Supplement to "High-Frequency Liquidity in the Chinese Stock Market: Measurements, Patterns, and Determinants"

Chaoyi Zhao[123], Yufan Chen[1], Lintong Wu[1], Yuehao Dai[1],
Ermo Chen[1], Lan Wu[145], and Ruixun Zhang[1456]

[1]Department of Financial Mathematics, School of Mathematical Sciences, Peking University
[2]Sloan School of Management, Massachusetts Institute of Technology
[3]Laboratory for Financial Engineering, Massachusetts Institute of Technology
[4]Laboratory for Mathematical Economics and Quantitative Finance, Peking University
[5]Center for Statistical Science, Peking University
[6]National Engineering Laboratory for Big Data Analysis and Applications, Peking University

November 20, 2024

A Trading Rules for the Shenzhen Stock Exchange

We introduce the trading rules for the Exchange[1] which is related to the re-matching procedure in Appendix C.

A.1 Trading Market

Trading instruments such as stocks, mutual funds, bonds, and warrants can be traded on the Exchange. The Exchange is open for trading from Monday to Friday and closed on public holidays and other days as pre-announced by the Exchange. Auction trading of securities begins on each trading day with an opening call auction between 9:15-9:25, followed by continuous trading between 9:30-11:30 and 13:00-14:57, and finally the closing call auction between 14:57-15:00. The Exchange's trading system will not accept cancellation of any auction orders between 9:20-9:25 and 14:57-15:00 on each trading day. After the close of each trading day, unexecuted orders will be removed by the system.

A.2 Orders

Investors may place a limit order or market order through members for securities trading. The tick size of the quote price of an order is RMB 0.01. Orders are matched and executed according to the principle of **price-time priority**, which means priority is given to a higher buy order over a lower buy order, and a lower sell order is prioritized over a higher sell order. The order sequence which is arranged according to the time when the Exchange trading system receives the orders determines the priority of trading for the orders with the same prices.

A **limit order** is an instruction given by an investor to a member to buy a particular security at a specified price or lower, or to sell at a specified price or higher.

A **market order** is an instruction given by an investor to a member to buy or sell a particular security at the current market price.

The Exchange accepts the following types of market orders in line with market conditions:

[1]See `http://www.szse.cn/English/rules/siteRule/P020181124401737559498.pdf` for the full version.

1. Opposite-side Best Price. Opposite-side Best Price is an order whose quote price is set at the best price on the opposite side in the order book at the time the order is routed into the Exchange trading system.

2. Same-side Best Price. Same-Side Best Price is an order whose quote price is set at the best price on the same side in the order book at the time the order is routed into the Exchange trading system.

3. Five Best Orders Immediate or Cancel. Five Best Orders Immediate or Cancel is executed in sequence against the current five best prices on the opposite side. In case that part of the order cannot be executed, the unfilled part of the order shall be canceled automatically.

4. Immediate or Cancel. Immediate or Cancel is an order that is executed in sequence against all the orders on the opposite side in the order book at the time the order is routed into the Exchange trading system. In case that part of the order cannot be executed, the unfilled part of the order shall be canceled automatically.

5. Fill or Kill. Fill or Kill is an order that must be executed in its entirety against all the orders on the opposite side in the order book at the time the order is routed into the Exchange trading system, otherwise the entire order shall be canceled automatically.

In the raw data, only Same-side Best Prices have a unique type label "U" in the order file, and the other four types of market orders share a common type label "1". We distinguish different types of market orders by analyzing their trade events as follows.

- If a market order's type is recorded as "U" in the order file, we identify it as a Same-side Best Price. We regard it as a limit order with a quote price of the best price on the same side.
- If a market order has execution events with a unique price and does not have any cancellation event right after its submission, we identify it as an Opposite-side Best Price. We regard it as a limit order with a quote price of the best price on the opposite side.
- If a market order has execution events with five distinct prices and a cancellation event right after its submission, we identify it as a Five Best Orders Immediate or Cancel. We regard it as a limit order with a quote price of the 5th best price on the opposite side following an immediate cancellation of its unexecuted part.
- If a market order is canceled right after its submission, we identify it as a Fill or Kill. We directly ignore it since we can only identify it when it is killed.
- Otherwise, we identify it as an Immediate or Cancel. We regard it as a limit order with a quote price of the price limit on the opposite side following an immediate cancellation of its unexecuted part.

A.3 Auction

Auction trading of securities is conducted either as a call auction or a continuous auction.

Call auction (9:15-9:25 and 14:57-15:00) refers to the process of one-time centralized matching of buy and sell orders accepted during a specified period of time. The execution price in a call auction shall be determined based on the following principles: (1) The price that generates the greatest trading volume; (2) The price which allows all the buy orders with a higher bid price and all the sell orders with a lower ask price to be executed; (3) The price which allows at least all the buy orders at the identical price or all the sell orders at the identical price to be executed. The orders that are not executed during the opening call auction automatically enter the continuous auction.

Continuous auction (9:30-11:30 and 13:00-14:57) refers to the process of continuous matching of buy and sell orders on a one-by-one basis. The execution price in a continuous auction shall be determined based on the following principles: (1) Where the highest bid price matches the lowest ask price, such price shall be taken as the execution price; (2) Where the bid price is higher than the lowest ask price currently available in the order book, such lowest ask price shall be taken as the execution price; (3) Where the ask price is lower than the highest bid price currently available in the order book, such highest bid price shall be taken as the execution price.

The Exchange imposes a daily price up/down limit of 10% on the trading of shares during the continuous auction. Orders whose quote price exceeds the valid price range are not accepted instantly for auction but will line up in the Exchange trading system.

A.4 Special Rules on the ChiNext Market

The Exchange enacted some special rules on the ChiNext Market since 24th August, 2020. For trading of ChiNext-listed stocks, the valid price range of limit orders during the continuous trading session shall meet both of the following requirements: (1) The bid price shall not be higher than 102% of the best bid price; (2) The ask price shall not be lower than 98% of the best ask price. Orders whose quote price exceeds the valid price range are not accepted instantly for auction but will line up in the Exchange trading system. When their quote price falls within the valid price range with the fluctuation of prices, the Exchange trading system will take out such orders automatically for auction.

B Notations for Limit Order Book

We formulate a precise description of the trading in the Exchange.

Definition A.1 (Order)**.** An order $x = (p_x, v_x, t_x)$ submitted at time t_x with price p_x and size $v_x > 0$ ($v_x < 0$) is a commitment to sell (buy) up to $|v_x|$ units of the traded asset at a price no less than (no greater than) p_x.

We call x a buy order if $v_x < 0$, and a sell order if $v_x > 0$. We denote by $\mathcal{X}(t)$, $\mathcal{X}^{\mathrm{b}}(t)$, and $\mathcal{X}^{\mathrm{a}}(t)$ the collection of all orders, all buy orders, and all sell orders submitted before t, respectively. To be specific,

$$\mathcal{X}(t) := \{x : t_x < t\}, \quad \mathcal{X}^b(t) := \{x : t_x < t, v_x < 0\}, \quad \mathcal{X}^a(t) := \{x : t_x < t, v_x > 0\}.$$

Meanwhile, we define $\mathcal{X} := \mathcal{X}(\infty)$, $\mathcal{X}^{\mathrm{b}} := \mathcal{X}^{\mathrm{b}}(\infty)$ and $\mathcal{X}^{\mathrm{a}} := \mathcal{X}^{\mathrm{a}}(\infty)$ the collection of all orders, all buy orders and all sell orders in a given trading day.

After submission, an order will face two kinds of events: execution and cancellation. Here we define these events formally.

Definition A.2 (Event)**.** An event is defined as a 5-tuple $(x^{e,b}, x^{e,a}, v_e, p_e, t_e)$, where $x^{e,b}$ and $x^{e,a}$ are the bid and sell orders involved in this event and t_e is the time when the event occurs.

- Execution event: if $x^{e,b}, x^{e,a} \neq \emptyset$, then the buy order $x^{e,b}$ and the sell order $x^{e,a}$ are matched and executed at price p_e with v_e units of the stocks. If $t_{x^{e,b}} > t_{x^{e,a}}$, we say the event is bid-initiated, otherwise we say it is ask-initiated.

- Cancellation event: If $x^{e,a} = \emptyset$, then the buy order $x^{e,b}$ cancels v_e units of orders; if $x^{e,b} = \emptyset$, then the sell order $x^{e,a}$ cancels v_e units of orders.

For order x, we define the event set and the execute event set related to x as

$$\mathcal{E}_x = \begin{cases} \{e : x^{e,b} = x\}, & x \in \mathcal{X}^b \\ \{e : x^{e,a} = x\}, & x \in \mathcal{X}^a \end{cases}, \quad \mathcal{E}_x^e = \begin{cases} \{e : x^{e,b} = x, x^{e,a} \neq \emptyset\}, & x \in \mathcal{X}^b \\ \{e : x^{e,a} = x, x^{e,b} \neq \emptyset\}, & x \in \mathcal{X}^a \end{cases},$$

respectively. Besides, we define the death time of x as $T_x = \max_{e \in \mathcal{E}_x} t_e$. After time T_x, the order x disappears from the market because it is either totally executed or canceled.

Definition A.3 (Limit Order Book, LOB)**.** An LOB $\mathcal{L}(t)$ is the set of all active orders in a market at time t. To be specific, $\mathcal{L}(t) = \{x : t_x \leq t < T_x\}$.

The active orders in an LOB $\mathcal{L}(t)$ can be partitioned into the set of active buy orders, $\mathcal{B}(t)$, for which $v_x < 0$, and the set of active sell orders, $\mathcal{A}(t)$, for which $v_x < 0$. An LOB can then be considered as a set of queues, each of which consists of active buy or sell orders at a specified price.

Definition A.4 (Bid/Ask-side Depth)**.** The bid-side depth and ask-side depth available at price p and at time t are

$$n^b(p,t) := \sum_{\{x \in \mathcal{B}(t): p_x = p\}} v_x, \quad n^a(p,t) := \sum_{\{x \in \mathcal{A}(t): p_x = p\}} v_x,$$

respectively.

Definition A.5 (Bid/Ask i Price/Size)**.** Let $P^b(t) := \{p_x : x \in \mathcal{B}(t)\}$ and $P^a(t) := \{p_x : x \in \mathcal{A}(t)\}$. For $i = 1, 2, \ldots, |P^b(t)|$, we define the bid i price at time t, $p_i^b(t)$, as the i-th largest element in $P^b(t)$, and call $n_i^b(t) := n^b(p_i^b, t)$ the bid i size at time t. For $i = 1, 2, \ldots, |P^a(t)|$, we define the ask i price at time t, $p_i^a(t)$, as the ith smallest element in $P^a(t)$, and call $n_i^a(t) := n^b(p_i^a(t), t)$ the ask i size at time t.

