (the size of trades) on the dependent variable (the stock) price is not constant in trade size, but decreasing, as is evident by calculating the partial derivative of the independent variable relative to the dependent one:

$$\frac{\partial \text{Price impact}}{\partial \text{Trade size}} = \frac{\beta}{\text{Trade size}} \tag{6}$$

Therefore, we report marginal effects, i.e. the price impact per unit of trade. To facilitate the interpretation of our results we also report the distribution of trade size for the different liquidity segments:

Table 4: Distribution of trade value per liquidity segment. Trade size expressed in thousands of Bahts.

Trade value	(1) 1st perc.	(2) 10th perc.	(3) 25th perc.	(4) 50th perc.	(5) 75^{ch} pert.	(6) 90th perc.	(7) 99th perc.
ES(< 10)	7.97	9.15	17.45	47.38	166.25	538.65	4252.50
ES($10, 25$)	2.37	7.10	14.00	52.60	179.38	687.69	6100.00
ES($25, 50$)	0.37	2.17	8.52	29.50	108.12	384.16	3494.40
ES($50, 75$)	0.17	1.42	3.95	17.13	65.70	195.00	1987.50
ES($75, 90$)	0.25	1.45	4.23	18.56	70.50	274.70	2469.15
ES(> 90)	0.10	0.61	2.53	11.30	46.00	128.32	1170.00
All	0.24	2	7.15	25.65	96.4	343.65	3300.00

Evidently, trade size is increasing in liquidity, i.e. tighter spreads (typically found on larger stocks) are associated with larger trades. Trade size spans from hundreds of Bahts (especially when spreads are high – see Table 4 Column (1)), to few millions of Bahts (Table 4, Column (7)). Results of the estimation of Equation 5 are summarised in Figure 5 and briefly commented on below.

---

[23]It is debated what should be considered as a 'single trade'. The reason is that large marketable orders might execute against several smaller resting limit orders. In our setting we consider each match as a single trade. Lillo et al. (2002) however collapse all sequential trades originating from the same order as a single, larger order. As a robustness check, we perform the analyses of this and the next section using the categorisation proposed by Lillo et al. (2002). The results are overall comparable.

[24]Lim and Coggins (2005) estimate different models for buyer- and seller-initiated trades, assessing the impact of buys on best ask prices, and the impact of sells on best bid prices. Our price impact function differs in being positive when prices move in the same direction of trades (be they buys or sells) and in being negative when prices move against trades (which happens in roughly 13% of the sample). As such, it is appropriate to run our regressions without differentiating between buys and sells.

Figure 5: Price impact of increasing the trade size by 1000 Baht. Marginal effects based on the estimation of Equation 5, measuring the price impact of increasing the trade size by 1000 Baht when trade size is equal to the median for each category.
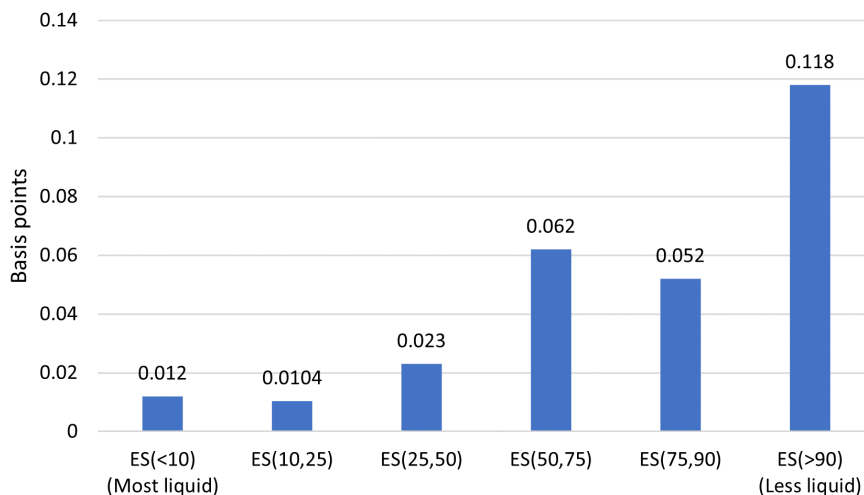
Our results suggest that the marginal price impact of trade size is higher when liquidity is low, consistently with the literature (Lim and Coggins, 2005). The marginal effect of increasing the trade size by 1000 Baht when the trade size is at the median for that segment is over ten times larger when estimated in the segment ES($>$ 90) than it is when estimated in the segment ES($<$ 10). This implies that when liquidity is low, even small trades (in absolute terms) can have a non-negligible price impact. In other words, when liquidity is low, small trades also convey relevant information. For example, in the segment ES($>$ 90) a 5,000 Baht trade increase can lead to a 1 bps price impact increase (when the trade size is equal to 25,650 Baht). To have the same price impact in the high liquidity segment ES($<$ 10), one should observe a 12,500 Baht trade increase in trade size (when the trade size is equal to 25,650 Baht).

We move on to explore whether trades convey different information depending on which market participants trade with each other. Our database flags the investor category for each sides of each trade, and hence allows us to identify which traders contribute the most to price formation in our sample. To perform this analysis, we categorise each trade into one of the six possible traders' combinations: retail to retail (R-to-R); foreign institution to foreign institution (F-to-F); domestic institution to domestic institution (D-to-D); retail to foreign institution (R-to-F); retail to domestic institution (R-to-D); domestic institution to foreign institution (D-to-F).[25] Table 5 provides a breakdown.

---

[25] Note that the main models do not identify the buyer or the seller and we ignore which side triggered the trade. We run a differentiated model for buys and sells as a robustness check, and briefly comment on it later in the section.

Table 5: *Trades by investors' category. Small cap: Market capitalisation ≤ 25th percentile. Large cap: Market capitalisation ≥ 75th percentile.*

| | (1) | | (2) | | (3) | |
| | Whole sample | | Small cap | | Large cap | |
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
R-to-F	646,455	31.66%	7,467	6.74%	478,057	36.87%
R-to-R	639,324	31.31%	101,658	91.72%	234,920	18.12%
R-to-D	293,844	14.39%	1,494	1.35%	201,464	15.54%
F-to-F	241,185	11.81%	157	0.14%	208,680	16.10%
D-to-F	190,380	9.32%	60	0.05%	150,951	11.64%
D-to-D	30,809	1.51%	0	0.00%	22,400	1.73%
Total	2,041,997.00	100.00%	110,836.00	100.00%	1,296,472	100.00%

As evident from the table, retail investors take at least one side in the vast majority of trades (77.36%) (Column (1)). As already noted in Section 3.4, Figure 1, market participation is not uniform across the sample, with retail investors essentially dominating the market in small-cap stocks (see Column (2)), and foreign and domestic institutions trading the most in large-cap stocks (see Column (3)). Yet, it must be noted that even in the large cap segment retail investors take at least one side in most trades: 70.53%. Therefore, while they initiate trades in only 15% of the cases (see Figure 1 in Section 3.4) they still have an important role as liquidity suppliers in large cap stocks.

