

Order Book Queue Hawkes Markovian Modeling

Philip Protter *

Department of Statistics, Columbia University
and

Qianfan Wu

Department of Finance, Indiana University Bloomington
and

Shihao Yang[†]

Department of Industrial and Systems Engineering,
Georgia Institute of Technology

February 26, 2023

Order Book Queue Hawkes Markovian Modeling

Abstract

**The research of Dr. Philip Protter is partially supported by NSF Grant DMS-2106433.*

[†]Corresponding Author. Email: shihao.yang@isye.gatech.edu. The authors thank Dr. Dan Christina Wang for providing access to the LOBSTER data used in this study.

This article presents a Hawkes process model with Markovian baseline intensities for high-frequency order book data modeling. We classified intraday order book trading events into a range of categories based on their order types and the price change after their arrivals. In order to capture the stimulating effects between multiple types of order book events, we use multivariate Hawkes process to model the self- and mutually-exciting event arrivals. We also integrate a Markovian baseline intensities into the event arrival dynamic, by including the impacts of order book liquidity state and time factor on the baseline intensity. A regression-based non-parametric estimation procedure is adopted to estimate the model parameters in our Hawkes+Markovian model. To eliminate redundant model parameters, LASSO regularization is incorporated into the estimation procedure. Besides, model selection method based on Akaike Information Criteria is applied to evaluate the effect of each part of the proposed model. An implementation example based on real LOB data is provided. Through the example we studied the empirical shapes of Hawkes excitement functions, the effects of liquidity as well as time factors, the LASSO variable selection, and the explanation power of Hawkes and Markovian elements to the dynamics of order book.

Keywords: Hawkes Process, Order book modeling, Non-parametric estimation, Model selection

1 Introduction

Electronic Limit Order Book (LOB) is the list of electronic orders that a trading venue uses to record the interest of buyers and sellers in a particular financial instrument. The modern financial market has witnessed unprecedented increases in trading volume and frequency during the recent decades, with the global total value of stock traded escalating and the average stock holding period plunging significantly from 1990 to 2018 (The World Bank 2019, Maloney & Almeida 2019). Therefore, understanding the dynamics of LOB has become increasingly significant in the analysis of the liquidity, transaction cost, and regulation of the modern global financial market.

1.1 literature review

Hawkes process (Hawkes 1971, Hawkes & Oakes 1974), also called “self-exciting and mutually-exciting point process”, is a type of stochastic point process whose essential property is that the arrival/occurrence of any event will impact the arrival probability of further events. [Hawkes process has various applications in finance. For example, Chavez-Demoulin & McGill \(2012\) proposes a Hawkes process model to enhance the estimation of high-frequency stock trading Value-at-Risk \(VaR\) measures; Bacry et al. \(2013a,b\) propose a novel Hawkes process construction to capture the microstructure noise indicated by the jumps of asset prices, and further discussed the limit properties of the model. Bacry et al. \(2015\) and Hawkes \(2018\) have provided comprehensive reviews on Hawkes process applications in finance, especially in modeling LOB data, including related works in estimation procedure \(Kirchner 2017, Bacry & Muzy 2014a\), Hawkes process generalizations \(Blanc et al. 2017\), and model modifications of minor details \(Clements et al. 2015, Ferriani & Zoi 2020\). On](#)

the theoretical side, Jaisson & Rosenbaum (2015) provides the stability conditions and limiting distributions of Hawkes process when applied in finance.

As electronic LOB data is typically complex, large-scale, and high-frequency, Hawkes process has become increasingly popular in fitting LOB dynamics due to its ability to capture the complex stimulating effects between event flows. Plenty of research has utilized the Hawkes process to estimate LOB order flows. Among them, Hawkes process with exponential or power-law stimulating functions is among the most popular model specifications in recent works (Muni Toke & Pomponio 2012, Abergel & Jedidi 2015, Morariu-Patrichi & Pakkanen 2018, Kirchner 2017), and literature has discussed the strengths of each type (Lalouache & Challet 2016, Filimonov & Sornette 2015). Furthermore, in addition to using the parametric kernels (e.g. the exponential kernel), Kirchner (2017), Bacry & Muzy (2014a,b) introduce non-parametric estimation methods that enable the estimation of flexible kernel shapes and enhance computational efficiency.

Besides the self and cross-exciting property, LOB data can also be viewed as a chain of transitions from one state to another based on different price and order size levels as different types of order arrives. Based on this perspective, Markov models can be applied to LOB data, assuming the transition from one LOB state to another as an event arrives depends only on the state attained in the previous event. For example, Huang et al. (2015), Huang & Rosenbaum (2017) proposes simulation and analytical frameworks for Markov models on LOB data and demonstrates empirical findings on various LOB event types; Based on the “Queue-Reactive Model” proposed by Huang et al. (2015), Wu et al. (2019) discusses that the model performance can be boosted by adding the Hawkes stimulating components to the order arrival intensities. Besides, Omi et al. (2017) argues a Hawkes model with background rates relating to the time state of order arrival can achieve better model fit. Also, Morariu-Patrichi & Pakkanen (2018) estimates a state-dependent Hawkes model for LOB modeling in which the exponential excitement kernel depends on the process

that switches state whenever an event arrives.

1.2 our work and contribution

Most of the previous works mentioned above apply Hawkes process and Markov models independently to LOB data. The effectiveness of combining the two models in LOB modeling is not studied yet with non-parametric kernel. In this paper, we introduce a Hawkes+Markovian model to capture the dynamics of electronic LOB data. We apply a multi-dimensional Hawkes process to a comprehensive range of event types derived from LOB movements. In addition, by integrating the Markovian model on the LOB data, our model captures the intuition that the stimulating effects among events also depend on the liquidity state and time factor of the LOB right before event arrives. We then implement a regression-based non-parametric method (Kirchner 2017) for parameter estimation. The main idea of our estimation procedure is to allocate the event arrival sequences into a series of fixed-size bins of discretized short-period, and then obtain the Hawkes kernel estimators as a step function. In contrast to previous works (Kirchner 2017, Morariu-Patrichi & Pakkanen 2018) that do not consider order sizes, we take order sizes into account in the arrival sequence construction to better capture the stimulating effects of large and small orders. As the number of estimated parameters is large, we also incorporate LASSO regularization (Tibshirani 1996) in our model estimation. Furthermore, we propose a model selection method based on Akaike Information Criterion (AIC) to analyze the contribution of the Hawkes stimulation part, the Markovian part, and the LASSO part to model explanation power. Finally, after validating the estimation procedure on a fully simulated LOB dataset, we demonstrate an implementation example using real order book data.

Main contribution of this paper is summarised as follows: (1) We proposed a novel Hawkes + Markovian model combination for order book modeling, in which the baseline intensity of the Hawkes process depends on the liquidity state and time factor. (2) We

develop a non-parametric estimation procedure extended from Kirchner (2017), in which the Hawkes kernels and order book states can be collectively estimated using a regression-based method. The non-parametric method also removes any assumptions on the kernel shapes, such as the exponential. (3) We utilize LASSO regularization to enhance model performance and adopt an AIC-based method to evaluate model explanatory power. (4) We document empirical findings using real LOB data. For example, we find some Hawkes kernels do not exhibit convex monotonic decreasing shapes.

Distinction of our work from other Hawkes LOB modeling methods. We especially point out that Wu et al. (2019), Omi et al. (2017) are two very recent works that analyze the interaction between order book state and Hawkes process. Although Wu et al. (2019) also uses the intuition of order liquidities, its specification differs from ours since it uses a parametric exponential Hawkes kernel. It also doesn't include time factor estimation in the baseline intensity estimation. Furthermore, we use more detailed level-3 order book data on the U.S. market while Wu et al. (2019) uses level-1 order book data in the EU futures market. Omi et al. (2017)'s Hawkes specification also follows the exponential kernel. The events considered in this study are the filtered movements of the mid-price, which differs from our classification based on the tick-level orders on each side of the order book. Omi et al. (2017) documents the time-based baseline intensities estimated using a Bayesian method, which is also fundamentally different from our regression-based method.

The rest of this paper is organized as follows: section 2 demonstrates a brief mathematical introduction of order book data representations; section 3 introduces our proposed model, with detailed illustrations on event classification and model structures; section 4 provides a detailed description on the non-parametric model estimation as well as model selection; section 5 showcases the aggregated empirical results from an implementation example based on real LOB data; section 6 discusses and concludes.

2 Order book representation

The limit order book is mainly constructed by two elements. The first element is the shape of the order book, consisting of all the orders at which prices the market wants to buy (bid price) and the market wants to sell (ask price). The bid/ask prices form the bid/ask queues and must be multiples of the tick size, which is the measure of the minimum upward or downward movement of security prices (currently the tick size for all U.S exchanges is \$0.01). The second element is the center position of the order book between the best bid price (the highest price the market wants to buy) and the best ask price (the lowest price the market wants to sell). The center position of LOB is often referred to as the “reference price” and the distance between the best bid and best ask is referred to as “bid-ask spread”. The easiest way to approximate the reference price is to define it as the midpoint of the best bid and best ask, also known as the “mid-price”. The order book also includes information on the size of each order, which is the quantity of shares an order attempts to execute. See Figure 1 for an example.