In particular, we define the best bid price $b(t)$, best ask price $a(t)$, and mid-price $m(t)$ at time t as

$$b(t) := p_1^b = \max_{x \in \mathcal{B}(t)} p_x, \quad a(t) := p_1^a = \min_{x \in \mathcal{A}(t)} p_x, \quad m(t) := \frac{a(t) + b(t)}{2},$$

respectively.

Definition A.6 (Aggressive/Passive Orders)**.** We say an order $x = (p_x, v_x, t_x)$ is

(a) an aggressive buy order, if $v_x < 0$ and $p_x \geq a(t_x)$;

(b) an aggressive sell order, if $v_x > 0$ and $p_x \leq b(t_x)$;

(c) a passive buy order, if $v_x < 0$ and $p_x < a(t_x)$;

(d) a passive sell order, if $v_x > 0$ and $p_x > b(t_x)$.

According to the definition, aggressive orders are orders that can be executed immediately, while passive orders will stay in the limit order book for a while and wait for other aggressive orders to trigger execution events. Each transaction event is completed by an aggressive order and one or more passive orders.

Not all aggressive orders can be executed and disappear immediately. Actually, once an aggressive limit order is submitted, maybe a portion of its volumes will be traded immediately, while others will be left on the limit order book and waiting for later transactions. In our empirical studies, for aggressive orders, we regard their volumes traded immediately as aggressive orders, and regard the volumes left on the order book as passive orders. In other words, we divide these aggressive orders into "aggressive parts" and "passive parts".

C Order Book Reconstruction Based on Trading Rules

In this section, we briefly introduce the auction mechanism in the Exchange, and describe the detailed order book reconstruction procedures for each trading day and each stock. Figure A.1 summarizes the entire reconstruction procedures.

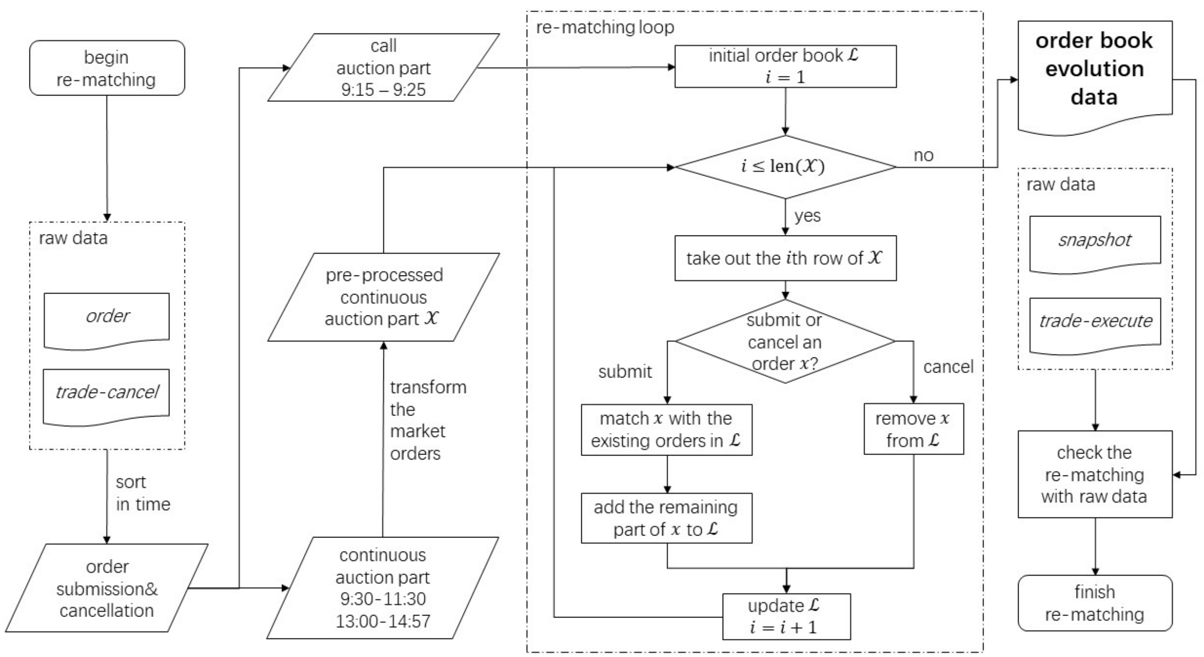


Figure A.1: Procedures of order book reconstruction.

There are two types of auctions in the Exchange: call auction and continuous auction. Call auction refers to the process of a one-time centralized matching of buy and sell orders submitted during a specified period of time, while continuous auction refers to the process of continuously matching buy and sell orders as they arrive. On each day, the trading of all stocks begins with an opening call auction between 9:15–9:25, followed by continuous auctions between 9:30–11:30 and 13:00–14:57, and ends with a closing call auction between 14:57–15:00.[2] In contrast to the US market, there is a 1.5-hour lunch break from 11:30 to 13:00

[2]See Article 2.4.2 in Shenzhen Stock Exchange (2016).

in the Chinese stock market. Market participants can place limit orders and cancellations during the call auction, while market orders are allowed in addition during the continuous auction periods.[3]

As instructed by the auction rules in the Exchange, we re-match the orders as follows. We first obtain the initial limit order book at the beginning of the continuous auction (9:30) by collecting all order submissions and cancellations during the opening call auction (9:15–9:25), and we match the remaining orders at 9:25. After the matching, all unmatched orders after the call auction form the order book at 9:30, when the continuous auction begins.

During the continuous auction (9:30–11:30 and 13:00–14:57), we process all order submissions and cancellations one by one in chronological order. For each new order submitted to the Exchange, we check whether it is possible to be executed immediately by matching with resting orders in the order book at a price within the 10% daily price limit.[4] The unexecuted part of the new order—if that exists—remains in the limit order book. Algorithm 1 summarizes the step-by-step procedure for processing limit order submissions in our empirical study, with the notations defined in Appendix B. In addition, for each order cancellation, we simply remove the corresponding order from the limit order book. With these steps, we update the limit order book every time there is a new order submission or cancellation. Because the limit order book changes only at these timestamps, we obtain a complete reconstruction of the evolution of the limit order book for the entire trading day.

Algorithm 1: Procedures for processing limit order submissions.

Input: A new order submission x at time t_x with price p_x and quantity v_x; The order book right before the submission: $\mathcal{L}(t_x-)$; The upper price limit $p_{\max}$ and lower price limit $p_{\min}$.
Output: The order book right after the submission: $\mathcal{L}(t_x)$.

```
if x is a buy order, i.e. v_x < 0 then
    for execution price p ranges from the best ask price in L(t_x−), a(t_x−), up to min{p_x, p_max} do
        for sell order y in L(t_x−) with quote price p_y = p and quantity v_y (y's are ordered by time priority) do
            Set the matched size v ← min{|v_x|, |v_y|};
            Remove the matched size from both the orders x and y, i.e., v_x ← v_x + v, v_y ← v_y − v;
            if y is all matched, i.e., v_y = 0 then
                Remove y from the current order book L(t_x−);
            if x is all matched, i.e., v_x = 0 then
                Break;
else if x is a sell order, i.e. v_x > 0 then
    for execution price p ranges from the best bid price in L(t_x−), b(t_x−), down to max{p_x, p_min} do
        for buy order y in L(t_x−) with quote price p_y = p and quantity v_y (y's are ordered by time priority) do
            Set the matched size v ← min{|v_x|, |v_y|};
            Remove the matched size from both the orders x and y, i.e., v_x ← v_x − v, v_y ← v_y + v;
            if y is all matched, i.e., v_y = 0 then
                Remove y from the current order book L(t_x−);
            if x is all matched, i.e., v_x = 0 then
                Break;
if x has an unmatched part, i.e., v_x ≠ 0 then
    Add x with quantity v_x to the level p_x of the order book L(t_x−).
Set the order book right after the submission L(t_x) ← L(t_x−).
```

Market orders need to be dealt with separately. The Exchange accepts five types of market orders,[5]

[3]However, not all cancellations can be accepted by the exchange. The Exchange's trading system does not accept cancellations during 9:20–9:25 and 14:57–15:00 on each trading day, see Article 3.3.1 in Shenzhen Stock Exchange (2016).

[4]The price limit is a special feature of the Chinese stock market. The Exchange requires that no trades can be executed outside the upper/lower price limit of 10% with respective to the previous closing price. For stocks under special treatment (ST) status, the limit is 5%.

[5]This includes Opposite-side Best Price, Same-side Best Price, Five Best Orders Immediate or Cancel, Immediate or Cancel, and Fill or Kill.

but the specific type of each market order is not recorded in the `order` file provided by the Exchange. Therefore, in order to match the market orders correctly, we retrieve the type of each market order from its corresponding execution and cancellation events, and then treat each market order as a combination of a limit order and an order cancellation.[6] Appendix A.2 of the Online Supplementary Material introduces additional details for identifying and transforming different types of market orders.

It is also worth noting that the Exchange enacted certain special rules on the ChiNext market,[7] which took effect on Aug. 24th, 2020. According to the special rules, the valid price range of limit orders for ChiNext-listed stocks during the continuous auction shall meet both of the following requirements: (1) a buy order's quote price shall not be higher than 102% of the best ask price; (2) a sell order's quote price shall not be lower than 98% of the best bid price.[8] Orders whose quote prices exceed the valid price range are not immediately accepted for auction, but rather lined up in the Exchange's trading system. When their quote prices fall within the valid price range as the stock price fluctuates, the Exchange automatically adds these orders for auction.[9] To address this special rule, we additionally maintain a "frozen book" to collect these orders temporarily on hold. Whenever the best bid or ask price changes, we check the frozen book to process orders whose prices fall into the valid price range.

After the steps above, we finally recover the limit order book for all levels throughout the day, and we verify our order book reconstruction results by checking that the execution events after the reconstruction and from the `trade` file match exactly.

D Details for the Price Impact Regression

This appendix gives the details for the price impact regression model (2), which is first proposed by Glosten and Harris (1988).

For the n-th execution, we denote its direction by d_n and its (directional) executed quantity by q_n. To be more specific, if this trade is bid-initiated, then $d_n = 1$ and $q_n > 0$; otherwise $d_n = -1$ and $q_n < 0$. We further assume that the n-th execution occurs at price p_n and, at the same time, that the consensus value of the stock is μ_n.

