



## High-frequency liquidity in the Chinese stock market: Measurements, patterns, and determinants<sup>☆</sup>

Chaoyi Zhao <sup>a,b,c</sup>, Yufan Chen <sup>a</sup>, \*<sup>\*</sup>, Lintong Wu <sup>a</sup>, Yuehao Dai <sup>a</sup>, Ermo Chen <sup>a</sup>, Lan Wu <sup>a,d,e</sup>, Ruixun Zhang <sup>a,d,e,f</sup>

<sup>a</sup> Department of Financial Mathematics, School of Mathematical Sciences, Peking University, China

<sup>b</sup> Sloan School of Management, Massachusetts Institute of Technology, United States of America

<sup>c</sup> Laboratory for Financial Engineering, Massachusetts Institute of Technology, United States of America

<sup>d</sup> Laboratory for Mathematical Economics and Quantitative Finance, Peking University, China

<sup>e</sup> Center for Statistical Science, Peking University, China

<sup>f</sup> National Engineering Laboratory for Big Data Analysis and Applications, Peking University, China

### ARTICLE INFO

#### JEL classification:

C55  
G12  
G14  
G15

#### Keywords:

Market liquidity  
Bid–ask spread  
Aggressive–passive imbalance  
Chinese stock market

### ABSTRACT

We explore a broad range of high-frequency liquidity measures for the Chinese stock market, based on a comprehensive tick-level dataset for stocks on the Shenzhen Stock Exchange (SZSE) with approximately 46.64 billion events in 2019–2021. We integrate the raw event-level data into a granular and continuous limit order book for each stock for the three years. We summarize their liquidity levels and key distributional properties. Hypothesis tests show that order interarrival times follow Weibull—not exponential—distributions, implying that Poisson flow is not an appropriate model for order flow in the Chinese stock market. We analyze the intraday and cross-sectional patterns of liquidity, and find novel intraday periodicities in liquidity at whole-minute frequencies such as 1-minute, 5-minute, and 10-minute. Finally, we propose the *aggressive–passive imbalance* (API), analogous to the order flow imbalance of Cont, Kukanov, and Stoikov (2014), and develop an order-based model of the change in bid–ask spread that sheds light on the universal mechanism of spread formation with respect to order flows. To the best of our knowledge, this is by far the most comprehensive study of market liquidity for the Chinese stock market in the literature.

### 1. Introduction

The role of market liquidity in empirical finance has grown rapidly over the last decade, influencing studies in microstructure, asset pricing, market efficiency, and corporate finance. While it has long been acknowledged that market liquidity has an impact on asset prices and risk management, questions remain about its measurements, determinants, and time-series and cross-sectional properties (Chai et al., 2021). The literature on market liquidity faces its greatest challenges in non-US markets because global intraday data is less widely available compared to the US despite its exponential growth in recent years (Fong et al., 2017).

<sup>☆</sup> Research funding from the National Key R&D Program of China (2022YFA1007900), the National Natural Science Foundation of China (12271013), and the Fundamental Research Funds for the Central Universities, China (Peking University) is gratefully acknowledged. We thank Zhiwei Yao, Yihao Zhou, and seminar and conference participants at the 2022 INFORMS Annual Meetings, the 2022 Bachelier Finance Society World Congress, the 2022 IFZ FinTech Colloquium, the 2022 China Society for Industrial and Applied Mathematics (CSIAM) Annual Conference, and 2023 China Finance Review International & China International Risk Forum (CFRI&CIRF) Joint Conference.

\* Correspondence to: Peking University School of Mathematical Sciences, 5 Yiheyuan Road, Beijing, 100871, China

E-mail address: [yufan\\_chen@pku.edu.cn](mailto:yufan_chen@pku.edu.cn) (Y. Chen).

Among international markets, the Chinese financial market is of particular importance not only because of its growing size, but also of its increasing centrality in the global financial system (Billio et al., 2022). As of 2020, the total value of China's stock market has climbed to a record high of more than USD 12.2 trillion, making it the second-largest in the world after that of the US.<sup>1</sup> Moreover, recent empirical evidence shows that the world has started to move from a unipolar to a multipolar financial system in which China plays an increasingly central role (McKibbin and Fernando, 2021; Billio et al., 2022).

The Chinese stock market has several unique features. First, each stock in the Chinese market is traded only on one exchange and there are no dark pools. Second, unlike developed markets that are dominated by institutional investors, the Chinese stock market is dominated by retail investors (Li and Wang, 2010), and there is a 1.5-hour lunch break in each trading day. Third, the degree of automated trading in the Chinese market, though has increased in recent years, is still relatively low compared to developed markets such as the US, which is partly due to the tightened regulations that were introduced after the 2015 Chinese stock market crash. These factors together make research on liquidity and market microstructure in China both interesting and challenging.

In this article, we explore the measurements, patterns, and determinants of high-frequency liquidity empirically using a comprehensive tick-level dataset from the Chinese stock market. We consider high-frequency liquidity measures widely used in the literature, including measures based on trading costs (such as bid–ask spread and weighted spread), execution efficiency (such as order interarrival time, order lifetime, and order execution ratio), and price impact (such as Kyle's Lambda (Kyle, 1985), and the price impact regression of Glosten and Harris (1988)). We document their liquidity levels, analyze their distributional properties as well as intraday and cross-sectional patterns, and provide a model that explains the formation mechanism of liquidity based on order flows. To the best of our knowledge, this is by far the most comprehensive study of liquidity for the Chinese stock market in the literature.

Our dataset is taken from the *Shenzhen Stock Exchange Historical Tick Data*, which consists of tick-level data for all 1,936 stocks traded on the Shenzhen Stock Exchange from Jan. 1st, 2019 to Dec. 31st, 2021. It provides all trades, quotes, cancellations, their timestamp and price information, and market snapshots for the first ten non-empty levels of the book on both sides every three seconds. Overall, we have approximately  $1,936 \text{ (stocks)} \times 730 \text{ (trading days)} \times 33,000 \text{ (average number of events per stock per day)} = 46.64 \text{ billion events}$ . Through meticulous integration of the raw event-level data, we reconstruct the order book at the timestamp of each event, the highest possible time resolution. The reconstructed order book contains the sequences of all orders, trades, and detailed market snapshots, including the price and size at each level of the limit order book. This comprehensive dataset enables us to compare a variety of widely used liquidity measures in the literature, analyze their distributional, time-series, and cross-sectional patterns, and explore their determinants. To study cross-sectional patterns, we also join this dataset with several standard company-level characteristics such as each stock's market capitalization, institutional versus retail ownership, and whether the stock is an index constituent.

First, we find that the proportional effective spread in the Chinese market is consistently lower than that in the US market, a potential result of centralized trading for all stocks in the Chinese market. In addition, we propose a methodology to estimate the distribution of the censored order lifetime data and find that order lifetimes follow the power-law, and that order interarrival times follow Weibull—not exponential—distributions, empirically confirming the intuition proposed by Abergel et al. (2016). This implies that the Poisson flow is not an appropriate model for order flow in the Chinese A-share market.

Second, we explore the intraday patterns of liquidity measures. In contrast to the well-known U-shaped pattern of trading volume and liquidity in developed markets, we observe a W-shaped pattern in our data because the Chinese stock market is closed for a 1.5-hour lunch break around noon. In addition, the W-shape varies among liquidity measures such as the bid–ask spread and Kyle's Lambda, which highlights that different measures capture different aspects of market liquidity. Furthermore, we document novel and subtle intraday periodicities in liquidity by estimating spectral density functions for detrended intraday liquidity measures. The periodic patterns at high frequencies are consistent over three years, with lower-frequency periodicities more pronounced in 2020. This is analogous to the price clustering effect such as the whole-dollar effect (Niederhoffer, 1965; Harris, 1991) and the stock-pinning effect on option strikes (Ni et al., 2005; Avellaneda et al., 2012). We find a clustering effect for liquidity along the time-series dimension rather than the price dimension.

Third, we conducted a detailed analysis and comparison of several cross-sectional patterns of liquidity measures across the three years in 2019–2021. Taking the Kyle's Lambda as an example, liquidity measures tend to be higher for stocks with higher market capitalization, higher prices, more institutional ownership, that are index constituents,<sup>2</sup> and not trading under the special treatment (ST) status.<sup>3</sup> Kyle's Lambdas generally decrease from 2019 to 2021, which reflects an overall improvement in liquidity during this period. Interestingly, “Work-from-home” stocks have the lowest Kyle's Lambdas in 2020 among the three years, implying that these stocks have particularly high liquidity during the COVID-19 period. These empirical patterns highlight differences in liquidity with respect to both the characteristics of the underlying assets and different time periods including the COVID-19 pandemic.

Finally, to understand what determines liquidity, we develop an order-based model of the bid–ask spread—arguably the simplest but most commonly used high-frequency liquidity measure for both academics and practitioners. We define the *aggressive–passive imbalance* (API)—similar in spirit to the order flow imbalance of Cont et al. (2014)—which measures the imbalance between aggressive orders and passive orders. We show that, over short time intervals, while OFI predicts changes in prices, changes in the

<sup>1</sup> Retrieved from World Bank in March 2022: <https://data.worldbank.org>.

<sup>2</sup> For example, stocks that are part of the CSI 300 Index, which consists of 300 A-share stocks listed on the Shanghai or Shenzhen Stock Exchange.

<sup>3</sup> The Shanghai and Shenzhen Stock Exchanges implement a stock listing rule that gives special treatment (ST) or delisting risk warning to the stocks of the listed companies with abnormal financial and nonfinancial conditions, in order to indicate the risks of the stock to investors.

bid–ask spread are mainly driven by API. The significance of API as a determinant of the bid–ask spread is robust across different stocks and over our entire sample period. This result highlights a universal mechanism of spread formation with respect to order flows. In addition, we estimate the impact of API on the direction and magnitude of changes in spread separately and provide a cross-sectional analysis of the degree of API's impact with respect to several company characteristics, including stock price, market capitalization, and whether the company is primarily held by institutional or retail investors. API characterizes the dynamic of spread from the perspective of trading urgency, which provides an alternative to the inventory-based perspective offered by studies such as Grossman and Miller (1988) and Bogousslavsky and Collin-Dufresne (2023).

The remainder of the paper is organized as follows. We provide a brief review of related literature in Section 2. Section 3 describes our data. Section 4 summarizes the measurements of high-frequency market liquidity. Section 5 explores intraday and cross-sectional patterns in market liquidity. Section 6 develops the API as an important determinant of market liquidity with respect to order flows. Section 7 concludes and we provide additional technical details in Appendix.

## 2. Literature review

Liquidity is one of the most important topics in the study of market microstructure (Madhavan, 2000), with important implications both from a theoretical perspective such as price formation (Glosten and Milgrom, 1985; Amihud and Mendelson, 1986; Datar et al., 1998) and a practical perspective such as the optimal execution of large orders (Bertsimas and Lo, 1998; Almgren and Chriss, 2001).

There exists a wide range of liquidity definitions and their corresponding measurements. Foucault et al. (2013) define liquidity as “the degree to which an order can be executed within a short time frame at a price close to the security’s consensus value”. According to this definition, liquidity should measure two aspects of trading: (1) whether executing an order is timely, and (2) whether the execution price is close enough to the fair value. Examples of high-frequency liquidity measures include bid–ask spread as used in Amihud and Mendelson (1986) and Hollifield et al. (2017), order book depth and price impact as used in Biais et al. (1995) and Adrian et al. (2017), and inventory cost for market makers as used in O’Hara and Zhou (2021). There is also a large literature on using low-frequency (such as daily) data to estimate high-frequency liquidity measures across equity and bond markets, both in the US and internationally.<sup>4</sup>

One specific dimension of liquidity is price impact, also known as “depth” or “resiliency” in the language of the foundational work by Kyle (1985). Since then, several models of market impact have been proposed, including the price impact regression of Glosten and Harris (1988) and the order flow imbalance of Cont et al. (2014). There is also empirical literature estimating Kyle’s Lambda using high-frequency data (Hasbrouck, 2009; Goyenko et al., 2009; Khandani and Lo, 2011). In addition, the price impact of market orders has been investigated by Lillo et al. (2003) and Farmer et al. (2004), and the significant impact of limit orders and cancellations has been documented by Eisler et al. (2012) and Hautsch and Huang (2012).<sup>5</sup>

The patterns of liquidity have been studied across stocks and bonds, and over different time periods, though primarily based on data in the developed markets. Chordia et al. (2000, 2001) and Moshirian et al. (2017) study market- and industry-wide commonality in liquidity. Cenesizoglu and Grass (2018) disentangle bid- and ask-side liquidity using 11 years of comprehensive New York Stock Exchange limit order book data to document several empirical facts. Lee and Chung (2022) show that hidden liquidity on lit venues of the US stock market has significant effects on various measures of market liquidity. Other studies look at how market liquidity behaves in crisis periods. For example, Spatt (2010) and Adrian et al. (2017) study market liquidity during and after the financial crisis of 2007–2008; Kirilenko et al. (2017) study intraday market intermediation and liquidity during the Flash Crash; Chung and Chuwonganant (2023) and O’Hara and Zhou (2021) focus on the stock market and corporate bond market during the COVID-19 crisis, respectively.

The relationship between liquidity and asset returns has been extensively studied (Amihud and Mendelson, 1986; Acharya and Pedersen, 2005; Chen et al., 2021). It is well-known both theoretically and empirically that liquidity premiums exist for stocks (Amihud and Mendelson, 1986; Datar et al., 1998; Pástor and Stambaugh, 2003), bonds (Lin et al., 2011), and in international capital markets (Chaiet et al., 2021). In addition, the impact of algorithmic and high-frequency trading on liquidity has also been explored.<sup>6</sup>

The literature on liquidity for the Chinese stock market include Narayan and Zheng (2010, 2011), and Ma et al. (2018), who have studied how liquidity interacts with stock returns. In these studies, liquidity is usually proxied by one or a few measures. Our article contributes to this literature by providing a comprehensive analysis of high-frequency liquidity in the Chinese stock market, its similarities and differences to the US market, and what mechanism determines the formation of bid–ask spread with respect to order flows.

<sup>4</sup> See, for example, Roll (1984), Lesmond et al. (1999), Hasbrouck (2009), Goyenko et al. (2009), Corwin and Schultz (2012), Schestag et al. (2016), Fong et al. (2017), and Abdi and Ranaldo (2017).

<sup>5</sup> Additional literature on price impact models include Madhavan et al. (1997), Bouchaud et al. (2003), Hasbrouck (2007), Bouchaud et al. (2009), and Mertens et al. (2022).

<sup>6</sup> See, for example, Hendershott et al. (2011), Clark-Joseph et al. (2017), Upson and Van Ness (2017), Chung and Chuwonganant (2018), Yao and Ye (2018), Roşu (2020), Baldauf and Mollner (2020), Bongaerts and Van Achter (2021), and Li et al. (2021).

**Table 1**  
Samples of raw data for Ping An Bank (stock code: 000001) on Jan. 2nd, 2020.

snapshot		order	
TradeDate	20200102	tradestate	20200102
OrigTime	20200102093500000	TransactTime	20200102093144900
SecurityID	000001	ApplSeqNum	939273
OfferPX1	16.81	SecurityID	000001
BidPX1	16.80	Price	16.74
OfferSize1	700	OrderQty	2500
BidSize1	297100	Side	1 (Bid)
...	...	OrderType	2 (Limit order)
OfferPX10	16.90		
BidPX10	16.71		
OfferSize10	436640		
BidSize10	29900		
trade – execute		trade – cancel	
tradestate	20200102	tradestate	20200102
tradetime	20200102093337000	tradetime	20200102093337100
ApplSeqNum	1316723	ApplSeqNum	1316934
SecurityID	000001	SecurityID	000001
BidApplSeqNum	1303835	BidApplSeqNum	1302992
OfferApplSeqNum	1316722	OfferApplSeqNum	–
Price	16.78	Price	–
TradeQty	7400	TradeQty	100
ExecType	F (done)	ExecType	4 (withdrawal)

### 3. Data

The complexity of high-frequency transaction data and its huge volume warrants a detailed description of the data used in this study. Sections 3.1 and 3.2 introduce the raw dataset and the processed data after order book reconstruction, respectively. Based on the processed data, we briefly show several stylized facts of the limit order book on sample stocks in the Shenzhen Stock Exchange in Appendix F.1 of the Online Supplementary Material.

#### 3.1. Raw data

Our raw data is obtained from the *Shenzhen Stock Exchange (SZSE) Historical Tick Data*,<sup>7</sup> which contains all stocks traded in the Shenzhen Stock Exchange<sup>8</sup> (hereinafter, the Exchange) for 730 trading days from Jan. 1st, 2019 to Dec. 31st, 2021. Specifically, we use three files in the SZSE Historical Tick Data:

- (a) **snapshot** file. This file comes from `snap_level_spot` in the database and records the prices and quantities for the first ten nonempty levels on both sides of the limit order book every three seconds;
- (b) **order** file. This file comes from `hq_order_spot` in the database and contains all information about submitted orders, including time, price, quantity, order type, and whether it is a buy or a sell order;
- (c) **trade** file. This file comes from `hq_trade_spot` in the database and contains all information about the executions and cancellations in the market. For each execution event, the file records the time, price, quantity, and both the buy and sell orders involved in this execution. For each cancellation event, the canceled quantity of the order is recorded.

Table 1 provides several segments of the raw data for Ping An Bank on Jan. 2nd, 2020 as an example. The **snapshot** sample contains a snapshot of the top ten nonempty levels of the limit order book at 9:35:00.000. It shows the first ten bid prices (BidPX1–BidPX10) and ask prices (OfferPX1–OfferPX10), as well as the order quantities remaining on the bid side (BidSize1–BidSize10) and the ask side (OfferSize1–OfferSize10) at these respective prices. The **order** sample is recorded when a buy limit order is submitted at the price of RMB 16.74 with 2,500 shares at 9:31:44.900, and the exchange attached a unique ID 939273 (see `ApplSeqNum`) to this order. As for the **trade** file, there are certain differences between executions and cancellations. The “execute” sample demonstrates an execution where the buy order with ID 1303835 and the sell order with ID 1316772 are matched at RMB 16.78 with a size of 7,400 at 9:33:37.000. The “cancel” sample demonstrates that the buy order with ID 1302992 is canceled with size 100 at 9:33:37.100, and there are no sell orders and prices involved in this cancellation event.

The comprehensive SZSE Historical Tick Data allows a detailed investigation of all trading activities for stocks listed on the Exchange, thanks to the unique features of the Chinese market that each stock is traded at only one exchange where it is listed, and there are no dark pools. Therefore, we can use the dataset to reconstruct the complete order book evolution of these stocks and study their liquidity.

<sup>7</sup> See <http://www.szse.cn/English/services/dataServices/index.html>.

<sup>8</sup> Appendix A of the Online Supplementary Material introduces the trading rules for the Shenzhen Stock Exchange.