To explore how the price impact function relates to trading between different market participants, we estimate the following regression models:

$$
\begin{aligned}
E\left[\text{Price impact}_{it} \mid X\right] &= \beta_0 + \sum_{j=1}^{6} \beta_j ln(\text{Trade Size}_{jit}) D_{jit} + \sum_{j=1}^{6} \alpha_j D_{jit} + \\
&+ \text{Stock fixed effects} + \text{Day fixed effects} + \gamma'\text{Controls}
\end{aligned}
\tag{7}
$$

Essentially, we allow the natural logarithm of trade size to interact with the trading participant dummies, and estimate these models for different levels of liquidity. This setting differs from and represents an enhancement on the existing literature which typically inferred the information advantage of different investor types by assessing their portfolios' performance. The model is estimated in tick-time, using the same set of control as in Equation 2 as well as stock and day fixed effects. Inference is performed on a variance-covariance matrix clustered by stock. Table 6 displays the regression results segmented by liquidity. The numbers reported are marginal effects at the median[26] for each trading participant category/liquidity segment.

---

[26]We calculated marginal effects at the sample means and the results (available upon request) are qualitatively similar.

Table 6: Price impact of trades by different investors' categories segmented by liquidity. Dependent variable: Price impact of a trade.

	(1) Full sample	(2) $ES(<10)$	(3) $ES(10, 25)$	(4) $ES(25, 50)$	(5) $ES(50, 75)$	(6) $ES(75, 90)$	(7) $ES(>90)$
Ln (Trade value):							
R-to-R	0.051***	0.00903***	0.019***	0.0308***	0.056***	0.046***	0.16***
	(0.140)	(0.071)	(0.179)	(0.134)	(0.185)	(0.212)	(0.564)
F-to-F	0.034***	0.0066***	0.0078***	0.017***	0.092***	0.077***	0.15***
	(0.248)	(0.077)	(0.136)	(0.164)	(0.211)	(0.462)	(0.691)
D-to-D	0.006***	0.0035**	0.0045***	0.0074***	0.012***	0.018	0.055***
	(0.246)	(0.377)	(0.227)	(0.227)	(0.386)	(1.303)	(1.087)
R-to-D	0.013*	0.0099**	0.0007	0.0072*	0.033***	0.031**	0.0209
	(0.178)	(0.205)	(0.235)	(0.151)	(0.095)	(0.323)	(0.766)
R-to-F	0.023***	0.015***	0.014***	0.032***	0.071***	0.051***	0.068*
	(0.164)	(0.082)	(0.159)	(0.132)	(0.167)	(0.166)	(0.392)
D-to-F	0.021***	0.0096***	0.0088***	0.017***	0.057***	0.065***	0.064**
	(0.246)	(0.095)	(0.171)	(0.177)	(0.274)	(0.366)	(0.640)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed affects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared (within)	0.131	0.135	0.136	0.114	0.077	0.100	0.251
R-squared (between)	0.731	0.425	0.001	0.634	0.013	0.198	0.074
Adjusted R-squared	0.131	0.134	0.136	0.114	0.077	0.100	0.251
Number of stocks	97	14	16	41	57	35	80
Average # of trades	20197.732	13658.714	18517.438	12082.561	8426.088	8547.971	2460.350
Observations	1,959,180	191,222	296,279	495,385	480,287	299,179	196,828

Over the whole sample, we find that trades between retail investors (R-to-R) are associated with the highest price impacts: a 1000 Bath increase in trade size at the median leads to a 0.05 bps increase in prices. Trades between foreign investors (F-to-F) also positively contribute to price discovery, though less than retail investors: a 1000 Bath increase in trade size at the median is associated with a price impact of 0.034 bps. Importantly, trades between foreign and retail investors (R-to-F) are the third most important contributor to price discovery: increasing trade size by 1000 Bath at the median has an impact of prices equal to 0.023 bps. Domestic institutions' trades have non-negligible impact prices, but only when they interact with foreign investors (D-to-F): a 1000 Bath increase in trade size at the median is associated with a price increase equal to 0.021 bps. Trades among domestic institutions and between domestic institutions and retail investors are instead associated with the lowest price impact.

The price impact of trades by different market participants is heterogeneous depending on liquidity levels, though we are able to isolate some clear patterns. For example, we find that D-to-D and R-to-D trades are the least impactful for almost all liquidity levels, while R-to-F trades are among the most impactful for all liquidity levels. F-to-F trades are associated with lower price impacts for high liquidity segments and become more impactful in mid-low liquidity segments. The segment $ES(25, 50)$ is perhaps the most representative of this general trend: R-to-F trades have the largest price impact (0.032 bps per 1000 Baht increase), followed by R-to-R trades (0.031 bps per 1000 Baht increase), and F-to-F trades (0.017 bps per 1000 Baht increase). D-to-D trades and R-to-D trades rank last.[27]

As a robustness check we performed the same regressions by differentiating between buyers and sellers

---

[27]We notice some exceptions, such as in the segment $ES(<10)$, where R-to-D trades place themselves in second position; or in segment $ES(75, 90)$, where D-to-F trades are the most impactful. This latter evidence is consistent with the findings reported in Table 2, where domestic institutions are shown to have sizeable price impacts in the segment $ES(75, 90)$.

in trades between different categories.[28] In these analyses between-participant categories (say, R-to-F) are split into two different sub-categories (in our example, R-to-F, where the buyer is retail and the seller is foreign; and F-to-R, where the buyer is foreign is the buyer and the seller is retail). The results appear to be quantitatively and qualitatively similar to the one reported above. R-to-R trades are the most informative for all liquidity segments; F-to-F trades are relatively more informative in low liquidity segments than they are in high liquidity segments; D-to-D and R-to-D trades are the least informative, the latter category no matter who buys or sells. Interestingly, we find that among R-to-F trades it does matter instead who buys or sells, with foreign buys from retail investors being the most informative type of trades for all liquidity segments, and conveying significantly more information than foreign sales to retail investors. This general pattern holds for different levels of liquidity.

From the evidence reported in this section we conclude that price discovery can largely be attributed to retail and foreign investors, and, most importantly, to the interaction between these two investor types. This findings adds to the literature studying whether international or domestic investors are more informed (Boehmer and Wu, 2008; Chan et al., 2007; Dvorak, 2005; Goncalves and Eid, 2017; Lee et al., 2004; Xu and Wan, 2015), showing that the interaction between different investor types, in our case foreign institutions and domestic retail investors, can also be important for the price discovery process. Finally we note that excluding Segment 1, trades by domestic institutions are impactful only when foreign institutions take the other side of the trade, consistently with domestic institutions being large buy-and-hold players.