First, we consider the order-processing costs. On the one hand, there is a constant part of execution cost per share for each execution. On the other hand, the execution costs may vary with the quantity of executions. For example, executions with a large quantity are likely to obtain better execution prices, which can be regarded as "quantity discounts." Thus, we assume that the price of the n-th execution p_n satisfies that

$$p_n = \mu_n + \gamma_0 d_n + \gamma_1 q_n, \tag{A.1}$$

where $\gamma_0 d_n$ is the constant part of the execution costs and $\gamma_1 q_n$ is the part of the execution costs varying with the quantity of executions. For example, a bid-initiated execution should bear a constant execution cost of γ_0, while an additional executed share will lead to an execution price increment of γ_1. We note that the quantity discount exists if $\gamma_1 < 0$.

Second, the variation of the consensus value, $\mu_n - \mu_{n-1}$, is affected by the information in the order flow,

[6] The order cancellation part exists for "Five Best Orders Immediate or Cancel", "Immediate or Cancel", and "Fill or Kill" market orders.

[7] ChiNext is a sub-market in the Exchange to facilitate the development of innovative and growth businesses. Firms listed on ChiNext are typically small- and medium-sized firms with high debt-to-asset ratios.

[8] See Article 2.1 in Shenzhen Stock Exchange (2020).

[9] See Article 2.4 in Shenzhen Stock Exchange (2020).

which leads to adverse selection costs. Both the side, d_n, and quantity, q_n, of the execution have influences on the variation of consensus value. Specifically:

$$\mu_n = \mu_{n-1} + \lambda_0 d_n + \lambda_1 q_n + \epsilon_n, \tag{A.2}$$

where ϵ_n is the random noise. For instance, a bid-initiated execution will lead to a consensus value increment of λ_0, and an additional executed share will lead to a consensus value increment of λ_1.

Finally, the consecutive execution price changes can be modeled by combining (A.1) and (A.2) as:

$$p_n - p_{n-1} = \lambda_0 d_n + \lambda_1 q_n + \gamma_0(d_n - d_{n-1}) + \gamma_1(q_n - q_{n-1}) + \epsilon_n, \tag{A.3}$$

which is the price impact regression model (2) that Glosten and Harris (1988) consider.

E Sample Stocks and Their Classification

In the empirical studies, we consider the stocks which are listed on the Shenzhen Stock Exchange before Sep. 30th, 2018, and are normally traded for more than 200 days in each year of 2019, 2020, and 2021.

Following Shenzhen Stock Exchange (2021), we analyze the liquidity characteristics for several stock classifications in each year, including Main Board, ChiNext, Special Treatment (ST) stocks, and sample stocks for Shenzhen Component Index (SZI Index, for short) and CSI 300 Index (in SZSE). Moreover, we divide the stocks into several categories based on firm size, institutional holding, and stock price. To be specific, in each year from 2019–2021, we allocate the stocks in three categories based on the breakpoints for the top 30%, middle 40%, and bottom 30% of the daily average market capitalizations of stocks (large scale stocks, medium scale stocks, and small scale stocks), daily average institutional holdings of stocks (high institutional holding stocks, medium institutional holding stocks, and low institutional holding stocks), and average transaction prices of stocks (high price stocks, medium price stocks, and low price stocks) in that year. Finally, to study the impact of COVID-19 on liquidity, we consider the "work-from-home" stocks, which are the components of the "Work-from-home" Index on Eastmoney.com.[10]

Table A.1 presents the basic information of the classifications.

[10]See `https://quote.eastmoney.com/bk/90.BK0912.html` for the details of the "Work-from-home" Index. Eastmoney.com is one of China's leading financial and investment portals, offering a wide range of financial services such as market data, stock analysis, fund management, and general financial news.

Table A.1: Basic information of the classifications of stocks.[1]

	Stock Number	Effective Trading Day[2]	Volume per Day (million)	Turnover per Day (million)	Average Price	Market Cap.[3] (billion)	Inst. Holding (%)
(1) 2019							
Total	1,936	237.96	13.27	138.58	13.99	7.73	34.99
Main Board	1,236	238.42	15.12	148.83	12.10	9.57	39.78
ChiNext	700	237.14	10.01	120.49	17.33	4.50	26.50
Shenzhen Component Index	454	239.95	26.17	307.82	17.42	22.10	46.91
CSI 300 Index (in SZSE)	110	240.93	39.52	570.07	25.82	54.50	56.94
ST Stocks	36	219.22	9.67	43.65	6.00	2.89	35.93
Small Scale Firms	581	236.96	5.04	55.36	14.69	1.51	20.81
Medium Scale Firms	774	237.49	11.47	92.38	10.83	3.58	34.18
Large Scale Firms	581	239.57	23.90	283.36	17.49	19.49	50.15
Low Inst. Holding	580	237.19	11.73	109.46	13.20	3.08	8.37
Medium Inst. Holding	772	237.58	14.76	136.89	12.95	5.64	34.32
High Inst. Holding	580	239.21	12.90	170.60	16.12	15.22	62.50
Low Price Stocks	581	237.29	19.50	93.32	5.06	5.30	33.09
Medium Price Stocks	774	238.32	12.78	128.85	10.21	6.42	36.98
High Price Stocks	581	238.15	7.69	196.81	27.96	11.91	34.24
Work-from-home	23	237.13	22.42	252.03	18.53	8.30	22.45
(2) 2020							
Total	1,936	236.01	18.59	234.82	17.01	10.76	36.03
Main Board	1,236	235.98	20.07	235.27	14.17	12.74	40.74
ChiNext	700	236.07	15.98	234.03	22.01	7.25	27.71
Shenzhen Component Index	460	238.32	34.08	550.35	26.83	32.24	49.55
CSI 300 Index (in SZSE)	104	240.04	52.49	1,060.38	40.85	84.34	57.42
ST Stocks	55	219.40	11.65	42.87	4.99	2.51	32.78
Small Scale Firms	581	235.28	8.17	74.21	12.50	1.87	24.61
Medium Scale Firms	774	235.70	16.09	142.25	12.72	4.41	34.47
Large Scale Firms	581	237.17	32.34	518.74	27.24	28.10	49.42
Low Inst. Holding	579	235.98	17.69	172.30	13.82	3.98	8.99
Medium Inst. Holding	772	235.24	21.35	258.79	15.72	8.84	35.62
High Inst. Holding	579	237.09	15.92	267.16	22.00	20.19	63.62
Low Price Stocks	581	234.96	26.19	129.00	5.03	5.69	33.17
Medium Price Stocks	774	236.54	17.50	188.20	11.09	7.33	36.07
High Price Stocks	581	236.36	12.44	402.74	36.86	20.39	38.83
Work-from-home	23	236.52	25.23	340.72	21.29	12.19	20.57
(3) 2021							
Total	1,936	238.31	18.48	264.16	18.20	13.47	35.72
Main Board	1,236	236.56	20.56	271.87	15.54	15.41	40.26
ChiNext	700	241.41	14.81	250.53	22.89	10.03	27.72
Shenzhen Component Index	456	239.63	33.26	640.70	33.99	42.02	49.01
CSI 300 Index (in SZSE)	109	240.13	48.67	1,346.29	58.56	107.08	55.33
ST Stocks	64	225.97	13.05	62.67	5.80	2.82	33.60
Small Scale Firms	581	237.95	8.14	66.97	10.65	2.03	24.16
Medium Scale Firms	774	238.37	15.70	138.50	12.59	4.80	34.00
Large Scale Firms	581	238.60	32.52	628.75	33.22	36.45	49.52
Low Inst. Holding	580	239.09	17.00	176.37	12.87	4.38	9.20
Medium Inst. Holding	772	237.88	20.00	265.31	16.61	10.46	35.32
High Inst. Holding	580	238.18	17.94	351.91	25.76	26.64	62.79
Low Price Stocks	581	237.35	23.76	115.22	4.95	5.60	31.90
Medium Price Stocks	774	238.50	18.51	196.49	10.99	7.76	34.44
High Price Stocks	581	239.02	13.15	503.24	41.06	28.94	41.23
Work-from-home	23	239.87	22.95	309.27	20.37	14.85	20.28

[1] All the columns are the average in the corresponding classifications except for the stock number.
[2] "Effective trading day" means the trading days without touching the high or low price limits.
[3] Also called "Negotiable Market Capitalization" in the Shenzhen Stock Exchange. See `http://www.szse.cn/English/siteMarketData/marketStatistics/overview/index.html`.

F Additional Empirical Results

In this section, we provide some additional empirical results on the liquidity measures defined in Section 4 for a few stocks on specific trading dates.

In particular, we take three different samples for illustration purposes, including Ping An Bank on Jan. 2nd, 2020 (Sample I), Ping An Bank on Feb. 3rd, 2020 (Sample II), and Changchun High and New Technology Industries on Jan. 2nd, 2020 (Sample III); see Table A.2. Note that Ping An Bank is the first-ever financial company listed on the Exchange with great market liquidity and a relatively low stock price (around RMB 16.00), while Changchun High and New Technology Industries has the largest average spread among all stocks in our dataset and a high stock price (around RMB 500.00). Jan. 2nd, 2020 is the first trading day of 2020, and Feb. 3rd, 2020 is the first trading day after the announcement of the COVID-19 lockdown in Wuhan, China, which is included to reflect extreme market conditions.

Table A.2: Information of sample data.

Sample	I	II	III
Stock Code	000001[1]	000001	000661[2]
Date	Jan. 2nd, 2020	Feb. 3rd, 2020	Jan. 2nd, 2020

[1] Ping An Bank
[2] Changchun High and New Technology Industries

F.1 Daily Summary Statistics for Single Stocks

Table A.3 reports the descriptive statistics for several variables of Sample I in Table A.2, including the 3-second mid-price returns, the quantities of orders, the price of orders, the (RMB) volume of all submitted orders in 3-second windows, and the (RMB) volume of all executed orders in 3-second windows. The minimum number of shares of an order is one,[11] while the maximum number of shares of an order is one million, which is the maximum number of shares for one order allowed by the Exchange.

In particular, we consider the empirical distributions for the following characteristics of the order flow and order book.

Size of Orders. Figure A.2a provides the fitted tail distribution of order sizes in our data based on both the log-normal and the power-law distributions. We find that the log-normal distribution captures the tails better than the power-law distribution. This is different from Abergel et al.'s (2016) results for developed markets which documented that "the distribution of order sizes is complex to characterize, and a power-law distribution is often suggested."

Size of Executed and Submitted Orders in 3-second Windows. In addition to the distribution for the size of a single order, we also fit the tail distribution for the total size of executed orders and submitted orders in 3-second windows, which are shown in Figure A.2b and Figure A.2c, respectively. Consistent with the results for single orders, the log-normal distribution performs better than the power-law when fitting the tail of order sizes in our data.

[11] This is not a multiple of 100 shares, the minimum allowed by the exchange, since the Exchange allows to sell existing positions of less than 100 shares via a single order.