**Table 2**

The reconstructed order book of Ping An Bank (stock code: 000001) at 11:00:00 on Jan. 2nd, 2020.

Level	Price	Quantity	Order Queue <sup>a</sup>
:	:	:	:
Ask20	17.10	537100	1000 200 500 500 100 ... 800 2000 900 200 4700
:	:	:	:
Ask10	17.00	2618876	3000 400 840 500 2000 ... 500 1000 1000 1000 300
:	:	:	:
Ask2	16.92	112300	700 100 200 1600 100 ... 1100 500 1100 1000 500
Ask1	16.91	42500	800 1800 500 3500 700 ... 100 800 300 6000 200
Bid1	16.90	5000	3000 2000
Bid2	16.89	8100	2000 300 1600 2700 1000 500
:	:	:	:
Bid10	16.81	261600	2000 300 1200 40000 12000 ... 80000 800 10000 1800 1200
:	:	:	:
Bid20	16.71	63200	2000 1000 400 300 2000 ... 100 500 3000 500 300
:	:	:	:

<sup>a</sup> Order queues are displayed by the quantities of the orders in the queue separated with “|”.

<sup>b</sup> Only the part in the box is available in the snapshot file.

**Table 3**

Information of an order  $x$  after the reconstruction.

Class	Notation	Meaning
Order	$p_x$	Quote price of $x$
	$\text{sign}(v_x)$	Side of $x$
	$ v_x $	Quote size of $x$
	$t_x$	Quote time of $x$
Full limit order book $(i = 1, 2, \dots)$	$p_i^b(t_x)$	Bid $i$ price right before $t_x$
	$n_i^b(t_x)$	Bid $i$ size right before $t_x$
	$p_i^a(t_x)$	Ask $i$ price right before $t_x$
	$n_i^a(t_x)$	Ask $i$ size right before $t_x$
Event	$\mathcal{E}_x$	Execution and cancellation events related to $x$

### 3.2. Data after order book reconstruction

The empirical studies in Sections 4–6 require more granular information from the limit order book than the 3-second snapshots with only ten nonempty levels provided by the snapshot file in the raw data. Hence, we reconstruct the sequence of order arrivals and trades by re-matching the orders according to the trading rules in the Exchange (Shenzhen Stock Exchange, 2016, 2020), with the detailed procedures outlined in Appendix C of the Online Supplementary Material. After the reconstruction, the limit order book is available for all nonempty levels at any time.

Table 2 provides the limit order book after order book reconstruction of Ping An Bank (stock code: 000001) at 11:00:00 on Jan. 2nd, 2020, which greatly enriches the original snapshot in the snapshot file. Our reconstructed limit order book contains not only the prices and quantities, but also the queue positions of all remaining orders at each level, which provides the most granular level of available information (see definitions of the limit order book in, for example, Harris (2003) and Abergel et al. (2016)). As a result, we are able to provide the snapshot of the order book right before the submission of each order—and its corresponding execution and cancellation events—during the continuous auction. Table 3 summarizes the available information of an example order after the reconstruction,<sup>9</sup> and these information are recorded in a new file, `order_based_data`.

### 4. Measurements of high-frequency market liquidity

Having established a granular view of the limit order book that contains information for all orders at all price levels at any time during a trading day, we are ready to measure the high-frequency liquidity in the Chinese stock market and document their distributions.

According to Foucault et al. (2013), liquidity is the degree to which an order can be executed within a short time frame at a price close to the security's consensus value. To determine whether an order can be executed at a better price, one can measure the liquidity based on trading cost, and to determine whether an order can be executed within a short time frame, one can measure the liquidity based on execution efficiency. In addition, price impact plays a crucial role in determining the trading cost of large orders. Therefore, we consider several liquidity measures grouped into the above three categories. The measures used in this paper are drawn from recent monographs of market liquidity (Foucault et al., 2013; Abergel et al., 2016) and the annual market quality reports of SZSE (Shenzhen Stock Exchange, 2021).

<sup>9</sup> For the details of the notations used in Table 3 and hereafter, see Appendix B of the Online Supplementary Material.

#### 4.1. Measures based on trading cost

The most intuitive and commonly used measure of the cost of a small trade is the quoted spread.

**Definition 1** (*Spread (Foucault et al., 2013)*). The (*quoted*) *spread* at time  $t$  is defined as the difference between the best ask price  $a(t)$  and the best bid price  $b(t)$  at time  $t$ , i.e.,

$$S(t) := a(t) - b(t).$$

If this spread is normalized by the midprice  $m(t) := (a(t) + b(t))/2$ , the average of best bid price and best ask price, one obtains the *proportional (quoted) spread*:

$$s(t) := \frac{S(t)}{m(t)} = \frac{a(t) - b(t)}{m(t)}.$$

In addition to the quoted spread, which measures the difference between the best bid and ask prices, another important measure of trading cost is the effective spread. While the quoted spread reflects the cost of trading in an ideal setting where trades occur at the best bid or ask price, the effective spread captures the actual cost incurred by traders due to market conditions, such as price improvement or slippage.

**Definition 2** (*Effective Spread (Foucault et al., 2013)*). The *effective spread* at time  $t$  is defined as twice the difference between the mid price  $m(t)$  and the trading price  $p(t)$  at time  $t$ , i.e.,

$$S^{\text{eff}}(t) := 2|p(t) - m(t)|.$$

If this spread is normalized by the midprice  $m(t)$ , one obtains the *proportional effective spread*:

$$s^{\text{eff}}(t) := \frac{S^{\text{eff}}(t)}{m(t)} = \frac{2|p(t) - m(t)|}{m(t)}.$$

This measure considers the distance between the transaction price and the midprice, thus offering a more realistic assessment of trading costs experienced by market participants.

The spread on larger orders can be gauged similarly, computing a weighted average bid-ask spread from the quotes posted. For instance, for the buy and sell limit orders posted at a given point in time, we use the following definition.

**Definition 3** (*Weighted-average Spread (Foucault et al., 2013)*). Denote the average execution price of a buy (sell) market order as  $\bar{a}(t, q)$  ( $\bar{b}(t, q)$ ), where a hypothetical trader places a market order at time  $t$  to buy (sell)  $q$ -RMB worth of the stock. Then, the *weighted-average spread* (wavg-spread, for simplicity) with a total stock worth of  $q$  at time  $t$  is defined as

$$S(t, q) := \bar{a}(t, q) - \bar{b}(t, q),$$

and the *proportional weighted-average spread* (proportional wavg-spread, for simplicity) is

$$s(t, q) := \frac{S(t, q)}{m(t)} = \frac{\bar{a}(t, q) - \bar{b}(t, q)}{m(t)}.$$

It is easy to see that  $S(t, q)$  and  $s(t, q)$  are monotonically increasing with respect to  $q$ .

#### 4.2. Measures based on execution efficiency

Here we introduce several measures based on execution efficiency from the perspectives of a single order and the entire market, respectively.

We begin with a few notations in addition to the `order` file and `trade` file introduced in Section 3.1 for each trading day. For an order  $x$  recorded in the `order` file, we denote its order submission time, quote price, and size as  $t_x$ ,  $p_x$ , and  $v_x$ , respectively. An event  $e$  recorded in the `trade` file can be either an execution or a cancellation. We denote the event time and the size of  $e$  as  $t_e$  and  $v_e$ , respectively. If  $e$  is an execution event, we further denote its execution price as  $p_e$ .<sup>10</sup>

From the perspective of a single order, the most direct measure of execution efficiency is the total length of time during which the order remains on the order book, weighted by the sizes of all trades and cancellations associated with this order.

**Definition 4** (*Lifetime of an Order (Abergel et al., 2016)*). The lifetime of an order  $x$  is defined as

$$\overline{T}_x := \frac{\sum_{e \in \mathcal{E}_x} v_e t_e}{v_x} - t_x,$$

where  $\mathcal{E}_x$  is the collection of events that the order  $x$  is involved in.

<sup>10</sup> For the rigorous definition of these notations, see Appendix B of the Online Supplementary Material.

In addition, the market is likely to be more liquid if orders arrive more frequently. This leads us to define the interarrival time between two consecutive orders.

**Definition 5 (Interarrival Time (Abergel et al., 2016)).** We define the *interarrival time* between an order  $x$  and its preceding order as the difference between  $t_x$  and its preceding order's submission time, which we denote by  $\Delta t_x$ <sup>11</sup>:

$$\Delta t_x := t_x - \max \{t_y : y \in \mathcal{X}, t_y < t_x\},$$

where  $\mathcal{X}$  is the collection of all submitted orders during the day.

From the aggregate market's perspective, we can measure the liquidity by the percentage of all submitted orders that are executed eventually. To do this, we first define several different cases in which an order can be executed. An order  $x$  is said to be executed at a single price if  $\#\{p_e : e \in \mathcal{E}_x\} = 1$ . An order  $x$  is said to be executed at a single time if  $\#\{t_e : e \in \mathcal{E}_x\} = 1$ . An order  $x$  is said to be executed at the best price if

$$\{p_e : e \in \mathcal{E}_x\} = \begin{cases} \{a(t_x)\}, & x \text{ is a sell order,} \\ \{b(t_x)\}, & x \text{ is a buy order,} \end{cases}$$

where  $a(t_x)$  and  $b(t_x)$  are the best ask price and best bid price at time  $t_x$ , respectively. We denote by  $\mathcal{X}^{\text{sp}}$ ,  $\mathcal{X}^{\text{st}}$ , and  $\mathcal{X}^{\text{bp}}$  the collections of orders executed at a single price, at a single time, and at the best price, respectively.

Now we are ready to define the execution ratio of all orders in a trading day.

**Definition 6 (Execution Ratio (Shenzhen Stock Exchange, 2021)).** The *total execution ratio* is defined as

$$r^e = 2 \times \frac{\sum_{e \in \mathcal{E}^e} v_e}{\sum_{x \in \mathcal{X}} v_x},$$

where  $\mathcal{E}^e$  is the collection of all execution events during the day. In addition, the execution ratio at a single price, at a single time, and at the best price is defined as

$$r^{\text{sp}} = \frac{\sum_{x \in \mathcal{X}^{\text{sp}}} v_x}{\sum_{x \in \mathcal{X}} v_x}, \quad r^{\text{st}} = \frac{\sum_{x \in \mathcal{X}^{\text{st}}} v_x}{\sum_{x \in \mathcal{X}} v_x}, \quad r^{\text{bp}} = \frac{\sum_{x \in \mathcal{X}^{\text{bp}}} v_x}{\sum_{x \in \mathcal{X}} v_x},$$

respectively.

There is a constant of 2 in the definition of  $r^e$  because, for an execution event  $e$ , both the buy and the sell orders involved are executed. It is easy to verify that all four execution ratios above are indeed between 0 and 1, and that  $r^e$  is the largest among the four execution ratios. Moreover,  $r^{\text{sp}} \geq r^{\text{bp}}$  because an order executed at the best price is also executed at a single price.

#### 4.3. Measures based on price impact

We consider two methods to measure liquidity based on price impact: Kyle's Lambda, as proposed by [Kyle \(1985\)](#), and the price impact regression of [Glosten and Harris \(1988\)](#).

*Kyle's Lambda* ([Kyle, 1985](#)). In the pioneering work of [Kyle \(1985\)](#), the reciprocal of Kyle's Lambda can be regarded as a measure of market depth. Subsequent empirical studies, such as those of [Hasbrouck \(2009\)](#), [Goyenko et al. \(2009\)](#), and [Khandani and Lo \(2011\)](#), estimate Kyle's Lambda using high-frequency data and view it as a liquidity measure. We follow [Goyenko et al. \(2009\)](#) who estimate Kyle's Lambda ( $\lambda$ ) using the following regression without intercept:

$$r_n = \lambda S_n + \epsilon_n, \tag{1}$$

where for the  $n$ th sampling period,  $r_n$  is the log-return of the stock,  $S_n$  is the signed square-root quantity defined by  $S_n = \sum_k \text{sign}(v_{kn}) \sqrt{|v_{kn}|}$ ,  $v_{kn}$  is the signed quantity of the  $k$ th execution in the  $n$ th sampling period, and  $\epsilon_n$  is the error term. A lower estimated  $\lambda$  indicates better liquidity.

*Price impact regression* ([Glosten and Harris, 1988](#)). Following the work of [Roll \(1984\)](#), many price impact regression models are proposed to capture the cost of providing liquidity caused by adverse selection, order-processing, and risky inventory holdings. In these models, price changes are regressed on contemporaneous and lagged measures of order flow, and the estimated coefficients can be regarded as liquidity measures based on price impact.

The exact specifications of the price impact regression depend on the assumptions for the sources of market illiquidity and the properties of the order arrival process. Here we follow the price impact regression of [Glosten and Harris \(1988\)](#) which decomposes the quote spread into the adverse selection cost component and the order-processing cost component. Specifically, the price change

<sup>11</sup> We have assumed without loss of generality that two orders are never submitted at exactly the same time. This is mostly true in practice for modern markets with nanosecond time resolution.

**Table 4**  
Liquidity measures in the price impact regression model.

Cost	Measure	Meaning
Order-processing cost	$\gamma_0$	The constant part of order-processing cost
	$\gamma_1$	The order-processing cost that varies with the execution quantity
Adverse selection cost	$\lambda_0$	The influence of the execution direction
	$\lambda_1$	The influence of the execution quantity

between two consecutive executions is modeled as<sup>12</sup>:

$$p_n - p_{n-1} = \lambda_0 d_n + \lambda_1 q_n + \gamma_0(d_n - d_{n-1}) + \gamma_1(q_n - q_{n-1}) + \epsilon_n, \quad (2)$$

where  $p_n$ ,  $d_n$ , and  $q_n$  are the price, direction, and quantity of the  $n$ th execution, respectively. In particular, if the execution is bid-initiated, then  $d_n = 1$  and  $q_n > 0$ ; otherwise  $d_n = -1$  and  $q_n < 0$ . In regression (2), the parameters  $\gamma_0$ ,  $\gamma_1$ ,  $\lambda_0$ , and  $\lambda_1$  can be regarded as liquidity measures based on the price impact, whose interpretations are summarized in Table 4.

In Appendix F of the Online Supplementary Materials, we provide some additional empirical results on the liquidity measures defined in this section for a few stocks on specific trading dates.

## 5. Patterns of high-frequency market liquidity

In this section, we explore the intraday and cross-sectional patterns of the liquidity measures outlined in Section 4, for all 1,936 stocks in our sample from 2019 to 2021.

### 5.1. Intraday patterns

We begin by focusing on the intraday patterns of liquidity measures by aggregating the entire year of data into a one-day time series. More specifically, for each stock, we group data based on the following rules:

- For `order_based_data`, we divide data into 3-second windows where each window contains all orders submitted within a fixed 3-second period, regardless of their trading dates. For example, we aggregate all orders submitted from 9:30:00 a.m. to 9:30:03 a.m. for the entire year into one group.
- For `snapshot`, similarly, we group data based on the time of the snapshot regardless of their trading dates. In other words, we aggregate all snapshots at 9:30:03 a.m. into one group, and each group contains exactly 730 snapshots (trading days in 2019–2021).

In Section 5.1.1, we study the intraday shape for several trading cost-based liquidity measures including the spread. We use Fourier analysis to further examine the intraday periodicity for average spread. In Section 5.1.2, we consider order lifetime, an execution efficiency-based liquidity measure. We use power-law distribution to model censored order lifetimes and propose a new method to estimate its parameter. The intraday pattern of the estimated parameter is also studied. For price impact-based liquidity measures, we primarily focus on their cross-sectional patterns in Section 5.2.

#### 5.1.1. Intraday shape and periodicity for spread

It is well known that liquidity measures and variables related to trading activity demonstrate U-shaped intraday patterns. For example, trading volume is significantly larger after market open and before market close compared to the rest of the day (Admati and Pfleiderer, 1988).

*W-shaped liquidity.* To understand intraday patterns for liquidity in the Chinese A-share market, we calculate different intraday quantiles for six spread measures: spread, proportional spread, 5 millions wavg-spread, 5 millions proportional wavg-spread, effective spread, and proportional effective spread as defined in Section 4. Specifically, for each stock, we calculate the cross-date average spread in each 3-second window to obtain a time series with 4740 data points. For each 3-second average, we calculate the cross-sectional 1%, 25%, 50%, 75%, and 90% quantiles for a given spread measure among all stocks. We show results for spread (Fig. 1(a)), proportional spread (Fig. 1(b)), effective spread (Fig. 1(c)), proportional effective spread (Fig. 1(d)), 5 millions wavg-spread (Fig. 1(e)), and 5 millions proportional wavg-spread (Fig. 1(f)).

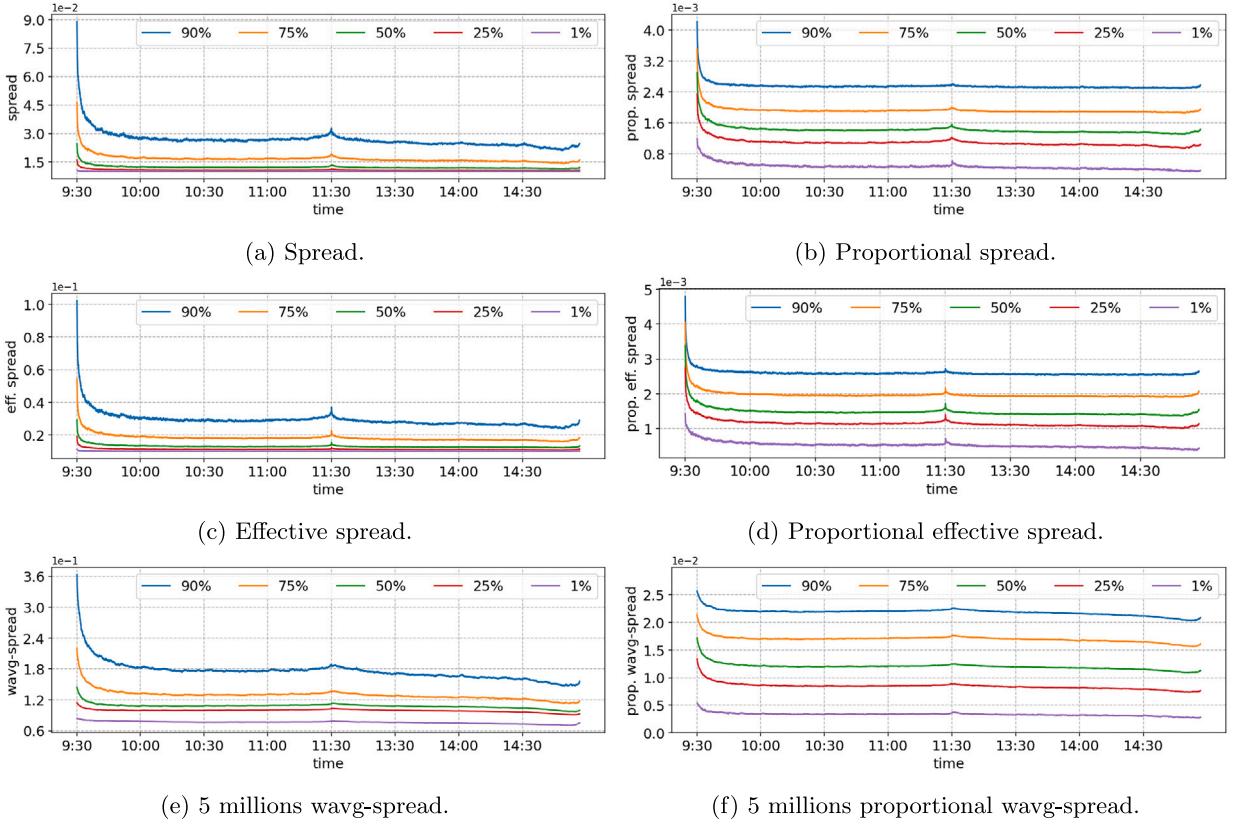
These results suggest a W-shaped pattern rather than the standard U-shaped pattern seen in other markets.<sup>13</sup> Taking spread as an example in Fig. 1(a), each quantile shows a high spread at the beginning of the day, and there exist additional peaks near 11:30 and 13:00. The spread also rises slightly immediately before the market close. These patterns W-shaped patterns are likely a result of the A-share market being closed for one and a half hours at noon for the lunch break.

