International Journal of Financial Engineering
Vol. 9, No. 4 (2022) 2250033 (32 pages)

DOI: 10.1142/S2424786322500335

World Scientific
www.worldscientific.com

Order types and natural price change: Model and empirical study of the Chinese market

Siyu Liu[*], Chaoyi Zhao[*,‡] and Lan Wu[†]

[*]*School of Mathematical Sciences*
Peking University, Beijing, P. R. China

[†]*LMEQF, School of Mathematical Sciences*
Peking University, Beijing, P. R. China

Received: 11 April 2022; Accepted: 30 June 2022
Published: 18 November 2022

Abstract

Order type plays an important role in algorithmic trading and is a key factor of price impact. In this paper, we propose a new framework for studying the discrete price change process, which focuses on the impacts of aggressive orders (market orders and aggressive limit orders) and cancelations. The price change process is driven by states and events of best quotes, and we define the event-based price change as the “natural price change” (NPC). Under the framework, we propose a heteroscedastic linear econometric model for the NPC to explore the impact of different types of orders on the price dynamics. To verify the usability of our model and explore the driving factors of price dynamics, we conduct a thorough empirical analysis for 786 large-tick stocks traded on the Shenzhen Stock Exchange. Empirical results statistically demonstrate that aggressive orders can introduce stronger impact on the NPC than cancelations. Meanwhile, splitting a big order into several small orders can lead to a larger NPC. Our framework can also be applied for the prediction of price change.

Keywords: Market microstructure; high-frequency data; bid-ask spread; natural price change; order type.

JEL Classification: C51, C58, G12, G14

[‡]Corresponding author.
Email address: [‡]zhaochaoyi@pku.edu.cn

1. Introduction

Exploring the nature of the price formation mechanism is an important and challenging research field of finance. Thanks to the recent development of electronic transaction platforms, scholars can analyze the price dynamics from a more profound perspective by accessing more sophisticated trade-by-trade data instead of level-1 (i.e., only the first level of the order book) and level-2 (i.e., the first five nonempty levels of the book on each side) snapshot data. These trade-by-trade data can help to calibrate and verify many complicated models, and in return, spawn a wide range of event-driven limit order book models. For example, Abergel *et al.* (2016) systematically summarized the modeling of limit order books. Furthermore, Abergel and Jedidi (2015) and Lu and Abergel (2018b) applied the modeling theories to practice, and explored the optimal order placing strategy of market makers.

This paper focuses on modeling the impact of order book events on price change. However, analyzing the high-frequency price change on different time scales may lead to different empirical results of price change. On a minute time-scale, Cont *et al.* (2014) proposed a single variable — the OFI (Order Flow Imbalance), defined as the sum of signed volumes of all incoming orders at the best quotes over a time interval — which has a strong power to explain the price change. Capponi and Cont (2020) further demonstrated that the cross-impact among stocks is unstable and insignificant, in comparison with the OFI. On the contrary, on a higher frequent time-scale, Mertens *et al.* (2022) argued that using a filtering approach can assist the modeling of price impact due to the discrete nature of price changes. Gould and Bonart (2016) employed the queue imbalance to model and predict the high-frequency price change, and Xu *et al.* (2018) made a further attempt via the multi-level order flow imbalance. In addition, Eisler *et al.* (2012), Lu and Abergel (2018a), Taranto *et al.* (2018a) and Taranto *et al.* (2018b) demonstrated that different types of orders have different impact on the price dynamics, and linear relationship exists between the order flow and price impact. Therefore, we propose to establish models that account for different types of orders to explain their impact on price change on a fair time-scale.

For the impact on price change, it is well-known that the order volume is an important feature correlated with price impact. However, Weber and Rosenow (2006) argued that it is not sufficient to fully explain large price changes if we only use order volume in the model. For instance, the short-term impact of orders on prices is also inevitably influenced by liquidity, especially in high-frequency trading. Hasbrouck (1991) also found that larger trades will lead to wider bid-ask spreads, which affect the market liquidity and the price impact. Hence, inspired by

Lu and Abergel (2018a), instead of volume, we take the change of the best quotes into consideration while modeling the impact of different orders on price.

When modeling asset prices, it is necessary to analyze the price change dynamics of large-tick stocks and small-tick stocks separately, because stocks with different price magnitudes relative to tick size have different trading patterns. Lots of existing literature studied the tick-to-price ratio. The studies found a strong relationship between the ratio and the dynamic of the stock (Angel, 1997; Lipson and Mortal, 2006). Biais *et al.* (1995) showed that the spread has significant influences on later events. Eisler *et al.* (2012) illustrated that large-tick stocks (i.e., stocks for which the relative tick is large) and small tick stocks have different patterns. It is documented "large-tick stocks are such that the bid-ask spread is almost always equal to one tick, whereas small tick stocks have spreads that are typically a few ticks" (Eisler *et al.*, 2012). Thus, it is essential to consider large and small tick stocks separately.

In this paper, we analyze the impact of different types of orders on the price change of large-tick stocks. For large-tick stocks, the bid-ask spread usually stays at one tick for a long time. By definition, price only changes when at least one of the best bid and the best ask price changes. Instead of dividing the whole price dynamic into several time intervals with the same durations, we propose a new method to determine the start and end time of a price change event based on the bid-ask spread. The new framework can separate different price change events clearly.

The main contributions of this paper are provided as follows:

(1) We propose a model for the impact of different types of orders on the price change. Our model considers the impact of different types of orders separately. Specifically, we separate all orders according to their order types (aggressive orders, passive orders, and cancelations) and the change of best quotes caused by them.
(2) We give a new definition of the price change events and the corresponding "natural price change". The new framework provides a new viewpoint to segment a price sequence into several price change events for the high-frequency price dynamics of large-tick stocks.
(3) We conduct an extensive empirical analysis on 786 stocks on the Chinese stock market for a period of three calendar months. Results demonstrate not only the differences among different types of orders but also the predictability and potential applications of our model.

The paper is organized as follows. In Sec. 2, we propose a new framework for the price change of large-tick stocks and specify a model for the impact of different types of orders on the price change. In Sec. 3, we describe our data and present relevant empirical results to demonstrate the effectiveness of our framework. The empirical facts of our model not only reveal the impact of different types of orders on price change but also show the predictability of our model. Section 4 concludes.

2. Model Setup

In this section, we propose a model of the trade-by-trade price change. The model aims to illustrate and explain the impact of each order on the price change dynamics at the transaction level.

Rydberg and Shephard (2003) proposed a trade-by-trade dynamic of mid-price as

$$p(t) - p(0) = \sum_{k=1}^{N(t)} \frac{1}{2} Z_k, \quad t > 0, \tag{1}$$

where $p(t)$ is the mid-price of a stock at time t, $N(t)$ is the number of orders from time 0 to t, and Z_k is the change of the best quote caused by the k-th order from time 0 to t. A positive (negative) Z_k implies that the mid-price is increasing (decreasing). A factor $\frac{1}{2}$ is included in the model since the mid-price is the arithmetic average of the best bid and the best ask. Model (1) is a natural expression of mid-price changes caused by orders, and model is determined by the process $N(t)$ and variables Z_k. Now we further investigate the two components.

First, we decompose the total number of orders $N(t)$ into $N_m(t)$, which represents the number of type-m orders from time 0 to t. Denote OT as the set of order types. Hence, $N(t) = \sum_{m \in \mathrm{OT}} N_m(t)$. We further denote the change of best quote caused by the k-th arrived type-m orders as $Z_{m,k}$. Therefore, Model (1) can be represented as

$$p(t) - p(0) = \sum_{m \in \mathrm{OT}} \sum_{k=1}^{N_m(t)} \frac{1}{2} Z_{m,k}. \tag{2}$$

Next, we count the number of ticks of best quote change introduced by each order. In this paper, we use one tick as the unit of price (e.g., in the Chinese market, one tick = 0.01RMB), and therefore, the prices and spreads are both positive integers. Thus, the price change can be represented as

$$p(t) - p(0) = \sum_{m \in \mathrm{OT}} \sum_{j=-\infty}^{\infty} \frac{1}{2} N_{m,j}(t) \cdot j, \tag{3}$$

where $N_{m,j}(t) := \sum_{k=1}^{N_m(t)} \mathbf{1}_{\{Z_{m,k}=j\}}$ is the number of type-m orders that lead to j tick(s) change of best quotes, from time 0 to t.

Model (3) describes the mechanism of price change, using the number of type-m orders that lead to j tick(s) change of best quotes from time 0 to t. The most direct way to analyze the impact of orders on price change is modeling each process $N_{m,j}(t)$, as lots of order book models. However, the relationship between $N_{m,j}(t)$ and price change is too complicated to solve in practice. We should think about the impact of various orders on price change deeply.

We will further modify the Model (3), to characterize the price change dynamics based on the order flow. To separate different orders clearly, we use the timestamps of new orders and cancelations arrive as a new *discrete* time scale, and name it as *order time*. Thereby the order time t corresponds to the natural time that the t-th order arrives.

In the remainder of this section, we describe basic quantities used in the model and rewrite Model (3). In Sec. 2.1, we redefine the concept of price change (the left-hand side of Model (3)) based on order flow. In Sec. 2.2, we rearrange the variables on the right-hand side of Model (3). Finally, in Sec. 2.3, we propose our complete model.

2.1. Response variable: Natural price change

Price change is an ambiguous concept for high-frequency data. Many scholars view the price change as the change of the mid-price over a fixed natural time interval, as this is a natural extension of the "profit and loss" of low-frequency investment. However, for high-frequency trading, the transaction is always executed at the best quotes, and the mid-price cannot be traded directly. Both the bid-ask spread and the mid-price need to be considered while defining the trade-by-trade price change. Moreover, for large-tick stocks, the bid-ask spread usually stays at one tick for a long time, without any mid-price change. Price change only occurs when at least one of the best bid and the best ask price changes. Therefore, for large-tick stocks, it is more natural to segment a price sequence into a series of price change events according to the spread size.

In this section, we propose our new definition of price change for large-tick stocks. First, we use the bid-ask spread to define the unit-spread state based on the trade-by-trade sequence. If the spread equals one tick, both buyers and sellers can trade at the best quotes immediately, and the price observed is "real" to some extent. Next, we define the price change event based on two neighboring unit-spread states. Finally, we give the definition of natural price change for a price change event.

We denote the best bid, the best ask, the spread and the mid-price at order time $t = 1, 2, \ldots$ as $\mathrm{bestbid}(t), \mathrm{bestask}(t), \mathrm{spread}(t),$ and $\mathrm{midprice}(t),$ respectively.

Definition 1 (Unit-Spread State). We define a stock to be in the *unit-spread state* at order time t, if the stock's bid-ask spread at t equals to 1.

If a stock is in the unit-spread state, the market will be in the most liquid state, since both buyers and sellers can trade at the best quotes immediately. Hence, the unit-spread state can be regarded as a certain kind of equilibrium.