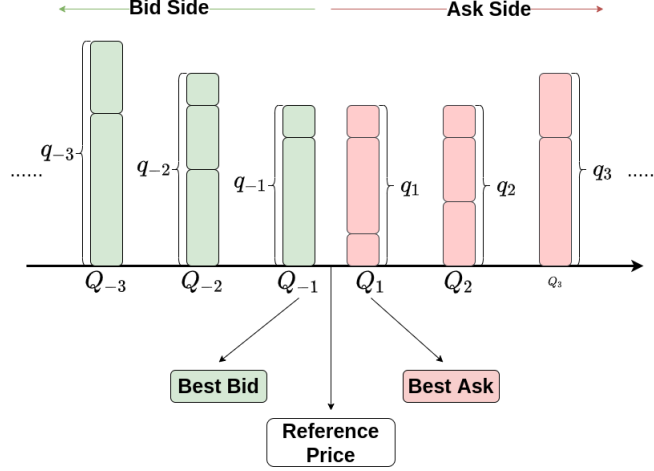


Figure 1: A simple order book representation. $(Q_{-1}, Q_{-2}, Q_{-3}) / (Q_1, Q_2, Q_3)$ represent the first, second, and third price level on the bid/ask side, respectively; $(q_{-1}, q_{-2}, q_{-3}) / (q_1, q_2, q_3)$ represent the number of orders on the first, second, and third price level on the bid/ask side, respectively.

Mathematical Representation Huang et al. (2015) and Huang & Rosenbaum (2017) demonstrate a comprehensive overview of the mathematical representations of LOB. Recall the reference price $p_{\text{ref}}(t)$ must lie strictly between the best bid and best ask. Let $q_i(t)$ be the size of the ask orders at price level Q_i that is the i -th tick strictly above $p_{\text{ref}}(t)$. Symmetrically, $q_{-i}(t)$ is the negation of size of the bid orders at price level Q_{-i} that is the i -th tick strictly below $p_{\text{ref}}(t)$.

Formally, let α be the single tick value, then $Q_1 := \min\{n\alpha : n\alpha > p_{\text{ref}}, n \in \mathbb{Z}\}$, $Q_i := Q_1 + (i-1)\alpha$, $\forall i > 1$. Similarly, $Q_{-1} := \max\{n\alpha : n\alpha < p_{\text{ref}}, n \in \mathbb{Z}\}$, $Q_{-i} := Q_{-1} - (i-1)\alpha$, $\forall i > 1$. See Figure 1 and its caption for an example.

The complete order book's shape at time t is an infinite vector for the current size at all prices $q(t) = [\dots, q_{-k}(t), \dots, q_{-1}(t), q_1(t), \dots, q_k(t), \dots]$ with $q_i \in \mathbb{Z}$ denoting the size at each price. Note $q_i < 0$ if these orders are bid orders and $q_i > 0$ if they are ask orders.

If $q_i = 0$ there is no orders at price level Q_i . The reference price, $p_{\text{ref}}(t)$, is often the mid-price, with some technicality detailed in Supplementary Material section A. The LOB information at time t is therefore fully represented by $[q(t), p_{\text{ref}}(t)]$, $t \geq 0$ (Huang et al. 2015, Huang & Rosenbaum 2017).

To restrict the dimensions of $[q(t), p_{\text{ref}}(t)]$, we consider only K limits on each side, and thus have now $q(t) = [q_{-K}(t), \dots, q_{-1}(t), q_1(t), \dots, q_K(t)]$, which we shall call “level- K order book”. Since q_i can be 0 if there’s no order at its price level, the best bid and best ask prices are defined as the nearest price levels to the reference price with non-empty orders sizes:

$$Q_{\text{best-bid}} = Q_{\max\{i:i<0 \text{ and } |q_i|\neq 0\}} , Q_{\text{best-ask}} = Q_{\min\{i:i>0 \text{ and } |q_i|\neq 0\}}$$

3 Model specifications

Following the framework we outlined above, we propose a specific event arrival dynamic for empirical modeling of the level- K order book.

We consider the LOB event arrival process as $6 \times (K + 1)$ dimensional, indicating $6 \times (K + 1)$ types of events are studied in the level- K order book. They are generally grouped into two categories: (1) order book events that do not change reference price, and (2) order book events that change the reference price. The way we classify LOB events is mainly extended from previous work on the “Queue-Reactive” LOB model (Huang et al. 2015, Huang & Rosenbaum 2017). Compared to the simpler LOB event classifications used in (Kirchner 2017, Morariu-Patrichi & Pakkanen 2018), our classification not only separates events belonging to each of the K order book levels, but also classifies events in more detailed groups when the reference price changes, enabling our proposed model to capture more complex dynamic of event stimulation. Table 1 illustrate order types with $K = 3$.

	No Price Change			Price Change	
	Level 1	Level 2	Level 3	Price Up	Price Down
Insertion	+1(i), -1(i)	+2(i), -2(i)	+3(i), -3(i)	p+(i)	p-(i)
Cancellation	+1(c), -1(c)	+2(c), -2(c)	+3(c), -3(c)	p+(c)	p-(c)
Trade	+1(t), -1(t)	+2(t), -2(t)	+3(t), -3(t)	p+(t)	p-(t)

Table 1: An illustration of $6 \times (K + 1)$ order types. The table shows the order types of $K = 3$, meaning limit and market orders above or below three ticks of the reference price are considered in the modeling. More details on the order types are explained in section 3.1 and 3.2.

3.1 order book events that do not change reference price

When studying the level- K order book, each order book queue can have the following events:

- a trade, or market order, which is denoted as (t)
- an insertion of new limit order, which is denoted as (i)
- a cancellation of existing limit order, which is denoted as (c)

For level- K order book, we have in total $2K$ number of queues:

- K ask queues, denoted as $+1, +2, \dots, +K$, which are the 1st tick above reference price, 2nd tick above reference price, ..., K -th tick above reference price, respectively.
- K bid queues, denoted as $-1, -2, \dots, -K$, which are the 1st tick below reference price, 2nd tick below reference price, ..., K -th tick below reference price, respectively.

Therefore, we have in total $3 \times 2K$ number of events that do not change reference price. For example, “+1($\mathbf{t}$)” denotes the event that a trade order arrives at the first ask queue, “-2($\mathbf{i}$)” denotes the event that a new order arrives at the second bid queue and is inserted, and “+3($\mathbf{c}$)” denotes the event that an existing order is canceled at the third ask queue.

3.2 order book events with reference price change

This section focuses on the modeling for another 6 types of events that shift the reference price. The reference price can increase one tick due to the following event:

- Trade at the best ask price that depletes the queue of best ask. The event of reference price increase due to trade is denoted as $\mathbf{p+}(\mathbf{t})$
- Cancellation of all orders at the best ask price. The event of reference price increase due to cancellation is denoted as $\mathbf{p+}(\mathbf{c})$
- Insertion of bid order at a higher price than the current best bid offer, which is only possible when the bid-ask spread is strictly larger than one tick. The event of reference price increase due to insertion is denoted as $\mathbf{p+}(\mathbf{i})$

On the flip side, the reference price can decrease one tick due to the following event:

- Trade at the best bid price that depletes the queue of best bid. The event of reference price decrease due to trade is denoted as $\mathbf{p-}(\mathbf{t})$
- Cancellation of all orders at the best bid price. The event of reference price decrease due to cancellation is denoted as $\mathbf{p-}(\mathbf{c})$
- Insertion of ask order at a lower price than the current best ask offer, which is only possible when the bid-ask spread is strictly larger than one tick. The event of reference price decrease due to insertion is denoted as $\mathbf{p-}(\mathbf{i})$

3.3 Hawkes+Markovian model of the order book

The order book event processes of dimension $6K + 6$ at time t are represented as $X_i(t)$, $i = 1, 2, \dots, 6K + 6$, with $X_1, X_2, \dots, X_{6K+6}$ representing the arrival process for the following events:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{6K+6} \end{pmatrix} = \begin{pmatrix} \underbrace{(-K(i), -K(c), -K(t), \dots, -1(i), -1(c), -1(t))^\top}_{3K} \\ \underbrace{(+1(i), +1(c), +1(t), \dots, +K(i), +K(c), +K(t))^\top}_{3K} \\ \underbrace{(p^-(t), p^-(c), p^-(i), p^+(t), p^+(c), p^+(i))^\top}_6 \end{pmatrix}_{(6K+6) \times 1}$$

Basically, $X_i(t)$ represents the cumulative size of the corresponding event. For example, $X_1(t)$ corresponds to event $-K(i)$, the insertion of limit order at the bid queue K -th tick below reference price; $X_{6K+6}(t)$ corresponds to event $p^+(i)$, price increase due to insertion of bid above current best bid.

3.3.1 Hawkes part

Define the instantaneous rate of event i 's arrival to be

$$\lambda_i(t) = \lim_{\Delta t \rightarrow 0+} \frac{\mathbb{E}(X_i(t + \Delta t) - X_i(t) | \mathcal{F}_t)}{\Delta t}$$

$\forall i = 1, 2, \dots, 6K + 6$, where $\mathcal{F}_t$ contains all information about order book queue, mid-price, liquidity state, and time factor up to time t . In particular, the complete (infinite-level) order book characterization process $[Q_t(t), q(t), p_{\text{ref}}(t)]$ (in Figure 1) is adapted to the filtration $\mathcal{F}_t$. Then the intensities of event i for a plain-vanilla multivariate Hawkes process model for LOB data can be represented as:

$$\lambda_i(t) = \eta_i + \sum_{j=1}^{6K+6} \int \phi_{ji}(t-s) dX_j(s), \quad \forall i = 1, 2, \dots, 6K + 6. \quad (1)$$

in which η_i represents the baseline intensity, $\phi_{ji}(\cdot)$ represents the Hawkes kernel for event j stimulating event i .