<sup>12</sup> We give the details for the price impact regression model (2) in Appendix D of the Online Supplementary Material. In addition, the model is closely related to the quoted spread. In particular, Glosten and Harris (1988) demonstrated that the quoted spread at the time of the  $n$ th execution is

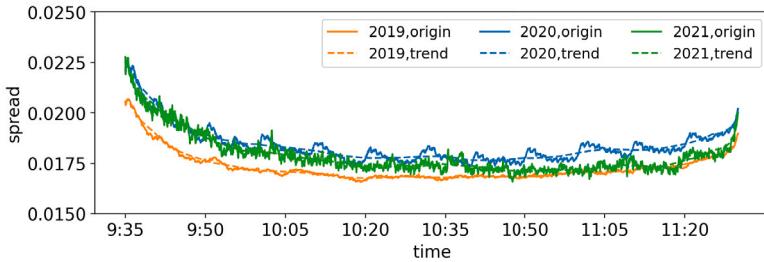
$$2(\lambda_0 + \lambda_1 |q_n| + \gamma_0 + \gamma_1 |q_n|),$$

where  $2(\lambda_0 + \lambda_1 |q_n|)$  and  $2(\gamma_0 + \gamma_1 |q_n|)$  are the adverse selection cost component and the order-processing cost component, respectively.

<sup>13</sup> One may also refer to this as a double-U-shaped pattern; see, for example, Tian and Guo (2007).



**Fig. 1.** Intraday time-series plot for different quantiles of spread measures across all stocks from 2019 to 2021.

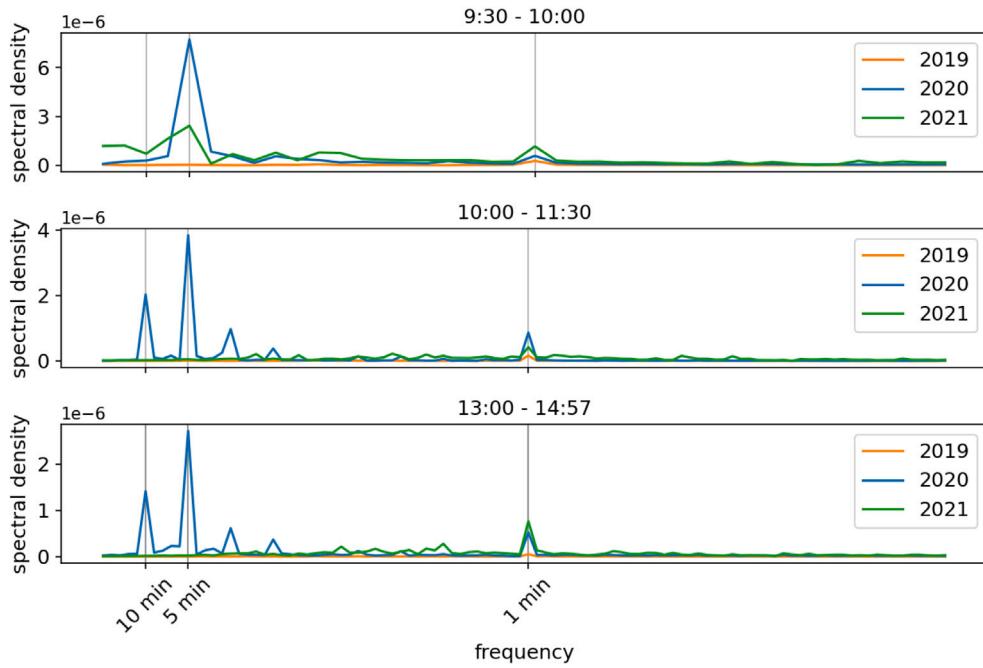


**Fig. 2.** Cross-sectional average spread over all stocks from market open to lunch break in 2019, 2020, and 2021.

*Intraday periodicity.* In addition to the W-shaped patterns, we discover more granular intraday periodicities in these spread measures. Fig. 2 shows the intraday time series of cross-sectional average spread of all stocks in each 3-second window in 2019, 2020, and 2021, as well as a trend estimated by a rolling window average. The time series exhibits spiked periodicity, where the average spread climbs up to the peak quickly and then slowly decays over time, a phenomenon consistent with observations in [Brouard and Nikiforov \(2014\)](#). In particular, low-frequency periodicities are more pronounced in 2020.

To quantitatively investigate the intraday periodicity, we estimate the spectral density of detrended time series by the Fourier transform for sample autocovariances, which is a classical method in time series analysis ([Brockwell and Davis, 2009](#)). Specifically, we first detrend the raw time series using a rolling window of 100 data points (corresponding to 5 minutes in natural time) and then estimate the autocovariance function for three time intervals separately: 9:30 to 10:00, 10:00 to 11:30, and 13:00 to 14:57. Finally, we implement Fourier transform on the autocovariance function to estimate the spectral density of the detrended time series.<sup>14</sup>

<sup>14</sup> Estimating the spectral density via the Fourier transform of the autocovariance function assumes that the original time series can be decomposed into the sum of cosine functions. Though the trigonometric basis functions may not be ideal to capture the asymmetry around spikes in the time series, the Fourier



**Fig. 3.** Estimated spectral density for detrended time series of average spread during 9:30–10:00 (top panel), 10:00–11:30 (middle panel) and 13:00–14:57 (bottom panel) in 2019, 2020, and 2021.

Because the accuracy of autocovariance function estimation depends on the length of the time series, and the three time intervals we choose have different lengths, we estimate the autocovariance with different lags for different series:

- 9:30–10:00: time-series length is 600 and the max lag is 200 for the estimated autocovariance function;
- 10:00–11:30: time-series length is 1800 and the max lag is 500 for the estimated autocovariance function;
- 13:00–14:57: time-series length is 2340 and the max lag is 500 for the estimated autocovariance function.

**Fig. 3** shows the estimated spectral density function at different frequencies. Overall, the periodic patterns at 1 minute are consistent across three years, with the spectral densities in 9:30–10:00 and 13:00–14:57 increasing from 2019 to 2021. This implies that more trading activities occur at the 1-minute frequency.<sup>15</sup> Wu et al. (2022) observed similar periodic patterns for trading volume, and they suggest that these patterns may be attributed to two factors: first, investors' tendency to execute trades at round minutes; second, the common practice in algorithmic trading with repeated instructions at fixed time increments. Similar arguments have also been put forth by Broussard and Nikiforov (2014) and Muravyev and Picard (2022).

More interestingly, **Fig. 3** shows more significant periodic patterns at 5- and 10-minute frequencies in 2020. From 9:30 to 10:00, there exists a strong periodicity at 5 minutes. In 10:00–11:30 and 13:00–14:57, the strongest frequency is at 5-minute, followed by 10-minute.<sup>16</sup> A possible explanation for this phenomenon is that retail investors, who typically trade less frequently than institutional investors, were more active during the COVID-19 pandemic (Ozik et al., 2021).

#### 5.1.2. Censoring and periodicity for order lifetime

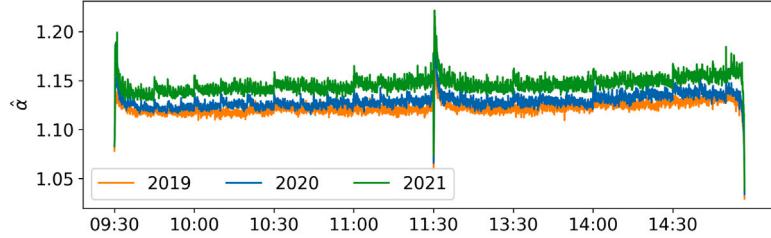
We study the intraday pattern for order lifetimes by fitting a power-law distribution for order lifetime data in each 3-second window. We then study periodicity in order lifetimes through the estimated parameters of the power-law distribution.

**Fitting censored order lifetimes.** The observed order lifetimes are censored because orders are automatically canceled every day at market close. The specific censored thresholds are determined by the difference between each order's submission time and market close. Thus, simple estimation based on the observed data leads to systematic biases. We propose a novel method below to deal with this problem.

analysis is still a simple and effective tool for us to rigorously uncover periodicities.

<sup>15</sup> We also observe similar patterns of intraday periodicity in both the number of trades and the trading volumes.

<sup>16</sup> The estimated spectral density function is robust with regard to how we split the time period. We also observe similar patterns of intraday periodicity in both the number of trades and the trading volumes.



**Fig. 4.** Time-series of fitted parameters,  $\hat{\alpha}$ , of the power-law distribution for order lifetimes of Ping An Bank (stock code: 000001) in 2019, 2020, and 2021.

Following [Abergel et al. \(2016\)](#), we model order lifetimes,  $\bar{T}_x$ , for an order  $x$  by a power-law distribution whose probability density function  $f_{\bar{T}_x}(y)$  is given by:

$$f_{\bar{T}_x}(y) = ky^{-\alpha}, \quad y > a, \quad (3)$$

where  $a > 0$  is a pre-specified exogenous constant,  $\alpha > 1$  is an unknown parameter, and  $k = \frac{\alpha-1}{a^{-\alpha+1}}$  is an auxiliary parameter to guarantee that (3) is a proper probability density function. We denote the closing time of each day to be  $T$ . This model assumption is consistent with our empirical results in Appendix F.2.2. Denote the submission time of order  $x$  by  $t_x$  (and therefore, with censored threshold  $T - t_x$ ), then the likelihood function is given by:

$$L(\alpha) = \prod_{a < \bar{T}_x < T - t_x} \frac{\alpha-1}{a^{-\alpha+1}} \bar{T}_x^{-\alpha} \prod_{\bar{T}_x = T - t_x} \frac{\alpha-1}{a^{-\alpha+1}} \frac{(T - t_x)^{-\alpha+1}}{\alpha-1} = \prod_{a < \bar{T}_x < T - t_x} \frac{\alpha-1}{a^{-\alpha+1}} \bar{T}_x^{-\alpha} \prod_{\bar{T}_x = T - t_x} \frac{\bar{T}_x^{-\alpha+1}}{a^{-\alpha+1}}, \quad (4)$$

and the maximum likelihood estimate of  $\alpha$  under censored data is given by:

$$\hat{\alpha} = 1 + \frac{\sum_{a < \bar{T}_x < T - t_x} 1}{\sum_{a < \bar{T}_x} \ln \bar{T}_x - \sum_{a < \bar{T}_x} \ln a}. \quad (5)$$

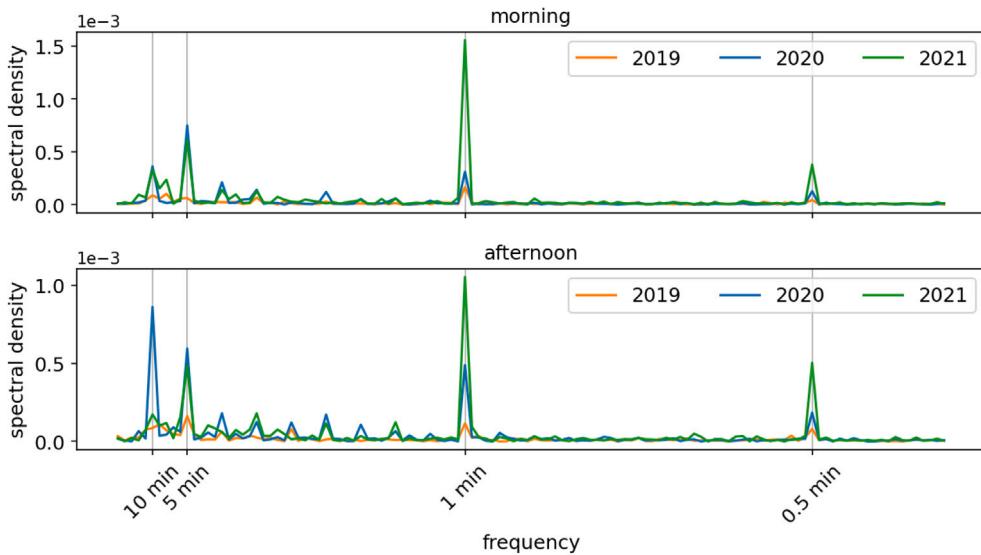
Parameter  $\alpha$  in power-law distribution captures the heaviness of its tail. A larger value of  $\alpha$  represents thinner tails, which imply higher levels of execution efficiency in the market. Similarly, our maximum likelihood estimate  $\hat{\alpha}$  in (5) is negatively correlated with the “log-weighted average” of order lifetimes. In other words, a dataset with shorter observed order lifetimes leads to larger values of estimated  $\hat{\alpha}$ .

For each stock, we group order lifetime data for all passive orders in the entire year into 3-second windows based on their submission times. For each group, we perform maximum likelihood estimation according to (5), where  $a$  is set to be 100 ms.<sup>17</sup> Fig. 4 shows the time series of estimated  $\hat{\alpha}$  for Ping An Bank as an example. There are three sharp drops in the estimated  $\hat{\alpha}$  at the morning market open, the afternoon market open, and the afternoon market close, and  $\hat{\alpha}$  climbs up rapidly in 20 seconds immediately after the first two drops. This is a result of large order submission volumes at prices far from the market price right after the morning and the afternoon market open, many of which remain in the order book until the market close. However, after a short period of price discovery (approximately 20 seconds), most orders are submitted at a price relatively close to the market price, and therefore market liquidity in terms of execution efficiency quickly recovers to a high level.

*Intraday periodicity.* Market liquidity based on execution efficiency, as measured by the estimated time series of  $\hat{\alpha}$ , shows similar intraday periodicity and asymmetry around spikes as spread measures in Section 5.1.1. We apply the same spectral analysis directly to the estimated  $\hat{\alpha}$  for Ping An Bank. We split the time series into morning and afternoon sessions and estimate their spectral densities separately, as shown in Fig. 5. In 2019, only weak periodic patterns at 5-, 1- and 0.5-minute frequencies exist, and they become more significant at 10-, 5-, 1-, and 0.5-minute frequencies in 2020. After that, the periodicities at 1- and 0.5-minute frequencies are more pronounced in 2021, while the pattern at 10-minute frequency becomes weaker.

Furthermore, we perform the same analysis for all 1,936 stocks in our sample. The two strongest frequencies for each sequence are summarized in Table 5. In 2020, most stocks have strong 10-minute and 5-minute frequency densities, and the 10-minute frequency density in the afternoon is stronger than that in the morning. This again confirms the whole-minute temporal effect in the transactions—and therefore market liquidity—of the Chinese A-share market, which is universal across all individual stocks in our sample. However, in 2019 and 2021, most stocks exhibit highest frequencies of 5 minutes and 1 minute, respectively, both higher than in 2020. As discussed in Section 5.1.1, this may be due to the increase in retail trading activity during the COVID-19 pandemic.

<sup>17</sup> In fact, different  $a$  gives rather similar results, and thus we choose  $a = 100$  ms for simplicity.



**Fig. 5.** Estimated spectral density for detrended time series of average  $\hat{\alpha}$  in the morning (top panel) and afternoon of Ping An Bank (stock code: 000001) in 2019, 2020, and 2021.

**Table 5**  
A breakdown of the first and second strongest frequencies for  $\hat{\alpha}$  across all stocks.

	2019				2020				2021			
	1st	num.	2nd	num.	1st	num.	2nd	num.	1st	num.	2nd	num.
Morning	5 min.	1,261	10 min.	442	10 min.	1,021	5 min.	961	1 min.	1,785	5 min.	944
	10 min.	254	2.5 min.	333	5 min.	935	10 min.	838	5 min.	167	0.5 min.	592
	Others	566	Others	1,306	Others	125	Others	282	Others	129	Others	545
Afternoon	5 min.	1,097	10 min.	418	10 min.	1,564	5 min.	1,313	1 min.	1,672	0.5 min.	539
	10 min.	294	2 min.	300	5 min.	330	10 min.	313	5 min.	131	10 min.	491
	Others	690	Others	1,363	Others	187	Others	455	Others	278	Others	1,051

## 5.2. Cross-sectional patterns

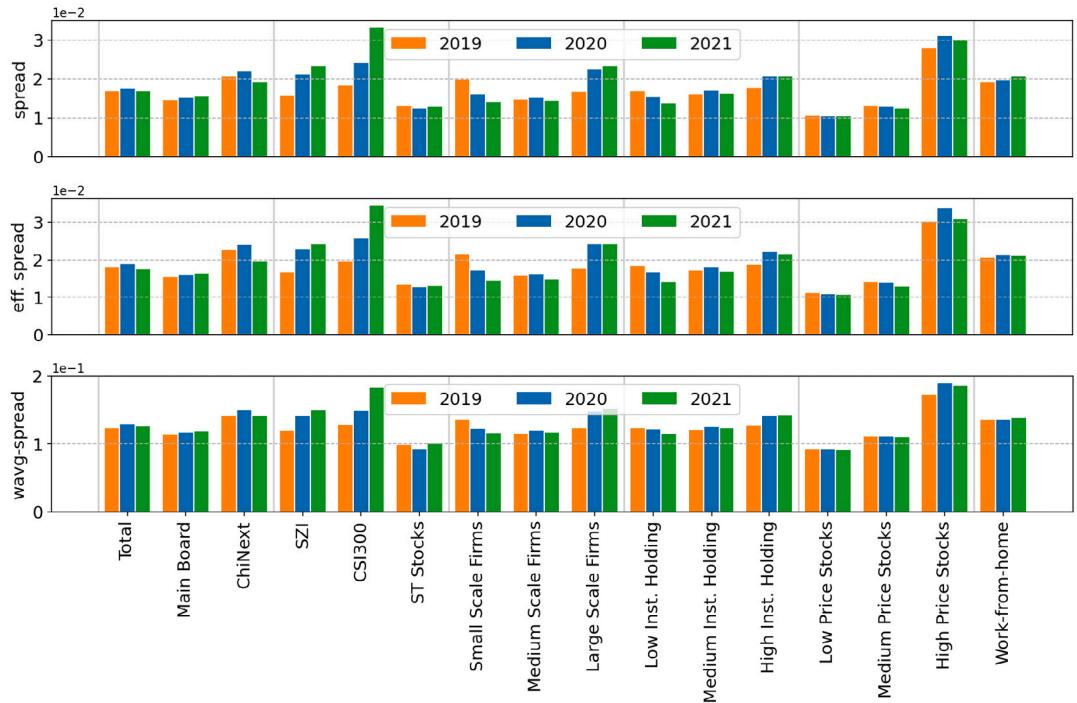
We now turn our attention to cross-sectional patterns of different liquidity measures and investigate whether certain stock and company characteristics are associated with higher or lower liquidity levels. Appendix E of the Online Supplementary Material provides details for the categorization of stocks based on stock and company characteristics.