4.3 Relation between permanent price impact and liquidity

While a spread-based indicator of price impact is a legitimate measure of the information content of a trade, it is liable to depend on the arbitrarily chosen length of the time interval (Riordan and Storkenmaier, 2012) and to capture the information content of trades only over a limited time horizon. We therefore use our data to construct a more sophisticated indicator of the information content of stock trades, the widely used Hasbrouck (1991a,b) measure based on VAR estimation (Barclay and Hendershott, 2003; Benos and Sagade, 2016; Brandt and Kavajecz, 2004; Dufour and Engle, 2000; Hendershott et al., 2011; Marsh and Payne, 2012; Menkhoff and Schmeling, 2010; Riordan and Storkenmaier, 2012).

In the Hasbrouck (1991a,b) framework, estimation of the information content of trades is based on a VAR model as set out below:

$$r_{it} = \sum_{j=0}^{T} \alpha_{ij} x_{it-j} + \sum_{j=1}^{T} \beta_{ij} r_{it-j} + u_{it}^r \tag{8}$$

$$x_{it} = \sum_{j=1}^{T} \gamma_{ij} x_{it-j} + \sum_{j=1}^{T} \delta_{ij} r_{it-j} + u_{it}^x \tag{9}$$

Where i identifies the stock/day (i.e. we run equations 8 and 9 for each stock/day). We follow the literature and estimate the Hasbrouck model in tick-time (rather than clock-time), each entry t representing a different trade event (Barclay and Hendershott, 2003; Benos and Sagade, 2016; Hasbrouck, 1991a). T is the chosen number of lags. r_{it} is the percentage change in the quote midpoint between one trade and the subsequent one, while x_{it} is the trade direction (equal to 1 if the trade is buyer-initiated, equal to -1 one if the trade is seller-initiated). That is, trades and quotes are jointly determined based on contemporaneous and past values of each other. This is intuitive, as traders observing a series of buys (sells) in a row are likely to adjust their quotes up (down); likewise, traders observing an increase (decrease) in the price of a stock are more likely to buy (sell) that particular stock.

---

[28]Results are available upon request.

It must be noted that a VAR model that allows contemporaneous relations among its variables is in fact a structural VAR, as noted in Menkhoff and Schmeling (2010). Being a system of simultaneous equations, a structural VAR needs identifying restrictions for the estimators to be interpreted in a causal fashion (Ronayne, 2011; Stock and Watson, 2001). As evident from Equations 8 and 9, in the Hasbrouck VAR model (Hasbrouck, 1991a,b) identification is ensured by allowing for a contemporaneous influence of trades on quote midpoints but imposing that trade direction does not depend on contemporaneous quote midpoint changes; i.e. trades respond only to past quote midpoints. The model also imposes that shocks are not jointly and serially correlated, which leads to contemporaneous trades and quotes revisions to be also uncorrelated. These assumptions ensure identification in the Hasbrouck (1991a,b) framework.[29]

Upon estimation, the SVAR model can be inverted to obtain a Vector Moving Average representation.[30] The VMA representation for returns is:

$$r_{it} = \sum_{j=1}^{\infty} a_{ij} u_{it-j}^r + \sum_{j=1}^{\infty} b_{ij} u_{it-j}^x \tag{10}$$

The information content of stock trades is calculated as the sum of the impacts of trade innovations on quote revisions, as set out formally below (Hasbrouck, 1991a):

$$b_{it}(u_{i0}^x) = \sum_{t=0}^{M} E[r_{it}|x_{i0}] \tag{11}$$

Where M is a chosen time horizon (in our case, 100 trades), and 0 the initial value. This value is the cumulative impulse response function, found by inverting the SVAR model to obtain a Vector Moving Average representation and summing up the cumulative impact of trade innovations u^x on quote revisions (Hasbrouck, 1991b, page 577). Through the VMA representation and the calculation of impulse response functions it is also possible to estimate the contribution of trades (private) and quotes (public) information on the variance of returns. Noting that r_{it} is a zero-mean process (Hasbrouck, 1991a, p 592), that $E(u_{(t-j)}^r * u_{(t-j)}^x) = 0 \; \forall j$, and that shocks are serially uncorrelated, the variance of returns can be written as:

$$\sigma_{iw}^2 = (\sum_{j=0}^{\infty} \alpha_{ij})^2 \sigma_{ir}^2 + (\sum_{j=0}^{\infty} \beta_{ij})^2 \sigma_{ix}^2 \tag{12}$$

Where w is the shock to the permanent component of prices using the (Hasbrouck, 1991b) notation (Benos and Sagade, 2016, p 19). Equation 12 describes the relative contribution of private information extracted from trade behaviour and on public information[31] inferred from the limit order book to price impact.

We estimate one SVAR model for each stock/day. We follow the literature and as a general rule estimate the model using 10 lags (Barclay and Hendershott, 2003; Benos and Sagade, 2016; Riordan and

---

[29]Using Hasbrouck's words, "there is a strong presumption of causality running from trades to quote revisions" (Hasbrouck, 1991a), as sufficiently large marketable orders can deplete the depth of the order book at the best available price and therefore cause the best quote to simultaneously change (Benos and Sagade, 2016). Traders, however, will only react to public information that is already available.

[30]Where invertibility is a plausible assumption, as noted in (Hasbrouck, 1991b, page 576) Hasbrouck. Invertibility requires stationarity. Bid/ask prices are integrated of order one (Engle and Patton, 2004), and so must be quote midpoints, a linear combination of bid/ask prices. The percentage change in quote midpoints should therefore be stationary. As per trades, these are represented by an indicator equal to one in case of a buy, and equal to minus one in case of a sell, i.e. a bounded variable by definition. Unit root tests conducted on randomly sampled stock/days confirm these conjectures. We therefore trust that the VAR model is investible.

[31]Private information in this context should not be interpreted as the practice of using information not available to the public to manipulate the market (insider trading), illegal in most jurisdictions including Thailand.

Storkenmaier, 2012). Some stock/days however do not have enough trading events (read: observations) to fit a model with ten lags. We therefore estimate SVAR models with five or three lags when the number of observations is not sufficient. After removing outliers, we are left with 1675 stock/day estimations, amounting to 96% of our sample. We retain a database of stock/day variables, namely the permanent price impact of trades, calculated as in Equation 10, as well as trade and quote correlated information, calculated as in Equation 12. To perform our analyses we supplement this database with the average effective spread per stock/day, as effective spreads will be used to segment the market for different liquidity levels. Table 7 below contains descriptive statistics for the Hasbrouck measures:

Table 7: Hasbrouck measures - descriptive statistics. Effective spreads, price impacts, quote and trade correlated information: Basis points.