Definition 2 (Type I Price Change Event). For a stock in the unit-spread state at two order times t^{start} and t^{end}, if we have

(1) $t^{\text{end}} \geq t^{\text{start}} + 2$, and
(2) $\text{spread}(t) > 1$, when $t^{\text{start}} < t < t^{\text{end}}$,

then we say the stock experienced *a type I price change event during* $(t^{\text{start}}, t^{\text{end}})$. The prefix "type I" is often omitted.

From the above definition, if a stock is in the unit-spread state at order time t^{start}, next escape from the state, and finally back to the unit-spread state at order time t^{end}, then we say it experienced a type I price change event during this period. Either the best bid or the best ask (or both) should move during a type I price change event.

Note that there is a constraint in our definition of price change event: $t^{\text{end}} \geq t^{\text{start}} + 2$, i.e., there exist at least one order time between t^{start} and t^{end} such that the spread at that time is strictly greater than one. However, there is a special type of orders, namely the *aggressive limit orders*, or *marketable limit orders*. These are limit orders priced to meet or better the opposing quote, see Harris and Hasbrouck (1996), and Peterson and Sirri (2002). These limit orders will hit the opposing orders whose prices are better than or equal to them. Those orders that cannot be further matched will be left on the limit order book and wait for new orders. Aggressive limit orders may introduce the following type II price change events.

Definition 3 (Type II Price Change Event). For a stock in the unit-spread state at two order times t^{start} and t^{end}, and also satisfies

(1) $t^{\text{end}} = t^{\text{start}} + 1$, and
(2) $\text{midprice}(t^{\text{start}}) \neq \text{midprice}(t^{\text{end}})$,

then we say it experienced *a type II price change event during* $(t^{\text{start}}, t^{\text{end}})$.

Based on the concepts of price change events, we propose a new definition of price changes.

Definition 4 (Natural Price Change, NPC). For a stock, for a (type I or II) price change event during $(t^{\text{start}}, t^{\text{end}})$, we define its *natural price change* (*NPC*) *during* $(t^{\text{start}}, t^{\text{end}})$ as

$$\text{NPC} = \text{midprice}(t^{\text{end}}) - \text{midprice}(t^{\text{start}}).$$

Definition 4 is applicable to both type I and type II events. For type II price change events, we use t2NPC to represent the NPC (the prefix "t2" means type II).

Next, we introduce the concepts of up- and down-events.

Definition 5 (Up- and Down-Event). For a stock, for a (type I or II) price change event during $(t^{\text{start}}, t^{\text{end}})$, we say it is an *up-event* if it satisfies one of the following two conditions:

(1) its $\text{NPC} > 0$,
(2) its $\text{NPC} = 0$ and

$$\max_{t^{\text{start}} < t \leq t^{\text{end}}} (\text{bestask}(t) - \text{bestask}(t^{\text{start}})) \geq - \min_{t^{\text{start}} < t \leq t^{\text{end}}} (\text{bestbid}(t) - \text{bestbid}(t^{\text{start}})).$$

Otherwise, we say it is a *down-event*.

We give an empirical example about NPC in Appendix A. Here, we provide a simple example in Fig. 1 to illustrate the concepts intuitively. There is a (type I) price change (up-)event occurred from order time $t^{\text{start}} = 1$ to $t^{\text{end}} = 10$, since the bid-ask spreads are one tick at both time, and greater than one tick at other order times between. The mid-price is 400.5(ticks) at order time 1, and 404.5(ticks) at order time 10. Therefore, its $\text{NPC} = 4$ (ticks) for this event.

We will use NPC (Definition 4) as the response variable of our model (the left-hand side of Eq. (3)). That is because, as we have emphasized, for large-tick stocks, the mid-price in the unit-spread state is more credible. Therefore, it is more natural to segment a price sequence into several price change events according to the unit-spread state, and the NPCs are the corresponding price change of the price change events.

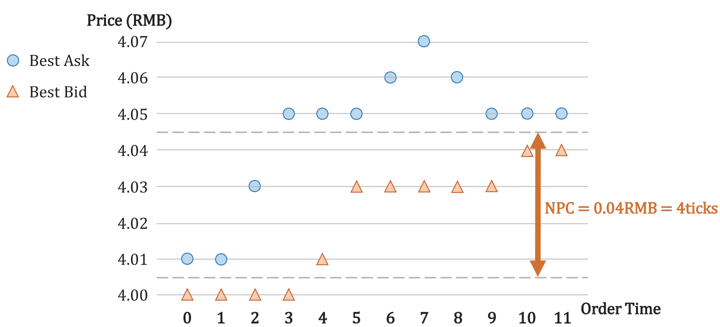


Fig. 1. A simple example of type I price change event.

2.2. Explanatory variables: Aggressive orders and cancelations

In this section, we specify the explanatory variables in our model. We first reclassify limit orders and market orders into two types as follows:

- *aggressive orders*: market orders, and aggressive limit orders (limit buy orders with prices higher than or equal to the best ask, and limit sell orders with prices lower than or equal to the best bid).[1]
- *passive orders*: passive limit orders (limit buy orders with prices lower than the best ask, and limit sell orders with prices higher than the best bid).

Hence, aggressive orders will trade immediately, while passive orders have to wait for other aggressive orders to hit them. Generally speaking, aggressive limit orders have similar effect on the order book as market orders. Hence, to analyze the impact of different orders on price changes, in this paper, we put market orders and aggressive limit orders together. For clarity, we specify the order types we consider in this paper as

$$\mathrm{OT} = \{\text{aggressive orders, passive orders, and cancelations}\}.$$

In our model, explanatory variables are the numbers of *aggressive orders* and *cancelations*, without considering passive orders. There are three main reasons. The first reason is the logic that the aggressive orders and passive orders have different functions. The aggressive orders and cancelations are what indeed break the original equilibrium (the unit-spread state), whereas the passive orders usually repair the order book and pull the bid and ask prices back to the unit-spread states. Figure 2 explains the logic in more detail. Suppose that the original best bid and best ask are 4.00RMB and 4.01RMB, respectively (and the stock is thus in the unit-spread state). An aggressive buy order is executed, the best ask price increases to 4.04RMB, and a price change event begins. The next unit-spread state is achieved until new passive limit orders fill in the gap between the best bid and the new best ask. Therefore, it is the aggressive orders and cancelations that counts, while passive limit orders only stitch those wounds of the order book.

The second reason is due to the fact that each order will end up with a cancelation or a transaction, and each transaction is corresponding to an aggressive order and several passive orders. Hence, to simplify our model, we only use aggressive orders to represent the transactions.

[1]Once an aggressive limit order is submitted, 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 later empirical studies, for an aggressive limit order, we regard its volumes traded immediately as an aggressive order, and regard the volumes left on the order book as a passive order. In other words, we will split each aggressive limit order into an "aggressive part" and a "passive part".

Fig. 2. An example of price change event.

The third reason is that it is improper to include all orders while using linear models, since the observed price process can be replicated perfectly if all orders are used. As for the example in Fig. 2, the aggressive buy order leads to a change of 0.03RMB in the best ask, and the passive limit sell order leads to a change of -0.03RMB. The NPC of this price change event is 0, which can be directly calculated by $(0.03 + (-0.03))/2$.

Not all aggressive orders and cancelations are included in our model. We only consider orders changing the best quotes. In fact, there are many aggressive orders which do not change the best quotes since they can not remove enough passive orders, and there are also many cancelations not at the best quotes. Although these aggressive orders and cancelations contribute to the price change, for the simplicity of our model, we do not use them, and only include orders changing the best quotes.

2.3. Complete model

Since all transactions must be closed within a day, the time series models of order book usually only focus on the dynamics within a day, half-day, or several hours. Similarly, for a given stock, our model considers the sequence of price change events within a given time horizon, with a maximum of one trading day.

We first give some notations. For a particular stock, suppose that there are n price change events during a specific period. Then, for the i-th price change event of the stock, $i = 1, 2, \ldots, n$, let us denote[2]

- M^j: Aggressive orders (market orders or marketable limit orders) who changed opposing best quote for $|j|$ tick(s), where j is a nonzero integer. A positive j corresponds to aggressive buy orders, and a negative j corresponds to aggressive sell orders.

[2]For convenience, we omit the subscript of the stock in these notations.

- C^j: Cancelations who changed their own side's best quote for $|j|$ tick(s), where j is a nonzero integer. A positive j corresponds to sellers' cancelations, and a negative j corresponds to buyers' cancelations.
- NM_i^j: Number of M^j orders in event i.
- NC_i^j: Number of C^j orders in event i.

Now we propose our model. For a price change event i, we express its NPC as follows:

$$\begin{aligned}\mathrm{NPC}_i = &\sum_{j=1}^{\infty} j \cdot (\beta_{\mathrm{M}}^{+j}\mathrm{NM}_i^{+j} + \beta_{\mathrm{M}}^{-j}\mathrm{NM}_i^{-j} + \beta_{\mathrm{C}}^{+j}\mathrm{NC}_i^{+j} + \beta_{\mathrm{C}}^{-j}\mathrm{NC}_i^{-j}) \\ &+ \sqrt{K_i}\varepsilon_i, \quad i = 1, 2, \ldots, n, \end{aligned} \tag{4}$$

where β_{M}^{+j}, β_{M}^{-j}, β_{C}^{+j}, and β_{C}^{-j} are unknown parameters for $j = 1, 2, \ldots$, $\varepsilon_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$, and

$$K_i = \sum_{j=1}^{\infty} (\mathrm{NM}_i^{+j} + \mathrm{NM}_i^{-j} + \mathrm{NC}_i^{+j} + \mathrm{NC}_i^{-j}) j^2, \quad i = 1, 2, \ldots, n. \tag{5}$$

In other words, our model is a *heteroscedastic* linear econometric model. This model is similar to the model suggested in Eq. (3), but for the left-hand side, we use the new concept of price change; and for the right-hand side, we consider different types of orders.

In Sec. 2.3.1, we use a latent price model to explain the idea behind the Eq. (4). Then, in Sec. 2.3.2, we propose an example.

2.3.1. *Explanation of the model*

We use the following framework to explain the idea behind our model, i.e., Eq. (4). For a stock, for a type I price change event i during $(t^{\text{start}}, t^{\text{end}})$, we define the corresponding *latent price process* $p(t)$ from t^{start} to t^{end} as follows: (As a reminder, the time t is order time, not natural time.)