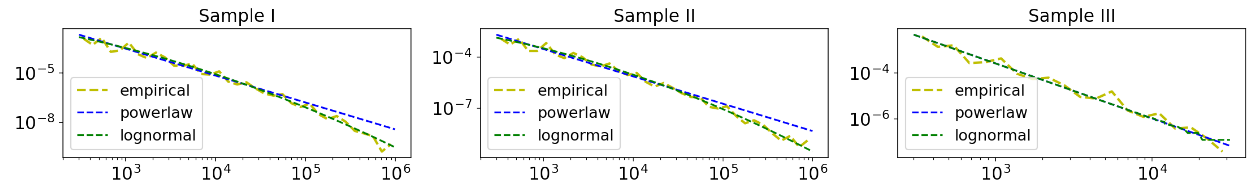
Table A.3: Summary statistics of order flow variables for Sample I.

	3s return (%)	order size v_x (shares)	order price p_x (RMB)
sample number	4739	94498	94498
mean	0.0003	3924.81	16.81
std	0.0400	19104.56	0.32
min	-0.6546	1	14.81
1% quantile	-0.1187	100	14.81
25% quantile	0.0000	300	16.76
50% quantile	0.0000	800	16.85
75% quantile	0.0000	2000	16.90
99% quantile	0.1198	60000	17.74
max	0.4792	1000000	18.10
	total number of orders in 3s	total sizes of orders in 3s (share)	total volume of orders in 3s (RMB)
sample number	4740	4740	4740
mean	19.93	78246.2	1312759.9
std	32.39	166448.0	2792518.8
min	1	100	1684
1% quantile	4	2300	38773.3
25% quantile	10	15200	254210.5
50% quantile	14	31400	527944.0
75% quantile	22	75700	1265115.8
99% quantile	104	691483	11700967.7
max	1620	5952531	99498571.3
	total number of trades in 3s	total sizes of trades in 3s (shares)	total volume of trades in 3s (RMB)
sample number	4740	4740	4740
mean	14	31647.4	531823.4
std	29.1	82532.8	1386500.4
min	0	0	0
1% quantile	0	0	0
25% quantile	4	2500	41993.8
50% quantile	7	7900	133685.5
75% quantile	15	24600	413767.5
99% quantile	108	373537	6254954.1
max	1086	2084895	35226973.5

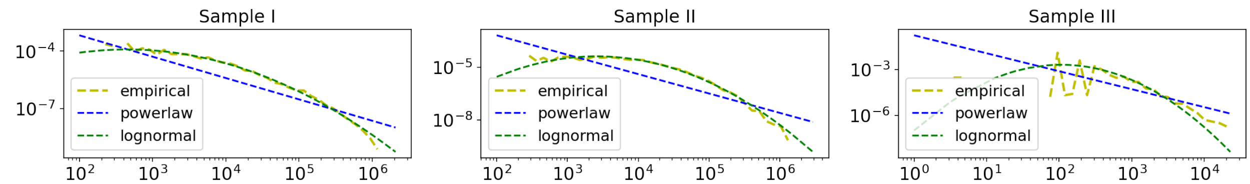
Average Shape of the Limit Order Book. Abergel et al. (2016) find that the average price gap between two consecutive levels is a constant between the five best bid and ask levels using data in developed markets. Figure A.3 verifies this observation using our dataset. In particular, we find that both Samples I and III satisfy this property, but not Sample II. This is because an overwhelming volume of investors panicked and liquidated their stock positions at the outbreak of COVID-19 on Feb. 3rd, 2020, which leads to a nearly empty order book on the bid side. This causes the abnormal phenomenon that the average best ask price is even lower than the average best bid price for Sample II.

F.2 Liquidity Distribution for Single Stocks

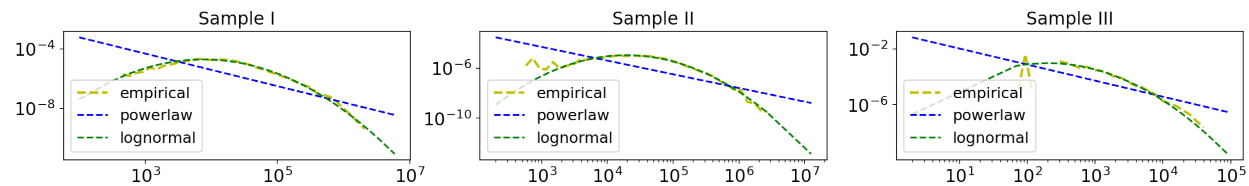
We analyze several useful distributional patterns for various liquidity measures, based on the three stock-day samples in Table A.2 as an example.



(a) Order sizes.



(b) Total sizes of executed orders in 3-second windows.



(c) Total sizes of submitted orders in 3-second windows.

Figure A.2: Empirical and fitted probability densities for the tail of order sizes, total sizes of executed orders in 3-second windows, and total sizes of submitted orders in 3-second windows, in log scale.

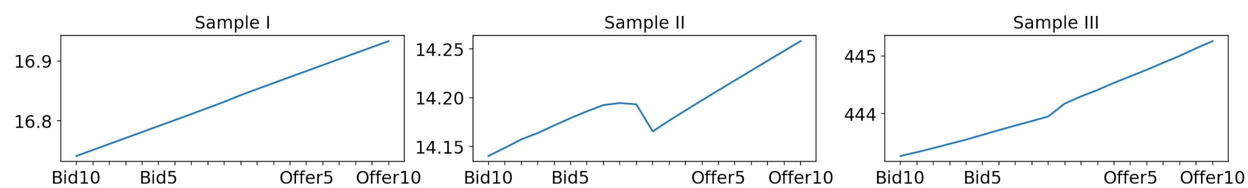


Figure A.3: Average shapes of order books.

F.2.1 Descriptive Statistics

Table A.4 reports the descriptive statistics of liquidity measures introduced in Section 4. Table A.4a presents the descriptive statistics of liquidity measures based on trading cost sampled every three seconds for Sample I. For this sample, the spread, $S(t)$, and effective spread, $S^{\text{eff}}(t)$, is RMB 0.01 (tick size) for more than 75% of the sample. The sample quantiles of wavg-spread, $S(t,q)$, increase as q increases. The proportional spread and the propotional wavg-spread perform similarly to spread and wavg-spread, respectively.

Inspired by Foucault et al. (2013), we consider the marginal transaction cost, defined as the partial derivative of $s(t,q)$ with respect to q, as an indicator for market depth, which is shown in Figure A.4. For a given q, smaller values of $\frac{\partial s(t,q)}{\partial q}$ imply deeper order books and hence better liquidity. We also observe that the curve of $\frac{\partial s(t,q)}{\partial q}$ for Ping An Bank moves upward from Jan. 2nd (Sample I) to Feb. 3rd (Sample II), implying that extreme market conditions reduce the market depth and liquidity. On the other hand, Changchun High and New Technology Industries (Sample III) has higher stock prices and lower levels of liquidity. As shown in Figure A.4, the marginal transaction cost, $\frac{\partial s(t,q)}{\partial q}$, decreases as q increases, because the marginal effect of executing an additional RMB diminishes as q increases.

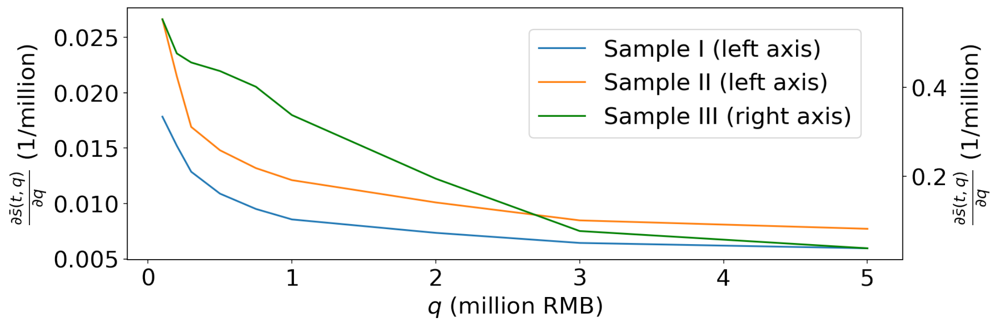


Figure A.4: Estimated average $\frac{\partial s(t,q)}{\partial q}$ on sample data.

Figure A.5 presents the daily time-series of average proportional spread for Ping An Bank in 2020, which fluctuates around 78 bpsthroughout the year. The average proportional spread experiences a clear peak on Feb. 3rd, 2020, and is then maintained at a high level from March to June. After June, the proportional spread declines slowly. These patterns reflect the rapid plunge of the Chinese stock market on Feb. 3rd, 2020 due to the negative sentiment from COVID-19.

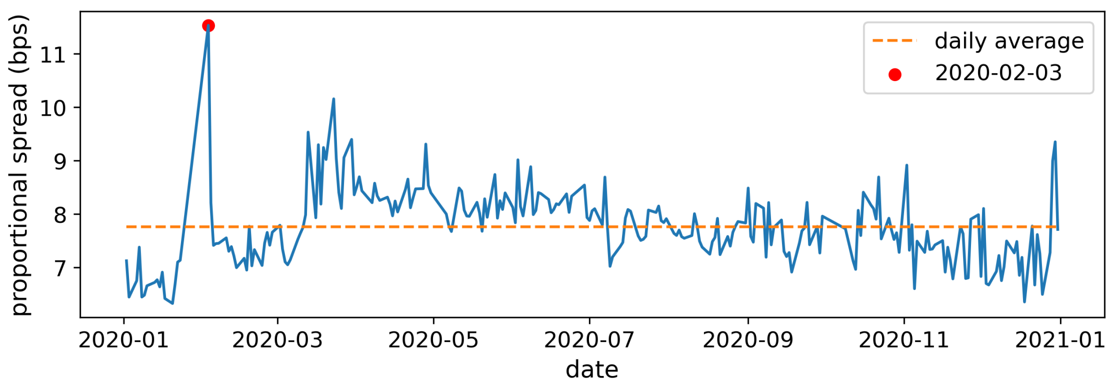


Figure A.5: Time-series of daily average proportional spread for Ping An Bank (stock code: 000001) in 2020.

Tables A.4b and A.4c show the descriptive statistics of liquidity measures based on execution efficiency for Sample I. On this day, there are over 90,000 orders submitted with an average interarrival time of around 0.15 seconds. More than 25% of these orders have zero order lifetimes, which implies that they were executed immediately after hitting the order book. On the other hand, the maximum order lifetime reaches 14,220 seconds, which is the length of a full trading day. These orders are submitted at the market open and stay in the order book until market close. The total execution ratio r^{e}, shown in Table A.4c, indicates that around 80% of all orders submitted within a trading day can be executed, which is much higher than a typical large-cap stock in the US market. Meanwhile, the execution ratio at a single price is around 64%, and the execution ratio at a single time and the execution ratio at the best price are both around 50%.