	(1)	(2)	(3)
	Mean	Median	Std. dev.
Price impact function (iIRF)	12.27	9.33	11.074
Total volatility (std. dev.)	19.53	16.51	13.63
Quote correlated information (std. dev.)	17.08	13.83	12.54
Trade correlated information (std. dev.)	8.21	6.104	7.16
Trade correlated information (% total)	0.201	0.14	0.17
Observations	1675		

The Hasbrouck measure of price impact is the cumulative impulse response function of trade-related shocks to stock returns. In the Hasbrouck (1991a,b) framework it is a measure of the private information content of stock trades. According to our estimates, on average a trade has a permanent price impact equal to 12 bps, roughly half of the average price impact of trades calculated over a five-minute horizon (see Table 1). This result suggests that the impact of trades on prices over a five minute interval is likely to be biased upwards by the presence of transitory shocks, which the better specified Hasbrouck model manages instead to strip out (see (Benos and Sagade, 2016, Equation 13). Trade correlated information, the contribution of trades of all sizes to the volatility of the efficient component of prices, is 8.21 bps (note that, we report standard deviations rather than variances to allow comparability like in Riordan and Storkenmaier (2012)). Trade correlated information is 20% of the total volatility of the efficient component of prices, meaning that 20% of price volatility is attributable to trades (and hence, 80% to quotes). Past research also reports trades to have a proportionally lower contribution to price discovery than public information arising from quotes (Riordan and Storkenmaier, 2012).

We investigate how price discovery relates to stock liquidity. To do so we run ANOVA models as formalised below:

$$E[Pricediscovery_{it}|X] = \sum_{j=1}^{6} \alpha_j L_{jit} \qquad (13)$$

Where i indicates the stock, t indicates the trading day, L_{jiy} are dummies for different liquidity segments (identified as in the previous sections) and price discovery measures are: permanent price impact of trades, trade correlated information, quote correlated information and the relative contribution of trades to total price volatility. Results are summarised in Figure 6, and briefly commented on below.[32]

Figure 6 shows that the price impact function is monotonically increasing in spread size, with the permanent price impact of a trade as little as roughly 3.9 bps in the higher liquidity segment, and as

---

[32]For a full display of the results and a more complete commentary we refer to the Supplementary Materials, Section 4.

Figure 6: Hasbrouck measure per liquidity level
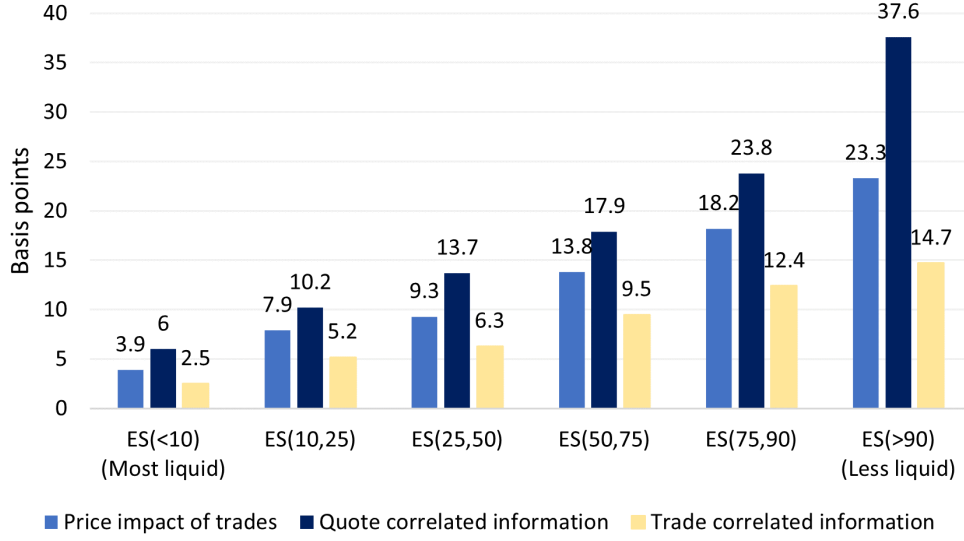
large as over 23.3 bps in the lower liquidity segment. This result is consistent with the overall findings of the paper, and with the literature. Unsurprisingly, efficient price volatility also monotonically increases with spreads, with both trade- and quote-correlated information monotonically increasing as liquidity goes down.

To assess the relative contribution of different types of trading participants to price discovery, we estimate an "augmented" Hasbrouck VAR model where trades are attributed to different trading participants, considering the six possible combinations, as in Section 4.2. We follow the approach of Benos and Sagade (2016), who estimate a similar model to study the contribution to price discovery of high frequency traders with different trading behaviours. To the best of our knowledge, however, this is the first time the Hasbrouck model has been used to study the heterogeneous contribution of investor types to price discovery. Our model is as follows:

$$
\begin{aligned}
r_{it} =& \sum_{j=0}^{T} \alpha_{ij} x_{it-j}^{R-to-R} + \sum_{j=0}^{T} \gamma_{ij} x_{it-j}^{F-to-F} + \sum_{j=0}^{T} \delta_{ij} x_{it-j}^{D-to-D} + \sum_{j=0}^{T} \epsilon_{ij} x_{it-j}^{R-to-D} + \\
&+ \sum_{j=0}^{T} \theta_{ij} x_{it-j}^{R-to-F} + \sum_{j=0}^{T} \sigma_{ij} x_{it-j}^{D-to-F} + \sum_{j=1}^{T} \beta_{ij} r_{it-j} + u_{it}^{r}
\end{aligned}
\tag{14}
$$

$$
\begin{aligned}
x_{it}^{CAT} =& \sum_{j=0}^{T} \pi_{ij}^{CAT} x_{it-j}^{R-to-R} + \sum_{j=0}^{T} \rho_{ij}^{CAT} x_{it-j}^{F-to-F} + \sum_{j=0}^{T} \tau_{ij}^{CAT} x_{it-j}^{D-to-D} + \sum_{j=0}^{T} \phi_{ij}^{CAT} x_{it-j}^{R-to-D} + \\
&+ \sum_{j=0}^{T} \omega_{ij}^{CAT} x_{it-j}^{R-to-F} + \sum_{j=0}^{T} \lambda_{ij}^{CAT} x_{it-j}^{D-to-F} + \sum_{j=1}^{T} \Delta_{ij} r_{it-j} + u_{it}^{x}
\end{aligned}
\tag{15}
$$

Where CAT is either R-to-R, F-to-F, D-to-D, R-to-F, R-to-D or D-to-F, i.e. Equation 15 is a compact notation to indicate six different equations in the system. The variables x_{it}^{CAT} identify trades between different categories (i.e. R-to-R, F-to-F, etc.) and are equal to one for a buy, equal to minus one for a sell, and equal to zero if it is not a trade in that category.