3.3.2 Markovian part

To further extend the plain-vanilla multivariate Hawkes process, we integrate two factors derived from LOB queues to the baseline intensity (intercept) η_i of the Hawkes process:

- **The current LOB liquidity state:** The liquidity state includes the number of existing orders on the associated price of the event, the number of orders on the best bid/ask price, and the bid-ask spread.
- **Time clustering:** Figure 2 indicates that the trading frequency for real LOB data differs between time groups throughout trading hours. The last 30 minutes have significantly more activities, and the first 30 minutes see slightly higher activities, while other time periods during the day seem to be tranquil.

Therefore, we model the baseline event arrival intensities η_i with a Markovian structure that depends on the current liquidity state and time clustering:

$$\eta_i(t) = M_i(l_i(t)) + \Theta_i(t) \tag{2}$$

where $M_i(\cdot)$ and $\Theta_i(\cdot)$ denote specified functions on liquidity state $l_i(t)$ and time t .

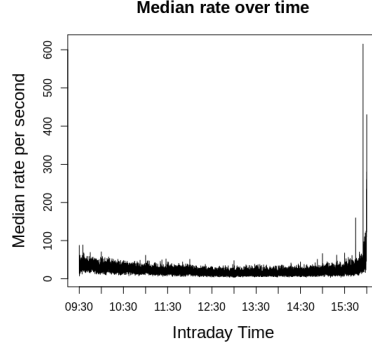


Figure 2: An illustration of the order arrival rate for LOB data. The figure demonstrates median order arrival rates per second for 20 trading days from 1/2/2019 to 1/31/2019 based on LOB data for Apple.Inc.

Next, we shall define the liquidity state $l_i(t)$. Note $l_i(t)$ only changes if an event arrives.

For all events that do not change the reference price (i.e., $-K(i)$, $-K(c)$, $-K(t)$, $\dots$, $+K(i)$, $+K(c)$, $+K(t)$), the liquidity state $l_i(t)$ is intuitively $q_{k_i}(t)$, the accumulated order size right before time t on the k_i -th price level where event i belongs. For example, events $+1(i), +1(c), +1(t)$ have $k_i = 1$ and $l_i(t) = q_1(t)$.

Then we consider the events changing the reference price. For events $p-(t)$ and $p-(c)$, $l_i(t)$ is defined as the queue size at the best bid price. For events $p+(t)$ and $p+(c)$, $l_i(t)$ is defined as the queue size at best ask price; this is because the queue at the best bid/ask must be either consumed or canceled for these event types. For events $p+(i)$ and $p-(i)$, we consider order insertions within the bid-ask spread, and therefore the bid-ask spread is considered as the liquidity state $l_i(t)$ for these two types of events. $p+(i)$ and $p-(i)$ are the only two types of events whose liquidity state is based on price (bid-ask spread) while the liquidity state of the rest of the events is based on existing order size at the associated price.

Formally, we have the mathematical definition:

$$l_i(t) = \begin{cases} q_{k_i}(t), & i \in \{-K(i), -K(c), \dots, +K(i), +K(c), +K(t)\} \text{ (size of the corresponding queue)} \\ q_{\text{best-bid}}(t), & i \in \{p-(t), p-(c)\} \text{ (size of the best-bid queue)} \\ q_{\text{best-ask}}(t), & i \in \{p+(t), p+(c)\} \text{ (size of the best-ask queue)} \\ Q_{\text{best-ask}}(t) - Q_{\text{best-bid}}(t), & i \in \{p+(i), p-(i)\} \text{ (bid-ask spread)} \end{cases}$$

where k_i denote the queue to which event i belongs to.

3.3.3 final form of Hawkes+Markovian combined model

Our final model on the instantaneous rate $\lambda_i(t)$ is a combination of Hawkes part Eq.(1) and Markovian part Eq.(2):

$$\lambda_i(t) = M_i(l_i(t)) + \Theta_i(t) + \sum_{j=1}^{6K+6} \int \phi_{j,i}(t-s) dX_j(s), \quad \forall i, j = 1, 2, \dots, 6K+6 \quad (3)$$

4 Estimation procedure

Given the above hybrid Hawkes+Markovian model, we employ a non-parametric regression-based approach to estimate the model parameters. Our approach is extended from Kirchner (2017) as well as Bacry & Muzy (2014a).

Inspired by Kirchner (2017) and Bacry & Muzy (2014a), we approximate the whole intensity function as a standard vector-valued linear autoregressive time series. The estimation procedure discretizes a continuous point process into multiple fixed-size bins on the time domain and thereby fits a vector autoregression model to the discretized samples.

To outline the intensity function estimation, we first define discretization bin-size Δ and maximum support s . The maximum support represents the maximum duration (in seconds) during which the arrival of one event can have stimulating effects on the arrival of

other events. The bin-size Δ defines the short period of time during which the Hawkes self- and cross-exciting function stay unchanged. Take $s = 20$ seconds, $\Delta = 0.25$ seconds as an example: this setting indicates the maximum duration that an event i can stimulate the intensities of future event arrivals is 20 seconds, while event i 's self- and cross-stimulating functions stay the same within each 0.25 seconds short period. That is to say, the 80 of 0.25-second windows combined define the 20-second stimulating horizon for event i and its stimulating function shrinks to 0 beyond 20 seconds after event i 's arrival.

Given an appropriate choice of bin-size Δ and maximum support s , a continuous Hawkes excitement function between two events can be approximated by a piece-wise constant function of $p = \lfloor s/\Delta \rfloor$ steps, with each step standing for the constant function value over the short period Δ . To complete the estimation, we need to apply some smoothing methods over the estimated point-wise function. We adopt the cubic smoothing spline in our method. The cubic smoothing spline is a smoothing technique such that the curve spanning each data interval is represented by a cubic polynomial. The cubic smoothing spline is achieved by minimizing the curvature of the smoothed function and is not required to pass all data points. The overall illustration of the estimation method is demonstrated in 3.

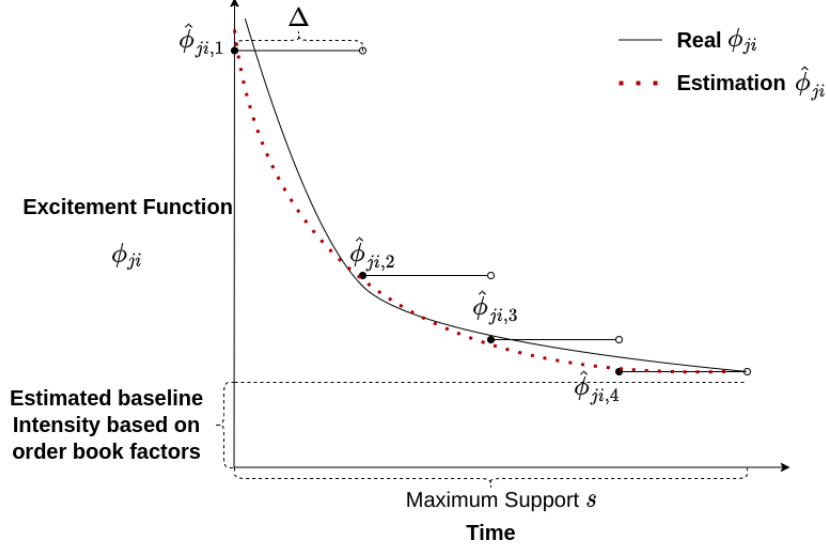


Figure 3: An example illustration on Hawkes excitement function estimation. The example demonstrates the estimation for a continuous function ϕ_{ji} . Suppose $s = 4\Delta$ and therefore $p = 4$. The estimation idea is that ϕ_{ji} can be approximated by $\hat{\phi}_{ji} = (\hat{\phi}_{ji,1}, \hat{\phi}_{ji,2}, \hat{\phi}_{ji,3}, \hat{\phi}_{ji,4})$. $\hat{\phi}_{ji,1}$ represents the fixed function value over $[0, \Delta]$, $\hat{\phi}_{ji,2}$ represents the fixed function value over $[\Delta, 2\Delta]$Though the main objective is to obtain the discrete estimators $\hat{\phi}_{ji}$, we can further fit a cubic smoothing spline to the discrete estimators to derive a continuous estimated function as shown by the red dotted line.

4.1 bin construction

To obtain our proposed discretized estimators, the data needs to be discretized under the (s, Δ) -framework as the first step. The estimation requires allocating LOB event arrival and LOB state sequences into fixed-size bins with length Δ over the time horizon and counting the number of realizations in each bin. Afterward, these bin-count sequences can be used as sufficient statistics to obtain model estimates.

Recall in section 3.3, we denote the order book event process at time t as $X_i(t)$, where i

is index for events. For LOB data, each types of event outlined in section 3.1 and 3.2 comes with an order size indicating the quantity of shares the LOB event attempts to execute. For example, a $+1(i)$ event with order size 100 indicates the event at the first ask queue to insert 100 shares.