### 5.2.1. Measures based on trading cost

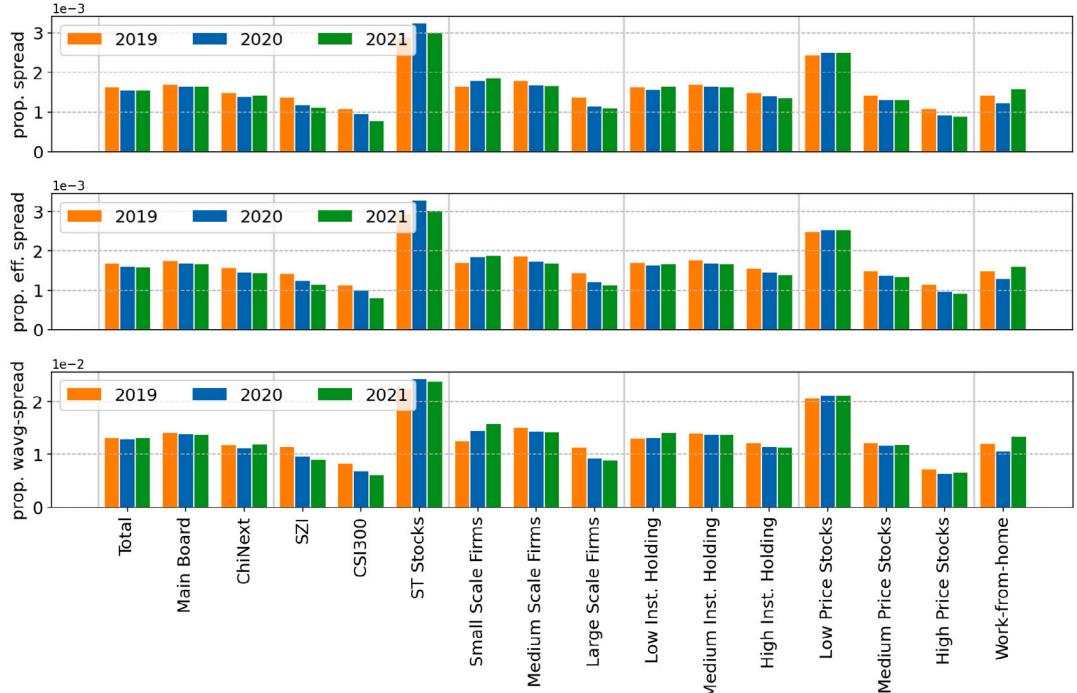
Fig. 6 shows the cross-sectional patterns of six measures based on trading costs defined in Section 4.1: average spread, average proportional spread, average effective spread, average proportional effective spread, and average 5 millions proportional wavg-spread in different categories of stocks. We find that special treatment (ST) stocks have the largest average proportional spread, average proportional effective spread, and average 5 millions proportional wavg-spread, which indicates a lack of liquidity. By contrast, constituents of the CSI 300 Index (in SZSE) have very small average proportional spread, average proportional effective spread, and average 5 millions proportional wavg-spread, which indicates their high liquidity.

Another interesting phenomenon is that stocks with higher sizes (in 2020, and 2021), institutional holdings, and prices have larger spreads yet smaller proportional spreads. The opposite dependences of the spread and the proportional spread on the stock prices suggest the existence of a low-price effect in China. In particular, nearly 52% of the stocks in our sample have average close prices lower than RMB 10 yuan and, therefore, their spreads are usually one tick. These stocks are sometimes referred to as large tick assets (Dayri and Rosenbaum, 2015), and their spreads remain one tick as the stock prices increase, leading to lower proportional spreads.

When comparing the results over the three years, we observe that the average effective spread, average proportional effective spread, and average 5 millions proportional wavg-spread of the constituents of the CSI 300 Index (in SZSE) and the SZI Index are all decreasing from 2019 to 2021. This trend highlights the improvement in liquidity of these high-quality stocks in the Chinese market from the perspective of trading costs. Furthermore, all three average proportional spreads for “Work-from-home” stocks in 2020 are the smallest among the three years, reflecting improved liquidity for these stocks during the COVID-19 lockdown period.

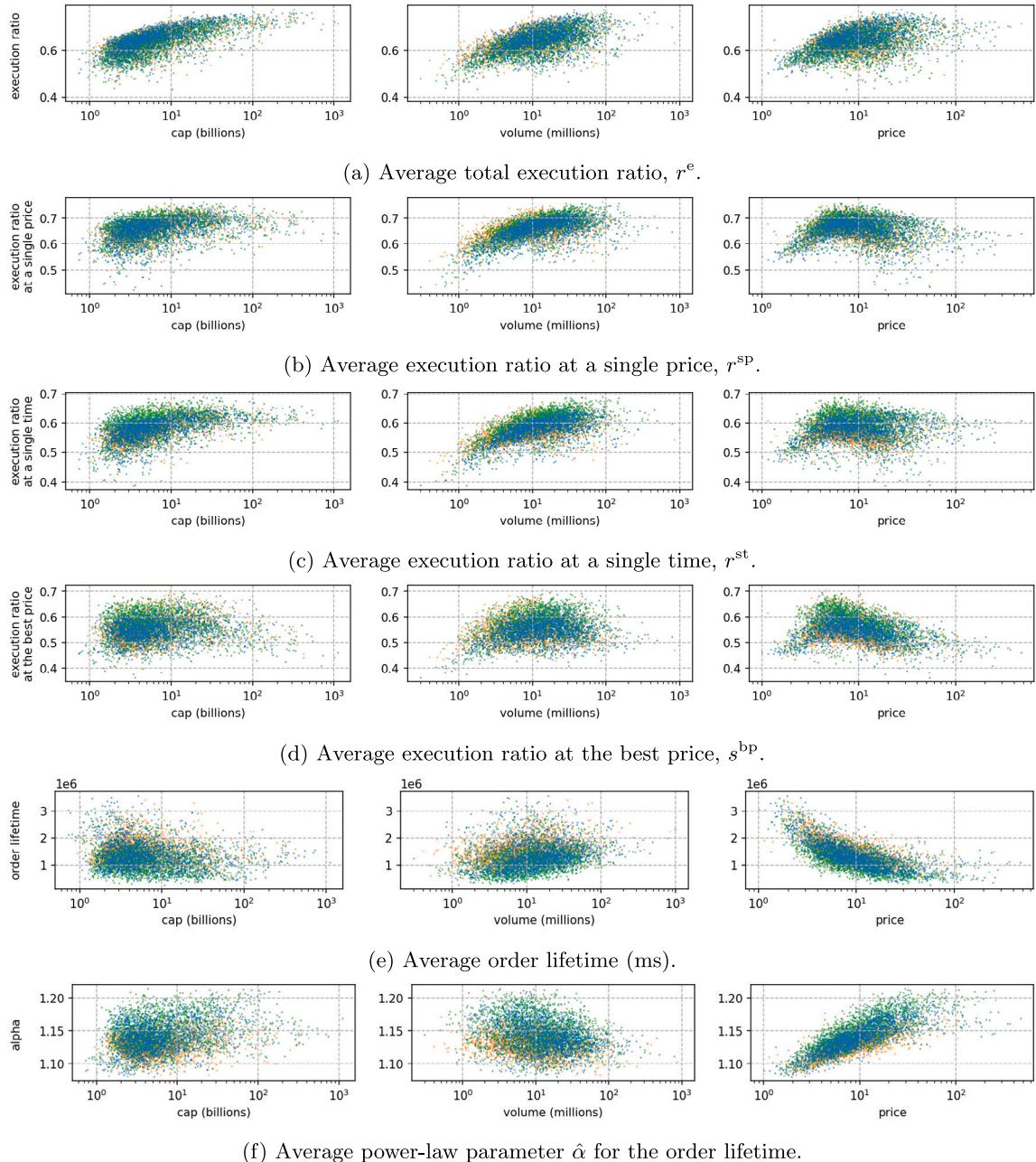


(a) Spread (top panel), effective spread (middle panel), and 5 millions wavg-spread (bottom panel).



(b) Proportional spread (top panel), proportional effective spread (middle panel), and 5 millions proportional wavg-spread (bottom panel).

Fig. 6. Cross-sectional patterns of liquidity measures based on the trading cost in different categories of stocks in 2019, 2020, and 2021.



**Fig. 7.** Scatter plots of different measures based on execution efficiency versus market value, average volume, and average price for all stocks in 2019 (orange), 2020 (blue), and 2021 (green). Each data point represents one stock. The y-axis represents a particular liquidity measure, and the x-axis represents a particular company characteristic. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, we note that the Chinese market generally exhibits lower average proportional effective spreads compared to the US market, and the average proportional effective spread of CSI 300 stocks in the Chinese market is converging towards that of S&P 500 stocks in the US market. In particular, the middle panel of Fig. 6(b) shows that the average proportional effective spread across most stocks in the Chinese market remains consistently below 20 bps each year from 2019 to 2021. This is lower than the level in the US market, which is documented to be 21 bps during August and September, 2019 (Adams et al., 2024, Table 1) and 25.8 bps from Jan. 21st to Jun. 11th, 2020 (Ozik et al., 2021, Table 1). In addition, the average proportional effective spreads for CSI 300 components and high price stocks in the Chinese market decrease from 11 bps in 2019 to 8 bps in 2021. In contrast, during the same period, the average proportional effective spreads for S&P 500 stocks increase from 4 to 6 bps (Chinco and Sammon, 2024, Fig. C.2). The lower proportional effective spreads in the Chinese market may partly reflect the advantages of its centralized trading structure.

### 5.2.2. Measures based on execution efficiency

We examine the cross-sectional patterns for a wide range of liquidity measures based on execution efficiency, as defined in Section 4.2. For each stock we study the relationship of these measures and three stock characteristics—average market value, average volume, and average price. Fig. 7(a) shows the scatter plots of average total execution ratio  $r^e$  versus three characteristics of 1,936 stocks in 2019–2021. The same relationships are shown for  $r^{sp}$ ,  $r^{st}$ , and  $r^{bp}$ , in Figs. 7(b), 7(c), and 7(d), respectively. Most stocks have execution ratio  $r^e$  between 50% and 75%, execution ratio at a single price  $r^{sp}$  between 45% and 70%, execution ratio at a single time  $r^{st}$  between 40% and 65%, and execution ratio at the best price  $r^{bp}$  between 40% and 65%.

The average total execution ratio,  $r^e$ , the average execution ratio at a single price,  $r^{sp}$ , and the average execution ratio at a single time,  $r^{st}$ , each have a similar relationship with the average market value and average volume, which is consistent from 2019 to 2021. Overall, larger stocks (with larger market value) and more actively traded stocks (with larger volume) have higher  $r^e$ ,  $r^{sp}$ , and  $r^{st}$ . For stocks with market values less than 10 billion RMB, most of their execution ratios  $r^e$  are around 65%, while for stocks with market values larger than 10 billion RMB, their execution ratios lie between 70% and 75%. However, this relationship becomes less pronounced in the average execution ratio at the best price,  $r^{bp}$ . On the other hand, the relationships between stock price and these four measures have different patterns. Indeed, as the average stock price increases from near 0, these four execution ratios also increase, but when the average stock price rises to above RMB 10, these execution ratios begin to decline.

In addition, Fig. 7(e) shows the scatter plots of average order lifetimes as compared to the market value, volume, and average stock price for all 1,936 stocks. The average order lifetimes of all stocks lie between 200 and 3,000 seconds and most of them concentrate between 500 and 1,500 seconds. The average order lifetime decreases as the average price increases, which suggests that stocks with higher prices tend to have better liquidity. Fig. 7(f) presents the same scatter plot for average estimated power-law parameter  $\hat{\alpha}$  for order lifetime. Nearly all stocks have an average  $\hat{\alpha}$  of between 1.08 and 1.20. Moreover, it suggests that  $\hat{\alpha}$  is positively correlated with the average stock price and negatively correlated with the order lifetime, which confirms the intuition mentioned at the end of Section 5.1.2. Meanwhile, most stocks with an average price greater than RMB 10 have an average  $\hat{\alpha}$  larger than 1.13, and vice versa.

Finally, Fig. 8 examines the cross-sectional patterns for measures based on execution efficiency in different categories of stocks. Special treatment (ST) stocks have the lowest  $r^e$ ,  $r^{sp}$ ,  $r^{st}$ , and  $r^{bp}$ , the highest order lifetime, and the lowest estimated  $\alpha$ , indicating a lack of liquidity for ST stocks. Meanwhile, constituents of the CSI 300 Index (in SZSE) and the SZI Index have the highest execution ratio, together with the lowest lifetime and the highest estimated  $\alpha$ , which indicates that they have better liquidity. Furthermore, Fig. 8(b) show that the price of stocks is negatively related to the order lifetime and positively related to  $\hat{\alpha}$ .

Importantly, Fig. 8 shows that liquidity in the Chinese market generally improves in 2019–2021 from the perspective of execution efficiency. First, the execution ratios at a single price, at a single time, and at the best price in 2021 are almost all the highest among the three years. In addition, the order lifetime is getting shorter and  $\hat{\alpha}$  is getting larger from 2019 to 2021 across all classes of stocks, except for ST stocks. These findings collectively indicate an improvement in liquidity in the Chinese market during this period, as observed in the liquidity measures based on execution efficiency.

### 5.2.3. Measures based on price impact

*Kyle's Lambda.* We implement the regression (1) for each stock on each trading day from 2019 to 2021 and calculate the average daily Kyle's Lambda for each stock in each year.

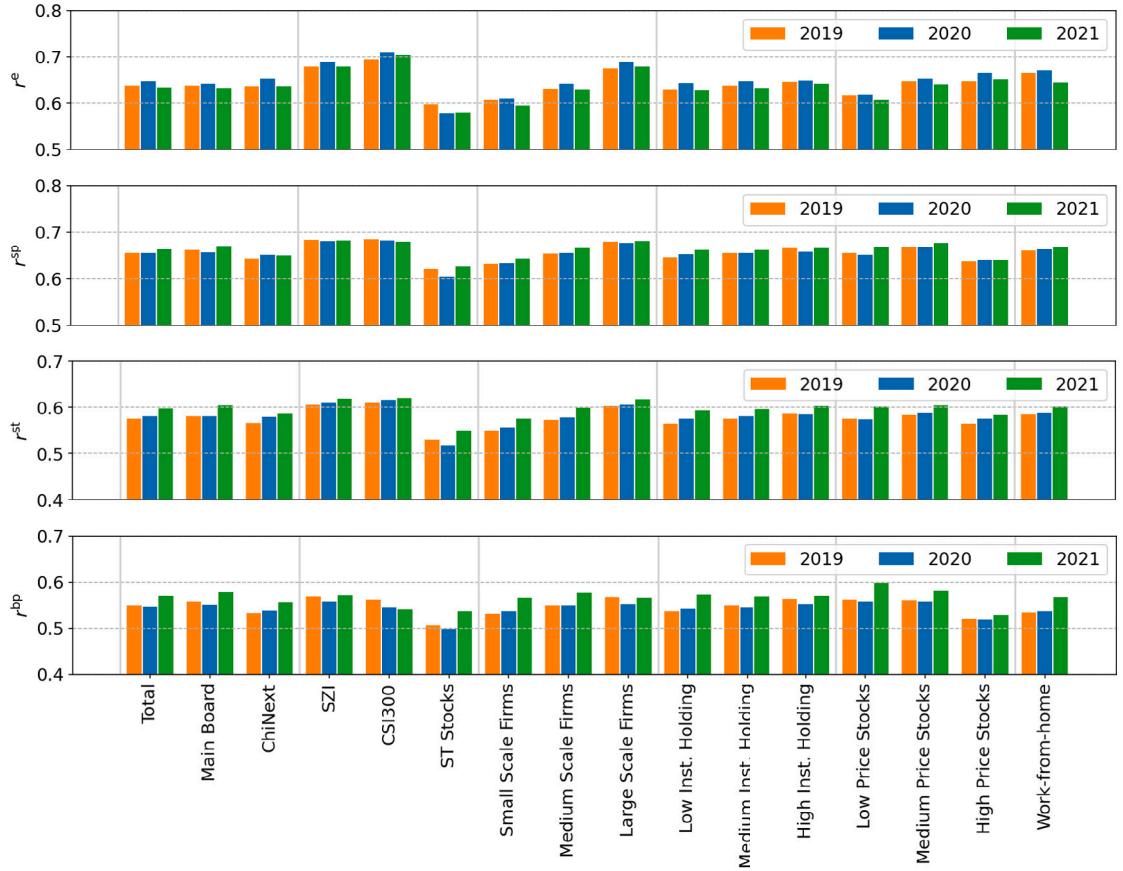
Fig. 9 summarizes the average daily Kyle's Lambda for different categories of stocks. First, Kyle's Lambda of constituents of the CSI 300 Index (in SZSE) and the SZI Index are smaller than that of stocks in other categories, while special treatment (ST) stocks have a larger average Kyle's Lambda than the average level. This indicates that the constituents of the CSI 300 Index (in SZSE) and the SZI Index are more liquid, while the ST stocks have less liquidity. Moreover, the average Kyle's Lambda of large-scale firms is smaller than that of small-scale firms, which suggests that stocks become more liquid as the scale of a company increases. In addition, with the increase of institutional holding, the value of Kyle's Lambda decreases slightly. This indicates that the liquidity of stocks with high institutional holdings is slightly better than that of low institutional holdings. Finally, as the stock price increases, the value of Kyle's Lambda decreases, which suggests that stocks with a high price are more liquid than those with a low price.<sup>18</sup>

Fig. 9 also highlights an improvement in market liquidity in the Chinese stock market over the three years, as reflected by a decline in Kyle's Lambda. Specifically, Kyle's Lambdas for all categories, except for "Work-from-home" stocks, decrease steadily from 2019 to 2021, which suggests a reduced price impact of trades. In addition, Kyle's Lambda for "Work-from-home" stocks is the lowest in 2020, likely due to increased attention on these stocks during the COVID-19 period.