We estimate one SVAR model per stock/day. It must be noted that not all stock/days feature trades in all six trading participant categories; for example trades in smaller stocks typically take place between

individual investors only (R-to-R). As such, we leave the estimation flexible, with the system comprised by two to seven equations depending on the stock/day. Depending on the number of observations the model is estimated with ten, five or three lags, to cater for the presence of stocks with lower trading activity. After removing outliers, we are left with 1661 stock/day estimates, amounting 95% of the stock/days in the sample.

For each stock day and trading category, we estimate the permanent price impact function as below:

$$b_{it}^{CAT}(u_{i0}^x) = \sum_{t=0}^{M} E\left[r_{it}|x_{i0}^{CAT}\right] \tag{16}$$

Equation 16 is a compact notation to indicate that each trading day can have up to six different permanent price impact functions, one for each combination of participants. Figure 7 graphically shows the price impact functions for different trading participant categories on a stock/day where all trading participants trading combinations are present:

Figure 7: Permanent price impact functions per trading participant on a selected stock day
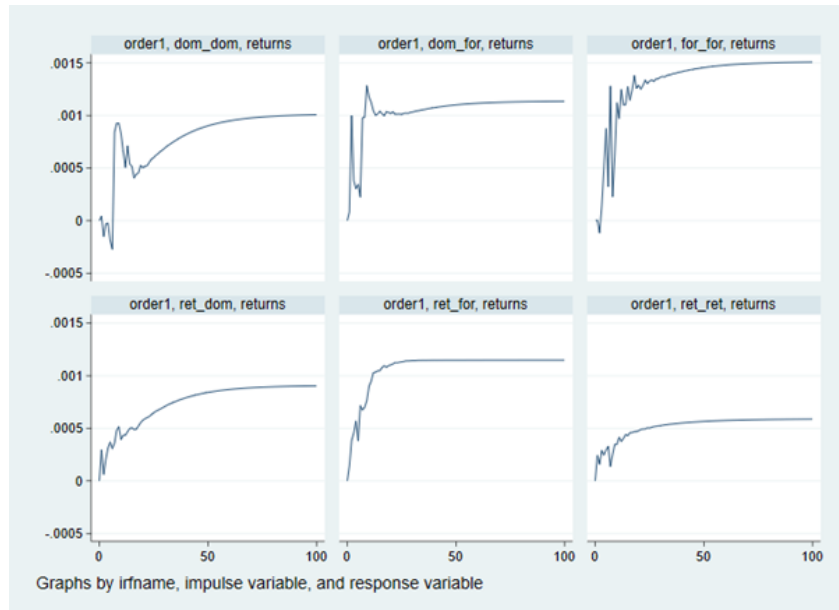
To provide evidence on the contribution of different trading participants to price discovery, and on their relation with liquidity, we run an ANOVA model for each trading participant category putting liquidity segments on the right-hand-side. Formally:

$$E\left[\text{Price discovery}_{it}^{CAT}|X\right] = \sum_{j=1}^{6} \alpha_j L_{jit} \tag{17}$$

Where i indicates the stock, t indicates the trading day, L_{ijt} are dummies for different liquidity segments (identified as in the previous sections). Results are displayed in Table 8 below.

We comment on Table 8 starting from the high liquidity segment, ES(< 10). When effective spreads are lower than their 10th percentile, price discovery seems to be dominated by D-to-D trades: a D-to-D trade is associated with a permanent price impact equal to 5 bps (Panel 1, Column (5)). R-to-R trades are associated with the lowest permanent price impact (one R-to-R trade leads to a 3.4 bps price impact, (Panel 1, Column (1)), and F-to-F trades lie in be-tween (one F-to-F trade leads to a 4 bps price impact,

Table 8: Price impact function by trading participant category and liquidity segment. The table reports marginal effects at the sample medians, for each category. Clustered Standard errors in parentheses.* p< 0.1, ** p < 0.05, *** p < 0.01.

	Panel 1					
	R-to-R		F-to-F		D-to-D	
	(1)	(2)	(3)	(4)	(5)	(6)
	Price impact	Price impact per 1000 Bath	Price impact	Price impact per 1000 Bath	Price impact	Price impact per 1000 Bath
ES(< 10)	3.417***	0.017	4.026***	0.014	5.004***	0.006
	(0.868)		(1.277)		(0.908)	
ES(10, 25)	7.656***	0.061	7.801***	0.040	7.775***	0.013
	(0.679)		(1.063)		(0.824)	
ES(25, 50)	8.958***	0.147	9.630***	0.144	9.023***	0.023
	(0.547)		(0.978)		(0.824)	
ES(50, 75)	13.853***	0.208	13.151***	0.150	9.514***	0.024
	(0.546)		(1.027)		(0.949)	
ES(75, 90)	18.701***	0.449	12.516***	0.208	10.328***	0.021
	(0.700)		(1.435)		(1.308)	
ES(> 90)	22.997***	0.715	10.879***	0.237	7.667*	0.027
	(0.950)		(3.260)		(4.393)	
F-test	392.618		68.669		67.534	
Adj. R^2	0.586		0.273		0.333	
Obs.	1661		1082		798	

	Panel 2					
	R-to-D		R-to-F		D-to-F	
	(1)	(2)	(3)	(4)	(5)	(6)
	Price impact	Price impact per 1000 Bath	Price impact	Price impact per 1000 Bath	Price impact	Price impact per 1000 Bath
ES(< 10)	4.071***	0.011	3.676***	0.016	3.994***	0.009
	(0.937)		(1.179)		(1.113)	
ES(10, 25)	7.663***	0.036	8.474***	0.065	8.059***	0.026
	(0.778)		(0.921)		(0.940)	
ES(25, 50)	7.694***	0.063	10.138***	0.229	9.131***	0.070
	(0.716)		(0.750)		(0.894)	
ES(50, 75)	10.545***	0.083	15.049***	0.276	10.860***	0.056
	(0.788)		(0.755)		(1.031)	
ES(75, 90)	10.772***	0.124	17.889***	0.496	12.529***	0.090
	(1.119)		(1.014)		(1.391)	
ES(> 90)	11.697***	0.086	25.501***	0.778	-1.531	-0.009
	(2.753)		(1.514)		(3.680)	
F-test	86.908		211.493		63.795	
Adj. R^2	0.329		0.446		0.282	
Obs.	1049		1567		959	

(Panel 1, Column (3)). These figures, however, do not take into account that different trading participant categories engage in trades of different sizes, with R-to-R trades typically characterised by smaller sizes. As extensively discussed in Section 4.2, the price impact function is an increasing monotonic function of trade size, therefore larger trades have by definition a larger price impact. To gauge a better picture of who contributes the most to price discovery, we divide the price impact function by the median daily trade size per participant category and liquidity segment, identifying the price impact per 1,000 Bath (equivalent to calculating a marginal effect, as in Section 4.2).[33] These are reported in the even columns.