Since all regular order book events $-K(i), -K(c), -K(t), \dots, +K(i), +K(c), +K(t)$ are additive, big events can be considered as the sum of the same events with smaller sizes in a short period of time. As an intuitive example, one insertion order of size 100 shares is assumed to be equivalent to two insertion orders of size 50 shares happening at the same time and the same price. However, the events $p-(t), p-(c), p-(i), p+(t), p+(c), p+(i)$ are not additive because the reference price change caused by these event shift the LOB queue distribution. Specifically, the queue size value q_k can shift to its neighbours when the reference price (center of LOB distribution) goes up and down. Therefore, the events causing reference price changes cannot be thought simply as the sum of same event with smaller size. Overall, the size of LOB event is modeled in a dichotomous approach for X_i : for all regular additive events, X_i increases by its order size at event arrival; for all non-additive events, X_i by 1 at event arrival, treating as pure point processes.

We can then construct a series of fixed-size bins over the total time horizon $(0, T]$, and sum the realizations of X_i during each bin to obtain the bin-count sequences, as illustrated in Figure 4. Formally, for $t \in (0, T]$, for some bin-size $\Delta > 0$, we construct the $(6K + 6)$ -dimensional bin-count, liquidity state, and time factor sequences as:

$$\begin{aligned} B_k^{(\Delta)} &= \left(B_{i,k}^{(\Delta)} \right)_{i=1}^{6K+6}, \quad B_{i,k}^{(\Delta)} := X_{(i)}((k-1)\Delta, k\Delta], \\ l_k^{(\Delta)} &= \left(l_{i,k}^{(\Delta)} \right)_{i=1}^{6K+6}, \quad l_{i,k}^{(\Delta)} := l_i((k-1)\Delta), \\ t_k^{(\Delta)} &= \left(t_{i,k}^{(\Delta)} \right)_{i=1}^{6K+6}, \quad t_{i,k}^{(\Delta)} := (k-1)\Delta, \end{aligned}$$

where $i = 1, 2, \dots, 6K + 6$, $k = 1, 2, \dots, n$, and $n := \lfloor T/\Delta \rfloor$.

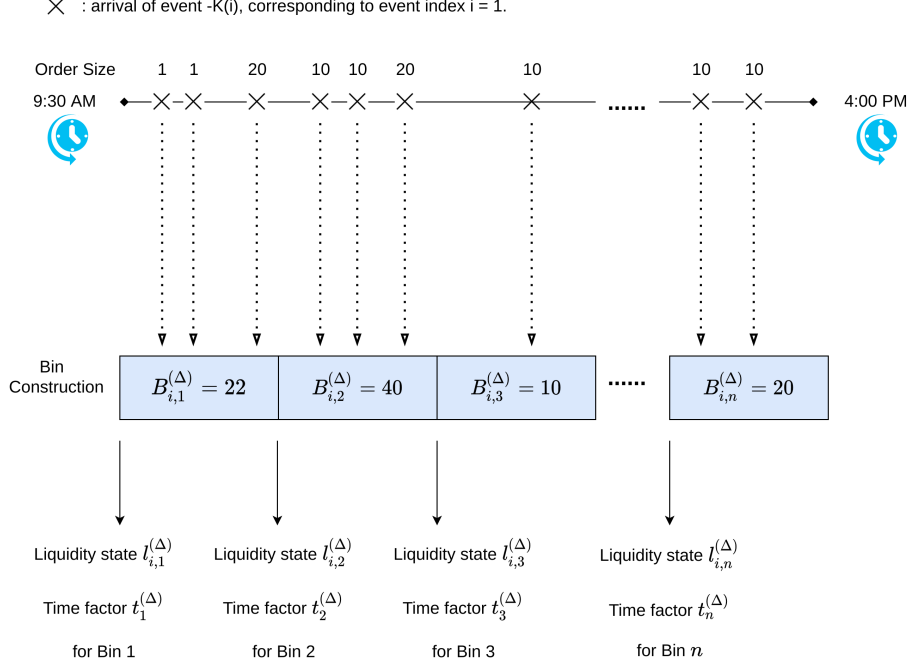


Figure 4: An illustration example on bin-count sequence construction for event $-K(i)$. Order size at the cross mark are allocated into bins $1, 2, \dots, n$ based on event arrival time.

4.2 bucketize LOB state factors

As mentioned in section 3.3, we assume that the baseline intensity function of the proposed model is controlled by functions based on LOB liquidity state $l_i(t)$ and intraday time t . In the previous section, we have discussed discretizing and transforming sequences $l_i(t)$ and t into bin-based sequences $l_{i,k}^{(\Delta)}$ and $t_{i,k}^{(\Delta)}$ for a selected bin-size Δ . Consequently, we have indeed transformed $M_i(l_i(t))$ and $\Theta_i(t)$ into $M_i(l_{i,k}^{(\Delta)})$ and $\Theta_i(t_{i,k}^{(\Delta)})$ for estimation purpose.

In order to approximate the non-parametric function $M_i(\cdot)$ and $\Theta_i(\cdot)$, we use the same discretization idea and treat $l_{i,k}^{(\Delta)}$ and $t_{i,k}^{(\Delta)}$ as categorical variables to obtain the linear parameters. Therefore, $M_i(\cdot)$ and $\Theta_i(\cdot)$ become step function that depends only on the specified categories of the discretized liquidity state and time factor. We use the following

method for bucketization:

- Liquidity state $l_{i,k}^{(\Delta)}$: we bucketize $l_{i,k}^{(\Delta)}$ into 10 categories $\{L_1, L_2, \dots, L_{10}\}$ ¹ with $L_1 = [0, 100), L_2 = [100, 200), \dots, L_{10} = [900, +\infty)$, which represents the number of existing orders on the corresponding price queue (except for events $\mathbf{p}+(\mathbf{i})$, $\mathbf{p}-(\mathbf{i})$) right before event arrives. For example, $l_{i,k}^{(\Delta)} \in L_1$ means that the number of orders on the queue is between 0 and 100 shares. For the special case of events $\mathbf{p}+(\mathbf{i})$, $\mathbf{p}-(\mathbf{i})$, the $l_{i,k}^{(\Delta)}$ represents the bid-ask spread in market price (section 3.3), and thus the group $l_{i,k}^{(\Delta)} \in L_1$ means that the bid-ask spread is between \$0 and \$0.01.
- Time factor $t_{i,k}^{(\Delta)}$: major U.S. electronic stock exchanges trades from 9:30 am ET to 4:00 pm ET. As illustrated in Figure 2, the LOB event arrival frequency exhibits time clustering effects at the beginning and ending intraday 30-minutes window (9:30 - 10:00 am ET and 3:30 - 4:00 pm ET). Therefore, we construct 1-minute categories at the beginning and ending 30-minutes to better capture the event arrival volatility. For the rest of the period between 10:00 am ET to 3:30 pm ET, we bucketize time $t_k^{(\Delta)}$ into 5-minute categories since the event arrivals are more tranquil. Consequently, we construct 126 time categories consisting of 60 1-minute categories and 66 5-minute categories, denoted as $\{T_1, T_2, \dots, T_{126}\}$.

4.3 non-parametric estimation

After defining the bin-count sequence and the LOB state/time categories, we fit a vector-valued autoregression model to the bin-count sequences to obtain the Hawkes excitement functions, together with the Markovian liquidity state and time factor parameters. For

¹These categories apply to all event types $\forall i = 1, \dots, 6K + 6$, and therefore the i subscript is eliminated from the $\{L_1, L_2, \dots, L_{10}\}$ notation.

each event-to-event pair, the auto-regression model is implemented with lag $p := \lfloor s/\Delta \rfloor$ on the bin-count sequence data $B_k^{(\Delta)}$ as well as the Markovian liquidity state and time factor sequence $l_k^{(\Delta)}$ and $t_k^{(\Delta)}$ across all event types.

Kirchner (2017) has provided a detailed framework for autoregressive non-parametric estimation for Hawkes process with constant baseline intensity. Our estimation is an extension of Kirchner's methods with modifications that the baseline intensity is varying and controlled by $l_i(t)$ and t . An illustration of the estimation procedure is given in Figure 5.

The mathematical representation details on the estimation procedure is given in Supplementary Material section B. As shown Eq.(4) in the Supplementary Material section B, $\hat{\Phi}^{(\Delta,s)} \in \mathbb{R}^{(6K+6) \times ((6K+6)p+10+126)}$ represents all the parameters of our model for a choice of bin-size Δ and maximum support s . To summarize, the estimation procedure for our proposed model estimates: A total of $(6K+6) \times (6K+6)$ Hawkes excitement functions $\phi(\cdot)$, each as a step function with $p = \lfloor s/\Delta \rfloor$ distinct levels (See Figure 3); A total of $6K+6$ liquidity state functions $M(\cdot)$, each as a step function with 10 distinct levels on the 10 liquidity state baskets; A total of $6K+6$ time factor functions $\Theta(\cdot)$, each as a step function with 126 distinct levels on the 126 time factor baskets.