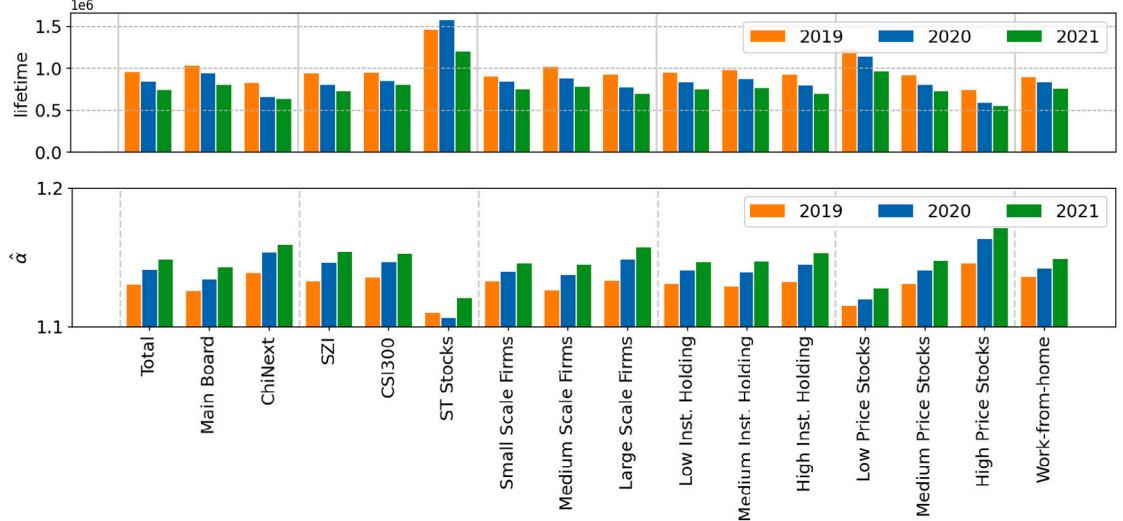
*Price impact regression.* We summarize the annual estimates for each group of stocks, which include the order-processing estimates ( $\gamma_0$ ,  $\gamma_1$ ) and adverse selection estimates ( $\lambda_0$ ,  $\lambda_1$ ). Because the estimated parameters depend on the magnitude of stock prices, we normalize these estimates either by a stock's average spread or average price in the following analysis.

First, Fig. 10(a) shows the histogram of  $\gamma_0$ , the constant execution cost per share, normalized by the spread of each category of stocks. The normalized  $\gamma_0$  are clustered around 0.4 in each year of 2019–2021, which indicates that the constant execution cost per share of each order is slightly less than half of the spread. In this sense,  $\gamma_0$  can be regarded as a liquidity measure similar to the spread.

<sup>18</sup> Overall, these conclusions are consistent with the observations in Shenzhen Stock Exchange (2021, Figure 3 and Figure 8).



(a) Total execution ratio  $r^e$  (first panel), execution ratio at a single price  $r^{sp}$  (second panel), execution ratio at a single time  $r^{st}$  (third panel), and execution ratio at the best price  $r^{bp}$  (fourth panel).



(b) Order lifetime (top panel) and power-law  $\hat{\alpha}$  for the order lifetime (bottom panel).

Fig. 8. Cross-sectional patterns of execution efficiency for different categories of stocks in 2019, 2020, and 2021.

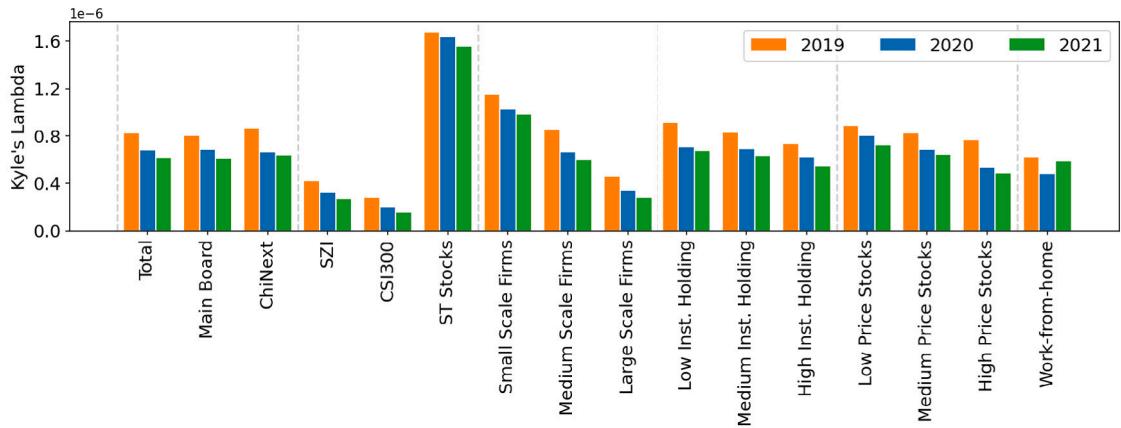
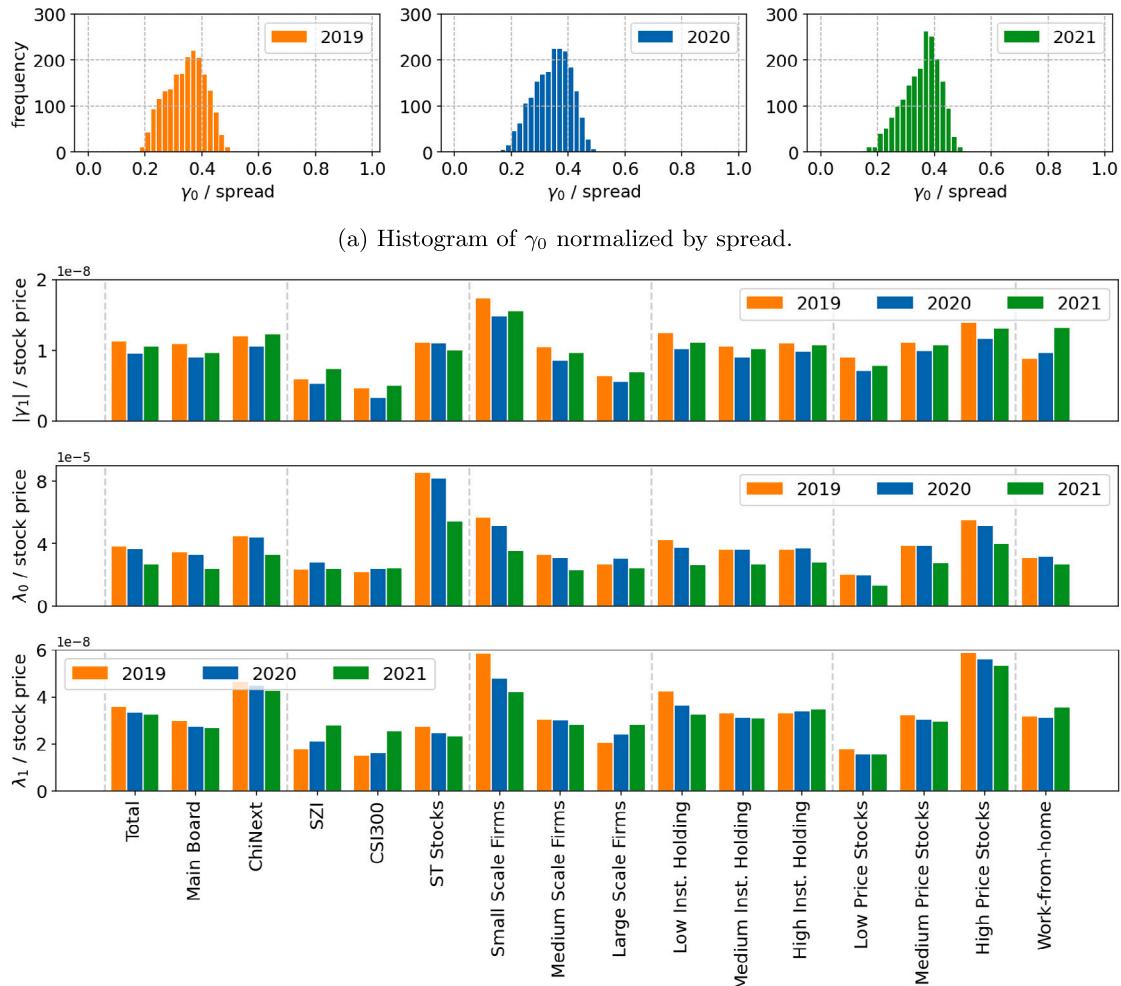


Fig. 9. Average Kyle's Lambda for different categories of stocks in 2019, 2020, and 2021.



(b)  $|\gamma_1|$  (top panel),  $\lambda_0$  (middle panel), and  $\lambda_1$  (bottom panel) normalized by stock price for each category of stocks.

Fig. 10. Estimated coefficients in the price impact regression for all stocks in 2019, 2020, and 2021.

In addition, we analyze the absolute value of  $\gamma_1$  normalized by the stock price. This value represents the “quantity discount rate” that market makers offer to attract large orders, with a larger “quantity discount rate” indicating poorer liquidity for the corresponding stock.<sup>19</sup> The top panel of Fig. 10(b) shows the “quantity discount rates” for different groups of stocks. The “quantity discount rates” for constituents of the CSI 300 Index (in SZSE) and the SZI Index are very small, and the “quantity discount rates” decrease significantly as the size of the firm increases. These findings, along with the results from Kyle’s Lambda, suggest that less liquid stocks tend to have higher “quantity discount rates”.

Finally, we examine the adverse selection estimates,  $\lambda_0$  and  $\lambda_1$ , normalized by stock price to assess the sensitivity of the consensus price to order flow information. Larger normalized estimates indicate greater price sensitivity and poorer liquidity. The middle and bottom panels of Fig. 10(b) show that the normalized  $\lambda_0$  and  $\lambda_1$  are largest in ChiNext, the normalized  $\lambda_0$  for ST stocks is higher than most categories, and the stocks of smaller firms and higher prices tend to have larger normalized  $\lambda_0$  and  $\lambda_1$ . These findings indicate that these stocks generally experience higher adverse selection and lower liquidity. However, the normalized  $\lambda_0$  and  $\lambda_1$  values show a general decrease across the whole market from 2019 to 2021, implying an overall improvement in liquidity.

#### 5.2.4. Common patterns and differences

From the cross-sectional analysis above, we conclude that different boards in the exchange exhibit different levels of liquidity. Measures based on both trading cost and execution efficiency suggest that the ChiNext is the most liquid among the three boards. For example, it has the smallest proportional spread, the highest execution ratio, and the shortest order lifetime. However, from the perspective of market impact, the prices of ChiNext stocks are more sensitive to the information contained in the order flow, which indicates their relatively poor liquidity.

Nearly all the measures we used indicate that the special treatment (ST) stocks are the least liquid across all the categories—they have the largest proportional spread, the lowest execution ratio, and the longest order lifetime. By contrast, the constituents of the CSI 300 Index (in SZSE) and the SZI Index have smaller proportional spread, higher execution ratio, shorter order lifetime, smaller Kyle’s Lambda, and smaller price impact regression estimates as compared to other categories, which demonstrates their superior liquidity.

Meanwhile, we examine the relationship between certain characteristics of stocks and their liquidity measures. Specifically, stocks for large-scale firms tend to have lower proportional spread, higher execution ratio, shorter order lifetime, smaller Kyle’s Lambda, and smaller normalized price impact regression estimators than the stocks for small-scale firms. These results suggest that stocks for larger-scale firms are more liquid.

In addition, our results reveal that stocks with high institutional holdings have slightly smaller proportional spread, higher execution ratio, shorter order lifetime, and smaller Kyle’s Lambda, implying better liquidity.

Furthermore, the price of a stock is also related to its liquidity. We observe that high-price stocks have smaller proportional spread and shorter order lifetimes compared to low-price stocks. This gives us the impression that high-price stocks may have better liquidity than low-price stocks from the perspective of trading costs and execution efficiency. However, high-price stocks have larger normalized price impact regression estimates than low-price stocks, which indicates relatively poor liquidity of the high-price stocks in terms of their market impact.

Finally, we note that market liquidity in the Chinese market improved significantly from 2019 to 2021 across nearly all measures examined in this section. Over these three years, proportional spreads decreased, execution ratios increased, order lifetimes shortened, price impacts lessened, and adverse selection became less pronounced in the whole market. In addition, “Work-from-home” stocks exhibited the smallest proportional spreads and Kyle’s Lambda in 2020, indicating increased trading activity and attention during the COVID-19 period.

## 6. Determinants of high-frequency market liquidity

We have explored the intraday and cross-sectional patterns of various liquidity measures comprehensively. However, to understand what determines the liquidity patterns, it is useful to study the mechanism of spread formation with respect to order flows from a high-frequency perspective.

For relatively longer time periods such as minutes, high-frequency liquidity has been well investigated, including through inventory models proposed by, for example, Grossman and Miller (1988) and Bogousslavsky and Collin-Dufresne (2023). In Appendix G, we follow the framework of Bogousslavsky and Collin-Dufresne (2023) to examine the relationship between high-frequency liquidity, volume, and volatility in order imbalance for the Chinese stock market.

In this section, we propose an alternative determinant of the quote bid–ask spread from the perspective of trading urgency based on high-frequency data. We show that, over short time intervals, quote bid–ask spreads are mainly driven by the aggressive–passive imbalance (API), which is defined as the imbalance between aggressive orders and passive orders. In Section 6.1, we define API and reveal the natural linear relationship between the API and bid–ask spread. In Section 6.2, we propose a discrete classification model for bid–ask spread based on the special properties of the change of spreads. Finally, in Section 6.3, we present the empirical results for the model and demonstrate the effectiveness of API as a determinant for change in spread.

<sup>19</sup> See Appendix F.2.2 of the Online Supplementary Materials for a more detailed discussion on the “quantity discount” intuition behind  $\gamma_1$ .

### 6.1. Aggressive–passive imbalance (API)

We first define API and then use a stylized model of the limit order book proposed by Cont et al. (2014) to illustrate the intuition behind the definition.

**Definition 7 (API).** For a specific stock, the aggressive–passive imbalance (API) of the stock from time  $t$  to time  $t + \Delta t$  ( $\Delta t > 0$ ) is defined as

$$\text{API}_{t,t+\Delta t} = \frac{(A_{t,t+\Delta t}^a + A_{t,t+\Delta t}^b) + (C_{t,t+\Delta t}^a + C_{t,t+\Delta t}^b) - (P_{t,t+\Delta t}^a + P_{t,t+\Delta t}^b)}{\bar{D}_t}, \quad (6)$$

where  $A_{t,t+\Delta t}^a$ ,  $A_{t,t+\Delta t}^b$ ,  $P_{t,t+\Delta t}^a$ ,  $P_{t,t+\Delta t}^b$ ,  $C_{t,t+\Delta t}^a$ , and  $C_{t,t+\Delta t}^b$  are the total volumes of the following orders from time  $t$  to  $t + \Delta t$ , respectively (see Definitions A.2 and A.6 in Appendix B of the Online Supplementary Material for details): aggressive sell orders, aggressive buy orders, passive sell orders, passive buy orders, cancellations of sell orders, and cancellations of buy orders;  $\bar{D}_t$  is defined as the average of best  $K$  ( $K = 5$  in our empirical studies) levels of bid sizes and ask sizes at time  $t$ , i.e.,

$$\bar{D}_t = \frac{1}{2K} \left[ \sum_{i=1}^K n^b(p_i^b(t), t) + \sum_{i=1}^K n^a(p_i^a(t), t) \right]. \quad (7)$$

The concept of API laid out in Definition 7 represents the relative difference (imbalance) between aggressive orders and passive orders. Note that, for simplicity, the cancellations are also regarded as “aggressive” (the sign before  $C_{t,t+\Delta t}^a + C_{t,t+\Delta t}^b$  is plus), because they are also realized immediately. Higher volumes of aggressive orders and cancellations lead to larger values of API, while higher volumes of passive orders lead to smaller values of API.

With the help of API, we can use a linear model to explain the relationship between API and the change of bid–ask spread,  $S(t)$ , from time  $t$  to  $t + \Delta t$  for a specific stock. In particular, suppose that:

$$S(t + \Delta t) - S(t) = \alpha + \beta \text{API}_{t,t+\Delta t} + \varepsilon_{t,t+\Delta t}, \quad \varepsilon_{t,t+\Delta t} \sim N(0, \sigma^2 \Delta t), \quad (8)$$

where  $\alpha$ ,  $\beta$  and  $\sigma^2$  are unknown parameters. Here the parameter  $\beta$  represents the impact coefficient of API on spread change.

#### 6.1.1. Intuition for API

We now illustrate the intuition behind model (8) and API. In a similar vein to Cont et al. (2014), we consider a stylized model of the limit order book. Suppose that there is a *stylized order book* that satisfies the following properties in a short time interval  $[t, t + \Delta t]$ :

- (a) At time  $t$ , the depth<sup>20</sup> of the limit order book at each price level less than or equal to the best bid, and each price level greater than or equal to the best ask, equals to a constant  $D$ ;
- (b) During  $[t, t + \Delta t]$ , all new cancellations and aggressive orders only occur at the best bid and best ask;
- (c) During  $[t, t + \Delta t]$ , if the depth of the best bid (or ask) is less than  $D$ , then new passive orders will fill in the best bid (or ask) first; once the depth of the best bid (or ask) reaches  $D$ , the new passive orders will arrive 1 tick above (or below) the best quote, initializing a new best level.

By definition, the depth of each price level of the stylized order book will never exceed  $D$ . The best bid and best ask fluctuate as new orders and cancellations arrive successively.

Fig. 11 illustrates how the aggressive orders, passive orders, and cancellations drive the spread change of the stylized limit order book with  $D = 4$ . Parameter  $\delta$  represents the tick size of the order book, e.g.,  $\delta = \text{RMB } 0.01$  in China. During the time period  $[t, t + \Delta t]$ , Fig. 11(a) shows the impact of aggressive buy orders with volume  $A_{t,t+\Delta t}^b = 7$  on the order book. All aggressive buy orders arrive at the best ask, so the best ask increases by  $\lfloor 7/4 \rfloor \delta = 1\delta$  and the best bid does not change. Thus, the bid–ask spread will increase by  $\delta$ . Similarly, Figs. 11(b) and 11(c) show the impact of passive sell orders and cancellations, respectively. The impact of cancellations is similar to that of aggressive orders, not surprisingly. Therefore, as we have demonstrated, we regard cancellations as aggressive in the definition of API (6). Finally, Fig. 11(d) is an example that combines the aggressive orders, passive orders, and cancellations.