Likely because price impacts calculated using the Hasbrouck method are not biased by transitory shocks, the results reported in Table 8 show very stable and clear-cut patterns across liquidity levels,

---

[33]We perform the same analysis using volumes (and therefore calculating the price impact per share traded) and the results are equivalent.

suggesting that while different trading participants contribute heterogeneously to price discovery, their heterogeneous contribution is not affected by different liquidity levels. Excluding the segment ES($<$ 10), when R-to-R trades are the most informative, R-to-F trades are the most informative across all liquidity segments. These are followed by R-to-R trades and F-to-F trades, confirming that both foreign institutions and individual investors are the categories that influence on price discovery the most. This analysis confirms that trades between different participant types do convey information, thus contributing to the debate on whether it is foreign or domestic investors who possess an informational advantage. These findings might seem at odds with the result that in low liquidity segments, foreign investors' trade imbalances and trades are associated with lower price impacts, which we attributed to passive index trackers. This explanation is still plausible though. First of all, and as noted, 5-minute interval analyses are biased by transitory shocks. Secondly, while in high liquidity segments foreign investors trigger most of the trades, R-to-F trades are the most informative, but F-to-F trades score only third in terms of price informativeness, suggesting once more that informed retail investors are likely to take the opposite position on uninformed foreign traders. R-to-D and D-to-F trades follow, and D-to-D trades appear to be the least informative throughout the sample. In Segment 4, for example, the most informative trades are R-to-F, where the price impact per 1,000 Baht trade is 0.27 bps. These are followed by R-to-R trades (0.20 bps) and F-to-F trades (0.15 bps). D-to-D trades rank last, with a price impact equal to 0.024 bps. Finally, we note that price impacts (per trade and per Bath-traded) monotonically increase in spreads, thus confirming once more that when liquidity is low trades are more informative. The way different trading participant contribute to price discovery is however stable across different liquidity levels.

5 Discussion and policy implications of the findings

The first main take of the results is that different investor types contribute heterogeneously to price formation. Besides large cap stocks, however, their relative contribution to prices (i.e. who contributes more to price discovery) seems to be fairly stable across liquidity levels, possibly as a consequence of high levels of active retail participation. Importantly, we find that the two most informed categories, retail investors and foreign institutions, contribute to price discovery by trading with each other. These findings are overall consistent with the ones outlined by Phansatan et al. (2012), who show that international investors on SET have a 'market timing' information advantage, but retail investors have a superior ability in stock-picking. Somewhat surprisingly, though our paper confirms their result, Phansatan et al. (2012) find that domestic institutions and proprietary traders (assimilated into domestic institutions in our data) are poor at selecting their stocks, which leads to an overall bad performance in the long-run (see also Richards (2005)).

When it comes to small-cap stocks, one way to enhance the overall market liquidity would be to attract more uninformed investors, typically liquidity suppliers; in the case of SET, these would be domestic institutional investors. By interacting with active, informed retail investors, who likely demand for immediate execution thus snatching liquidity away, institutional investors would help contain the price impact of trades, contributing to spread reductions and enhancing price efficiency overall. Attracting institutional investors towards smaller-cap stocks might not be an easy task though, as large institutions often have investment mandates that are not compatible with the levels of liquidity of small-cap companies. In addition, the levels of sophistication of the institutional segment in many emerging and frontier markets might not be high enough to cater for investors' demand in small-cap companies. One way to achieve this would be through financial literacy: by outlining the benefits of fund investing in small-cap stocks, the exchange can help create a demand-driven push towards small-cap products. Lobbying with the regulator and engaging in talks with the relevant national investors would also be crucial to achieve this target.

We found that SET has a very balanced participation in large cap stocks, with all investors category

contributing to trading, and foreign institutions' aggressive imbalances being the most important predictor of stock prices, according to our regression results. We believe this healthy mix of trading participants is the ideal situation exchanges should be aiming for. In particular, we hope this research will highlight that the interaction between more and less (but still) informed categories is beneficial for the market. Much effort has been spent in the past trying to understand who is more informed between domestic and foreign investors. This research however emphasises how it might be even more important to focus on how the interaction between investors possessing different levels of information could be beneficial for price discovery.

We would like to emphasise that while some of this paper's findings do expand the external validity of the evidence on the topic (the negative relation between liquidity and price impacts; the heterogeneous contribution of investors to price formation), policy implications on how to attract liquidity to the market are market-specific as they depend on the particular mix of market participants an exchange has managed to achieve. We therefore encourage the performance of similar studies in other jurisdictions to expand individual markets' understanding of the levers they can pull to enhance liquidity in their market by levering on market participation.[34]

6   Conclusions

This paper investigates whether different types of trading participants (i.e. retail investors, domestic institutions, and foreign institutions) contribute heterogeneously to price formation, and adds to the literature by studying whether their trades' impact on prices is in turn dependent on liquidity. To the best of our knowledge no paper before has investigated the interaction between these three dimensions. Better understanding how these three dimension interact is however of relevance for exchanges willing to lever on their market participation to attract a balanced mix of informed and uninformed investors, thus further enhancing liquidity and ameliorating price formation.

Overall, we find that retail investors are the category that retains the highest degree of 'information', understood as a superior ability to interpret market trends and fundamentals. Retail investors' trade imbalances are the most sizeable and statistically significant predictor of five-minute returns for almost all levels of liquidity. We find however that order imbalances of foreign traders in in the high liquidity segment is the most significant predictor or five-minute returns, suggesting that foreign investors also contribute to price formation.

These results are confirmed when analysing short-run (five minute) and long-run (daily) price impact functions: we find that trades trades between retail and foreign investors (R-to-F) have the largest permanent impact on prices and among the largest five-minute impact on prices for all levels of liquidity. We find that trades between retail investors (R-to-R) have the second largest impacts on prices, both over a five-minute and a daily horizon, while trades between foreign investors (F-to-F) rank third in terms of price impacts. These combined findings suggest that the trading activity between informed traders may convey information to the market, contributing to the debate on information asymmetries between international and domestic investors.

Domestic institutions' order imbalances predict five minutes returns only for medium to low levels of liquidity, are the second largest contributor to price inefficiency when liquidity is medium/low. With some exceptions, trades involving domestic institutions rank lowest in terms of five-minute price informativeness for all liquidity levels, and they rank among the lowest for all liquidity levels in terms of daily price impacts. These results suggest that domestic institutions are likely to be largely buy-and-hold traders.

We find that lower liquidity is always associated with less efficient prices and a higher information content of trades, consistently with the literature. We also find that price impacts increase monotonically with illiquidity for all trading participant types. We find however that the contribution of trading

---

[34]For a general discussion on the topic, see Alderighi (2017)

participant types to price discovery does not show a particular relation with liquidity, and exceptions apply only when considering short-term returns predictability and price impact measures. When considering permanent price impacts, not biased by transitory shocks, trading participants' contribution to price formation is not dependent on liquidity, i.e. the most informative trades are R-to-F for all liquidity levels, followed by R-to-R and F-to-F trades. Trades involving domestic institutions are instead the least informative for all liquidity levels.