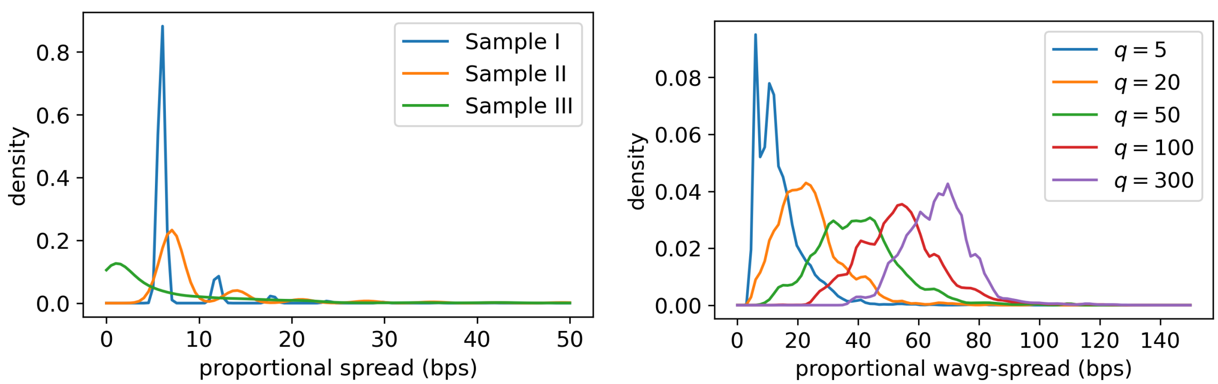
F.2.2 Distributions of Liquidity Measures

In this section, we study the distribution of liquidity measures for the three samples in Table A.2.[12]

[12]We skip the analysis of execution ratio in this section because it is only one number for each stock on a given day.

Proportional spread. Figure A.6a gives the kernel density estimation of the proportional spread. The curve of the density function of Ping An Bank moves rightward from Jan. 2nd to Feb. 3rd, which indicates that the proportional spread is larger right after the outbreak of COVID-19, and the liquidity is poorer under extreme market conditions. In addition, the proportional spread is lower for Changchun High and New Technology Industries because its price is higher than Ping An Bank.

Proportional wavg-spread. We calculate the proportional wavg-spread under different levels of shock. Specifically, we set $q = 5$, 20, 50, 100, and 300 ($\times 10^4$), and Figure A.6b shows the kernel density estimation for Sample I. The figure illustrates that, as q goes large, the shape of the curve moves rightward and becomes smoother. We also observe that the curve is multimodal when q is small and that it turns into bell-shaped when q increases.



(a) Proportional spread for Sample I, II, and III.

(b) Proportional wavg-spread for Sample I.

Figure A.6: Kernel density estimation of spread measures.

Interarrival Time. Figure A.7 presents the empirical densities of interarrival times for all orders of Samples I, II, and III and fits the empirical densities with power-law, exponential, log-normal, and Weibull distribution. In these samples, the exponential distribution does not fit well, while the Weibull distribution performs best. We observe that the tails of interarrival times are heavier than the exponential distribution, and that both the Weibull and the log-normal distribution fit the sample much better than the exponential distribution. From these observations, we suggest that the interarrival time does not follow the exponential distribution. Hence, the Poisson process is not a good candidate to model the order flow in the Exchange (Abergel et al., 2016).

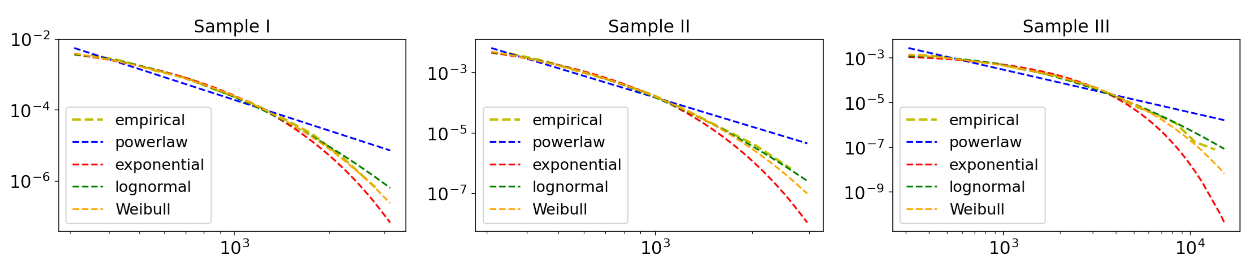
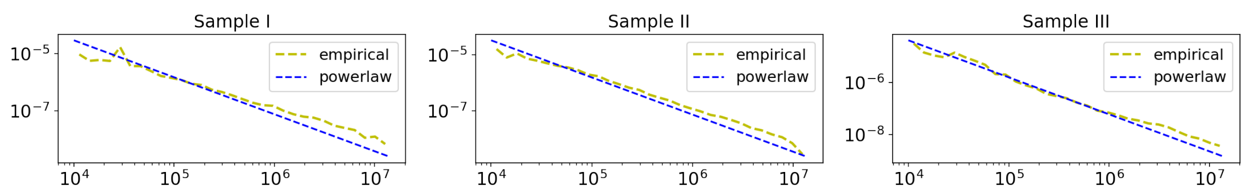


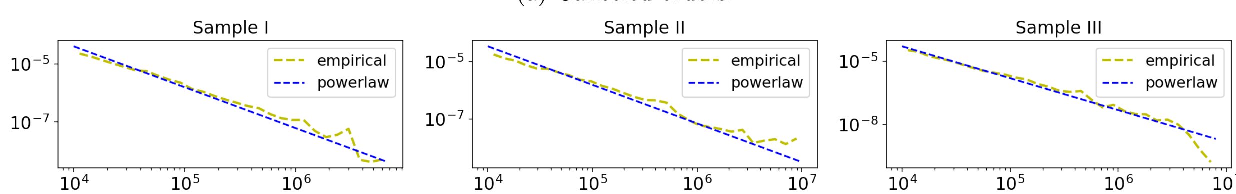
Figure A.7: Empirical and fitted densities for the tails of interarrival times (ms), in log scale.

To validate this conclusion, we conduct several hypothesis tests for the distribution of order interarrival times, and the results are reported in Table A.5 in Appendix F. First, the Kolmogorov-Smirnov tests on the three samples strongly reject the null hypothesis that interarrival times are exponentially distributed. Second, the chi-square goodness-of-fit tests confirm that interarrival times do not follow the exponential distribution, and both the Weibull and the log-normal distribution are shown to be good candidates. Moreover, we compare the sample quantiles with the quantiles of the fitted distribution, the results of which are shown in Figure A.11 in Appendix F. The Weibull distribution fits the sample tails best. Finally, likelihood ratio tests also confirm that interarrival times do not follow the exponential distribution, and we do not find statistically significant evidence that either the log-normal distribution or the Weibull distribution is more suitable than the other for modeling the interarrival time, suggesting that both of them are more reasonable candidates than the exponential distribution.

Order Lifetime. Abergel et al. (2016) concluded that the average order lifetime fits the power-law distribution. Figure A.8 confirms this fact empirically for our dataset. However, we also note that the tails of canceled orders and executed orders have different patterns. For canceled orders, the empirical densities are slightly heavier than the fitted densities on the tails, but this does not hold for the executed orders.



(a) Canceled orders.



(b) Executed orders.

Figure A.8: Empirical and fitted densities for the tails of order lifetime (ms), in log scale.

Kyle's Lambda. We analyze Kyle's Lambda of Ping An Bank for the whole year of 2020. First, in order to examine the intraday changes of Kyle's Lambda, we divide the data of 2020 into sub-samples per three minutes within the trading day to perform the regression (1), the results of which are shown in Figure A.9a. It can be seen that Kyle's Lambda after the market open and before the close is significantly smaller than other times, while the middle part is relatively stable. In other words, Kyle's Lambda follows an inverse U-shaped pattern over the course of a trading day. In addition, in order to examine the change in Kyle's Lambda throughout the year, we perform regression (1) on the data of each trading day, the results of which are shown in Figure A.9b.

Price Impact Regression. We first perform the regression (2) using all data of Ping An Bank in 2020, and the estimated coefficients are shown in Table A.4d. In addition, in order to examine the intraday patterns of each coefficient, we divide the transaction data of the whole year into sub-samples every three minutes

(a) Within a trading day.

(b) Over the whole year 2020.

Figure A.9: Trends of Kyle's Lambda for Ping An Bank (stock code: 000001) in 2020.

according to the execution time during the day, and perform linear regression estimation on (2). The results are displayed in Figure A.10.

We make a few interesting observations from Table A.4d and Figure A.10. First, γ_0 is very high immediately after the market open, and then fluctuates around 0.005. The latter can be regarded as the level of order-processing costs for most of the day. Second, γ_1 is almost always negative, which suggests that executions with large quantities are likely to obtain better execution prices. In other words, "quantity discounts" exist for Ping An Bank. Third, λ_0 and λ_1 are always positive, and they are higher in the first half hour after market open compared to the rest of the day.

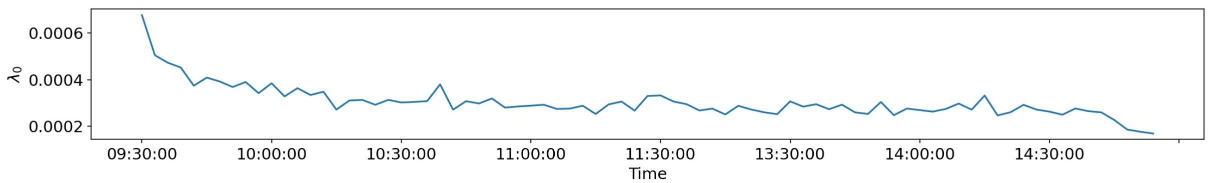
Finally, we summarize several key conclusions from our analysis of all liquidity measures above:

(a) The interarrival time does not follow the exponential distribution. Both the log-normal and the Weibull distribution have better goodness-of-fit for our sample.

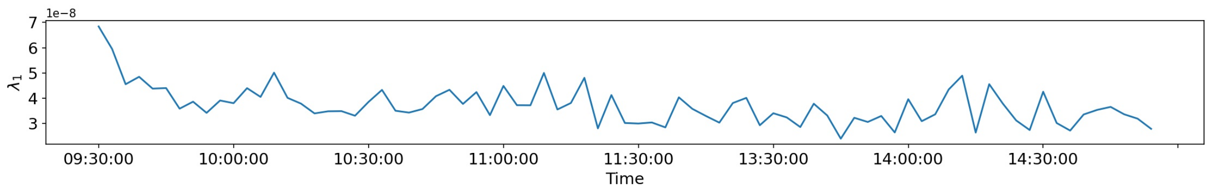
(b) A power-law decay describes the tail of order lifetimes reasonably well.