In this stylized model of limit order book, the change of best ask price  $a(t + \Delta t) - a(t)$  and best bid price  $b(t + \Delta t) - b(t)$  satisfy:

$$a(t + \Delta t) - a(t) = \delta \left[ \frac{C_{t,t+\Delta t}^a + A_{t,t+\Delta t}^b - P_{t,t+\Delta t}^a}{D} \right], \quad b(t + \Delta t) - b(t) = -\delta \left[ \frac{C_{t,t+\Delta t}^b + A_{t,t+\Delta t}^a - P_{t,t+\Delta t}^b}{D} \right].$$

Therefore, the spread change can be represented as

$$\begin{aligned} S(t + \Delta t) - S(t) &= [a(t + \Delta t) - b(t + \Delta t)] - [a(t) - b(t)] = [a(t + \Delta t) - a(t)] - [b(t + \Delta t) - b(t)] \\ &= \delta \left[ \frac{C_{t,t+\Delta t}^a + A_{t,t+\Delta t}^b - P_{t,t+\Delta t}^a}{D} \right] + \delta \left[ \frac{C_{t,t+\Delta t}^b + A_{t,t+\Delta t}^a - P_{t,t+\Delta t}^b}{D} \right]. \end{aligned}$$

Furthermore, according to (6), the definition of API, we have

$$S(t + \Delta t) - S(t) = \delta \cdot \text{API}_{t,t+\Delta t} + \varepsilon, \quad (9)$$

where  $\varepsilon$  represents the truncation error of the floor function. Eq. (9) provides an explanation of the linear model (8).

<sup>20</sup> See Definition A.4 in Appendix B of the Online Supplementary Material.

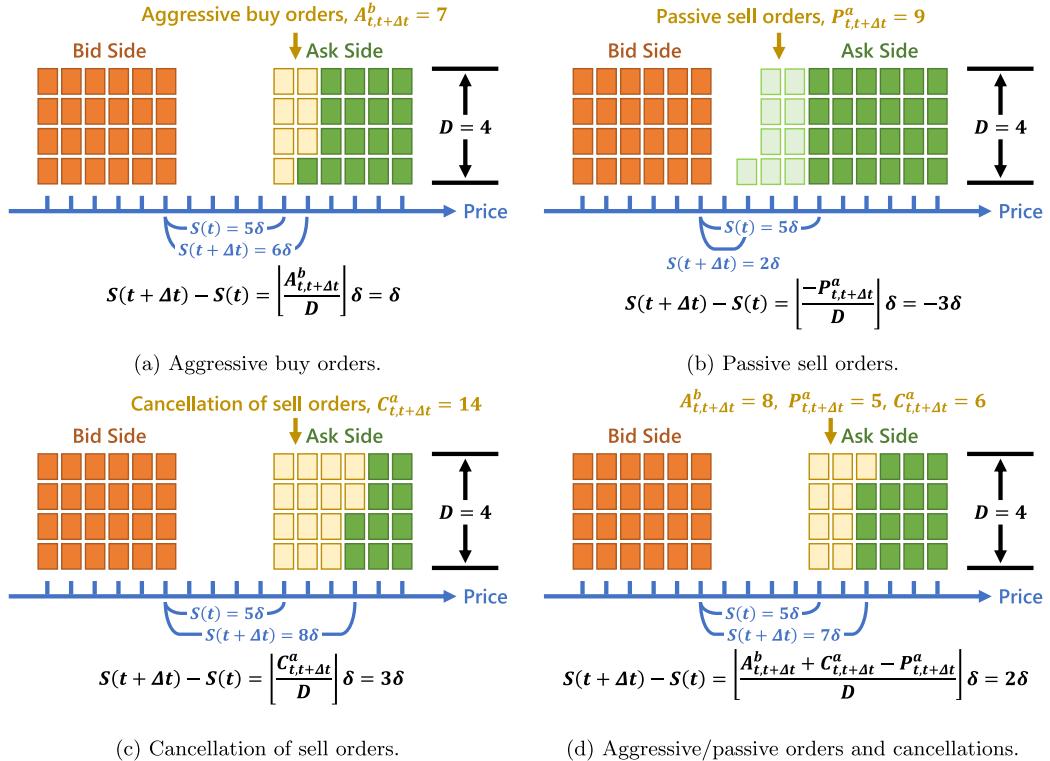


Fig. 11. An illustration of spread changes in the stylized limit order book.

Eq. (9) also sheds light on the meaning of  $\alpha$  and  $\beta$  in model (8). Under the stylized model, the following properties should hold:

$$\alpha \approx 0, \quad \beta \approx \delta. \quad (10)$$

In other words, the spread should not change in the absence of an imbalance of aggressive and passive orders ( $\alpha \approx 0$ ), and the impact of API on the spread change should be equal to the tick size of the stock ( $\beta \approx \delta$ ).

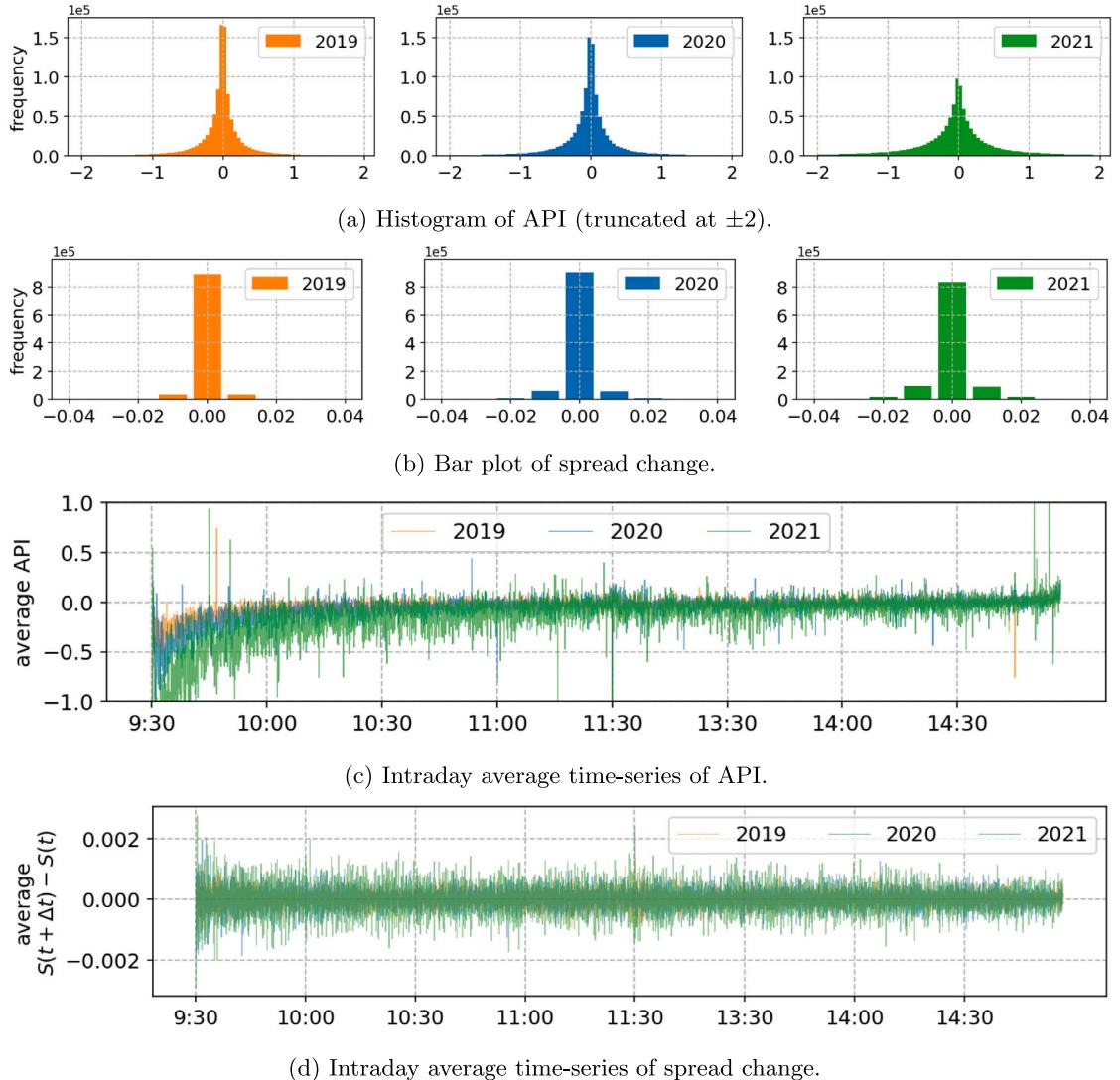
Cont et al. (2014) use this stylized model of limit order book to analyze the change in the mid-price. In particular, they introduce the concept of “order flow imbalance” (OFI), which is defined as the difference between the total volume of buy orders and sell orders, measuring the imbalance between the buy and sell sides of the market. Our definition of API (Definition 7) is inspired by OFI, and the main difference between our API model and Cont et al. (2014) OFI model is that we focus on characterizing the change in the spread, rather than the mid-price.

### 6.1.2. API and liquidity: Intraday patterns

Now we use our data to discover the empirical properties of API. For simplicity, in this section we only use the full-year data of Ping An Bank in 2019–2021 (the average stock price is approximately RMB 16.00) as samples and choose  $\Delta t = 3s$ .

We first explore the distribution of API and Fig. 12(a) shows its histogram. APIs are nearly symmetrically distributed and centered around 0. Panel (a) of Table 6 presents the summary statistics of API, whose sample average is negative, implying that the volume of passive orders is generally greater than the total volume of aggressive orders and cancellations. Meanwhile, the distribution of API has heavy tails. Panel (a) of Table 6 shows that the maximum of API is significantly larger than the 99% percentile, and the minimum of API is considerably smaller than the 1% percentile.

Next, we investigate the intraday patterns of API. For each year and a specific period  $[t, t + \Delta t]$ , we calculate the average API over all trading days in the year. This produces an *intraday* time series of average API with a length of 237 (minutes per day)  $\times 60$  (seconds per minute)/3 (seconds) = 4740. Fig. 12(c) shows the time series of “intraday” API, in which it can be seen that the API is negative at the open of the market, then increases to 0 in about one hour, and finally stays at around 0. This observation shows that, at the beginning of each trading day, the volume of passive orders is substantially larger than that of aggressive orders and cancellations. Investors quickly fill in vacancies in the limit order book using passive orders immediately after the market open, and then switch to a mixture of aggressive orders and passive orders throughout the day. By examining Fig. 12(c) carefully, one can also find that the API is slightly negative after the afternoon open (13:00). In addition, The API becomes positive before the morning close (11:30), and the afternoon close (14:57). Generally speaking, there are more passive orders than aggressive orders (a negative API) at the market open, and more aggressive orders than passive orders (a positive API) at the market close.

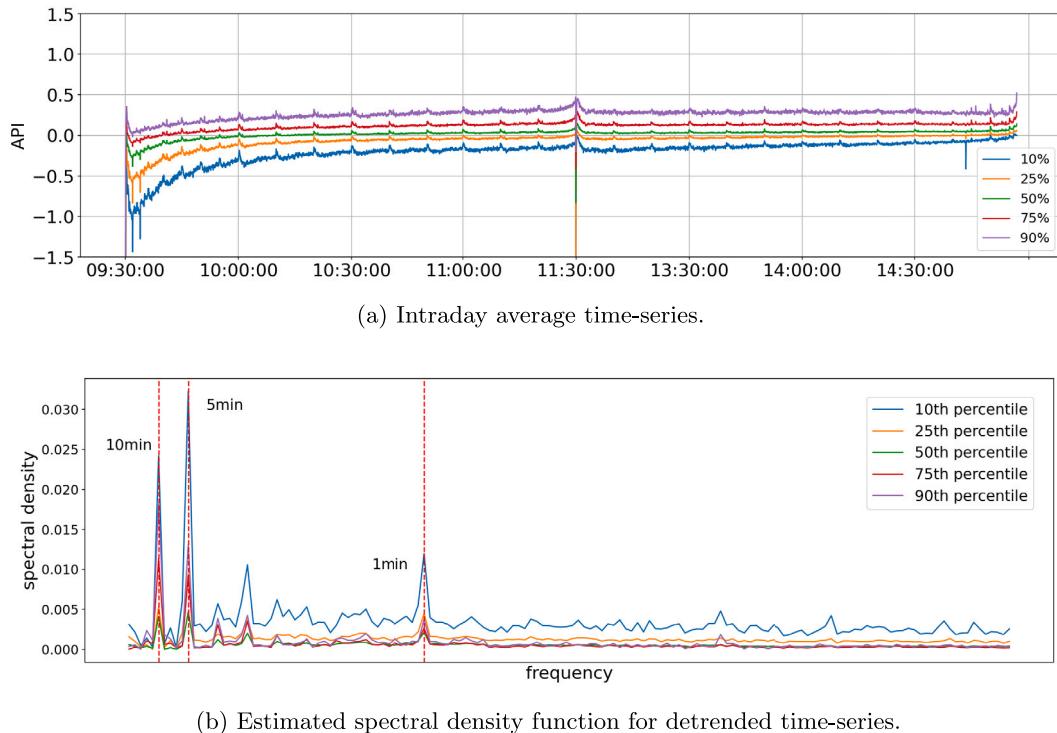


**Fig. 12.** Distribution and intraday time series of API and spread change  $S(t + \Delta t) - S(t)$ , Ping An Bank, 2019–2021,  $\Delta t = 3s$ .

**Table 6**

Distribution of API and spread change  $S(t + \Delta t) - S(t)$ , Ping An Bank, 2019–2021,  $\Delta t = 3s$ .

(a) Summary statistics of API.			(b) Counts of spread change $S(t + \Delta t) - S(t)$ .				
	2019	2020	2019	2020	2021		
Sample number	972,214	1,043,856	1,080,546	Sample number	971,973	1,042,480	1,080,304
mean	-0.0357	-0.0660	-0.1501	$\leq -0.04$	175	761	3,168
std	0.7036	1.1394	1.9964	$-0.03$	672	1,833	6,274
min	-54.4670	-179.6650	-180.9570	$-0.02$	3,780	8,431	20,429
1%	-2.0460	-3.1214	-5.8500	$-0.01$	38,024	60,985	95,223
25%	-0.1160	-0.1556	-0.3087	0.00	887,901	900,435	833,711
50%	-0.0054	-0.0099	-0.0190	0.01	36,118	58,045	90,167
75%	0.0803	0.1024	0.2006	0.02	4,195	8,876	20,913
99%	1.6971	2.3196	3.7862	0.03	828	2,178	6,775
max	187.3125	191.8961	277.4431	$\geq 0.04$	280	936	3,644



**Fig. 13.** Intraday average time series and the corresponding estimated spectral density function for different cross-sectional quantiles of API, 2020,  $\Delta t = 3\text{s}$ .

In fact, the intraday patterns of API are consistent with the W-shaped pattern of the spread, as we have illustrated in Section 5.1.1. Note that a negative API will lead to a narrower spread (better liquidity), and vice versa, as was explained in Section 6.1.1. Therefore, the negative API after the market open implies a shrinkage of spread (better liquidity), while the positive API before the market close means an expansion of spread (worse liquidity). It is the intraday shape of API that leads to the W-shaped spread.

The periodicity of the spread illustrated in Section 5.1.1 can also be explained via API. Taking 2020 as an example, Fig. 13(a) shows the time series of different quantiles of intraday average API across all 1,936 stocks, and Fig. 13(b) shows the corresponding estimated spectral density functions, which are calculated using the same methods as those employed in Section 5.1.1. We still find that the API series have strong periodicities at 10-minute, 5-minute, and 1-minute. This explains the source of the intraday periodicity of liquidity measures. Investors place aggressive and passive orders periodically, which leads to the periodicity of liquidity. This further develops the “liquidity begets liquidity” phenomenon first documented in Foucault et al. (2013). In fact, we find that the intraday pattern of API leads to “liquidity begets liquidity periodically” in the Chinese market.

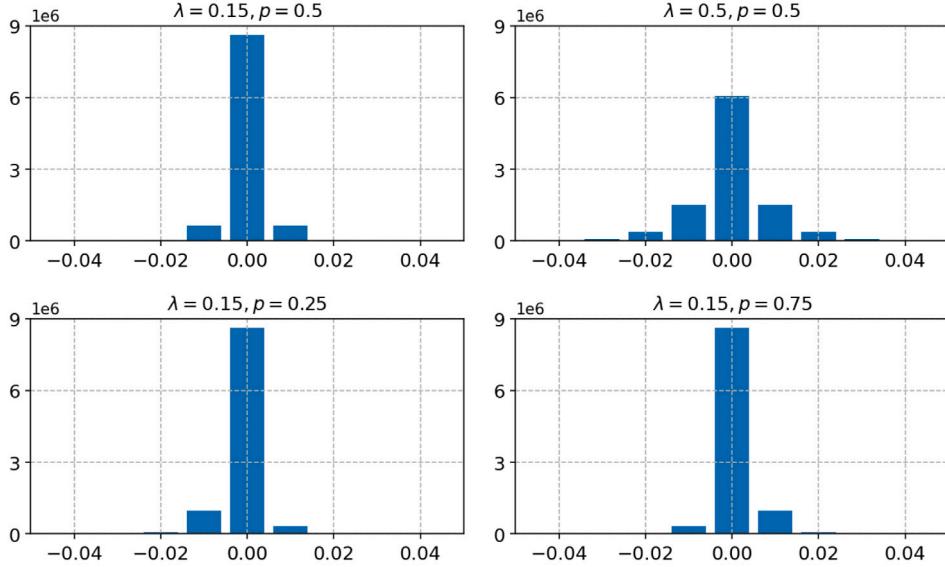
## 6.2. A discrete classification model for spread change based on API

In this section, we construct a more refined model of spread change to account for the fact that the values of spread and spread change are *discrete*. More precisely, the spread change can only take values of multiples of the tick size  $\delta$ . Section 6.2.1 studies the discrete nature of spread change, Section 6.2.2 proposes the discrete classification framework, and Section 6.2.3 discusses the model estimation.

### 6.2.1. Distribution of spread change

We begin by investigating the distribution of the spread change. Fig. 12(b) shows the bar plot of spread change,  $S(t + \Delta t) - S(t)$ , for Ping An Bank in 2019–2021. The values of  $S(t + \Delta t) - S(t)$  are not only discrete but that they also have a narrow range. Panel (b) of Table 6 exhibits the counting numbers of different values that  $S(t + \Delta t) - S(t)$  takes. The table shows that 99.86% of  $S(t + \Delta t) - S(t)$  are within  $-3\delta = -0.03$  and  $+3\delta = +0.03$ , and 88.11% values are zero. In addition,  $S(t + \Delta t) - S(t)$  is nearly symmetrically distributed.

Fig. 12(d) shows the intraday average time series of spread changes for Ping An Bank in 2019–2021. The method to plot the figure is similar to that shown in Fig. 12(c), i.e., the intraday curve of API. In comparison with the curve of API, though, the intraday curve of spread change is much more stable. The average spread change is negative just after the open of the market, but it is not as pronounced as the curve of API. The slightly negative spread change can be regarded as a result of the negative API at the beginning of the trading day, because the influx of passive orders leads to a narrower spread.