References

Admati, A. R. and Pfleiderer, P. (1988). A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies*, 1(1):3–40.

Aitken, M. and Comerton-Forde, C. (2005). Do reductions in tick sizes influence liquidity? *Accounting and Finance*, 45(2):171–184.

Aitken, M. and Frino, A. (1996). The accuracy of the tick test: Evidence from the Australian stock exchange. *Journal of Banking & Finance*, 20(10):1715–1729.

Aktas, N., de Bodt, E., and Van Oppens, H. (2008). Legal insider trading and market efficiency. *Journal of Banking & Finance*, 32(7):1379–1392.

Albuquerque, R., Bauer, G. H., and Schneider, M. (2009). Global private information in international equity markets. *Journal of Financial Economics*, page 29.

Alderighi, S. (2017). A note on how to enhance liquidity in emerging markets by levering on trading participants. *Economics Bulletin*, 37(4):2526–2532.

Alderighi, S., Cleary, S., and Varanasi, P. (2019). Do institutional factors influence cross-border portfolio equity flows? new evidence from emerging markets. *Journal of International Money and Finance*, 99:102070.

Alderighi, S., Gurrola-Perez, P., Lin, K., and Speth, B. (2021). Circuit breakers and other market safeguards. Technical report, World Federation of Exchanges.

Bae, K.-H., Stulz, R. M., and Tan, H. (2008). Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, 88(3):581–606.

Barber, B. M., Lee, Y.-T., Liu, Y.-J., and Odean, T. (2009). Just How Much Do Individual Investors Lose by Trading? *Review of Financial Studies*, 22(2):609–632.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Barclay, M. J. and Hendershott, T. (2003). Price Discovery and Trading After Hours. *Review of Financial Studies*, 16(4):1041–1073.

Benos, E. and Sagade, S. (2016). Price discovery and the cross-section of high-frequency trading. *Journal of Financial Markets*, 30:54–77.

Boehmer, E. and Wu, J. J. (2008). Order flow and prices. *SSRN Electronic Journal*.

Brandt, M. W. and Kavajecz, K. A. (2004). Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve. *The Journal of Finance*, 59(6):2623–2654.

Brennan, M. J. and Subrahmanyam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, 41(3):441–464.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-Frequency Trading and Price Discovery. *Review of Financial Studies*, 27(8):2267–2306.

Brown, A. and Yang, F. (2016). Market Liquidity and Price Efficiency: Evidence from Two Field Experiments. *SSRN Electronic Journal*.

Chakravarty, S., Gulen, H., and Mayhew, S. (2004). Informed Trading in Stock and Option Markets. *The Journal of Finance*, 59(3):1235–1257.

Chan, K., Menkveld, A., and Yang, Z. (2008). Information asymmetry and asset prices: Evidence from the china foreign share discount. *Journal of Finance*, 63(1):159–196.

Chan, K., Menkveld, A. J., and Yang, Z. (2007). The informativeness of domestic and foreign investors' stock trades: Evidence from the perfectly segmented Chinese market. *Journal of Financial Markets*, 10(4):391–415.

Chordia, T., Roll, R., and Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, page 20.

Chung, D. and Hrazdil, K. (2010). Liquidity and market efficiency: A large sample study. *Journal of Banking & Finance*, 34(10):2346 – 2357.

Cont, R., Kukanov, A., and Stoikov, S. (2014). The Price Impact of Order Book Events. *Journal of Financial Econometrics*, 12(1):47–88.

de Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Positive Feedback Investment Strategies and Destabilizing Rational Speculation. *The Journal of Finance*, 45(2):379.

Dufour, A. and Engle, R. F. (2000). Time and the Price Impact of a Trade. *The Journal of Finance*, 55(6):2467–2498.

Dvorak, T. (2005). Do Domestic Investors Have an Information Advantage? Evidence from Indonesia. *The Journal of Finance*, 60(2):817–839.

Easley, D., Hvidkjaer, S., and O'Hara, M. (2002). Is Information Risk a Determinant of Asset Returns? *The Journal of Finance*, 57(5):2185–2221.

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(4):529–551.

Engelberg, J. E., Reed, A. V., and Ringgenberg, M. C. (2012). How are shorts informed? *Journal of Financial Economics*, 105(2):260–278.

Engle, R. F. and Patton, A. J. (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets*, 7(1):1 – 25.

Ferreira, M. A., Matos, P., Pereira, J. P., and Pires, P. (2017). Do locals know better? A comparison of the performance of local and foreign institutional investors. *Journal of Banking & Finance*, 82:151–164.

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

Frijns, B., Gilbert, A., and Tourani-Rad, A. (2010). The dynamics of price discovery for cross-listed shares: Evidence from Australia and New Zealand. *Journal of Banking & Finance*, 34(3):498–508.

Frijns, B., Gilbert, A., and Tourani-Rad, A. (2015). The determinants of price discovery: Evidence from US-Canadian cross-listed shares. *Journal of Banking & Finance*, 59:457–468.

Frijns, B., Indriawan, I., and Tourani-Rad, A. (2018). The interactions between price discovery, liquidity and algorithmic trading for U.S.-Canadian cross-listed shares. *International Review of Financial Analysis*, 56:136–152.

Froot, K. A. and Seasholes, M. S. (2001). The portfolio flows of international investors. *Journal of Financial Economics*, page 43.

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(1):71–100.

Goncalves, W. and Eid, W. (2017). Sophistication and price impact of foreign investors in the Brazilian stock market. *Emerging Markets Review*, 33:102–139.

Grinblatt, M. (2000). The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of Financial Economics*, 55(1):43–67.

Hanke, M. and Weigerding, M. (2015). Order flow imbalance effects on the german stock market. *Business Research*, 8(2):213–238.

Hasbrouck, J. (1991a). Measuring the Information Content of Stock Trades. *The Journal of Finance*, 46(1):179–207.

Hasbrouck, J. (1991b). The Summary Informativeness of Stock Trades: An Econometric Analysis. *The Review of Financial Studies*, 4(3,):571–595.

Hasbrouck, J. (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. *The Journal of Finance*, 50(4):1175–1199.

Hasbrouck, J. (2007). *Empirical market microstructure: the institutions, economics and econometrics of securities trading.* Oxford University Press, Oxford ; New York. OCLC: ocm63472962.

Hasbrouck, J. (2010). The Best Bid and Offer: A Short Note on Programs and Practices. *SSRN Electronic Journal*.

Hasbrouck, J. and Saar, G. (2009). Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, 12(2):143–172.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4):646–679.