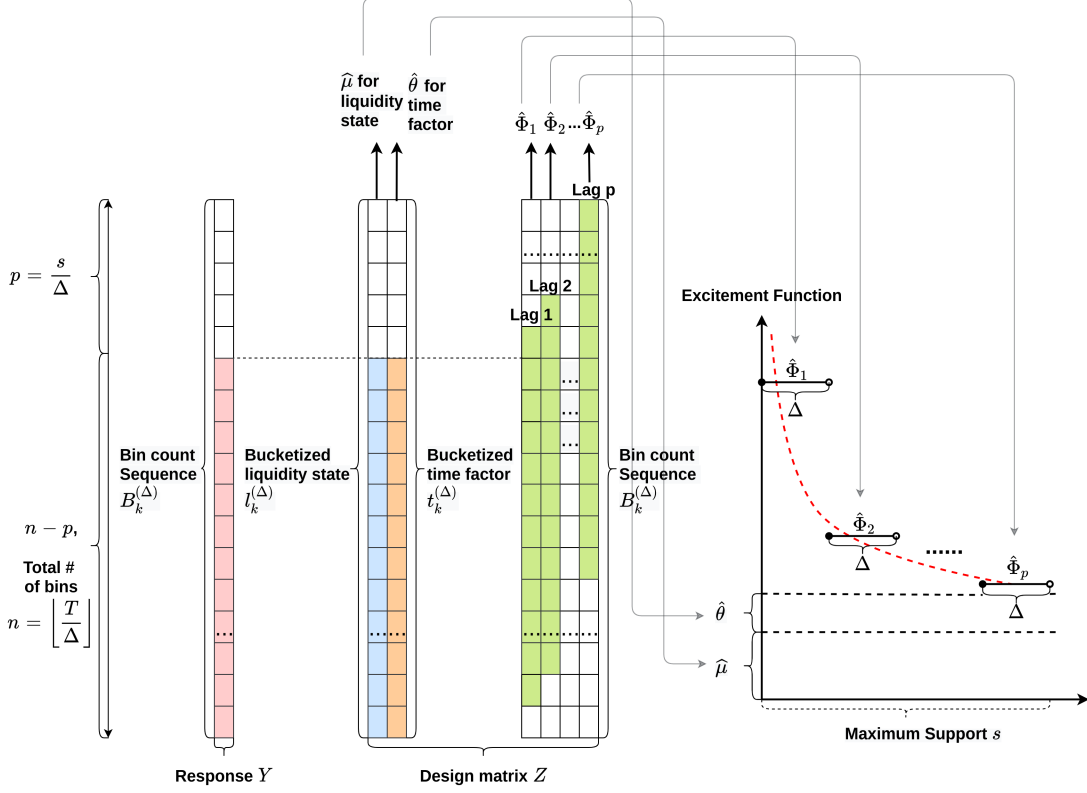


Figure 5: An illustration on the nonparametric estimation procedure. Over a time horizon $(0, T]$, given a bin-size Δ and a maximum support s , the parameters of our proposed model can be estimated using an autoregressive framework with lag $p = s/\Delta$. There are totally $n = \lfloor T/\Delta \rfloor$ bins. The autoregression response Y ranges from the $(p + 1)$ -th bin to the n -th bin from the bin-count sequence. The design matrix Z consists of two parts: the first part contains the bucketized liquidity state and time factor sequences controlling the regression intercept; the second part contains lag- p sequences constructed from the bin-count sequence. The estimated parameters $\hat{\mu}$ and $\hat{\theta}$ for liquidity state and time factor control the baseline intensity of the estimated excitement functions; Parameters $(\hat{\Phi}_1, \dots, \hat{\Phi}_p)$ are the estimates for the constant excitement function value during each short period of bin-size Δ .

4.4 LASSO regularization

Many previous works have discussed the exponential decaying shape of the LOB excitement functions (Muni Toke & Pomponio 2012, Abergel & Jedidi 2015, Morariu-Patrichi & Pakkanen 2018, Kirchner 2017). Therefore, We believe some of the estimators for the excitement function are likely to be zero, especially at the tail part when the support s is large and the bin-size Δ is small. As discussed in Remark 1 from Supplementary Material section B and section 3.3, for a single event type the OLS estimation procedure outputs a large $(6K + 6) \times p$ (recall $p = \lfloor s/\Delta \rfloor$) number of estimators for the Hawkes stimulating functions $\phi(\cdot)$ for a given combination of (s, Δ) . To fit linear models with such a large design matrix in a more robust way, we consider adding LASSO regularization to the non-parametric estimation of excitement function to shrink the estimate to zero.

On the other hand, we don't regularize any LOB liquidity state and time factor parameters $(\hat{\mu}, \hat{\theta})$ since we believe most of the redundant parameters tend to appear in the Hawkes excitement function part, while the liquidity state and time factor should have smooth and non-zero effect. The mathematical representation details of the LASSO is given in Definition 3 in Supplementary Material section C.

4.5 model selection using AIC (Akaike Information Criterion)

AIC is among the common approaches for linear model selection and diagnosis. Kirchner (2017) uses multivariate AIC to determine the ideal model fit using both simulated and real LOB data. In our proposed methodology, AIC can be used to evaluate the effects of different parts of the proposed model in section 3.3.3.

Given a selected bin size $\Delta = \Delta_0$, the estimation lag equals $p = \lfloor s/\Delta_0 \rfloor$ for any maximum support $s > 0$. Similarly, we denote the total number of the bins as $n_0 := \lfloor T/\Delta_0 \rfloor$. AIC calculation involves the total number of effective parameters (model degree-

of-freedom) in its formula. When the estimated parameters are not effective (exactly zero, or not estimable), they cannot be counted as degree-of-freedom. For example, a number of redundant estimators shrink to zero under LASSO regularization. Also, when there is no liquidity state observation in some of the bucketized categories among $(L_1, L_2, \dots, L_{10})$, the parameter $\hat{\mu}^{(\Delta, s)}$ will not be estimable. Consequently, these parameters are not part of the AIC calculation.

Denote the $(6K + 6) \times 1$ regression residual vector as $\hat{\mathbf{u}}_k^{(\Delta_0, p)} = (\hat{u}_{1,k}, \dots, \hat{u}_{6K+6,k})^\top$ for $\forall k = (p + 1), \dots, n_0$. The multivariate AIC is given as:

$$AIC^{(\Delta_0)}(p) = \log \left(\det \hat{\Sigma}^{(\Delta_0)}(p) \right) + \frac{2 \times d_e}{(n_0 - p)}, \quad \hat{\Sigma}^{(\Delta_0)}(p) = \sum_{k=p+1}^{n_0} (\hat{\mathbf{u}}_k^{(\Delta_0, p)})(\hat{\mathbf{u}}_k^{(\Delta_0, p)})^\top / (n_0 - p),$$

where d_e is the number of effective parameters involved in the estimation.

5 Implementation Example

To evaluate the validity of the proposed estimation procedure mentioned in the previous section, we first implement the estimation method in section 3 and section 4 on a fully-simulated level-one order book dataset. We document in the Supplementary Material section D that our proposed method can accurately recover the true Hawkes kernels and baseline intensity state variables. The purpose of the simulation is to ensure the validity of the method and thus we can apply it to real data with confidence.

We next apply our model to real order book data obtained from Lobster Data (<http://lobsterdata.com>). Order book data of Apple.Inc from 9:30 AM ET to 4:00 PM ET from 01/02/2019 to 01/31/2019 was used in our implementation. The data gives order book shapes and messages in milliseconds. We choose to model the level-3 order book, i.e., setting $K = 3$, so that there are $6K + 6 = 24$ events considered. They are:

$$\begin{pmatrix} \underbrace{(-3(i), -3(c), -3(t), -2(i), -2(c), -2(t), -1(i), -1(c), -1(t))^\top}_{9} \\ \underbrace{(+1(i), +1(c), +1(t), +2(i), +2(c), +2(t), +3(i), +3(c), +3(t))^\top}_{9} \\ \underbrace{(p-(t), p-(c), p-(i), p+(t), p+(c), p+(i))^\top}_{6} \end{pmatrix}_{24 \times 1}$$

The order book estimation for each single day can be considered as an independent realization of event arrivals. Since the support and bin size values are fixed throughout the days in consideration, we can obtain the unbiased estimator of the estimation parameters by taking the mean of each element from single-day parameters. The aggregated results across many days can smooth out single-day extreme values and therefore is a clear representation of estimation results over a certain period.

Suppose we have N days of order book data for a public company and the estimator $\bar{\Phi}_n^{(\Delta, s)}$ for each day with $n = 1, 2, \dots, N$, the aggregated estimator can be represented as:

$$\bar{\Phi}_{1:N}^{(\Delta, s)} := \frac{\sum_{n=1}^N \hat{\Phi}_n^{(\Delta, s)}}{N}$$

In addition, a small LASSO regularization parameter $\lambda_i = 0.0005$ has been implemented for our estimations, as a large parameter may drastically alter the estimation results. [We also apply cubic smoothing splines to the discretized points from the estimation for visualization purpose, in which the degree of freedom of the splines is chosen by leave-one-out cross-validation.](#) The main findings and all the observed features demonstrated in the rest of section 5 are based on 20-day aggregation method mentioned above.

5.1 main findings

Through our non-parametric estimation over 20-day LOB data of Apple.Inc, the main findings include:

- Most of the estimated Hawkes excitement functions exhibit [convex monotonic decreasing shapes](#) with trailing zeros, similar to the exponential and power-law kernels. However, there are some exceptions and most of them appear on market (trade) orders on higher levels (the 2nd and 3rd best bid/ask price) of the LOB queue.
- The estimated Hawkes excitement functions are similar with respect to insertion/deletion events on the 1st level of LOB queue (the 1st best bid/ask price).
- For most of the LOB events that do not change the price (reference price, which is the center of the order book), the event arrival intensities increase with the current order size in the order book queue.
- The arrival intensities for almost all events elevate during the beginning and ending 30-minutes window of daily trading hours. The intensity increases drastically before market close between 15:55 to 16:00 pm ET.
- Our qualitative result is not sensitive to the discretization size in time and the handling of the size of the order.
- Through model selection analysis using AIC (Akaike Information Criterion), the inclusion of the Hawkes excitement functions, the LOB liquidity state, and LOB time factor to the event arrival intensity all contributes to the improved model fitting, with the Hawkes excitement functions achieving the strongest improvement. Besides, reducing the bin-size (short discretization period of time during which the Hawkes excitement function are assumed to stay unchanged) and adding LASSO regularization to our estimation both contribute to better model fitting.