(c) Kyle's Lambda follows an inverse U-shaped intraday pattern.

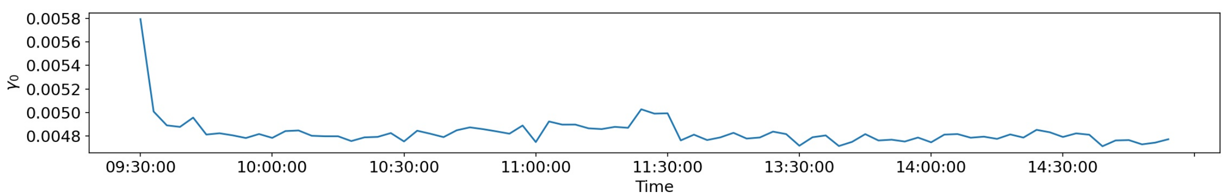
(d) The price impact regression shows that "quantity discounts" exist in the market, and that the impact of the transaction direction and quantity on the consensus price is larger during the first 30 minutes compared to the rest of the trading day.



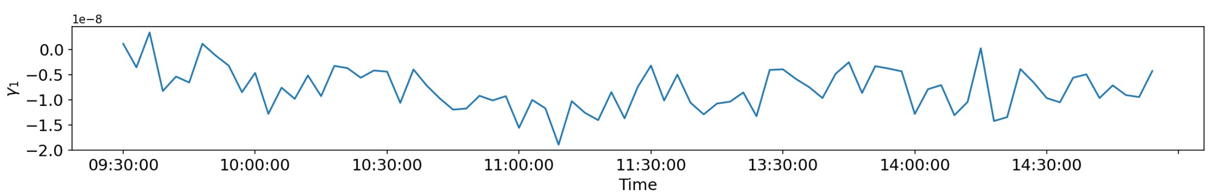
(a) λ_0.



(b) λ_1.



(c) γ_0.



(d) γ_1.

Figure A.10: Intraday trends for the estimated parameters of price impact regression in 2020.

Table A.4: Summary statistics and sample values of liquidity measures.

(a) Summary statistics of spread $S(t)$, effective spread $S^{\text{eff}}(t)$, wavg-spread $S(t,q)$, proportional spread $s(t)$, propotional effective spread $s^{\text{eff}}(t)$ and propotional wavg-spread $s(t,q)$, where $q = 1\times10^5$, 3×10^6, for Sample I.

	$S(t)$ (cents)	$S^{\text{eff}}(t)$ (cents)	$S(t, 1\times10^5)$ (cents)	$S(t, 3\times10^6)$ (cents)	$s(t)$ (bps)	$s^{\text{eff}}(t)$ (bps)	$s(t, 1\times10^5)$ (bps)	$s(t, 3\times10^6)$ (bps)
mean	1.153	1.296	2.955	11.095	6.857	7.711	17.588	65.997
std	0.543	0.961	1.611	1.851	3.242	5.728	9.610	11.068
min	1.000	1.000	1.000	6.196	5.902	5.900	5.901	36.608
1% quantile	1.000	1.000	1.000	6.860	5.905	5.906	5.908	40.522
25% quantile	1.000	1.000	1.902	9.883	5.925	5.925	11.306	58.772
50% quantile	1.000	1.000	2.682	11.125	5.940	5.943	15.950	66.233
75% quantile	1.000	1.000	3.599	12.188	5.985	5.991	21.417	72.555
99% quantile	3.000	5.000	8.286	16.718	17.991	29.931	49.448	99.744
max	14.000	24.000	17.080	22.757	83.334	142.686	101.665	135.457

(b) Summary statistics of interarrival time Δt_x and order lifetime $\bar{T}_x$ for Sample I.

	Δt_x (ms)	$\bar{T}_x$ (ms)
mean	150.47	1248180.91
std	212.86	3104990.94
min	0	0
1% quantile	0	0
25% quantile	20	0
50% quantile	70	11570
75% quantile	190	300550
99% quantile	1020	13808306
max	3130	14219900

(c) Sample value of total execution ratio $r^{\text{e}}(t)$, execution ratio at a single price $r^{\text{sp}}(t)$, at a single time $r^{\text{st}}(t)$, and at the best price $r^{\text{bp}}(t)$ for Sample I.

$r^{\text{e}}(t)$	$r^{\text{sp}}(t)$	$r^{\text{st}}(t)$	$r^{\text{bp}}(t)$
0.8089	0.6381	0.4948	0.5026

(d) Estimation results of Kyle's Lambda and price impact regression for Ping An Bank (stock code: 000001) in 2020.

model	notation	value
Kyle's Lambda	λ	5.8484×10^{-8}
Price Impact Regression	λ_0	3.2889×10^{-4}
	γ_0	4.8587×10^{-3}
	λ_1	3.8581×10^{-8}
	γ_1	-6.3239×10^{-9}

Table A.5 provides hypothesis testing results for the distribution of interarrival times, and Figure A.11 provides their tail QQ-plots, both of which are discussed in Section 5.2.2.

Table A.5: Hypothesis testing results for the distribution of interarrival times Δt_x.

(a) Kolmogorov–Smirnov (KS) test results. H_0 : interarrival times are exponentially distributed.

	Sample I	Sample II	Sample III
KS statistics	0.037*** (0.000)	0.046*** (0.000)	0.049*** (0.000)

(b) Chi-square goodness-of-fit test results for the distribution of interarrival times.

Chi-square Statistics	H_0 exponential	Weibull	log-normal
Sample I	26.278*** (0.002)	13.647 (0.135)	7.958 (0.538)
Sample II	37.478*** (0.000)	19.045** (0.025)	13.779 (0.130)
Sample III	93.61*** (0.000)	10.077 (0.344)	17.071** (0.048)

(c) Likelihood ratio (LR) test for different interarrival time distributions.

Log-likelihood Ratio	H_0/H_1 exponential/Weibull	exponential/log-normal	Weibull/log-normal
Sample I	-3.42*** (0.000)	-2.99*** (0.000)	-0.36 (0.720)
Sample II	-3.71*** (0.000)	-4.4*** (0.000)	-2.81*** (0.000)
Sample III	-7.22*** (0.000)	-5.28*** (0.000)	1.41 (0.160)

p-values in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

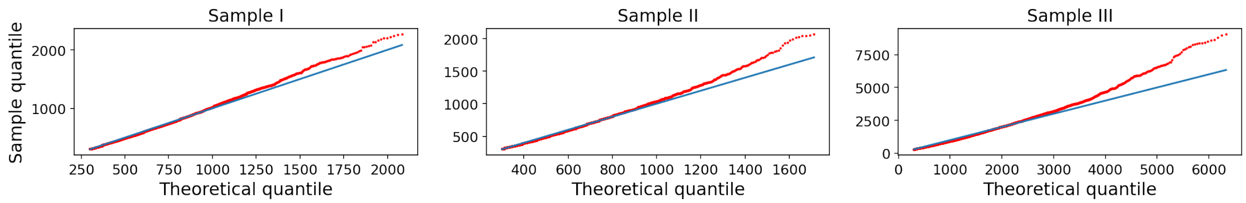
G Liquidity, Volume, and Order Imbalance Volatility

In Section 6, we use API to explain the change in quote bid-ask spreads over very short time intervals (3 seconds). For relatively longer time periods, such as on the minute scale, high-frequency liquidity has also been well investigated, including through inventory models proposed by, for example, Grossman and Miller (1988) and Bogousslavsky and Collin-Dufresne (2023). In this section, we follow the framework of Bogousslavsky and Collin-Dufresne (2023) to examine the relationship between high-frequency liquidity, volume, and the volatility in order imbalance for the Chinese stock market.

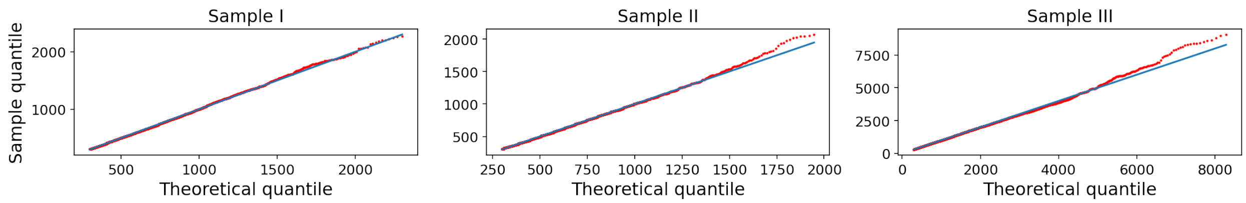
Bogousslavsky and Collin-Dufresne (2023) use high-frequency volume and volatility to explain the effective spread of stocks. Following their framework, for a given stock, we calculate its proportional effective spread defined by Definition 2. The effective spread over an interval is computed by summing the weighted effective spread associated with each execution over the interval, where the weight equals the RMB volume of the execution over the total RMB volume in the interval.

Bogousslavsky and Collin-Dufresne (2023) use the following panel regression to study the relationship between high-frequency liquidity, volume, and volatility based on daily data:

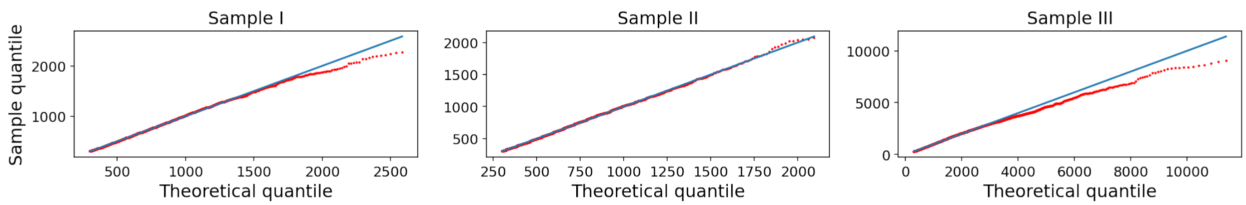
$$\ln s_{i,t}^{\text{Eff}} = \alpha_i + \beta_\tau \ln \tau_{i,t} + \beta_\sigma \ln \sigma_{i,t} + \text{controls} + \varepsilon_{i,t}, \tag{A.4}$$



(a) Exponential distribution.



(b) Weibull distribution.



(c) Log-normal distribution.