Hasbrouck, J. and Seppi, D. J. (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, page 29.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does Algorithmic Trading Improve Liquidity? *The Journal of Finance*, 66(1):1–33.

Huang, R. D. and Stoll, H. R. (1994). Market Microstructure and Stock Return Predictions. *Review of Financial Studies*, 7(1):179–213.

Huang, R. D. and Stoll, H. R. (1997). The Components of the Bid-Ask Spread: A General Approach. *Review of Financial Studies*, 10(4):995–1034.

Kang, J.-K. and Stulz, R. (1997). Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. *Journal of Financial Economics*, 46(1):3–28.

Kempf, A. and Korn, O. (1999). Market depth and order size. *Journal of Financial Markets*, page 20.

Khanna, K., Smith, M., Wu, D., and Zhang, T. (2009). Reconstructing the order book. Technical report, Stanford Unviersity Working Paper.

Korniotis, G. M. and Kumar, A. (2011). Do Older Investors Make Better Investment Decisions? *Review of Economics and Statistics*, 93(1):244–265.

Lang, M., Maffett, M., Omartian, J. D., and Silvers, R. (2020). Regulatory cooperation and foreign portfolio investment. *Journal of Financial Economics*, page S0304405X20301331.

Lao, L., Tian, S., and Zhao, Q. (2018). Will Order Imbalances Predict Stock Returns in Extreme Market Situations? Evidence from China. *Emerging Markets Finance and Trade*, 54(4):921–934.

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

Lee, Y.-I., Liu, Y.-J., Roll, R., and Subrahmanyam, A. (2004). Order imbalances and market efficiency: Evidence from the taiwan stock exchange. *Journal of Financial and Quantitative Analysis*, 39:327–341.

Likitapiwat, T. (2016). Algorithmic trading in an emerging market: Empirical study on the stock exchange of thailand. Technical report.

Lillo, F., Farmer, J. D., and Mantegna, R. N. (2002). Single Curve Collapse of the Price Impact Function for the New York Stock Exchange. *arXiv:cond-mat/0207428*. arXiv: cond-mat/0207428.

Lillo, F., Farmer, J. D., and Mantegna, R. N. (2003). Master curve for price-impact function. *Nature*, 421(6919):129–130.

Lim, M. and Coggins, R. (2005). The immediate price impact of trades on the Australian Stock Exchange. *Quantitative Finance*, 5(4):365–377.

Marsh, I. W. and Payne, R. (2012). Banning short sales and market quality: The UK's experience. *Journal of Banking & Finance*, 36(7):1975–1986.

Menkhoff, L. and Schmeling, M. (2010). Whose trades convey information? Evidence from a cross-section of traders. *Journal of Financial Markets*, 13(1):101–128.

Menkveld, A. and Wang, T. (2013). How do designated market makers create value for small-caps? *Journal of Financial Markets*, 16(3):571–603.

Narayan, P. K., Narayan, S., and Westerlund, J. (2015). Do order imbalances predict Chinese stock returns? New evidence from intraday data. *Pacific-Basin Finance Journal*, 34:136–151.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets*, 3(3):259–286.

O'Hara, M. (2003). Presidential Address: Liquidity and Price Discovery. *The Journal of Finance*, 58(4):1335–1354.

O'Hara, M. (2011). *Market microstructure theory*. Blackwell, Malden, Mass., repr. edition. OCLC: 839096576.

Otsubo, Y. (2014). International cross-listing and price discovery under trading concentration in the domestic market: Evidence from Japanese shares. *Journal of Empirical Finance*, page 16.

Pascual, R., Pascual-Fuster, B., and Climent, F. (2006). Cross-listing, price discovery and the informativeness of the trading process$. *Journal of Financial Markets*, page 18.

Pavabutr, P. and Sirodom, K. (2010). Stock splits in a retail dominant order driven market. *Pacific-Basin Finance Journal*, 18(5):427–441.

Phansatan, S., Powell, J. G., Tanthanongsakkun, S., and Treepongkaruna, S. (2012). Investor type trading behavior and trade performance: Evidence from the Thai stock market. *Pacific-Basin Finance Journal*, 20(1):1–23.

Potters, M. and Bouchaud, J.-P. (2003). More statistical properties of order books and price impact. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):133–140.

Reck, M. (2020). Xetra: the evolution of an electronic market. *Electronic Markets*, 30(1):69–73.

Rhee, S. G. and Wang, J. (2009). Foreign institutional ownership and stock market liquidity: Evidence from Indonesia. *Journal of Banking & Finance*, 33(7):1312–1324.

Richards, A. (2005). Big Fish in Small Ponds: The Trading Behavior and Price Impact of Foreign Investors in Asian Emerging Equity Markets. *Journal of Financial and Quantitative Analysis*, 40(1):1–27.

Riordan, R. and Storkenmaier, A. (2012). Latency, liquidity and price discovery. *Journal of Financial Markets*, 15(4):416–437.

Riordan, R., Storkenmaier, A., Wagener, M., and Sarah Zhang, S. (2013). Public information arrival: Price discovery and liquidity in electronic limit order markets. *Journal of Banking & Finance*, 37(4):1148–1159.

Ronayne, D. (2011). Which impulse response function? The warwick economics research paper series (twerps), University of Warwick, Department of Economics.

Stock, J. H. and Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

Subrahmanyam, A. and Zheng, H. (2016). Limit order placement by high-frequency traders. *Borsa Istanbul Review*, 16(4):185–209.

Tetlock, P. C. (2008). Liquidity and Prediction Market Efficiency. *SSRN Electronic Journal*.

Theissen, E. (2001). A test of the accuracy of the Lee/Ready trade classification algorithm. *Journal of International Financial Markets, Institutions and Money*, 11(2):147–165.

Vega, C. (2006). Stock price reaction to public and private information$. *Journal of Financial Economics*, page 31.

Visaltanachoti, N. and Luo, R. (2009). Order imbalance, market returns and volatility: evidence from Thailand during the Asian crisis. *Applied Financial Economics*, 19(17):1391–1399.

Wallace, D., Kalev, P. S., and Lian, G. (2019). The evolution of price discovery in us equity and derivatives markets. *Journal of Futures Markets*, 39(9):1122–1136.

Weber, P. and Rosenow, B. (2005). Order book approach to price impact. *Quantitative Finance*, 5(4):357–364.

WFE (2017). Enhancing retail participation in emerging markets. Technical report.

WFE and Oliver Wyman (2016). Enhancing the liquidity of emerging market exchanges. Technical report.

Xu, F. and Wan, D. (2015). The impacts of institutional and individual investors on the price discovery in stock index futures market: Evidence from China. *Finance Research Letters*, 15:221–231.

Yamamoto, R. (2012). Intraday technical analysis of individual stocks on the Tokyo Stock Exchange. *Journal of Banking & Finance*, 36(11):3033–3047.