The details of the observations listed above will be discussed in the rest of this section.

5.2 estimated excitement functions

Figure 6 illustrates the estimated excitement functions of insertion event at 1st ask (event $+1(i)$) stimulating insertion and cancellation at the 1st ask (event $+1(i)$ and $+1(c)$).

We observe the excitement function to have a time-decaying shape in general: the stimulation is highest in a very short time following the event arrival, and then gradually decaying to zero.

We also observe similarities between Fig.6(a) and Fig.6(b): the figures both exhibit decaying shapes with slight spikes around 5 and 13 seconds after event $+1(i)$ arrives. This observation suggests that the estimated excitement functions are very similar for the effect towards the insertion and cancellation at the 1st ask (i.e., effect towards $+1(i)$ and $+1(c)$). We have similar observation for 1st bid as well demonstrated in Supplementary Material section E, suggesting such stimulation behavior exists at the 1st level of the LOB queue.

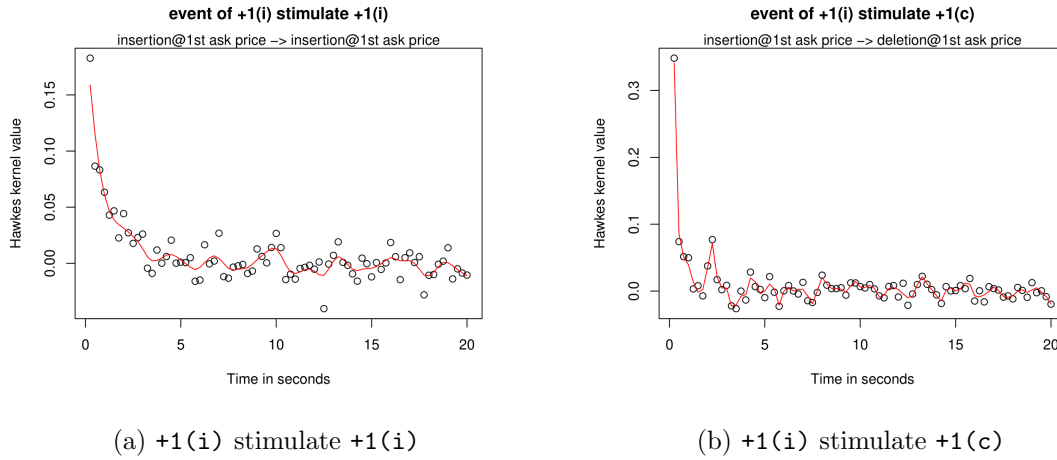


Figure 6: Aggregated Hawkes excitement function estimations under ($s = 20$ seconds, $\Delta = 0.25$ seconds) with LASSO regularization. The points illustrates the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

Further examination shows that this feature (similarity at the 1st ask/bid) becomes less obvious for the 2nd and 3rd levels of the LOB queue. One possible explanation of this feature is that market participants may use a sequence of insertion/deletion orders in a short period of time to drive up/down the price for specific purposes. Insertion/deletion on the 1st level is typically less risky than insertion/deletion on any higher levels especially when the short-term price movement is unpredictable.

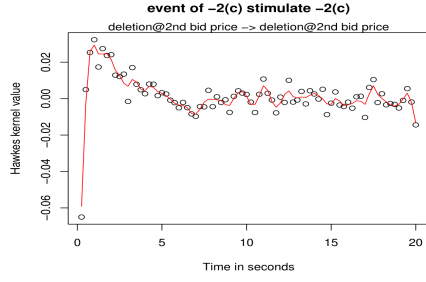
5.3 shapes of excitement functions

From the aggregated estimation results we observe that a large proportion of Hawkes excitement functions exhibit [convex monotonic decreasing shapes](#) with their function values converging to zero as time elapses on the x-axis ([similar to exponential and power-law kernels](#)). For example, we can see the excitement functions demonstrated in Figure 6 exhibit this feature. This observation also agrees with the existing literature that exponential-decaying shape features are commonly observed on LOB data and Hawkes process with exponential kernels is among the most popular models adopted by recent works (Muni Toke & Pomponio 2012, Abergel & Jedidi 2015, Morariu-Patrichi & Pakkanen 2018, Kirchner 2017).

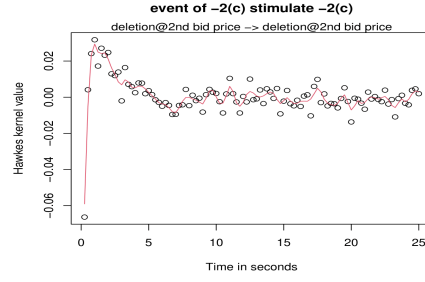
However, we have also observed that some Hawkes excitement functions exhibit other shapes. Figure 7 shows that the excitement functions for specific events can deviate from the [monotonic decreasing shape](#). These function shapes appear more frequently for market(trade) order events on higher levels of the LOB queue when they are the “stimulatee” part of the function, such as events $-3(t)$, $-2(c)$, and $+3(t)$. These “high-level” trade orders typically arrive when the bid-ask spread is large. One possible reason for this non-decaying feature is that, for a heavily traded stock like Apple.Inc, some high-frequency algorithmic market participants may use more complex execution and risk management strategies to deal with the increased risk brought by the enlarged bid-ask spread. There-

fore, our estimated Hawkes function can generate complicated shapes, possibly because of this layer of complexity. These non-decaying shape functions indicate the advantages of our non-parametric method since it might be problematic to adopt exponential or power law shape assumption for excitement functions throughout all events.

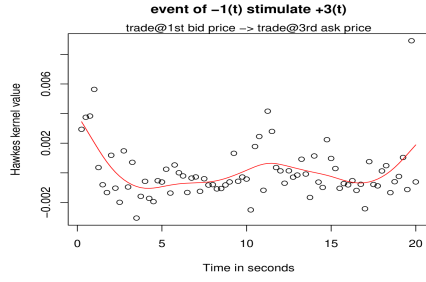
In Figure 7, we also present estimations under a longer maximum support $s = 25s$ to illustrate the stability of the estimation. We observe the excitement functions will move to zero as the maximum support increases.



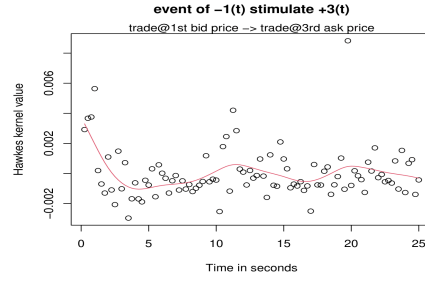
(a) $s = 20s$



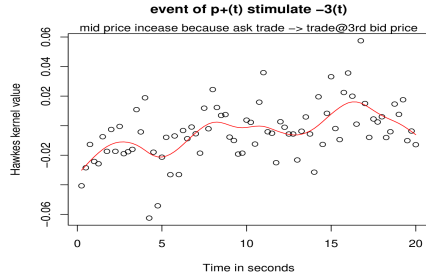
(b) $s = 25s$



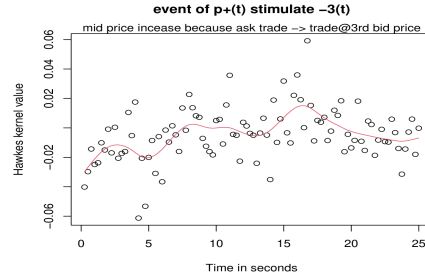
(c) $s = 20s$



(d) $s = 25s$



(e) $s = 20s$



(f) $s = 25s$

Figure 7: Aggregated Hawkes excitement function estimation under ($\Delta = 0.25$ seconds) with LASSO regularization. The subplots (a), (c), and (d) use maximum support $s = 20$ seconds. The subplots (b),(d), and (f) use maximum support $s = 25$ seconds. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

5.4 liquidity state

For liquidity state results, we have observed that for event types that do not change the reference price, the arrival intensities increase as the liquidity state increases for trade/cancellation events, as well as insertion events on the 1st level (i.e., $-3(c)$, $-3(t)$, $-2(c)$, $-2(t)$, $-1(i)$, $-1(c)$, $-1(t)$, $+1(i)$, $+1(c)$, $+1(t)$, $+2(c)$, $+2(t)$, $+3(c)$, $+3(t)$). An example is shown in Figure 8(a). This is consistent with the intuition that more trade/cancellation events are likely to happen when the number of existing orders on the corresponding queue is large, because the trade/cancellation are actions on existing orders. It also suggests that more insertion events at 1st level LOB are likely to happen when the number of existing orders on the 1st level is large, which could be potentially explained by the popularity of the stock – when the stock is popular, naturally the number of existing orders at 1st level LOB is high, yet market participants are willing to insert more orders at 1st level LOB, and vice versa.