**Fig. 14.** Histograms of the distribution defined by (11) versus parameters  $p$  and  $\lambda$ .

Taking the discrete feature of the spread change into consideration, we can use the following distribution to model the spread change. Suppose that

$$S(t + \Delta t) - S(t) \stackrel{d}{=} \delta \cdot B_{t,t+\Delta t} \cdot N_{t,t+\Delta t}, \quad (11)$$

where random variable  $B_{t,t+\Delta t}$  takes  $+1$  or  $-1$  with a probability of  $p_{t,t+\Delta t}$  and  $1 - p_{t,t+\Delta t}$ , respectively, random variable  $N_{t,t+\Delta t} \sim \text{Poisson}(\lambda_{t,t+\Delta t})$ , and  $B_{t,t+\Delta t}$  and  $N_{t,t+\Delta t}$  are independent. Parameters  $p_{t,t+\Delta t} \in (0, 1)$  and  $\lambda_{t,t+\Delta t} \in (0, +\infty)$  depend on the time interval  $[t, t + \Delta t]$ .

We now explain the meaning of (11). Random variable  $B_{t,t+\Delta t}$  controls the direction of spread change: if  $B_{t,t+\Delta t} = 1$ ,  $[S(t + \Delta t) - S(t)]/\delta$  will be a non-negative integer, while if  $B_{t,t+\Delta t} = -1$ ,  $[S(t + \Delta t) - S(t)]/\delta$  will be a non-positive integer. Random variable  $N_{t,t+\Delta t}$  controls the magnitude of spread change. Furthermore, it is easy to verify that, under the assumption of (11),

$$\begin{aligned} \mathbb{E}\left(\frac{S(t + \Delta t) - S(t)}{\delta}\right) &= (2p_{t,t+\Delta t} - 1)\lambda_{t,t+\Delta t}, \\ \mathbb{E}\left|\frac{S(t + \Delta t) - S(t)}{\delta}\right| &= \lambda_{t,t+\Delta t}, \\ \mathbb{E}\left(\frac{S(t + \Delta t) - S(t)}{\delta}\right)^2 &= \lambda_{t,t+\Delta t} + \lambda_{t,t+\Delta t}^2. \end{aligned}$$

To further illustrate (11), we conduct some simple simulations. Fig. 14 shows the bar plots of the distribution defined by (11) under different parameters  $p_{t,t+\Delta t}$  and  $\lambda_{t,t+\Delta t}$ . Here, for simplicity, we use a constant  $p_{t,t+\Delta t} = p$  and a constant  $\lambda_{t,t+\Delta t} = \lambda$ , neither of which relies on  $t$ . According to Fig. 14, if  $p = 0.5$ , the distribution will be symmetric about zero; if  $p > 0.5$ , there will be more mass greater than zero; and if  $p < 0.5$ , there will be more mass less than zero. Besides, the distribution will spread out as  $\lambda$  increases. Note that the bar plot given  $p = 0.5$  and  $\lambda = 0.15$  is very similar to Fig. 12(b), the bar plot of spread change. Therefore, (11) can be used to model the distribution of spread change.

### 6.2.2. Model setup

Now we link the distribution of spread change with API. The linear model in Section 6.1 demonstrates that a higher API leads to a higher probability of a positive spread change. To embody this intuitive relationship, suppose that the parameter  $p_{t,t+\Delta t}$  of  $B_{t,t+\Delta t}$  satisfies

$$p_{t,t+\Delta t} = \frac{e^{\gamma \cdot \text{API}_{t,t+\Delta t}}}{1 + e^{\gamma \cdot \text{API}_{t,t+\Delta t}}}, \quad (12)$$

where  $\gamma$  is an unknown parameter that does not depend on the time interval  $[t, t + \Delta t]$ . This expression is similar to the logistic regression. If  $\gamma > 0$ , a higher API will lead to a higher possibility of spread increase, and vice versa.

For simplicity, we further assume that the parameter  $\lambda_{t,t+\Delta t}$  with regard to  $N_{t,t+\Delta t}$  is not correlated with API and is time-homogeneous, i.e., there exists a constant  $\lambda$  such that  $\lambda_{t,t+\Delta t} = \lambda$  for all time intervals  $[t, t + \Delta t]$ . While these may not be true for all stocks, we still make these assumptions based on three facts. First, our empirical study shows that the API has a much weaker

correlation with the magnitude than the direction of the spread change. In particular, for the 2020 data of Ping An Bank and  $\Delta t = 3\text{s}$ , we have

$$\text{corr}(\text{API}, |S(t + \Delta t) - S(t)|) = -0.0097, \quad \text{corr}(\text{API}, \text{sign}(S(t + \Delta t) - S(t))) = 0.0598. \quad (13)$$

Second, by the intraday average time series of spread change (Fig. 12(d)), the magnitude of spread change has a stable pattern. Third, these assumptions can not only derive our key results regarding the impact of API on the direction of spread change but also maintain mathematical simplicity. In fact, we will see later in Section 6.2.3 that the parameters are easy to estimate under these assumptions. Indeed, one could easily extend our model to include a non-homogeneous  $\lambda$ .

We summarize our discrete classification model of spread change as follows:

$$\begin{aligned} S(t + \Delta t) - S(t) &= \delta \cdot B_{t,t+\Delta t} \cdot N_{t,t+\Delta t}, \\ \text{where } B_{t,t+\Delta t} &= \begin{cases} +1, & \text{with probability } p_{t,t+\Delta t}, \\ -1 & \text{with probability } 1 - p_{t,t+\Delta t}, \end{cases} \quad N_{t,t+\Delta t} \sim \text{Poisson}(\lambda), \\ \text{and } p_{t,t+\Delta t} &= \frac{e^{\gamma \cdot \text{API}_{t,t+\Delta t}}}{1 + e^{\gamma \cdot \text{API}_{t,t+\Delta t}}}. \end{aligned} \quad (14)$$

Here,  $\gamma \in (-\infty, +\infty)$  and  $\lambda \in (0, +\infty)$  are unknown parameters.

In comparison with Model (8), Model (14) can not only characterize the distribution of spread change but also create chances for us to analyze the direction and the magnitude of spread change separately. In particular,  $B_{t,t+\Delta t}$  represents the direction of spread change, and  $N_{t,t+\Delta t}$  represents the magnitude of spread change. The model has two parameters:  $\lambda$  and  $\gamma$ . Parameter  $\lambda$  represents the average magnitude of spread change (or liquidity change)  $S(t + \Delta t) - S(t)$ , and  $\gamma$  measures the impact of API on the direction of spread change, i.e., the sensitivity of the direction of liquidity change with respect to the API.

### 6.2.3. Estimation

Here we discuss the estimation of parameters  $\lambda$  and  $\gamma$  in (14). Given  $n$  samples  $(\text{API}_i, Y_i), i = 1, 2, \dots, n$ , where  $Y_i := \frac{S(t_i + \Delta t) - S(t_i)}{\delta}$ , the likelihood function of our model is

$$L(\lambda, \gamma) = \prod_{i: Y_i > 0} \left[ \frac{e^{\gamma \cdot \text{API}_i}}{1 + e^{\gamma \cdot \text{API}_i}} \frac{\lambda^{Y_i}}{Y_i!} e^{-\lambda} \right] \cdot \prod_{i: Y_i < 0} \left[ \left( 1 - \frac{e^{\gamma \cdot \text{API}_i}}{1 + e^{\gamma \cdot \text{API}_i}} \right) \frac{\lambda^{-Y_i}}{(-Y_i)!} e^{-\lambda} \right] \cdot \prod_{i: Y_i = 0} [e^{-\lambda}].$$

The corresponding log-likelihood function is

$$\ln L(\lambda, \gamma) = \sum_{i: Y_i > 0} [\gamma \cdot \text{API}_i] + \sum_{i: Y_i \neq 0} [-\ln(1 + e^{\gamma \cdot \text{API}_i}) + |Y_i| \ln \lambda] - n\lambda + C,$$

where constant  $C$  is not correlated with the unknown parameters. Thus, the maximum likelihood estimators (MLE)  $\hat{\gamma}$  and  $\hat{\lambda}$  should satisfy the following two equations:

$$\frac{\partial \ln L(\lambda, \gamma)}{\partial \lambda} = \frac{\sum_{i=1}^n |Y_i|}{\lambda} - n = 0, \quad (15)$$

$$\frac{\partial \ln L(\lambda, \gamma)}{\partial \gamma} = \sum_{i: Y_i > 0} \text{API}_i - \sum_{i: Y_i \neq 0} \frac{\text{API}_i e^{\gamma \cdot \text{API}_i}}{1 + e^{\gamma \cdot \text{API}_i}} = 0. \quad (16)$$

These equations are easy to solve because parameters  $\lambda$  and  $\gamma$  are separated. Eq. (15) implies  $\hat{\lambda} = \frac{\sum_{i=1}^n |Y_i|}{n}$ , and  $\hat{\gamma}$  can be solved by (16) numerically. One can easily check that (16) has a unique solution (including  $\pm\infty$ ) because the left-hand side is strictly decreasing wrt  $\gamma$ , the limit is non-negative as  $\gamma \rightarrow -\infty$  and non-positive as  $\gamma \rightarrow +\infty$ .

The MLE of  $\lambda$  is intuitive. Note that  $\hat{\lambda} = \frac{\sum_{i=1}^n |Y_i|}{n}$ , i.e., the average of the absolute values of spread changes. This is consistent with the meaning of  $\lambda$ : the average magnitude of spread change  $S(t + \Delta t) - S(t)$ . The MLE of  $\gamma$  is slightly complicated, but it is also consistent with the meaning of  $\gamma$ : the metric of API's impact on the direction of liquidity change. To strengthen our understanding of (16), let us consider two special cases: the direction of spread change is determined by the direction of API perfectly, and the direction of spread change is independent of API. For the former case, if  $\gamma = +\infty$ , we have

$$\sum_{i: Y_i > 0} \text{API}_i - \sum_{i: Y_i \neq 0} \frac{\text{API}_i e^{\gamma \cdot \text{API}_i}}{1 + e^{\gamma \cdot \text{API}_i}} = \sum_{i: Y_i > 0} \text{API}_i - \sum_{i: Y_i \neq 0, \text{API}_i > 0} \text{API}_i = \sum_{i: Y_i > 0} \text{API}_i - \sum_{i: Y_i > 0} \text{API}_i = 0.$$

Thus,  $\hat{\gamma} = +\infty$ . For the latter case, the MLE  $\hat{\gamma}$  will be approximately zero because if  $\gamma = 0$ ,

$$\sum_{i: Y_i > 0} \text{API}_i - \sum_{i: Y_i \neq 0} \frac{\text{API}_i e^{\gamma \cdot \text{API}_i}}{1 + e^{\gamma \cdot \text{API}_i}} = \sum_{i: Y_i > 0} \text{API}_i - \frac{1}{2} \sum_{i: Y_i \neq 0} \text{API}_i \approx \frac{\sum_{i: Y_i > 0} \text{API}_i - \sum_{i: Y_i < 0} \text{API}_i}{2} \approx 0.$$

### 6.3. Empirical results

We have already proposed two models regarding how API influences the bid-ask spread including the linear regression model (8), and the discrete classification model (14). In this section, we show the empirical results based on the data of the Shenzhen Stock Exchange. As before, we use data of all 1,936 stocks from 2019 to 2021. We still choose  $\Delta t = 3\text{s}$  and each stock has approximately  $730 \times 237 \times 60/3 = 3,460,200$  samples over three years.

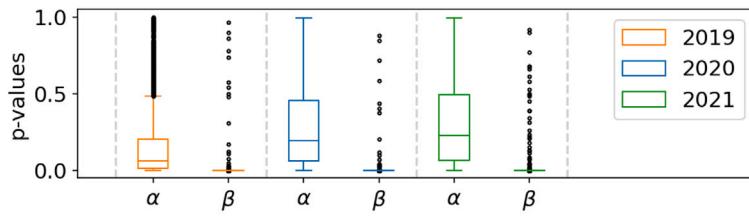


Fig. 15. Boxplots of  $\alpha$  and  $\beta$ 's p-values in the linear regression model. Results are based on 1,936 stocks in 2019, 2020, and 2021.

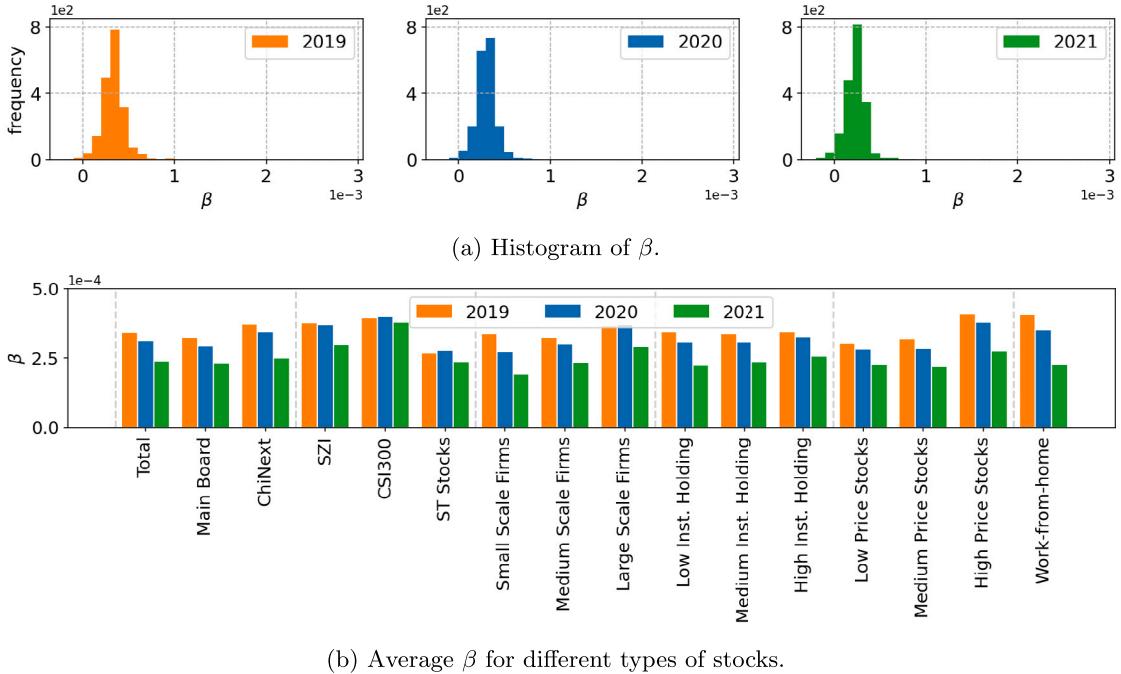


Fig. 16. Histogram and cross-sectional patterns of  $\beta$ 's in the linear regression model. Results are based on 1,936 stocks in 2019, 2020, and 2021.

### 6.3.1. Empirical results for the linear regression model

For each stock, we use the full-year data for each year of 2019–2021 to estimate the parameters  $\alpha$  and  $\beta$  in (8) via ordinary least squares. Fig. 15 shows the boxplots of  $\alpha$  and  $\beta$ 's p-values for all stocks during these three years. Almost all  $\alpha$ 's are insignificant at 5% level (i.e.,  $p$ -value  $> 0.05$ ), which suggests that we cannot reject  $\alpha = 0$  for most stocks. This is consistent with our illustration of the linear model based on the stylized limit order book (see (10)). On the other hand, almost all  $\beta$ 's are significant (i.e.,  $p$ -value  $< 0.05$ ), which is the parameter that we are more concerned with. This implies that API does have a significant impact on the change in spread.

Fig. 16(a) shows the histogram of  $\beta$ 's for all 1,936 stocks, and Table 7 presents the summary statistics of these  $\beta$ 's. Both the figure and the table demonstrate that the  $\beta$  for almost all stocks are positive. Therefore, the impact of API on spread change is positive. However, the  $\beta$ 's vary across stocks and are not close to  $\delta = \text{RMB } 0.01$  as was demonstrated in (10). This suggests the viewpoint proposed in Eisler et al. (2012) that the “real” tick sizes of the stocks may be different.

To further investigate the cross-sectional patterns of  $\beta$ , Fig. 16(b) shows the average  $\beta$  for different stock classes. We make the following observations:

- (a) The  $\beta$ 's for the constituents of the SZI Index and the CSI 300 Index (in SZSE) are much greater than other stocks. In contrast, the  $\beta$ 's for ST stocks are smaller compared to other stocks;
- (b) The  $\beta$  increases as the firm scale, the institutional holding, and the stock price increase.

It is also worth noting that the average  $\beta$ 's for different stock classes shown in Fig. 16(b) are similar to the cross-sectional patterns of spread shown in Fig. 6 of Section 5.2. Since the spread is highly dependent on the tick size of stocks, the two figures further demonstrate the meaning of  $\beta$ —the implied “real” tick size of stocks. Magnitudes of spreads are different since different stocks have different levels of tick sizes.

**Table 7**

Summary statistics of estimated parameters in the linear regression model and the discrete classification model. Results are based on 1,936 stocks in 2019, 2020, and 2021.

	API regression		API discrete			
	$\beta$ ( $\times 10^{-4}$ )	$R^2$ ( $\times 10^{-3}$ )	$\lambda$	$\gamma$	AUC <sub>1</sub>	AUC <sub>2</sub>
(a) 2019						
mean	3.38	4.14	0.47	0.17	0.56	0.58
std	1.50	2.46	0.71	0.18	0.02	0.02
min	-5.32	0.00	0.00	-0.00	0.49	0.46
1% quantile	0.08	0.07	0.01	0.02	0.53	0.55
25% quantile	2.69	2.40	0.10	0.06	0.55	0.57
50% quantile	3.33	3.99	0.24	0.12	0.56	0.58
75% quantile	3.96	5.59	0.58	0.21	0.57	0.60
99% quantile	8.01	10.75	3.12	0.83	0.61	0.64
max	24.07	27.93	11.01	3.76	0.66	0.68
(b) 2020						
mean	3.11	4.22	0.57	0.14	0.56	0.59
std	1.65	2.34	0.99	0.20	0.02	0.02
min	-1.60	0.00	0.00	-0.00	0.49	0.45
1% quantile	0.31	0.07	0.01	0.01	0.52	0.56
25% quantile	2.46	2.46	0.10	0.05	0.54	0.58
50% quantile	3.05	4.15	0.26	0.10	0.56	0.59
75% quantile	3.60	5.83	0.66	0.18	0.57	0.60
99% quantile	7.22	9.73	4.06	0.70	0.61	0.65
max	41.09	15.98	17.51	5.31	0.63	0.68
(c) 2021						
mean	2.31	3.35	0.57	0.12	0.56	0.59
std	1.49	2.26	1.10	0.17	0.02	0.02
min	-6.47	0.00	0.00	-0.01	0.47	0.45
1% quantile	-1.05	0.01	0.01	0.00	0.51	0.55
25% quantile	1.65	1.57	0.09	0.03	0.54	0.57
50% quantile	2.37	3.16	0.23	0.07	0.55	0.59
75% quantile	2.90	4.90	0.61	0.16	0.57	0.60
99% quantile	6.69	8.67	4.56	0.78	0.60	0.65
max	17.64	21.36	15.19	3.27	0.63	0.70

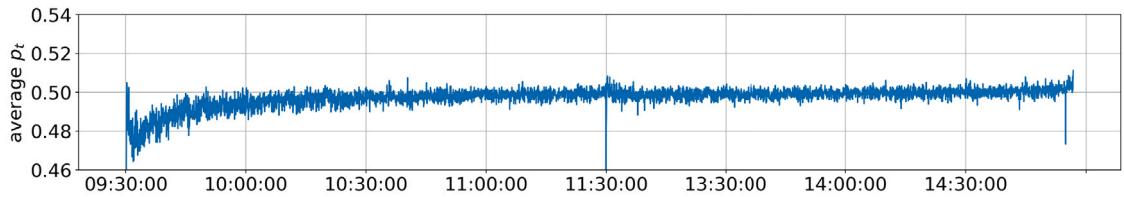


Fig. 17. Intraday average time series of  $p_{t,t+\Delta t} = \frac{e^{\gamma \cdot API_{t,t+\Delta t}}}{1+e^{\gamma \cdot API_{t,t+\Delta t}}}$  for Ping An Bank in 2020. ( $\Delta t = 3s$ ).