(i) For the start time $t = t^{\text{start}}$,

$$p(t^{\text{start}}) = \mathrm{midprice}(t^{\text{start}}). \tag{6}$$

(ii) For $t = t^{\text{start}} + 1, t^{\text{start}} + 2, \ldots, t^{\text{end}}$,

$$\begin{aligned} p(t) = p(t-1) &+ \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{M}^{+j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{M}}^{+j}(t) + \mathbf{1}_{\mathrm{M}^{-j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{M}}^{-j}(t)] \\ &+ \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{C}^{+j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{C}}^{+j}(t) + \mathbf{1}_{\mathrm{C}^{-j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{C}}^{-j}(t)], \end{aligned} \tag{7}$$

where the indicator function $\mathbf{1}_A(t)$ equals to 1 if the order arrives at order time t is an A order, and 0 otherwise; and

$$\tilde{\beta}_{\mathrm{M}}^{\pm j}(t) = \beta_{\mathrm{M}}^{\pm j} + \varepsilon_{\mathrm{M}}^{\pm j}(t), \quad \tilde{\beta}_{\mathrm{C}}^{\pm j}(t) = \beta_{\mathrm{C}}^{\pm j} + \varepsilon_{\mathrm{C}}^{\pm j}(t), \quad j = 1, 2, \ldots, \tag{8}$$

where $\beta_{\mathrm{M}}^{\pm j}$ and $\beta_{\mathrm{C}}^{\pm j}$ are constants w.r.t. time t, while $\varepsilon_{\mathrm{M}}^{\pm j}(t)$ and $\varepsilon_{\mathrm{C}}^{\pm j}(t)$ are i.i.d. and $N(0, \sigma^2)$ white noises. Note that only one of the indicator functions equal to 1 at order time t, since only one order arrives at order time t.

Intuitively, parameters $\beta_{\mathrm{M}}^{\pm j}$ and $\beta_{\mathrm{C}}^{\pm j}$ represent the average impact of different kinds of orders, which do not vary with time. This coincides with the fact we will illustrate in Sec. 3.2.3 that the features of price change do not vary among different trading times. Variables $\varepsilon_{\mathrm{M}}^{\pm j}(t)$ and $\varepsilon_{\mathrm{C}}^{\pm j}(t)$ are errors, whose effects to $p(t) - p(t-1)$ get larger as j increases, since on the right hand side of Eq. (7), $\tilde{\beta}_{\mathrm{M}}^{\pm j}(t)$ and $\tilde{\beta}_{\mathrm{C}}^{\pm j}(t)$ are all multiplied by j.

If we further require the latent price process satisfies

$$p(t^{\text{end}}) = \text{midprice}(t^{\text{end}}),$$

then, according to Eqs. (6) and (7), we immediately have

$$\begin{aligned}
&\text{midprice}(t^{\text{end}}) \\
&\quad = \text{midprice}(t^{\text{start}}) + \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{M}^{+j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{M}}^{+j}(t) + \mathbf{1}_{\mathrm{M}^{-j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{M}}^{-j}(t)] \\
&\qquad + \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{C}^{+j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{C}}^{+j}(t) + \mathbf{1}_{\mathrm{C}^{-j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{C}}^{-j}(t)].
\end{aligned}$$

Therefore, using the definition of NPC (Definition 4), we have

$$\begin{aligned}
\text{NPC}_i = &\sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{M}^{+j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{M}}^{+j}(t) + \mathbf{1}_{\mathrm{M}^{-j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{M}}^{-j}(t)] \\
&+ \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{C}^{+j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{C}}^{+j}(t) + \mathbf{1}_{\mathrm{C}^{-j}}(t) \cdot j \cdot \tilde{\beta}_{\mathrm{C}}^{-j}(t)].
\end{aligned}$$

Furthermore, by Eq. (8),

$$\begin{aligned}
\text{NPC}_i = &\sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{M}^{+j}}(t) \cdot j \cdot (\beta_{\mathrm{M}}^{+j} + \varepsilon_{\mathrm{M}}^{+j}(t)) + \mathbf{1}_{\mathrm{M}^{-j}}(t) \cdot j \cdot (\beta_{\mathrm{M}}^{-j} + \varepsilon_{\mathrm{M}}^{-j}(t))] \\
&+ \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} \sum_{j=1}^{\infty} [\mathbf{1}_{\mathrm{C}^{+j}}(t) \cdot j \cdot (\beta_{\mathrm{C}}^{+j} + \varepsilon_{\mathrm{C}}^{+j}(t)) + \mathbf{1}_{\mathrm{C}^{-j}}(t) \cdot j \cdot (\beta_{\mathrm{C}}^{-j} + \varepsilon_{\mathrm{C}}^{-j}(t))]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{\infty} \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} j \cdot [\beta_{\text{M}}^{+j} \mathbf{1}_{\text{M}^{+j}}(t) + \beta_{\text{M}}^{-j} \mathbf{1}_{\text{M}^{-j}}(t) + \beta_{\text{C}}^{+j} \mathbf{1}_{\text{C}^{+j}}(t) + \beta_{\text{C}}^{-j} \mathbf{1}_{\text{C}^{-j}}(t)] \\
&\quad + \sum_{j=1}^{\infty} \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} j \cdot [\mathbf{1}_{\text{M}^{+j}}(t) \varepsilon_{\text{M}}^{+j}(t) + \mathbf{1}_{\text{M}^{-j}}(t) \varepsilon_{\text{M}}^{-j}(t) + \mathbf{1}_{\text{C}^{+j}}(t) \varepsilon_{\text{C}}^{+j}(t) \\
&\quad + \mathbf{1}_{\text{C}^{-j}}(t) \varepsilon_{\text{C}}^{-j}(t)] \\
&= \sum_{j=1}^{\infty} j \cdot (\beta_{\text{M}}^{+j} \text{NM}_i^{+j} + \beta_{\text{M}}^{-j} \text{NM}_i^{-j} + \beta_{\text{C}}^{+j} \text{NC}_i^{+j} + \beta_{\text{C}}^{-j} \text{NC}_i^{-j}) + \sqrt{K_i} \varepsilon_i,
\end{aligned}
$$

where K_i is defined as Eq. (5), and

$$
\begin{aligned}
\varepsilon_i &= \frac{1}{\sqrt{K_i}} \sum_{j=1}^{\infty} \sum_{t=t^{\text{start}}+1}^{t^{\text{end}}} j \cdot [\mathbf{1}_{\text{M}^{+j}}(t) \varepsilon_{\text{M}}^{+j}(t) + \mathbf{1}_{\text{M}^{-j}}(t) \varepsilon_{\text{M}}^{-j}(t) + \mathbf{1}_{\text{C}^{+j}}(t) \varepsilon_{\text{C}}^{+j}(t) \\
&\quad + \mathbf{1}_{\text{C}^{-j}}(t) \varepsilon_{\text{C}}^{-j}(t)] \\
&\sim N(0, \sigma^2).
\end{aligned}
$$

Hence, under these assumptions, Eq. (4) holds.

2.3.2. *Example*

Figure 3 gives a simple example of a price change event. The blue points represent the best ask, the red points represent the best bid, and the green points represent the prices of our latent price process model. We can see that the price change event starts at order time 0. At order times 1 and 2, two aggressive orders or cancelations pull the best ask price up, and at order times 3 and 4, two passive orders fill in the gap and shrink the spread. However, the event is not over yet since the spread at order time 4 is two ticks, which is not a unit-spread state. At order time 5, a new aggressive order or cancelation tears the gap, and two new passive limit orders

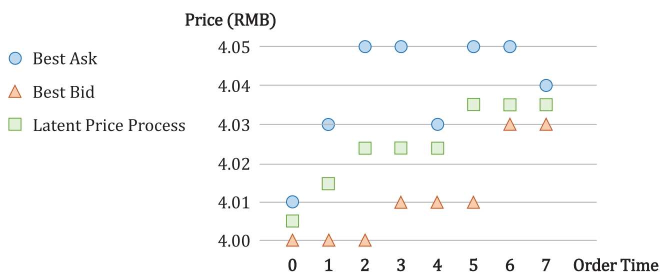


Fig. 3. A simple example of the latent price process. The blue points are best ask, and the red points are best bid. The green points are the latent price process, which is the same as the mid-price at order time 0 (beginning) and order time 7 (end).

stitch the wound at order times 6 and 7, when the stock finally arrives at the unit-spread state. Note that, only the aggressive orders and cancelations affect the latent price process (green dots) under our assumptions, and the function of passive limit orders is exposing the unobservable latent price process to us. The latent price process often deviates from the mid-price. However, the latent price process will finally go back to the mid-price since passive limit orders will eventually guide our stocks to the unit-spread state.

For the case in Fig. 3, if we assume the changes of latent price process at order times 1 and 2 are driven by aggressive orders, while at order time 5 driven by cancelations, then under our assumptions, we have

(1) At the beginning ($t^{\text{start}} = 0$) and the end ($t^{\text{end}} = 7$),

$$p(0) = \text{midprice}(0),$$
$$p(7) = \text{midprice}(7);$$

(2) During this period,

$$\begin{aligned}\text{NPC}_i = p(7) - p(0) &= 2\,\tilde{\beta}_{\text{M}}^{+2}(1) + 2\,\tilde{\beta}_{\text{M}}^{+2}(2) + 2\,\tilde{\beta}_{\text{C}}^{+2}(5)\\ &= 2\beta_{\text{M}}^{+2} + 2\beta_{\text{M}}^{+2} + 2\beta_{\text{C}}^{+2} + 2\varepsilon_{\text{M}}^{+2}(1) + 2\varepsilon_{\text{M}}^{+2}(2) + 2\varepsilon_{\text{C}}^{+2}(5)\\ &= 2\beta_{\text{M}}^{+2} + 2\beta_{\text{M}}^{+2} + 2\beta_{\text{C}}^{+2} + \sqrt{12}\varepsilon_i.\end{aligned}$$

This is consistent with the expression of Eq. (4).

3. Empirical Analysis

In order to verify the usability of the concepts and model proposed in Sec. 2, in this section, we conduct a thorough empirical analysis based on the Chinese A-share high-frequency trading dataset. First, in Sec. 3.1, we introduce our dataset. Next, in Sec. 3.2, we use the data to examine the basic statistics of the concepts, including unit-spread state, price change event, and the NPC. Then, in Sec. 3.3, we use the data to fit our model (4), and investigate the applicability and goodness-of-fit of the model. Finally, in Sec. 3.4, we propose a method to forecast the price change based on the NPC model and illustrate its effectiveness.

3.1. Data

The basic dataset in this paper consists of trades data, quotes data and snapshots data for all stocks traded on Shenzhen Stock Exchange[3] during three calendar

[3]Data were obtained from the Shenzhen Stock Exchange, see http://www.szse.cn/English/services/dataServices/index.html.

Table 1. Example of quotes data and trades data.

An Example of Quotes Data		An Example of Trades Data	
StockID	000012.SZ	StockID	000012.SZ
Timestamp	20191008:09:35:00:275497	Timestamp	20191008:09:37:00:354897
Volume	1500	Volume	1100
Price	5.48	Price	5.48
OrderID	751	OrderID	849
OrderKind	Limit	OrderKind	0
FunctionCode	B	FunctionCode	Cancel
—	—	BidOrderID	751
—	—	AskOrderID	0

months (October 2019 to December 2019, 60 trading days). Quotes data are information of all orders that traders submitted; trades data include both transactions and cancelations data[4]; and snapshots data are level-2 data for every 3 s.