On the other hand, the arrival intensities generally decrease as the liquidity state increases for the insertion events at 2nd and 3rd level (i.e., $+3(i)$, $-3(i)$, $+2(i)$, $-2(i)$). An example is demonstrated in Figure 8(b). The observation implies that market participants are less likely to insert orders on the 2nd and 3rd level of LOB when the number of existing orders is large on these levels. This behaviour is possibly due to the increased risk that the orders on the 2nd and 3rd level may fail to be filled as orders accumulate in these queues, and maybe the market participants focus more on the 1st level when such situation occurs.

Up to now we set the number of liquidity state to estimate as 10. Alternatively, we perform a sensitivity analysis by setting the number of liquidity state to be 5 (presented in Supplementary Material Figure S7 and S8). Not surprisingly, the observed pattern discussed above maintains even though the estimation with fewer liquidity state smooths out some across group variations among the liquidity state.

More examples demonstrating the above patterns and the sensitivity analysis are presented in Supplementary Material section F.

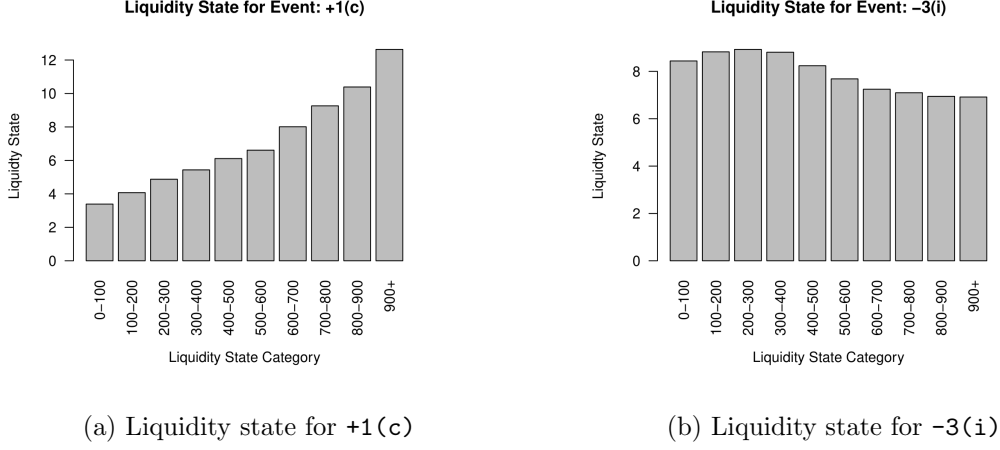


Figure 8: Aggregated estimation result for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds). In general, the arrival intensity increases/decreases as the liquidity state increases for event +1(c) and -3(i), respectively.

5.5 time factor

For estimation results of time factor, the order arrival intensities tend to be larger at the beginning and ending of the trading hours between 9:30 am and 4:00 pm (see Figure 9). This feature matches our initial arrival rate estimation in Figure 2 and reinforces our model assumption that more time factor categories should be constructed in the beginning and ending 30 minutes.

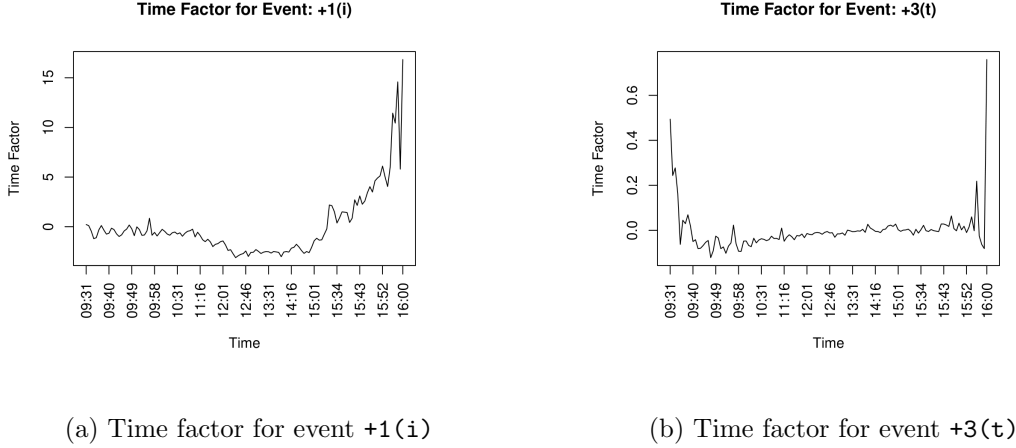


Figure 9: Aggregated estimation result for time factor between 9:30 am and 4:00 pm under ($s = 20$ seconds, $\Delta = 0.25$ seconds).

Figure 9 demonstrates the time factor pattern mentioned above for event $+1(i)$ and event $+3(t)$. From the estimation results we have also observed a significant intensity increase for many order types such as $p-(i)$, $p-(t)$, and $-1(t)$ at 15:55 pm and 15:59 am. These patterns possibly stem from large algorithmic trader's execution rules or the policy of stock exchange on last-minute order submission/cancellations.

5.6 kernel norms

This section presents the result of kernel norms $\|\phi_{j,i}\| = \int_0^\infty \phi_{j,i}(t)dt$ in Figure 10, which can be interpreted as the average impact (number) of event i stimulated by event j . All of the kernel norms are strictly less than one, implying the stability of the Hawkes process component (Bacry et al. 2016, 2013b). It is essential to point out that some of the kernel norms are negative, indicating the arrival of the stimulator decreases the arrival intensity of the stimulatee. These negative values imply the inhibition dynamics and are natural as

as result of non-parametric estimation, which has been discussed by Wu et al. (2019) and Bacry et al. (2016). Specifically, Wu et al. (2019) points out that any prior assumptions forcing the Hawkes kernel to be positive can lead to significantly biased estimations since the prior may distort the interacted stimulation and inhibition effects within order dynamics. Bacry et al. (2016) also shows that negative kernels will not affect the reliability of the estimation procedure as long as the realized intensity in Eq.(3) remains positive, which we have verified empirically.

Several patterns could be observed from the following heat map of the kernel norms. First, it is witnessed that events $(p-(i), p-(c), p-(t))$ that moves the reference price down overall have negative impacts on the arrival of limit order insertions above the reference price $(+1(i), +2(i))$. This pattern intuitively makes sense as selling limit order insertions may incur extra costs when the efficient price goes down, compared to the order book status before the reference price change. In symmetry, it is also observed that events $(p+(i), p+(c), p+(t))$ driving the reference price up in general have negative impacts on limit order insertions on the bid side (i.e. event $-1(i)$). Besides, we observe large positive impacts of high-level trade orders (event $+3(t)$ and $-3(t)$) on lower-level limit order insertions (event $+1(i)$ and $-1(i)$). This can be interpreted as that high-level orders may serves as a signal of liquidity at a certain timestamp and therefore more market participants insert orders at better prices in response to this signal.