Figure A.11: QQ-plot for the tails of the interarrival time.

where $s_{i,t}^{\text{Eff}}$, $\tau_{i,t}$, and $\sigma_{i,t}$ represent the proportional effective spread, daily intraday turnover, and volatility for stock i on day t, respectively, and $\varepsilon_{i,t}$ is the residual term. Here, turnover $\tau_{i,t}$ characterizes the daily trading volume of the stock, and volatility $\sigma_{i,t}$ is the realized volatility calculated using five-minute intraday midquote returns. The controls include calendar indicators for the day of the week and the previous day's market capitalization and price (in logs).

Based on (A.4), Bogousslavsky and Collin-Dufresne (2023) find that, in the US market, daily effective spreads are generally negatively related to volume and positively related to volatility. However, this pattern does not hold for large stocks, whose effective spreads are generally increasing in volume. To address this, they introduce the high-frequency order-imbalance volatility (HFOIV), and show that HFOIV can reconcile the empirical differences between large and small stocks.

The HFOIV is defined as the standard deviation of the five-minute share imbalance (as a fraction of shares outstanding) over the trading day. Bogousslavsky and Collin-Dufresne (2023) incorporate HFOIV as another independent variable in (A.4) and conduct the following panel regression:

$$\ln s_{i,t}^{\text{Eff}} = \alpha_i + \beta_\tau \ln \tau_{i,t} + \beta_\sigma \ln \sigma_{i,t} + \beta_{\text{HFOIV}} \ln \text{HFOIV}_{i,t} + \text{controls} + \varepsilon_{i,t}. \tag{A.5}$$

By conducting (A.5) for the US stock market, Bogousslavsky and Collin-Dufresne (2023) find that the HFOIV is strongly positively associated with effective spread. In addition, for large stocks, the relationship between volume and effective spread becomes strongly negative once the HFOIV is included as an independent variable.

We now apply Bogousslavsky and Collin-Dufresne's (2023) framework to our Chinese stock market dataset. Table A.6 shows the summary statistics of the variables used in regression (A.5) for our three-year dataset. In particular, for each stock, we first compute the annual time-series average of each variable, and then calculate the mean, standard deviation, minimum, maximum, and quantiles of these average values across all stocks.

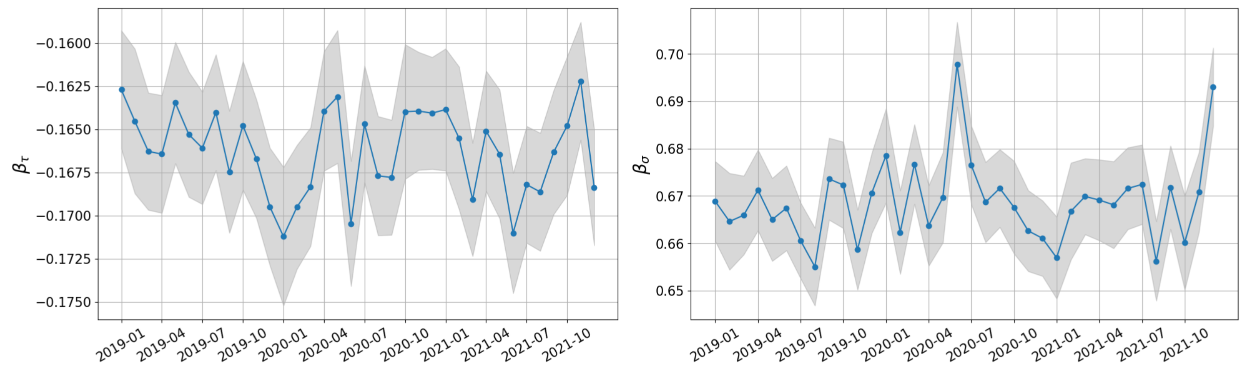
In Table A.6, Panel A shows the summary statistics for all stocks, while Panels B and C present the statistics for large and small stocks, respectively. Following Bogousslavsky and Collin-Dufresne (2023), we classify a stock as large (small) if its market capitalization in the previous year falls within the top 20% (bottom 20%). Panels B and C illustrate that, in general, large stocks tend to have a smaller effective spread, lower turnover, lower volatility, and smaller HFOIV compared to small stocks.

Table A.6: Summary statistics for variables in the regression model (A.5) from 2019 to 2021.

Variable	Proportional Effective Spread (bps)			Turnover (%)			Volatility (%)			HFOIV (bps)		
Year	2019	2020	2021	2019	2020	2021	2019	2020	2021	2019	2020	2021
Panel A: All stocks												
mean	21.78	21.47	19.55	2.95	3.63	3.40	1.84	1.99	1.77	0.82	0.95	0.76
std	9.78	16.87	13.68	2.69	3.02	3.10	0.37	0.45	0.50	0.68	0.69	0.61
min	4.38	3.73	4.12	0.00	0.09	0.11	0.84	0.89	0.59	0.04	0.04	0.04
1% quantile	7.99	7.10	6.18	0.31	0.41	0.41	1.08	1.17	0.99	0.10	0.12	0.12
25% quantile	15.35	13.78	12.45	1.22	1.66	1.44	1.61	1.72	1.46	0.37	0.51	0.39
50% quantile	20.18	18.28	16.51	2.14	2.85	2.48	1.81	1.96	1.68	0.63	0.79	0.59
75% quantile	25.63	24.72	22.71	3.72	4.62	4.19	2.02	2.20	1.99	1.04	1.19	0.90
99% quantile	57.28	64.51	56.82	12.52	16.06	16.37	2.77	3.10	3.13	3.21	3.61	2.99
max	123.76	367.71	266.32	30.55	28.60	28.60	5.61	8.31	9.52	6.78	7.48	6.69
Panel B: Large stocks												
mean	16.52	14.61	13.18	2.26	2.60	2.41	1.70	1.95	1.80	0.54	0.61	0.51
std	7.04	6.53	6.21	2.81	2.02	2.13	0.33	0.40	0.49	0.64	0.46	0.52
min	4.38	3.73	4.12	0.00	0.09	0.11	0.84	1.00	0.79	0.04	0.04	0.04
1% quantile	6.46	5.90	5.02	0.24	0.24	0.29	1.03	1.11	0.93	0.07	0.07	0.06
25% quantile	12.04	10.52	9.24	0.86	1.16	1.04	1.46	1.67	1.46	0.20	0.28	0.22
50% quantile	14.75	12.83	11.65	1.46	2.07	1.78	1.69	1.93	1.75	0.36	0.47	0.35
75% quantile	19.61	16.76	15.20	2.54	3.49	2.94	1.89	2.20	2.07	0.64	0.81	0.60
99% quantile	40.31	37.50	39.19	14.19	9.21	10.48	2.59	2.94	3.28	3.38	2.23	2.57
max	54.70	46.09	43.22	30.55	16.17	16.73	2.74	3.33	3.57	5.40	2.70	5.02
Panel C: Small stocks												
mean	26.12	30.00	26.77	3.35	3.54	3.36	1.93	2.05	1.88	1.04	1.11	0.91
std	11.81	26.98	23.40	2.39	2.47	2.69	0.35	0.53	0.61	0.65	0.69	0.59
min	11.61	10.71	8.72	0.00	0.37	0.35	1.12	1.31	0.93	0.12	0.19	0.25
1% quantile	13.49	12.72	10.42	0.52	0.48	0.47	1.33	1.43	1.16	0.23	0.29	0.29
25% quantile	19.52	18.31	16.53	1.62	1.78	1.53	1.72	1.77	1.55	0.59	0.66	0.55
50% quantile	23.29	24.02	21.53	2.68	2.97	2.55	1.90	1.97	1.75	0.87	0.95	0.74
75% quantile	29.27	31.69	30.37	4.25	4.59	4.13	2.09	2.22	2.05	1.27	1.36	1.05
99% quantile	68.97	149.80	68.36	11.24	12.36	14.20	2.76	3.84	4.66	3.25	3.76	3.24
max	123.76	367.71	266.32	14.56	19.28	17.71	5.14	6.88	6.90	4.24	7.43	5.53

For each month from Jan. 2019 to Dec. 2021, we use daily data for all trading days in this month and all stocks to run regression models (A.4) and (A.5). Figures A.12 and A.13 show the estimated values and 95% confidence intervals of the regression coefficients for (A.4) and (A.5), respectively. From these figures, we observe that daily effective spreads are generally negatively related to volume ($\beta_\tau < 0$) and positively related to volatility ($\beta_\sigma > 0$). Furthermore, after incorporating the HFOIV, we find that the effective spread is also positively correlated with the HFOIV ($\beta_{\text{HFOIV}} > 0$), and the coefficient for volume (β_τ) is significantly reduced. These findings align with those of Bogousslavsky and Collin-Dufresne (2023), suggesting similarities between the Chinese and US stock markets. -

We further conduct the two regressions separately for large and small stocks, and the results are shown in



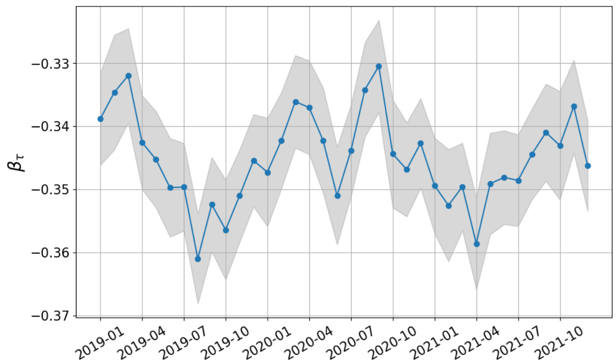
(a) β_τ.

(b) β_σ.

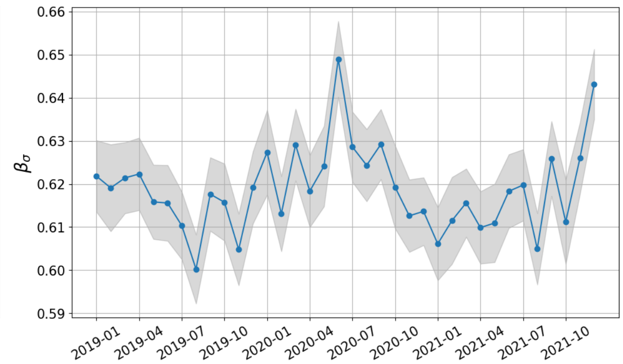
Figure A.12: Estimated coefficients for the regression model (A.4) across all stocks from 2019 to 2021.

Figures A.14–A.15 and Figures A.16–A.17, respectively. Interestingly, we find that the patterns we observed in the regressions for all stocks (Figures A.12–A.13) hold consistently for both large and small stocks. This is different from Bogousslavsky and Collin-Dufresne's (2023) findings, where large stocks have different patterns compared to small stocks.