Table 1 shows examples of quotes data and trades data, respectively. The entries of Quotes data consist of a stock ID, a timestamp, the quote's volume, the price, the order ID, the type of the order (OrderKind), and a function code (B=buy, S=sell). For the trades data, except for the elements mentioned above, it also includes information for the bid order ID and the ask order ID, which both refer to orders that lead to the transaction. For example, the quotes data in the table is a limit buy order of stock 000012.SZ quoted at about 9:35 a.m., its volume is 1500 and the associated price is 5.48RMB. The unique ID of this order is 751. The example of trades data is a cancelation of order 751 at 9:37 a.m.. We can find that only a volume of 1100 is canceled, and thus a volume of 400 has already been traded (since the volume of that order was 1500 at first). This cancelation with a volume of 1100 also has a unique ID 849.

Taking into account the features of the Chinese market, we introduce some data cleaning procedures to pre-process our data set. In particular, all data in 5 min after the opening time and 5 min before the closing time are removed, i.e., only data from 9:35 to 11:30 a.m. and 13:00 to 14:55 p.m. Beijing time (we say it *valid time*) are preserved as we only consider how price changes in normal market condition. Besides, the price limits of the Chinese market may lead to a lack of liquidity. Therefore, if one day a stock reaches limit-up or limit-down, or if one side of its order book is empty, we remove all the data of the corresponding stock for that day.

[4] In the data provided by the Shenzhen Stock Exchange, transactions and cancelations are collectively called "trades data".

To ensure the reliability of our dataset, we use quotes data and trades data to reproduce the time series of order books according to the principle of price and time priority. Then, we compared them with the observed snapshot data. Stocks whose average matching ratios are 90% above are reserved. Since we only consider large-tick stocks, we only reserve stocks whose proportions of durations of unit-spread state are greater than 40%.

Our data set consists of trading information for 786 large-tick stocks over 60 trading days, with a total of 599,484,160 orders. The average daily numbers of orders for different stocks have a maximum of 65,153 and a minimum of 728, with an average of 12,711, which means that, on average, there are 12,711 orders quoted for every stock each day.

3.2. Preliminary data analysis

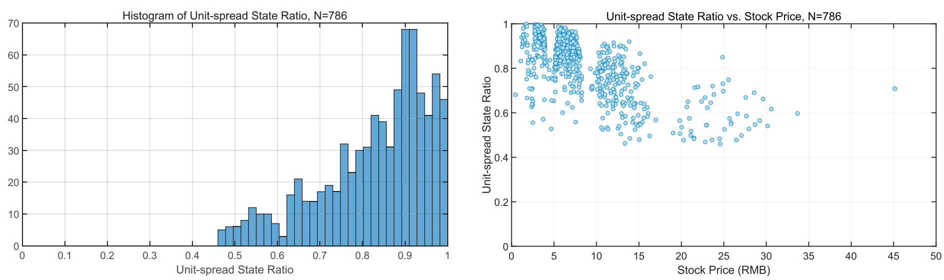
3.2.1. *Proportion of the unit-spread state*

To demonstrate the suitability of our definition for the unit-spread state, similar to Cont and De Larrard (2013) did, we calculate the *unit-spread state ratio* for each stock during the 60 trading days we study. It is defined as

$$\text{unit-spread state ratio of stock } i = \frac{\text{unit-spread state's duration of stock } i}{\text{valid time's duration of stock } i},$$

where the unit-spread state's duration is the summation of the natural duration that the stock is in the unit-spread state. The ratio represents the proportion of unit-spread state during the whole (valid) period, which can be considered as a measure of liquidity, since it is widely acknowledged as the bid-ask spread characterizes liquidity to some extent.

Figure 4 are empirical results about the unit-spread state ratio. Figure 4(a) is the histogram of different stocks' unit-spread state ratio calculated. We observe that



(a) Histogram of unit-spread state ratio

(b) unit-spread state ratio vs. Stock price

Fig. 4. Empirical facts of the unit-spread state ratio: (a) shows the distribution of stocks' unit-spread state ratio, and (b) shows the relationship between the unit-spread state ratio and stock price.

the ratio lies between 0.4 and 1, and three-quarters stocks are in the unit-spread state over 75% of the whole duration. Our result indicates that most large-tick stocks are in the unit-spread state in the majority duration of each trading day. Stocks with different price levels may present different patterns.

Figure 4(b) is the scatter plot of the unit-spread state ratio vs. the stock price. In general, the larger the stock prices, the lower the unit-spread state ratios. This phenomenon can be well-explained in the real world. For instance, the market makers would struggle to conclude transactions at both the best bid and the best ask prices to arbitrage. However, their arbitrage profit depends on the trade-off between the bid-ask spread and the transaction cost. A higher stock price leads to a higher transaction cost, which drives market makers to quote more conservatively. Therefore, the unit-spread state would be rare for stocks with high prices.

3.2.2. *Frequency of price change event*

To reveal the fact that price change events are common in practice, it is also important to analyze the frequency of price change events for different stocks.

Figure 5 shows daily average numbers of type I price change events and type II price change events. The red bar represents the daily total trading volumes of the Chinese market. Under our data set, we can see that the daily number of price change events for both types are within the range of 120 and 260. In general, there are more type II price change events than type I's. On average, there are 324 price change events daily for each stock, and 254,467 events daily for all stocks in total. Type I events account for 46% of all events. The correlation coefficient of type I's

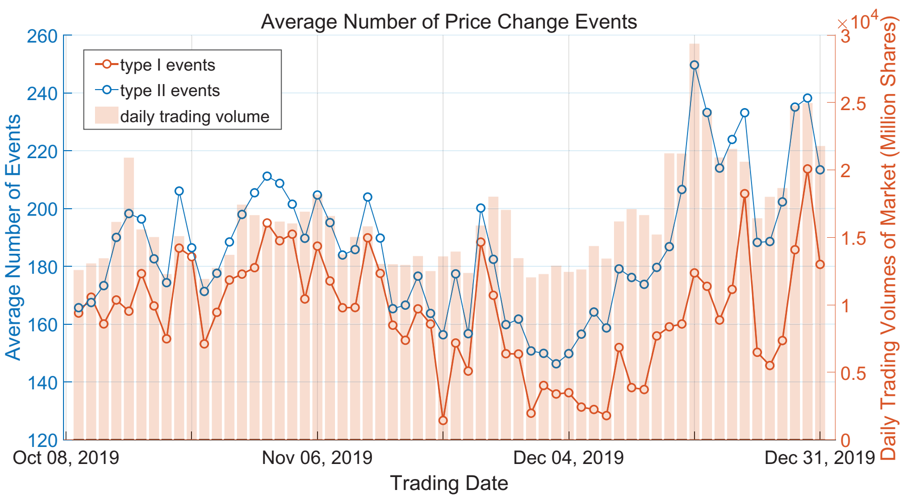


Fig. 5. The average number of price change events versus trading date.

Table 2. Groups used in Fig. 6, grouped by number of price change events.

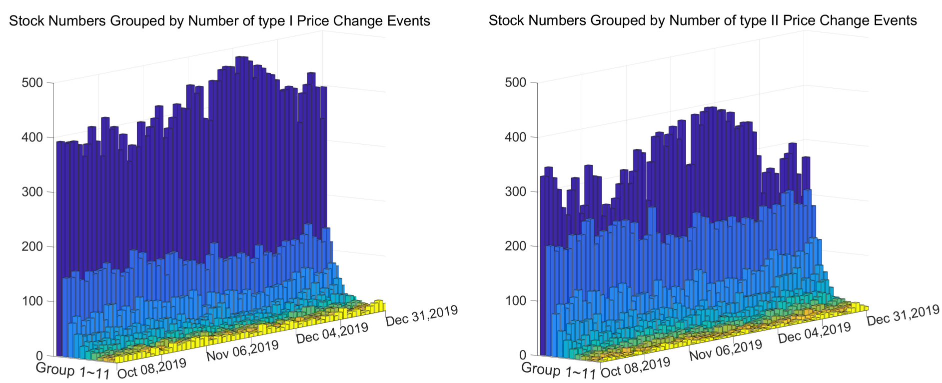
Group	1	2	3	$\cdots$	10	11
Number of type I or II events	[0,100)	[100,200)	[200,300)	$\cdots$	[900,1000)	[1000, $+\infty$)

and type II's series is 0.7873, and both of them are highly correlated with the daily trading volume of the Chinese market. It implies that our definitions of price change events are suitable, since larger trading volume should lead to more price change events.

To further investigate, for each day, we divide all stocks into 11 groups according to the number of both types of price change events, and the numbers of events for stocks in the k-th group are in the interval $[100(k-1), 100k)$ for $k = 1, \ldots, 10$. Numbers of events for stocks in group 11 are more than or equal to 1,000 (see Table 2). Figure 6 shows how stock numbers vary with trading dates and groups (number of events), type I and type II, respectively. For each day, both stock numbers decline with the number of events rapidly, and about 50% of the stocks have a number of price change events lower than 100. Both figures rarely change alongside the axis of trading dates, which means that the feature of the number of price change events is rather stable.

3.2.3. *Staistical analysis about NPC*

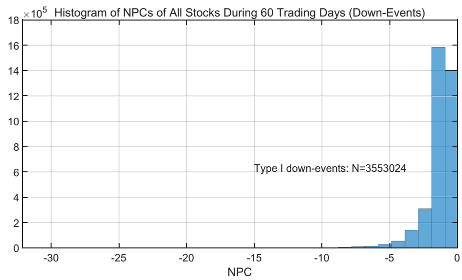
Based on our data set, one could provide some basic empirical facts and statistical properties about the NPC. Figure 7 shows histograms of NPCs of all stocks during the 60 trading days we study.



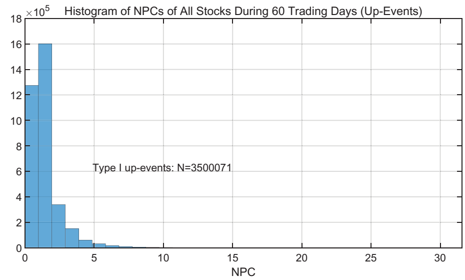
(a) type I

(b) type II

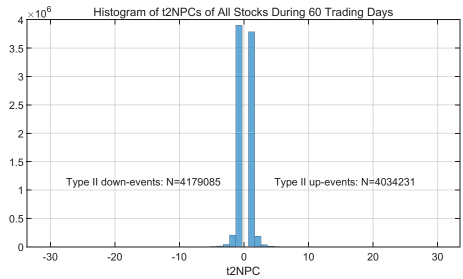
Fig. 6. Stock numbers versus stock group (grouped by number of type I and type II events, see Table 2) and trading date.