The  $R^2$  for the linear regression model is also shown in Table 7. Although the parameter  $\beta$ 's are significant, the largest  $R^2$  among all stocks in 2019–2021 is less than 3%, which implies that the goodness-of-fit is not satisfactory. This is our motivation to consider the discrete classification model instead of the linear regression model.

### 6.3.2. Empirical results for the discrete classification model

Now, for each stock, we use all data in 2019–2021 to estimate the parameters  $\lambda$  and  $\gamma$  in the discrete classification model (14). The estimation method is given in (15) and (16).

We first take the estimation results of Ping An Bank as an example. The estimated  $\gamma = 0.2055$  is greater than zero, which implies that a positive API will lead to a higher probability of spread expansion than spread shrinkage. The estimated  $\lambda = 0.1663$ , which represents the average magnitude of spread change. In addition, Fig. 17 presents the “intraday” average time series of  $p_{t,t+\Delta t}$ , which is obtained by calculating the intraday time series of  $p_{t,t+\Delta t}$  for each trading date first and then averaging the results of all dates for the same time period  $[t, t + \Delta t]$ . The trend of the figure is similar to Fig. 12(c), the intraday average time series of API, except that it is much smoother. We observe that the average probability of increasing spread is between 0.46 and 0.5 at the market open and then rises to 0.5 in around an hour.

We then consider cross-sectional patterns. Figs. 18(a) and 18(b) show the histogram of  $\lambda$ 's and  $\gamma$ 's for all 1,936 stocks, respectively, and Table 7 contains their summary statistics. Both the figures and the table show that the  $\gamma$  for almost all stocks are positive in 2019–2021. In other words, a positive API will lead to a higher probability of spread expansion (i.e., worse liquidity).

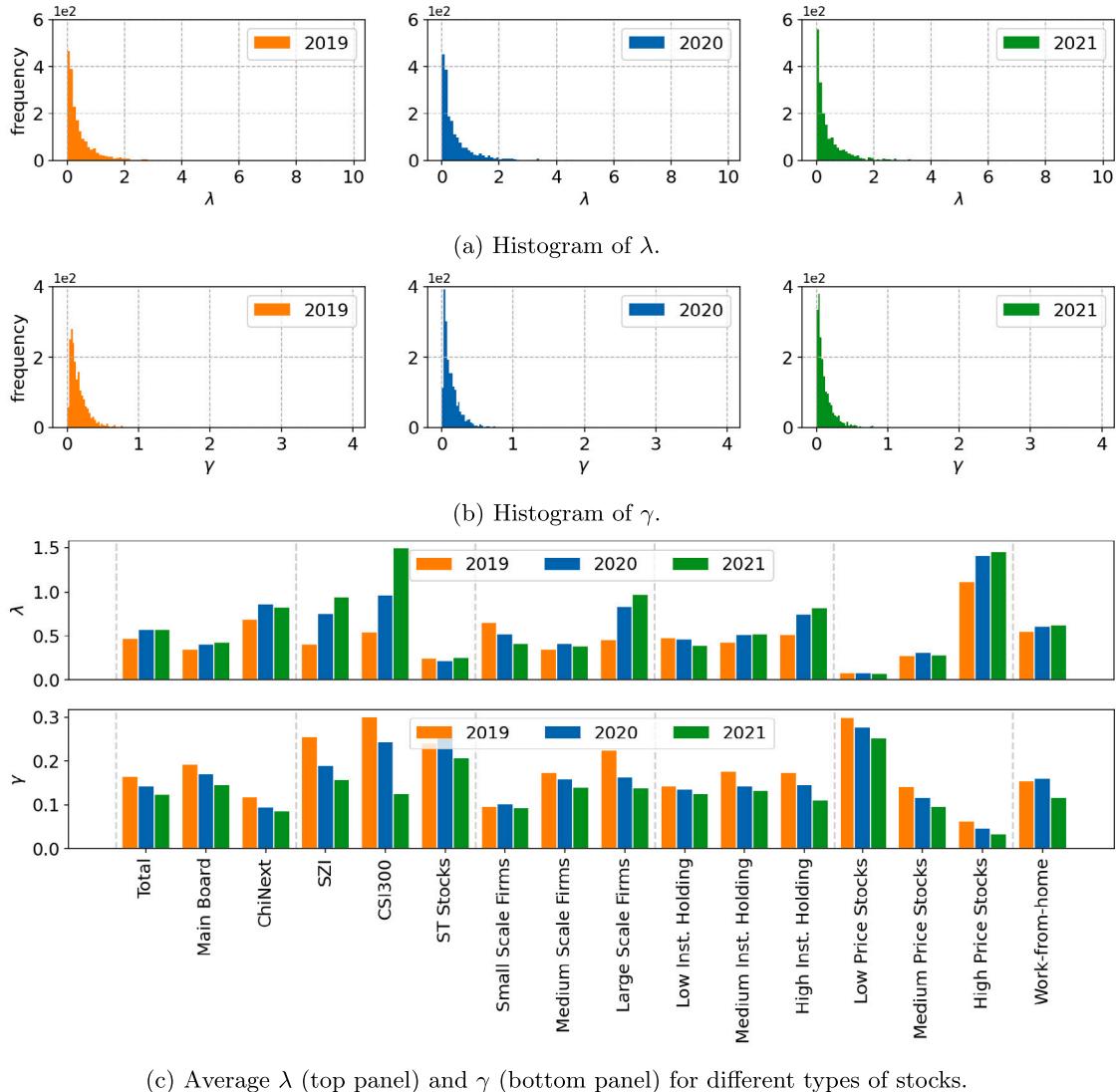
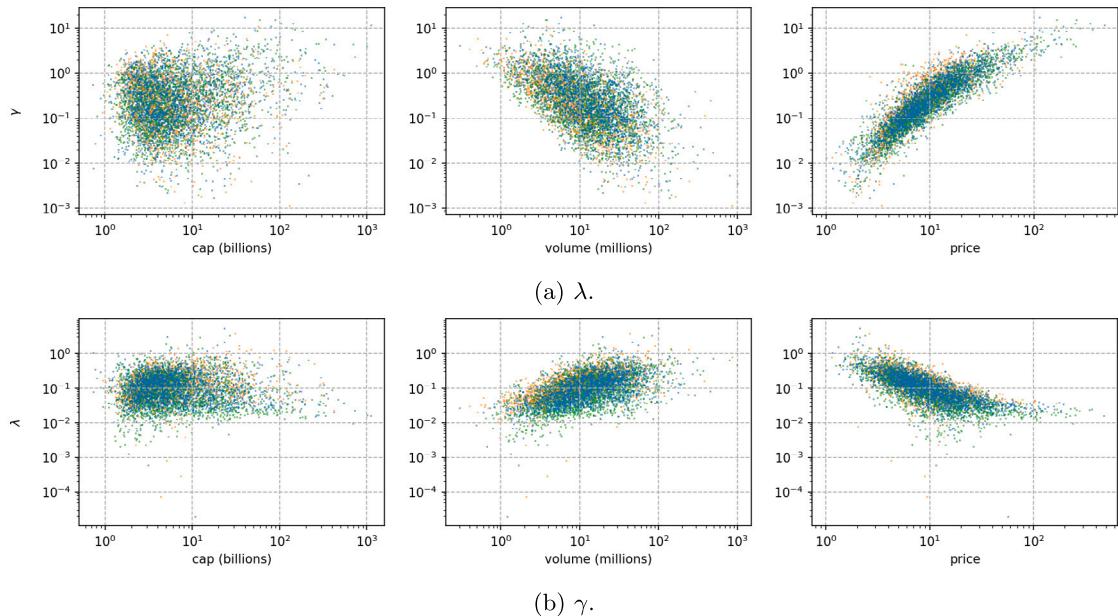
(c) Average  $\lambda$  (top panel) and  $\gamma$  (bottom panel) for different types of stocks.Fig. 18. Histogram and cross-sectional patterns of  $\lambda$ 's and  $\gamma$ 's in the discrete classification model for 1,936 stocks in 2019, 2020, and 2021.

Fig. 18(c) shows the average  $\lambda$ 's and  $\gamma$ 's for different categories of stocks in each year. Interestingly, the results are slightly different from the results presented in Section 6.3.1. On the one hand, we find that the  $\lambda$ 's are positively correlated with the stock price, the firm scale, and the institutional holding. This is intuitive because  $\lambda$  measures the magnitude of spread change, and the stocks of large-scale firms and with large institutional holdings usually have high prices, which lead to a high magnitude of spread (Eisler et al., 2012). This result is also consistent with the cross-sectional patterns of  $\beta$  (Fig. 16(b) in Section 6.3.1) and spread (Fig. 6 in Section 5.2).

On the other hand, we notice that  $\gamma$  is negatively correlated with the stock price and appears to not be correlated with firm scale and institutional holding. Because  $\gamma$  measures the sensitivity of the direction of spread change with regards to API, our empirical results imply that the API has a weaker impact on the direction of liquidity change for higher-priced stocks. This result is also intuitive since the liquidity of high-price stocks should be far more stable and the imbalance of aggressive and passive orders should have less of an impact on the liquidity of these stocks.

In addition, Fig. 19 shows the scatter plot of  $\lambda$  and  $\gamma$  versus the market values, average volumes, and average prices of all stocks from 2019 to 2021. It further demonstrates that the  $\lambda$  and  $\gamma$  are negatively and positively correlated with the stock price, respectively. The pattern for average volume is thus reversed. Meanwhile, neither parameter is significantly correlated with the market value of stocks. These correlations are consistent for the three years 2019, 2020, and 2021.

Finally, to study the goodness-of-fit of the discrete classification model, we calculate the area under the receiver operating characteristic (ROC) curve (AUC) of the classification. Because the AUC is only defined for binary classification problems, we consider



**Fig. 19.** Scatter plots of  $\lambda$  and  $\gamma$  versus market capitalization, average volume and average price for 1,936 stocks in 2019 (orange), 2020 (blue), and 2021 (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the following two binary classification methods:

- (1) We classify the spread change as non-negative if  $API_{t,t+\Delta t} > 0$ , and classify the spread change as negative if  $API_{t,t+\Delta t} < 0$ .
- (2) We classify the spread change as positive if  $API_{t,t+\Delta t} > 0$ , and classify the spread change as non-positive if  $API_{t,t+\Delta t} < 0$ .

We denote the corresponding AUCs for both methods by  $AUC_1$  and  $AUC_2$ . The summary statistics of AUCs for all stocks are shown in Table 7. For the binary classification method (1), the average AUC is 0.56 during 2019–2021, whereas for method (2), it is 0.58 in 2019 and 0.59 in both 2020, and 2021. Since both values are greater than 0.5, these results demonstrate that using API to classify spread changes performs better than making a random guess.

In conclusion, our API model provides a supplement to the empirical facts of various liquidity measures shown in Section 4, highlighting the mechanism through which liquidity is *determined* by order flows. The linear model (8) empirically reveals that the imbalance between aggressive orders and passive orders, i.e., the API, determines the change of liquidity (spread). The intraday patterns of API explain the W-shaped liquidity and its intraday periodicity. Furthermore, the discrete classification model (14) demonstrates that the impact of API on the direction of liquidity change is negatively correlated with the stock price, but not the company scale and the level of institutional ownership. Meanwhile, stocks with higher stock prices, firm scales, or institutional ownership show higher magnitudes of liquidity change in response to API.

## 7. Conclusion

In this article, we provide a comprehensive analysis of market liquidity in the Chinese stock market based on the *SZSE Historical Tick Data* of the Shenzhen Stock Exchange in 2019–2021. In total, we have processed approximately  $1,936$  (stocks)  $\times$   $730$  (trading days)  $\times$   $33,000$  (average number of events per stock per day) =  $46.64$  billion events and reconstructed the evolution of the limit order book with the most granular information possible. This enables us to compare different liquidity measures systematically, analyze their time-series and cross-sectional patterns, and study their determinants. Our comprehensive dataset with the reconstructed order book also opens up new avenues for microstructure-related research in the Chinese stock market.

We study intraday patterns of trading cost-based and execution efficiency-based liquidity measures, using the average spread and order lifetime as primary examples. Due to the lunch break in the Chinese stock market, the intraday shape of the average spread resembles a W-shape rather than the traditional U-shape in developed markets. Moreover, we also observe intraday periodicities in the time series of the average spread.

Moreover, we investigate the cross-sectional patterns of different liquidity measures. First, stocks with the special treatment status (ST) show worse liquidity as shown by their higher proportional spread (including the proportional effective spread and the proportional weighted average spread), lower execution ratios (including execution ratios at a single price, at a single time, and at the best price), higher order lifetimes, lower estimated  $\alpha$ 's, and larger Kyle's Lambdas compared to the average level of the

corresponding liquidity measures across all stocks. In addition, stocks in the CSI 300 Index show better liquidity indicated by their small proportional spread, high execution ratios, low order lifetimes, high estimated  $\alpha$ 's, and low Kyle's Lambdas. We also find that liquidity is correlated with firm size and the scale of stock prices. Stocks for large firms or with a high stock price tend to have a larger spread but a smaller proportional spread than other stocks. Both the order lifetime and Kyle's Lambda decrease as the scale of the firm or the stock price increases. Overall, these results highlight several stock characteristics that are associated with the level of liquidity.

Our analysis also reveals improvements in market liquidity in the Chinese market from 2019 to 2021. Almost all liquidity measures improve significantly over these three years across nearly all categories of stocks. Some distinctive trading dynamics during the COVID-19 period are also observed. For example, more significant intraday periodic patterns at 5- and 10-minute frequencies emerged in 2020, likely due to increased retail trading activities. In addition, "Work-from-home" stocks exhibit the highest liquidity based on measures of trading costs and price impact during the COVID-19 pandemic, implying increased attention on these stocks.

To explain the change in spread, we propose a function of the imbalance between aggressive and passive orders, which we define as the aggressive–passive imbalance (API). The empirical results demonstrate that the impact of API on spread change is statistically significant. Furthermore, the intraday patterns of liquidity, such as the W-shape and periodicity, can also be explained by the API. Much like the intuition of "liquidity begets liquidity" first proposed in Foucault et al. (2013), we discover that with the intraday patterns for API, "liquidity begets liquidity periodically" in the Chinese stock market.

Empirical results of a discrete classification model of the change in spread based on the API demonstrate that the impact of API on the direction of liquidity change is negatively correlated with the stock price. Meanwhile, stocks with higher prices, from larger firms, and with more institutional holdings demonstrate higher magnitudes of liquidity change. These implications from our model not only align with the empirical facts of liquidity measures but also shed light on the formation mechanism of high-frequency market liquidity. The API model provides a complement to the existing research on characterizing liquidity from an alternative perspective that is different from the inventory-based approaches proposed by studies such as Grossman and Miller (1988) and Bogousslavsky and Collin-Dufresne (2023).

Overall, our comprehensive study provides the measurements, patterns, and determinants for high-frequency liquidity, which we hope will serve as an empirical benchmark for liquidity in the Chinese stock market whilst also providing the impetus for more in-depth analysis of market liquidity and other microstructure topics.

#### CRediT authorship contribution statement

**Chaoyi Zhao:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Yufan Chen:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Lintong Wu:** Writing – review & editing, Writing – original draft, Visualization, Methodology. **Yuehao Dai:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology. **Ermo Chen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology. **Lan Wu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Funding acquisition. **Ruixun Zhang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Funding acquisition.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.pacfin.2025.102681>.

#### References

- Abdi, F., Ranaldo, A., 2017. A simple estimation of bid–ask spreads from daily close, high, and low prices. *Rev. Financ. Stud.* 30 (12), 4437–4480.
- Abergel, F., Anane, M., Chakraborti, A., Jedidi, A., Toke, I.M., 2016. Limit Order Books. Cambridge University Press.
- Acharya, V.V., Pedersen, L.H., 2005. Asset pricing with liquidity risk. *J. Financ. Econ.* 77 (2), 375–410.
- Adams, S.W., Kasten, C., Kelley, E.K., 2024. How free is retail? Trading costs with zero commissions. *J. Bank. Financ.* 165, 107226.
- Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. *Rev. Financ. Stud.* 1 (1), 3–40.
- Adrian, T., Fleming, M., Shachar, O., Vogt, E., 2017. Market liquidity after the financial crisis. *Annu. Rev. Financ. Econ.* 9, 43–83.
- Almgren, R., Chriss, N., 2001. Optimal execution of portfolio transactions. *J. Risk* 3, 5–40.
- Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid–ask spread. *J. Financ. Econ.* 17 (2), 223–249.
- Avellaneda, M., Kasyan, G., Lipkin, M.D., 2012. Mathematical models for stock pinning near option expiration dates. *Comm. Pure Appl. Math.* 65 (7), 949–974.
- Baldauf, M., Mollner, J., 2020. High-frequency trading and market performance. *J. Finance* 75 (3), 1495–1526.
- Bertsimas, D., Lo, A.W., 1998. Optimal control of execution costs. *J. Financial Mark.* 1 (1), 1–50.
- Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *J. Finance* 50 (5), 1655–1689.
- Billio, M., Lo, A.W., Pelizzon, L., Getmansky Sherman, M., Zareei, A., 2022. Global Realignment in Financial Market Dynamics. SAFE Working Paper.
- Bogousslavsky, V., Collin-Dufresne, P., 2023. Liquidity, volume, and order imbalance volatility. *J. Finance* 78 (4), 2189–2232.
- Bongaerts, D., Van Achter, M., 2021. Competition among liquidity providers with access to high-frequency trading technology. *J. Financ. Econ.* 140 (1), 220–249.