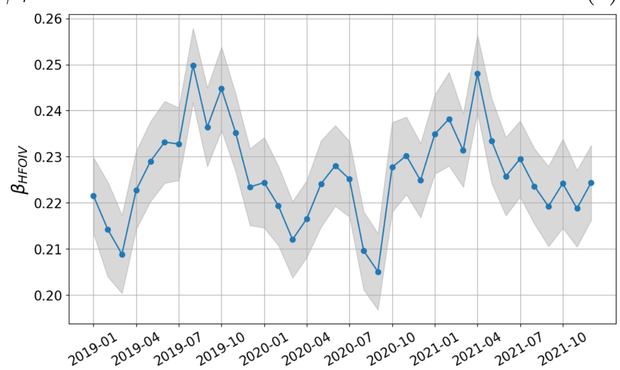
We provide several remarks that may help explain the observed differences for large stocks traded in the Chinese and US markets. First, Bogousslavsky and Collin-Dufresne (2023) use US stock high-frequency data from the Trades and Quotes (TAQ) dataset, which does not provide trade directions—buy and sell. Therefore, Bogousslavsky and Collin-Dufresne (2023) apply the Lee–Ready algorithm (Lee and Ready, 1991) to infer trade directions. In contrast, our dataset explicitly provides trade directions, potentially leading to different outcomes between the two markets. Second, the time periods studied differ significantly: our analysis covers 2019–2021, while Bogousslavsky and Collin-Dufresne (2023) focuses on 2002–2017. In fact, as discussed Bogousslavsky and Collin-Dufresne (2023, Section C), the positive relationship they observe between effective spread and volume for large stocks does not hold for the last couple of years of their sample. Bogousslavsky and Collin-Dufresne (2023, Figure 3) show that this relationship generally turns negative after 2014, aligning with our findings from 2019 to 2021. Finally, differences in trading rules, currencies, and participant characteristics between the two markets may also contribute to the discrepancies between Bogousslavsky and Collin-Dufresne's (2023) results and our findings.



(a) β_τ.

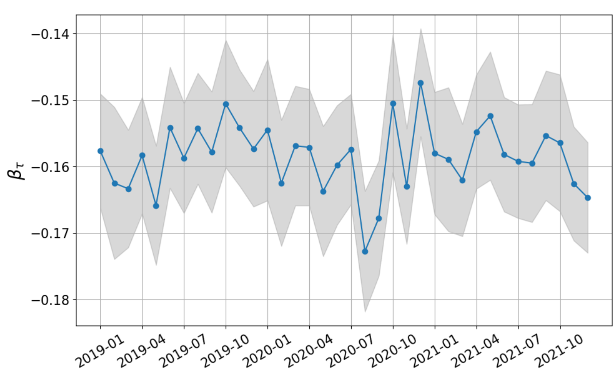


(b) β_σ.

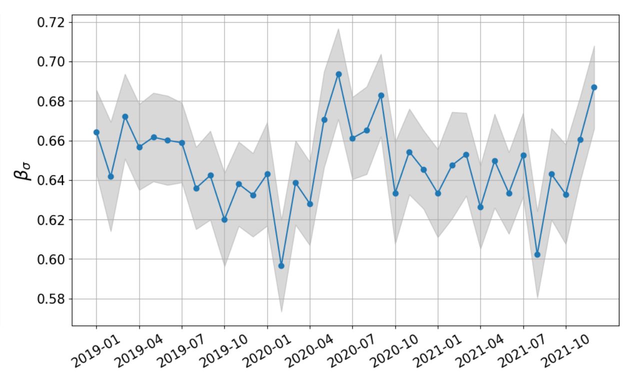


(c) β_{HFOIV}.

Figure A.13: Estimated coefficients for the regression model (A.5) across all stocks from 2019 to 2021.

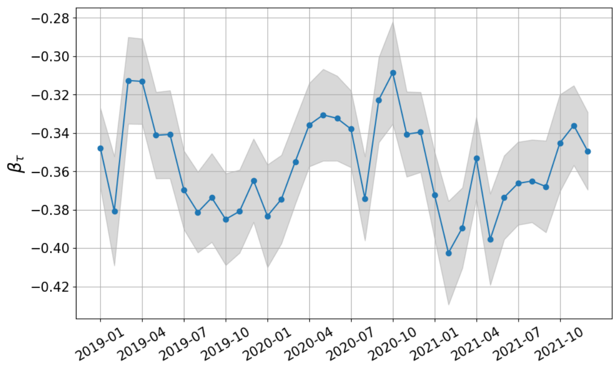


(a) β_τ.

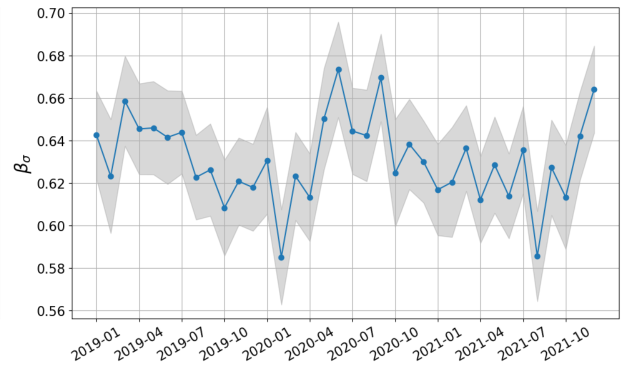


(b) β_σ.

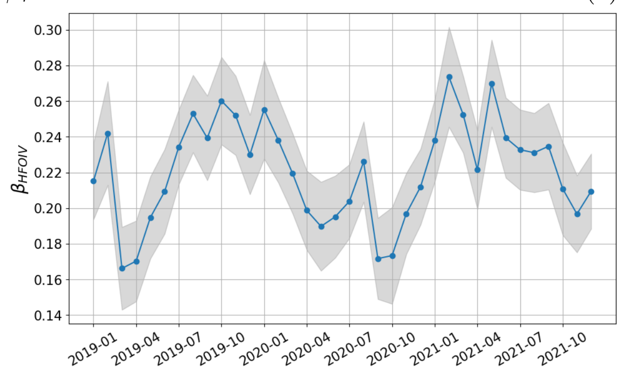
Figure A.14: Estimated coefficients for the regression model (A.4) across large stocks (stocks with top 20% market capitalization) from 2019 to 2021.



(a) β_τ.

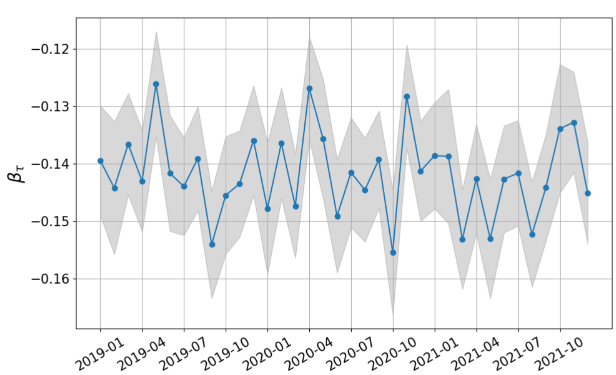


(b) β_σ.

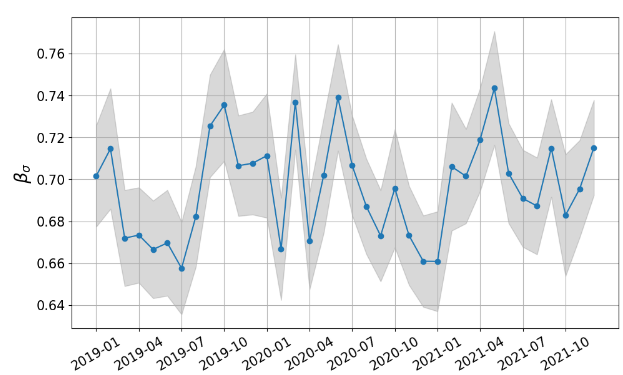


(c) β_{HFOIV}.

Figure A.15: Estimated coefficients for the regression model (A.5) across large stocks (stocks with top 20% market capitalization) from 2019 to 2021.

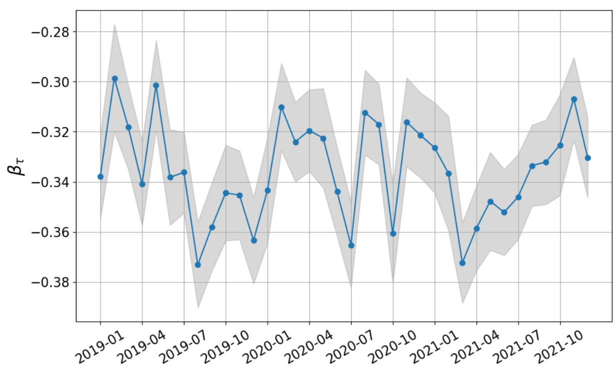


(a) β_τ.

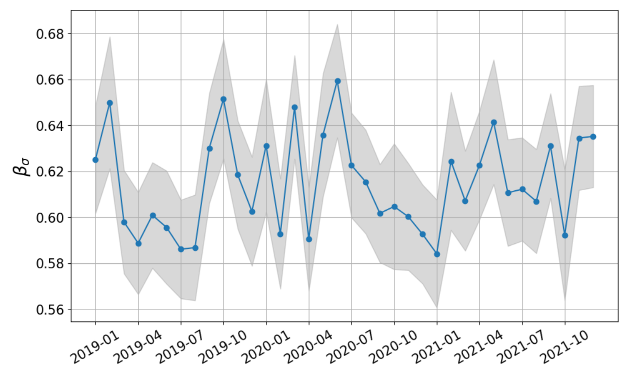


(b) β_σ.

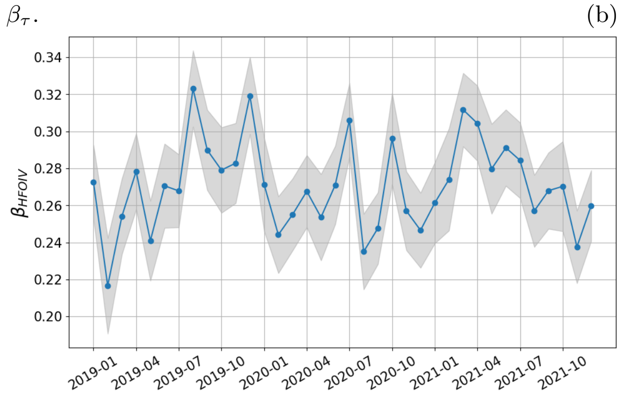
Figure A.16: Estimated coefficients for the regression model (A.4) across small stocks (stocks with bottom 20% market capitalization) from 2019 to 2021.



(a) β_τ.



(b) β_σ.



(c) β_{HFOIV}.

Figure A.17: Estimated coefficients for the regression model (A.5) across small stocks (stocks with bottom 20% market capitalization) from 2019 to 2021.