(a) histogram of NPC (down-events)



(b) histogram of NPC (up-events)



(c) histogram of t2NPC

Fig. 7. Histograms of all price change events' NPC for all stocks in 60 trading days. All figures' x-axes are truncated at ± 30.

Figures 7(a) and 7(b) are histograms of NPCs of all type I price change events. For both up- and down-events, NPC can be equal to 0, and it is improper to merge the histograms together. Figure 7(a) corresponds to down-events (NPC ≤ 0), and Fig. 7(b) corresponds to up-events (NPC ≥ 0). The density at zero implies the fact that a large amount of price changes are due to the noises rather than natural variations of supply and demand. Figure 7(c) is the histogram of t2NPCs of all type II price change events.

We list several empirical facts which are refined from these figures as follows:

(1) There are more down-events than up-events for both type I and type II events for our dataset.
(2) The distributions of up- and down-events are not symmetric. For instance, the proportion of down-events with NPC $= 0$ is greater than the one of up-events with NPC $= 0$.
(3) The distributions of type I and type II events are also not the same. Compared with NPCs of type I events, the distribution of t2NPCs of type II events is much narrower, which means that it is more concentrated on 0 nearby (although they cannot be 0).

Fact 7 may be due to the market environment of the period we study. Fact 7 can be explained by the "T+1" trading rule of the Chinese market. Traders will be much more prudent while deciding whether to buy stocks or not as they cannot sell them immediately once they buy-in. Hence, they will buy stocks with larger volumes and more orders when they have good expectations, and the larger volume and more orders will lead to more significant price change. Therefore, the distribution of up-events has a heavier tail than down-events. Fact 7 is due to the nature of aggressive limit orders. The majority of quotes of aggressive limit orders do not exceed best quotes too much, and thus the price changes caused by them will not be large.

We then analyze the behavior of the NPCs and t2NPCs in terms of trading dates. Figure 8 is about the daily average |NPC| and |t2NPC|. We can see that the average of |t2NPC| (1.0898) is greater than the one of |NPC| (0.9504). The values do not vary greatly among different trading days.

Similar to previous, we also separate all stocks into different groups according to the total absolute value of NPCs and t2NPCs respectively. The k-th group has a total value in interval $[200(k-1), 200k)$ for $k = 1, \ldots, 10$, and the total absolute value of group 11 is more than or equal to 2,000 (see Table 3). Figures 9(a) and 9(b) illustrate the number of stocks for different trading dates and stock groups (separated by NPC and t2NPC, respectively), which are similar to Figs. 6(a) and 6(b). These figures tell that the prices of most stocks in our data set change with lower frequencies and lower magnitudes. Whereas these features do not vary with the trading date, which means that these patterns are stable from the perspective of time.

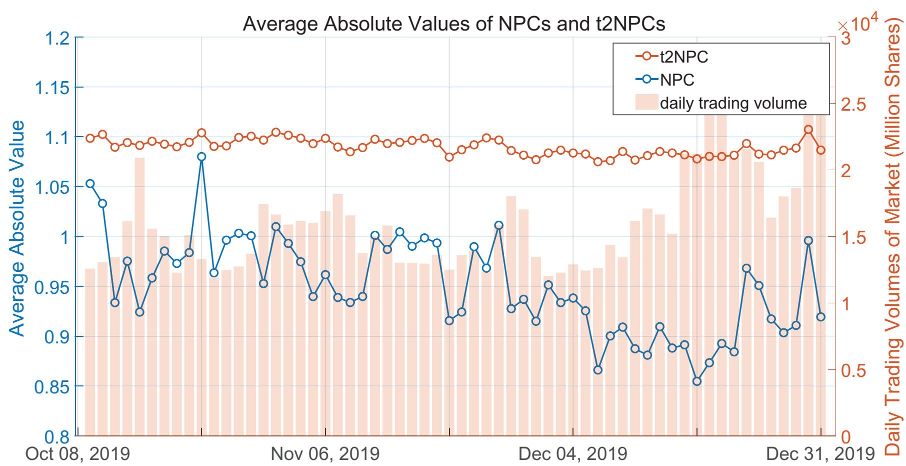
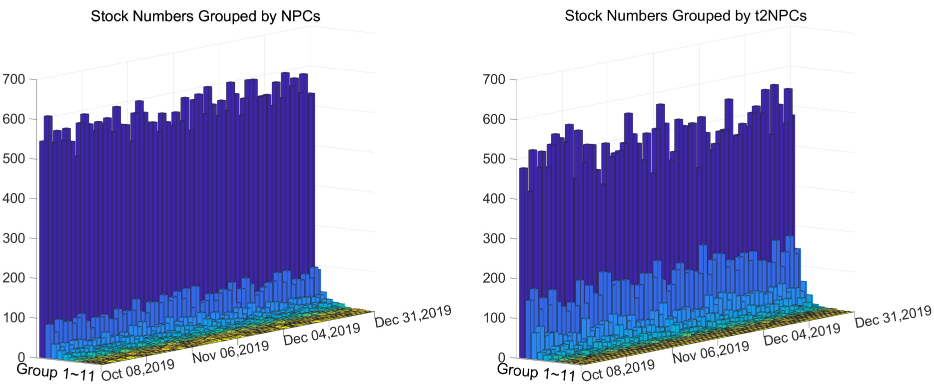


Fig. 8. Average of |NPC| and |t2NPC| versus trading date.

Table 3. Groups used in Fig. 9, grouped by total absolute value of NPCs or t2NPCs.

Group	1	2	3	···	10	11
Total absolute value	[0,200)	[200,400)	[400,600)	···	[1800,2000)	[2000, $+\infty$)



(a) NPC (b) t2NPC

Fig. 9. Stock numbers versus stock group (grouped by total value of |NPC|s and |t2NPC|s, see Table 3) and trading date.

Each type II price change event is only caused by one aggressive limit order. The order completely determines the price change. We are not interested in these events and therefore only consider the case with type I events hereinafter.

3.3. NPC model fitting and empirical results

In this section, we illustrate how to estimate the model (4) and present the corresponding results. To estimate the parameters in our model, the infinite series is truncated at $j = 10$ in the following empirical analysis. We use ridge regression and the weighted least square method to estimate this model. The reason we use ridge regression is that the linear model may not be estimable, since the rareness of large j orders leads to the sparseness of our data. In general, the ridge regression is robust, and the unobserved orders will not affect the overall estimation and hypothesis testings of other parameters.

The model is estimated for up-events and down-events respectively. For each stock, we run a regression for its all up-events during the trading period, and run one for all down-events. Figures 10 and 11 give the estimated values and p-values of $\beta_{\mathrm{M}}^{\pm j}$ and $\beta_{\mathrm{C}}^{\pm j}$ for up-events, and Figs. 12 and 13 for down-events.

The parameters in our model have strong practical meanings — the average impact of different types of orders. For example, parameter β_{M}^{+1} is significant since

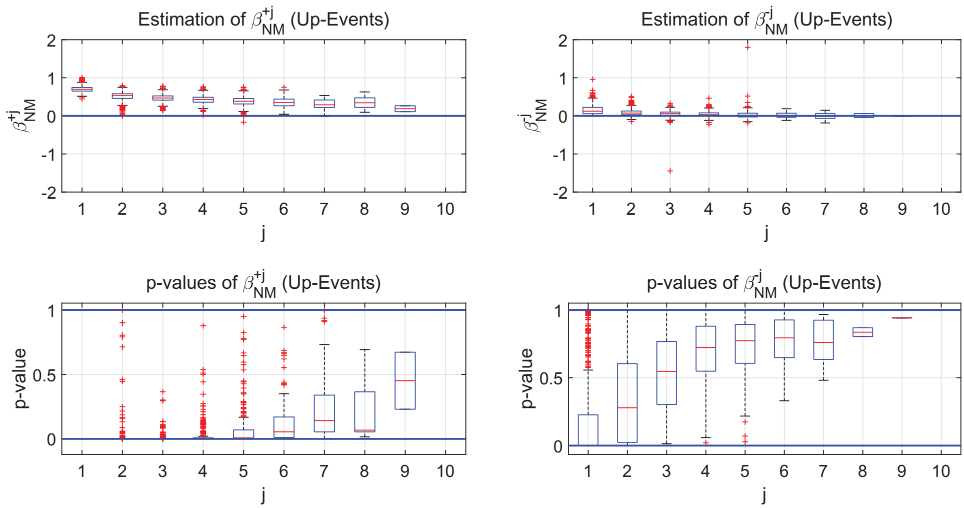


Fig. 10. The estimated values (upper) and p-values (lower) of β^{+j}_{NM} (left) and β^{-j}_{NM} (right) estimated by all up-events during the trading period, and each point represents a stock.

most stocks give a rather small p-value, see Fig. 10. The average of β^{+1}_{M} is 0.7, that is to say, aggressive order that pulls the best ask price up for 1 tick will drive the latent price process $p(t)$ up for 0.7 unit roughly, i.e., it will contribute to 70% of the natural price change, accompanied with 30% noises.

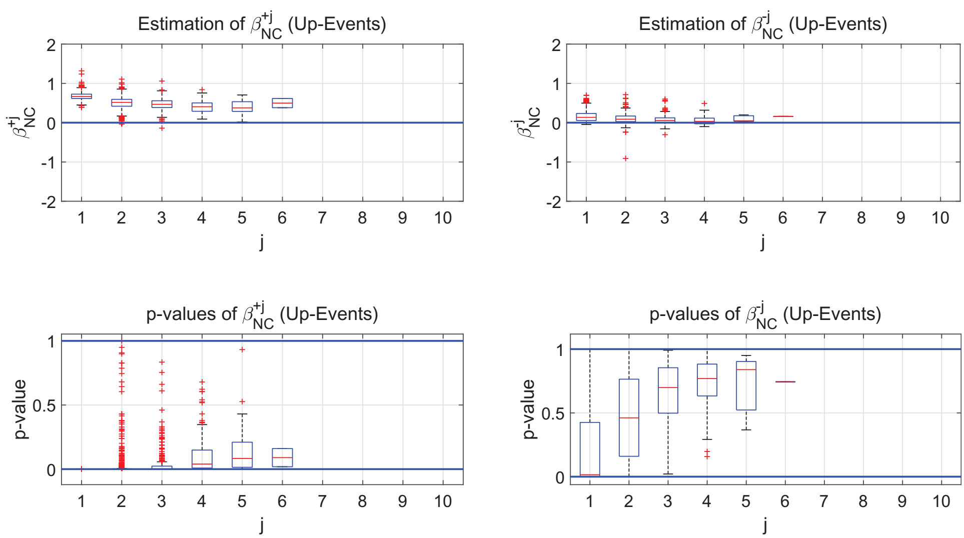


Fig. 11. The estimated values (upper) and p-values (lower) of β^{+j}_{NC} (left) and β^{-j}_{NC} (right) estimated by all up-events during the trading period, and each point represents a stock.

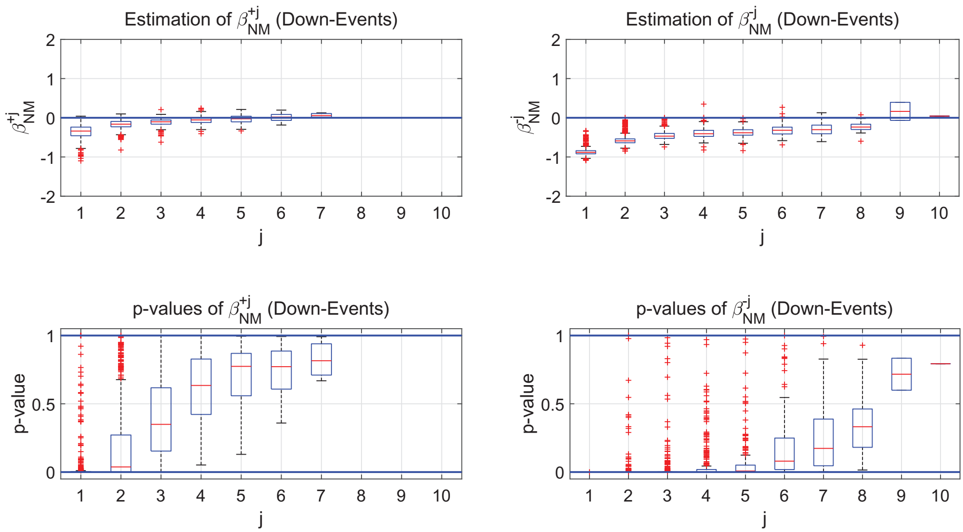


Fig. 12. The estimated values (upper) and p-values (lower) of β^{+j}_{NM} (left) and β^{-j}_{NM} (right) estimated by all down-events during the trading period, and each point represents a stock.

Several facts can be found from these figures:

(1) Figures of p-values show that the significance of all parameters decrease as j increases.
(2) The values of $|\beta|$'s for aggressive orders are usually bigger than the ones for cancelations.
(3) The values of $|\beta|$'s are usually smaller for larger $|j|$'s.

Fact 13 is reasonable for our data set since larger j has less samples. Fact 13 tells us that the impact of aggressive orders is stronger than cancelations. Fact 13 means that an investor can lead to a greater natural price change if he or she split a large order into several small orders.

To make Fact 13 and Fact 13 more persuasive, for both up- and down-events, we conduct F-tests with $H_0 : A \neq B$ for $A, B \in \{\beta^{\pm j}_{\mathrm{M}}, \beta^{\pm j}_{\mathrm{C}}, j = 1, 2, \ldots, 40\}$. Figures 14 and 15 are the p-values of the tests for up- and down-events, respectively. The larger the p-values (more insignificant), the darker the green color of the cells. From both figures, we can find that most β's are significantly different, which illustrates Fact 13 and Fact 13. In addition, one can also find that p-values usually go large for large $|j|$'s. In fact, Fig. 16 shows the number of nonzero samples for different parameters. There are only a few nonzero samples for large $|j|$'s, so it is logical that the p-values for large $|j|$'s are not significant.

We further explain Fact 13. It is consistent with the fact that "price impact is a concave function of the trade size" (Hasbrouck, 1991). Assume we have an order

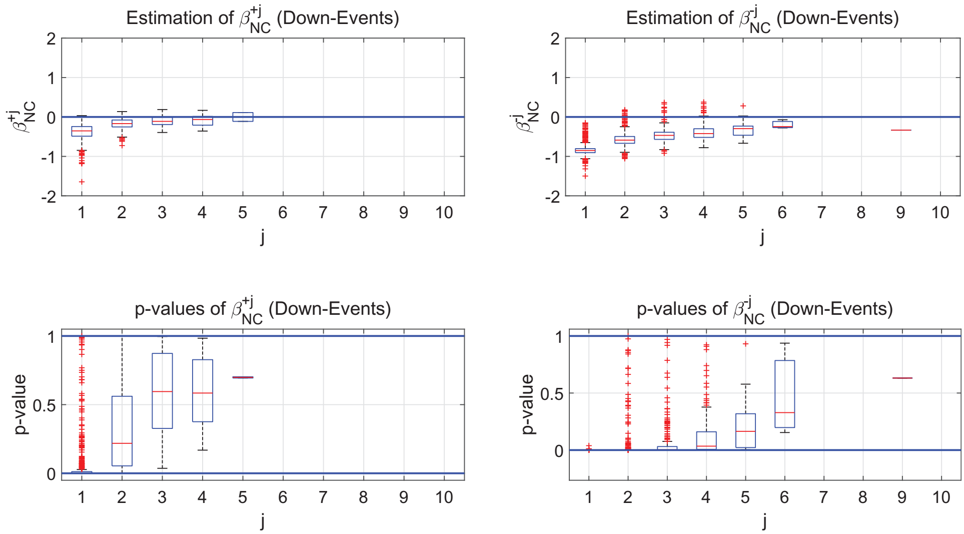


Fig. 13. The estimated values (upper) and p-values (lower) of β^{+j}_{NC} (left) and β^{-j}_{NC} (right) estimated by all down-events during the trading period, and each point represents a stock.

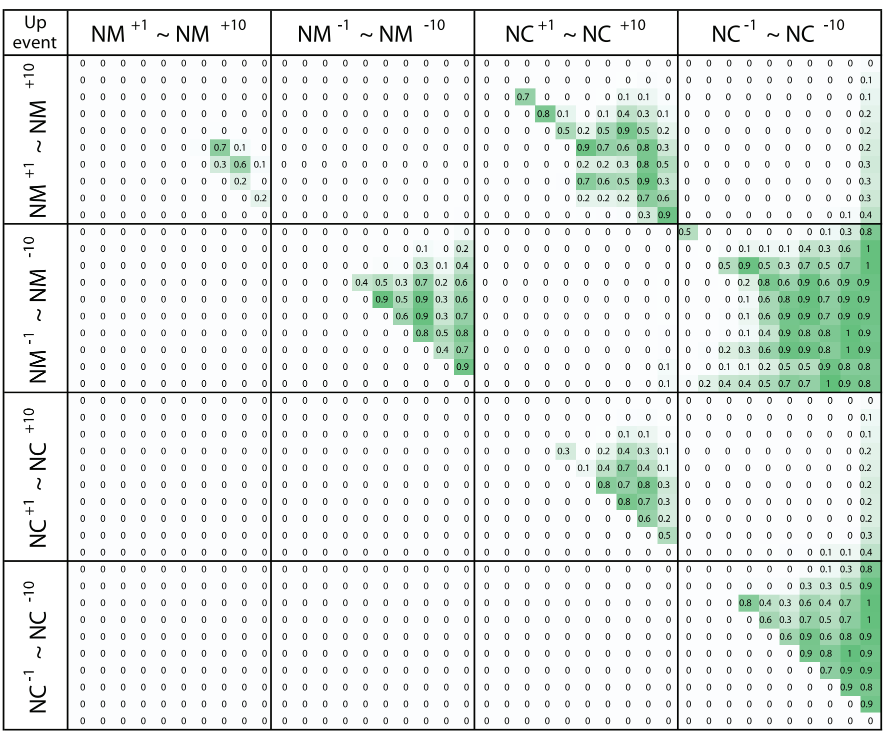


Fig. 14. The p-values of testing $A \neq B$ for all up-events, where $A, B \in \{\beta^{\pm j}_{\mathrm{M}}, \beta^{\pm j}_{\mathrm{C}}, j = 1, 2, \ldots, 40\}$.

Down event	NM^{+1} ~ NM^{+10}									
NM^{+1} ~ NM^{+10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0.3
	0	0	0	0	0	0	0	0.1	0.2	0.6
	0	0	0	0	0.1	0.1	0.4	0.3	0.4	0.9
	0	0	0	0	0	0.6	0.9	0.6	0.6	1
	0	0	0	0	0	0	0.9	0.8	0.8	0.9
	0	0	0	0	0	0	0	0.8	0.7	0.9
	0	0	0	0	0	0	0	0	0.9	0.8
	0	0	0	0	0	0	0	0	0	0.8
	0	0	0	0	0	0	0	0	0	0
NM^{-1} ~ NM^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
NC^{+1} ~ NC^{+10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
NC^{-1} ~ NC^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0

Down event	NM^{-1} ~ NM^{-10}									
NM^{+1} ~ NM^{+10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
NM^{-1} ~ NM^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0.1	0.3	0
	0	0	0	0	0	0	0	0	0.8	0.3
	0	0	0	0	0	0	0	0	0	0.3
	0	0	0	0	0	0	0	0	0	0
NC^{+1} ~ NC^{+10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
NC^{-1} ~ NC^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0

Down event	NC^{+1} ~ NC^{+10}									
NM^{+1} ~ NM^{+10}	0	0	0	0	0	0	0	0	0.1	0.4
	0	0	0	0	0	0.2	0.1	0.7	0.5	1
	0	0	0.3	0.1	0.4	0.8	0.4	1	0.7	0.8
	0	0	0	0.9	0.7	0.6	0.8	0.8	0.9	0.7
	0	0	0	0.3	0.3	0.4	1	0.6	1	0.7
	0	0	0	0.2	0.2	0.3	0.9	0.6	0.9	0.6
	0	0	0	0.4	0.4	0.4	1	0.6	1	0.6
	0	0	0	0.3	0.3	0.3	0.8	0.5	0.9	0.6
	0	0	0.1	0.4	0.4	0.3	0.8	0.5	0.9	0.6
	0	0.2	0.5	0.8	0.8	0.7	1	0.7	1	0.7
NM^{-1} ~ NM^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0.1
	0	0	0	0	0	0	0	0	0.1	0.1
	0	0	0	0	0	0	0	0	0.1	0.1
	0	0	0	0	0	0	0	0	0.2	0.2
NC^{+1} ~ NC^{+10}	0	0	0	0	0	0	0	0	0	0.3
	0	0	0	0	0	0.1	0.1	0.6	0.4	0.9
	0	0	0	0.1	0.2	0.6	0.4	0.9	0.7	0.9
	0	0	0	0	0.8	0.7	0.7	0.8	0.9	0.7
	0	0	0	0	0	0.8	0.7	0.8	0.8	0.8
	0	0	0	0	0	0	0.6	0.9	0.8	0.8
	0	0	0	0	0	0	0	0.7	1	0.7
	0	0	0	0	0	0	0	0	0.8	0.9
	0	0	0	0	0	0	0	0	0	0.7
	0	0	0	0	0	0	0	0	0	0
NC^{-1} ~ NC^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0

Down event	NC^{-1} ~ NC^{-10}									
NM^{+1} ~ NM^{+10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0.2
	0	0	0	0	0	0	0	0	0	0.4
	0	0	0	0	0	0	0	0	0	0.6
	0	0	0	0	0	0	0	0	0.1	0.8
	0	0	0	0	0	0	0	0	0.1	0.9
	0	0	0	0	0	0	0	0	0.1	0.8
	0	0	0	0	0	0	0	0	0.2	0.9
	0	0	0	0	0	0	0	0	0.3	1
	0	0	0	0	0	0	0	0.1	0.3	0.8
NM^{-1} ~ NM^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0.2	0	0	0	0	0	0
	0	0	0	0.2	0.1	0	0	0	0	0
	0	0	0	0	0.4	0.3	0	0	0	0
	0	0	0	0	0	0.7	0.1	0.1	0	0
	0	0	0	0	0	0.1	0.4	0.4	0.1	0
	0	0	0	0	0	0.2	0.4	0.4	0.1	0
	0	0	0	0	0	0	0.8	1	0.4	0.1
NC^{+1} ~ NC^{+10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0.2
	0	0	0	0	0	0	0	0	0	0.4
	0	0	0	0	0	0	0	0	0	0.6
	0	0	0	0	0	0	0	0	0	0.6
	0	0	0	0	0	0	0	0	0	0.5
	0	0	0	0	0	0	0	0	0.2	0.8
	0	0	0	0	0	0	0	0	0.2	0.6
	0	0	0	0	0	0	0.2	0.2	0.4	0.9
	0	0	0	0	0	0.1	0.2	0.2	0.3	0.6
NC^{-1} ~ NC^{-10}	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0.2	0	0	0	0
	0	0	0	0	0	0	0.1	0.1	0	0
	0	0	0	0	0	0	0	0.8	0.3	0.1
	0	0	0	0	0	0	0	0	0.4	0.1
	0	0	0	0	0	0	0	0	0	0.4
	0	0	0	0	0	0	0	0	0	0

Fig. 15. The p-values of testing $A \neq B$ for all down-events, where $A, B \in \{\beta_{\mathrm{M}}^{\pm j}, \beta_{\mathrm{C}}^{\pm j}, j = 1, 2, \ldots, 40\}$.

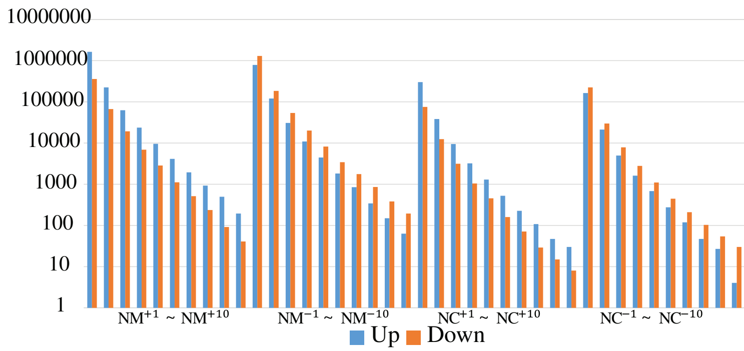


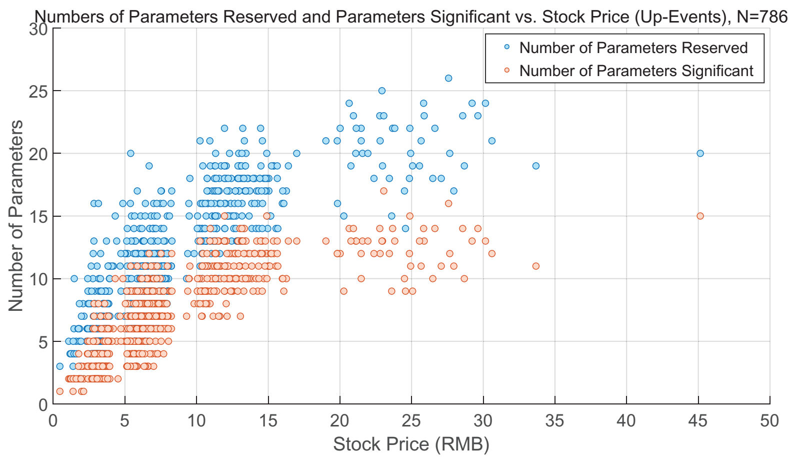
Fig. 16. $\{A_i : A_i \neq 0\}$, where $A \in \{\mathrm{NM}^{+j}, \mathrm{NM}^{-j}, \mathrm{NC}^{+j}, \mathrm{NC}_i^{-j}\}$.

book in which the number of shares (depth) at each price level beyond the best bid and ask is equal.[5] Then, for an order, its corresponding $|j|$ will almost be proportional to its size. The decreasing $|\beta|$'s imply $|j \cdot \beta^{|j|}|$ is a concave function of $|j|$,

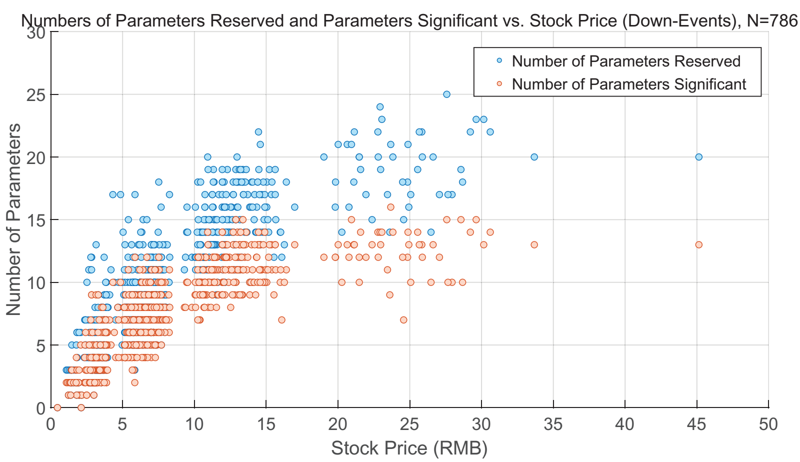
[5]This is similar to the assumption in Cont *et al.* (2014).

so is a concave function of the order size, which illustrates the fact. Although actual order books have complex dynamics, the simple assumption can help us understand Fact 13.

Orders with directions opposite to the direction of price change will not obviously change the latent price process $p(t)$ (since these parameters estimated are almost zero, see upper right graphs of Figs. 10 and 11, and upper left graphs of Figs. 12 and 13). For those opposing orders who change best quotes for 1 tick, i.e., $j = 1$, we discover that their contribution to price change has the same direction as



(a) up-events



(b) down-events

Fig. 17. Scatter plots about the number of parameters in our model are reserved (blue point) and the number of parameters significant (red point) versus with stock price, and each point represents a stock.

the price change. This can be explained from the perspective of market participants. For instance, during an up-event, if a seller chooses to sell more stocks by aggressive orders at the very beginning, then the seller's passive limit orders near the end of events will be reduced, which increases the NPC. However, if these aggressive sell orders' volumes are very large, and they change the best bid for many ticks, we conjecture that they would introduce a negative impact.

Figures 17(a) and 17(b) are scatter plots showing the number of parameters reserved (blue point) and the number of parameters significant (red point) versus the stock price (for up- and down-events, respectively) for every stock. Both of them are positively correlated with the stock price, and a simple regression implies that an increase of 1RMB in stock price is associated with approximately an increase of 0.71 units in parameters reserved for up-events (0.73 for down-events) and with approximately an increase of 0.47 units in parameters significant for up-events (0.46 for down-events). As a result, about 10 parameters are reserved, and about 7.3 parameters are significant on average.

The R^2 of our estimation is calculated, which is defined as

$$R^2 = 1 - \frac{\sum_i (\mathrm{NPC}_i - \widehat{\mathrm{NPC}}_i)^2}{\sum_i (\mathrm{NPC}_i - \overline{\mathrm{NPC}})^2},$$

where $\widehat{\mathrm{NPC}}_i$ is the estimated value, and $\overline{\mathrm{NPC}}$ is the average of NPC_i's. Negative R^2 may also appear since our model does not include the intercept term. For up-events, the median R^2 of the NPC model is about 19.07%, and for down-events, it is about 7.69% (see Fig. 18). The results show that up-events and down-events for large-tick stocks in the Chinese market have strong differences.

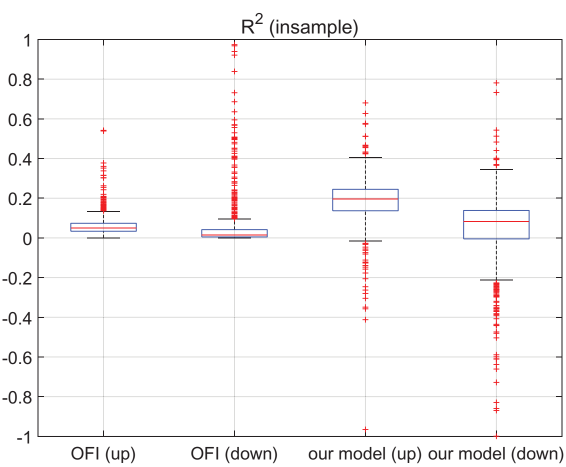


Fig. 18. Box plot of R^2, and each point represents a stock.

For the convenience of comparison, we also run the regression proposed in Cont *et al.* (2014)

$$\text{NPC} = \alpha + \beta \cdot \text{OFI} + \varepsilon.$$

The results of R^2 are also shown in Fig. 18. The median R^2 of the OFI model for up-events is about 5.08%, and for down-events, it is about 1.51%. The interpretation of the OFI model is not ideal for large-tick stocks in the Chinese market.

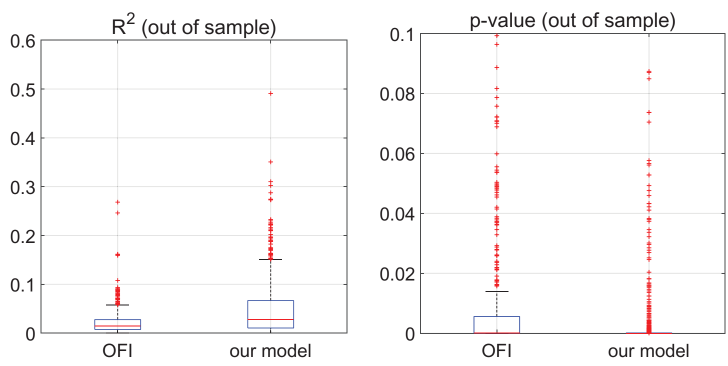
3.4. Application

The empirical results of model (4) in Sec. 3.3 show a strong link between the impact of different orders and the price change. As an application, we use our model to predict future price movements in this section.

We divide a total number of 60 trading days into a training set and a test set. We use the first 40 trading days (training set) to fit our heteroscedastic linear model and estimate the coefficients, and use the following 20 days (test set) to test its predictability. In the test set, for a price change event, we can use our model to give an estimation of the NPC. Meanwhile, we can also observe the real NPC of this price change event. Then, after the event, we use the difference between the estimation of NPC and the real NPC to predict the price change in the following 100 event times. More formally, we run the regression

$$\text{midprice}(t+100) - \text{midprice}(t) = \alpha + \beta \cdot (\text{estimation of NPC}_t - \text{real NPC}_t) + \varepsilon.$$

For comparison, similar to Cont *et al.* (2014), we use the OFI model to predict the price change in the following 100 event times. The R^2 and p-values of both models are shown in Fig. 19.



(a) R^2 (b) p-value

Fig. 19. Box plot of R^2 and the slope's p-value for prediction. Each point represents a stock.

The median R^2 is 1.49% for the OFI model and 2.81% for our model. The slope's p-value is also much more significant for our model. The results show that our model performs better than the OFI model for large-tick stocks in the Chinese market.

4. Conclusions

In this paper, we propose a new framework for studying the discrete price change process, which focuses on the impact of aggressive orders and cancelations. The framework is customized to the price dynamics of large-tick stocks. The process is driven by states and events of best quotes, and the corresponding event-based price change is called the NPC. We believe that the NPC is a better concept to study the impact of order types on price changes.

We conduct an extensive empirical analysis on 786 large-tick stocks traded on the Shenzhen Stock Exchange. The preliminary empirical studies show that our new concepts, price change event and NPC, have stable patterns across stocks and trading dates, and can capture the nature of "price change" to some extent. The fitting results demonstrate that aggressive orders may introduce stronger impact on price change than cancelations. This implies that the main driving factors of the real price change in the A-share market are aggressive orders. Meanwhile, splitting a big order into several small orders can lead to a greater natural price change. This conclusion is not novel from common sense, but it is rarely confirmed via statistical modeling. Our model also performs better than traditional models in terms of the prediction of future price movements.

Overall, these findings provide a novel perspective on the price change behaviors of large tick stocks. Our paper provides a new framework for high-frequency data analysis, which can incent the study of order book dynamic models and empirical studies related.

Declarations of Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Appendix A. Empirical Example of NPC

To demonstrate the intuition of our definitions, we give sample paths of the best bid and the best ask processes of real transaction data in Fig. A.1. The stock is Ping

Fig. A.1. Five price change events from 10:10:59 to 10:11:05 a.m. on October 17, 2019 (Beijing time) for stock "Ping An Bank" (000001.SZ). The orange dots show how best ask prices change with order time, and the blue dots represent best bid prices. Event A, B, D and E (green background color) are type I events, while C (yellow background color) is a type II event.

An Bank (000001.SZ), which is traded on Shenzhen Stock Exchange. The sample path begins at 10:10:59 and ends at 10:11:05 a.m., October 17, 2019 (Beijing time), and it lasts about 6 s (natural time). The horizontal axis of the figure represents the order time rather than the natural time. The corresponding order time begins at 23992 and ends at 24093, and there are 102 orders during this period in total.

Five price change events are observed during this period. Event A, B, D and E are type I events, while C is a type II event. The value of NPC for these five events are reported in Table A.1. The original data of Fig. A.1 can be seen in Table A.2.

Take event A and event C as examples. For price change event A, the mid-price is 1694.5 at the beginning, and 1691.5 at the end. Hence, it is a down-event. Its NPC equals to $1691.5 - 1694.5 = -3$. For price change event C, the mid-price is 1692.5 at the beginning, and 1694.5 at the end. No order time exists between t^{start} and t^{end}, and hence it is a type II up-event. Its t2NPC equals to $1694.5 - 1692.5 = 2$.

Table A.1. NPC of five events in Fig. A.1.

Event ID	t^{start}	t^{end}	Event Type	Background Color	Direction	NPC
A	23992	24023	type I	green	down	−3
B	24028	24040	type I	green	up	1
C	24040	24041	type II	yellow	up	2
D	24041	24071	type I	green	down	−2
E	24079	24093	type I	green	down	−1

Table A.2. Data of Fig. A.1.

Stock: Ping An Bank (000001.SZ), Trading Date: October 17, 2019 (Beijing time)							
Order Time	Best Bid	Best Ask	Timestamp	Order Time	Best Bid	Best Ask	Timestamp
23992	1694	1695	101059208156	24043	1692	1694	101102016186
23993	1691	1695	101059208161	24044	1692	1694	101102016186
23994	1691	1694	101059208161	24045	1692	1694	101102016186
23995	1691	1694	101059208161	24046	1692	1694	101102016190
23996	1691	1694	101059208161	24047	1692	1694	101102016190
23997	1691	1694	101059208161	24048	1692	1694	101102016686
23998	1691	1694	101059208192	24049	1692	1694	101102077250
23999	1691	1694	101059208192	24050	1692	1694	101102157154
24000	1691	1694	101059208193	24051	1692	1694	101102269496
24001	1691	1694	101059208193	24052	1692	1694	101102269676
24002	1691	1694	101059268354	24053	1691	1694	101102399668
24003	1691	1694	101059328689	24054	1691	1694	101102581222
24004	1691	1694	101059328989	24055	1691	1694	101102581571
24005	1691	1694	101059408405	24056	1691	1694	101102764881
24006	1692	1694	101059408410	24057	1691	1694	101102886078
24007	1692	1695	101059528475	24058	1691	1694	101103340283
24008	1692	1695	101059831323	24059	1691	1694	101103340936
24009	1692	1695	101059831386	24060	1691	1694	101103340942
24010	1692	1695	101059891233	24061	1691	1694	101103340945
24011	1692	1695	101059891281	24062	1691	1694	101103340945
24012	1692	1695	101059891316	24063	1691	1694	101103341086
24013	1692	1695	101059891408	24064	1691	1694	101103341088
24014	1692	1695	101100217087	24065	1691	1694	101103400929
24015	1689	1695	101100217144	24066	1691	1694	101103522955
24016	1689	1695	101100217155	24067	1691	1694	101103523871
24017	1689	1695	101100217155	24068	1689	1694	101103525208
24018	1689	1695	101100217155	24069	1689	1694	101103642348
24019	1689	1695	101100399427	24070	1692	1694	101103765369
24020	1689	1692	101100400519	24071	1692	1693	101103842968
24021	1689	1692	101100523454	24072	1692	1693	101104145999
24022	1689	1692	101100583158	24073	1692	1693	101104262222
24023	1691	1692	101100909920	24074	1692	1693	101104262523
24024	1691	1692	101101031733	24075	1692	1693	101104263202
24025	1691	1692	101101031983	24076	1692	1693	101104326234
24026	1691	1692	101101032142	24077	1692	1693	101104326407
24027	1691	1692	101101032244	24078	1692	1693	101104512224
24028	1691	1692	101101032534	24079	1692	1693	101104512595
24029	1692	1695	101101032591	24080	1689	1692	101104512611
24030	1692	1695	101101032591	24081	1689	1692	101104512611
24031	1692	1695	101101032591	24082	1689	1692	101104512611
24032	1692	1695	101101032591	24083	1689	1692	101104512611

Table A.2. (*Continued*)

Stock: Ping An Bank (000001.SZ), Trading Date: October 17, 2019 (Beijing time)							
Order Time	Best Bid	Best Ask	Timestamp	Order Time	Best Bid	Best Ask	Timestamp
24033	1692	1695	101101149574	24084	1689	1692	101104512611
24034	1692	1695	101101533128	24085	1689	1692	101104512655
24035	1692	1695	101101774011	24086	1689	1692	101104826760
24036	1692	1695	101101774048	24087	1689	1692	101105009342
24037	1692	1694	101101774116	24088	1689	1692	101105009586
24038	1692	1694	101101894994	24089	1689	1692	101105009689
24039	1692	1694	101101895045	24090	1689	1692	101105070861
24040	1692	1693	101102016165	24091	1689	1692	101105071077
24041	1694	1695	101102016182	24092	1689	1692	101105632200
24042	1692	1695	101102016186	24093	1691	1692	101105763366

References

Abergel, F, M Anane, A Chakraborti, A Jedidi and IM Toke (2016). *Limit Order Books*. Cambridge University Press.

Abergel, F and A Jedidi (2015). Long-time behavior of a hawkes process–based limit order book. *SIAM Journal on Financial Mathematics*, 6(1), 1026–1043.

Angel, JJ (1997). Tick size, share prices, and stock splits. *The Journal of Finance*, 52(2), 655–681.

Biais, B, P Hillion and C Spatt (1995). An empirical analysis of the limit order book and the order flow in the paris bourse. *The Journal of Finance*, 50(5), 1655–1689.

Capponi, F and R Cont (2020). Multi-asset market impact and order flow commonality. Available at SSRN 3706390.

Cont, R and A De Larrard (2013). Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1), 1–25.

Cont, R, A Kukanov and S Stoikov (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1), 47–88.

Eisler, Z, J-P Bouchaud and J Kockelkoren (2012). The price impact of order book events: Market orders, limit orders and cancelations. *Quantitative Finance*, 12(9), 1395–1419.

Gould, MD and J Bonart (2016). Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity*, 2(2), 1650006.

Harris, L and J Hasbrouck (1996). Market vs. limit orders: The superdot evidence on order submission strategy. *Journal of Financial and Quantitative Analysis*, 31(2), 213–231.

Hasbrouck, J (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1), 179–207.

Lipson, ML and S Mortal (2006). The effect of stock splits on clientele: Is tick size relevant? *Journal of Corporate Finance*, 12(5), 878–896.

Lu, X and F Abergel (2018a). High-dimensional hawkes processes for limit order books: Modeling, empirical analysis and numerical calibration. *Quantitative Finance*, 18(2), 249–264.

Lu, X and F Abergel (2018b). Order-book modeling and market making strategies. *Market Microstructure and Liquidity*, 4(1–2), 1950003.

Mertens, LP, A Ciacci, F Lillo and G Livieri (2022). Liquidity fluctuations and the latent dynamics of price impact. *Quantitative Finance*, 22(1), 149–169.

Peterson, M and E Sirri (2002). Order submission strategy and the curious case of marketable limit orders. *Journal of Financial and Quantitative Analysis*, 37(2), 221–241.

Rydberg, TH and N Shephard (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1(1), 2–25.

Taranto, DE, G Bormetti, J-P Bouchaud, F Lillo and B Tóth (2018a). Linear models for the impact of order flow on prices. i. history dependent impact models. *Quantitative Finance*, 18(6), 903–915.

Taranto, DE, G Bormetti, J-P Bouchaud, F Lillo and B Tóth (2018b). Linear models for the impact of order flow on prices. ii. the mixture transition distribution model. *Quantitative Finance*, 18(6), 917–931.

Weber, P and B Rosenow (2006). Large stock price changes: volume or liquidity? *Quantitative Finance*, 6(1), 7–14.

Xu, K, MD Gould and SD Howison (2018). Multi-level order-flow imbalance in a limit order book. *Market Microstructure and Liquidity*, 4(3–4), 1950011.