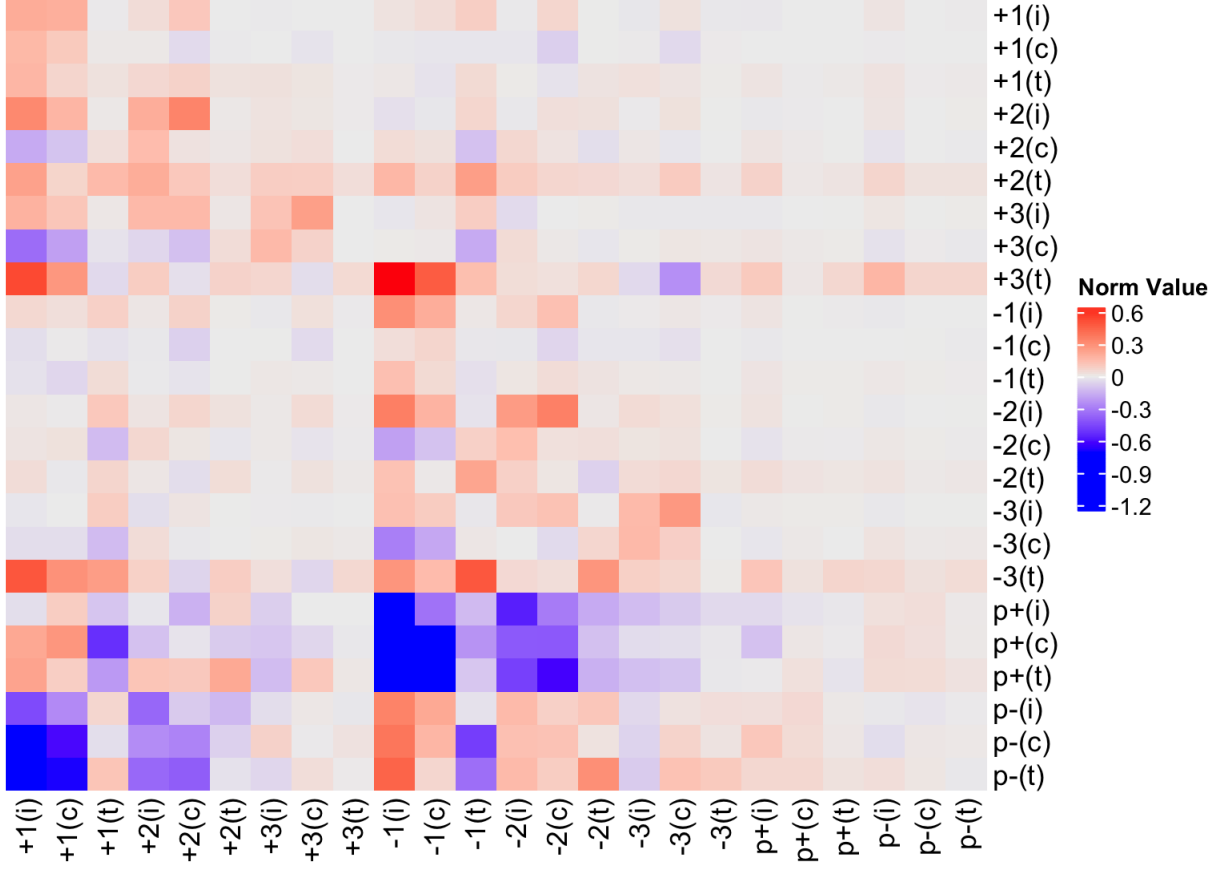


Figure 10: Kernel norms of the estimation. The heatmap illustrates the estimated kernel norm $||\phi_{j,i}|| = \int_0^\infty \phi_{j,i}(t)dt$ of the estimated model with maximum support $s = 20$ seconds and bin-size $\Delta = 0.25$ seconds. All orders are considered to have size one and a LASSO regularization parameter $\lambda = 0.0005$ is used in the estimation. The x-axis represents the stimulatee and the y-axis represents the stimulator. The color transition from blue to red indicates the Hawkes norm value transition from negative to positive, which has been similarly illustrated in Wu et al. (2019).

5.7 sensitivity analysis

So far our results have been presented using the bin count construction method based on order sizes, a fixed bin-size $\Delta = 0.25$ seconds, maximum support $s = 20s$, and the LASSO loss function. This section provides a sensitivity analysis of these model assumptions to explore estimation robustness. The analysis consists of the following parts:

- Estimation results when the size of order is ignored in bin count construction, assuming all orders to have size 1 in the estimation. (see Supplementary Material section G)
- Estimation results when the LASSO regularization is removed. (see Supplementary Material section H)
- Estimation results when the bin-size is increased from 0.25 seconds to 0.5 seconds. (see Supplementary Material section I)
- Estimation results when maximum support is extended from 20 seconds to 25 seconds. (see Supplementary Material section J)
- The number of liquidity state is reduced from 10 to 5. (see Supplementary Material section F)

Generally speaking, our estimation is robust for the above sensitivity tests. We have obtained qualitatively similar results as our original setting when we remove the LASSO regularization, enlarge the bin-size, [extend the maximum support](#), and [change the number of liquidity states](#). After ignoring the order size, the estimation scales tend to decrease and the excitement function becomes smoother and less volatile. However, the observations given from section 5.2 to section 5.5 still hold.

5.8 model selection result

In this section, we validate the added explanation power for different model parts using the AIC model selection method illustrated in section 4.5. We consider the following model specifications for model selection: **Model ①**: liquidity state only; **Model ②**: time factor only; **Model ③**: liquidity state + time factor; **Model ④**: Hawkes only; **Model ⑤**: Hawkes + LASSO (LASSO parameter 0.0005); **Model ⑥**: liquidity state + time factor + Hawkes; **Model ⑦**: Liquidity state + time factor + Hawkes + LASSO (LASSO parameter 0.005). **Model ⑦** is the model we mainly proposed and discussed in previous sections (section 3.3.3 and section 4). The "+LASSO" notation is used to illustrate that LASSO regularization is used in model estimations as discussed in section 4.4. Figure 11 demonstrate AICs for the 7 models mentioned above for Apple.Inc on 2019-01-03.

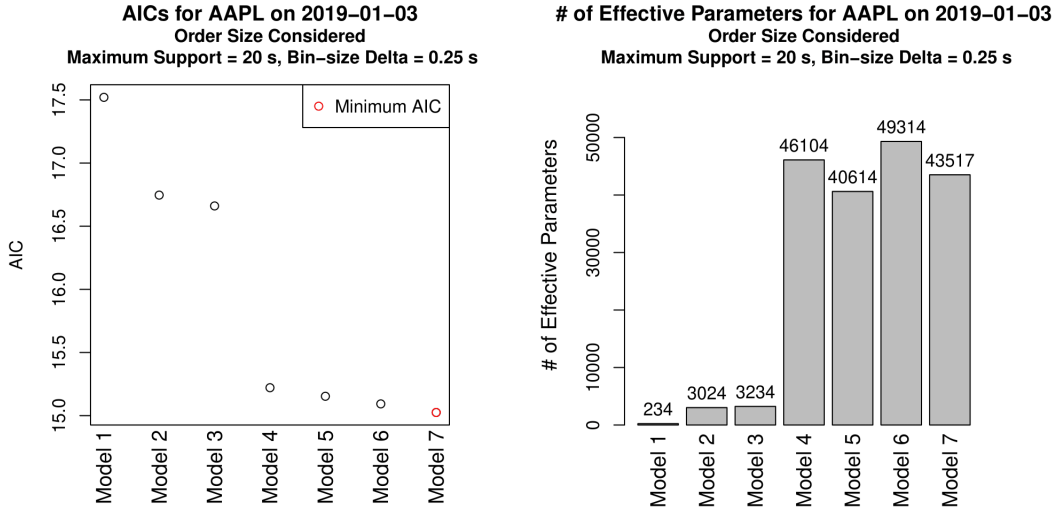


Figure 11: AICs and number of effective parameters for seven model types for Apple.Inc on 2019-01-03 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.25$ seconds. Order size is considered to construct bin count sequence.

We conclude that for Apple.Inc on 2019-01-03, **Model ①/②/③** generally have much

higher AICs with much smaller number of effective parameters than the rest of the models.

As expected, **Model ④** outperforms **Model ①/②/③**, indicating the Hawkes excitement functions have stronger explanation power than the combination of liquidity state and time variables. Furthermore, the LASSO models (**Model ⑤/⑦**) with a small regularization parameter $\lambda_i = 0.0005$ generate even smaller AICs compared to **Model ④** and **Model ⑥**, respectively. **Model ⑦** apparently dwarfs all other models by including liquidity state, time factor, Hawkes kernels, and LASSO all together.

We have calculated the AICs for the 7 types of model across all the 20 trading days from 2019-01-02 to 2019-01-31. The AIC comparison and difference between different model types is demonstrated in the following Table 2.

AIC Difference	Min	1st Quantile	Median	Mean	3rd Quantile	Max	# of days with decreased AIC
④ - ③ ¹	-4.09	-1.19	-0.58	-0.89	-0.16	0.09	18 out of 20
⑥ - ④ ²	-0.31	-0.24	-0.23	-0.22	-0.18	-0.13	20 out of 20
⑤ - ④ ³	-0.12	-0.09	-0.08	-0.08	-0.07	-0.03	20 out of 20
⑤ - ③ ⁴	-4.15	-1.28	-0.63	-0.97	-0.23	-0.01	20 out of 20
⑦ - ⑥ ⁵	-0.13	-0.09	-0.08	-0.08	-0.07	-0.03	20 out of 20

Interpretations:

¹ The Hawkes part has stronger explanation power than the liquidity state and time factor part.

² Adding the liquidity state and time factor to the Hawkes part further improves explanation power.

³ Adding LASSO (LASSO parameter 0.0005) to the Hawkes part further improves explanation power.

⁴ The Hawkes part with LASSO (LASSO parameter 0.0005) generates stronger explanation power than the liquidity state and time factor part.

⁵ Adding LASSO (LASSO parameter 0.0005) further improves the explanation power of the model with the liquidity state, time factor, and Hawkes.

Table 2: AIC difference summary statistics of Apple. Inc from 2019-01-02 to 2019-01-31. Maximum Support $s = 20s$, bin-size $\Delta = 0.25s$. AIC has been adjusted for sample size so that it reflects the AIC per single sample.

Following the sensitivity analysis provided in section 5.7, we also present the model selection results for the estimation results when the size of order is ignored and when the bin-size is enlarged from 0.25s to 0.5s. The results are demonstrated in Supplementary Material section K.

In summary, Figure 11, Table 2, and Supplementary Information section K demonstrate

the following conclusions:

- In event arrival intensity modeling, the inclusion of the Hawkes stimulating function part, the liquidity state part, and the time factor part all contributes to decreased AIC. Moreover, the Hawkes part is more powerful than the liquidity state and time factor parts in terms of reducing AIC.
- Adding LASSO regularization with a small regularization parameter can effectively eliminate redundant parameters for the Hawkes stimulating function and thereby further reduce AIC.
- Under maximum support $s = 20$ seconds, smaller AIC is achieved when the bin-size Δ decreases to a shorter period.

The model selection results are consistent whether we consider the order sizes or not. This is intuitive since the two methods are just different ways to account for the size of orders based on the same LOB dataset.

A very small $\lambda_i = 0.0005$ is chosen throughout our implementations. We have also implemented the same model AIC selections under $\lambda_i = 0.001$ and $\lambda_i = 0.00025$, and achieved similar results, as shown in Supplementary Material Table S3 and Table S4. Therefore, we believe a choice of λ_i that regularizes 10% to 20% of parameters to zeros given a small choice of Δ (around 0.5 seconds) is enough to reduce estimation AIC. Any more advanced method for selecting the appropriate λ_i is beyond the scope of this paper and is left for future works.

6 Discussion and conclusion

6.1 non-parametric estimation

Instead of assuming the Hawkes kernels to follow parametric forms such as the [exponential kernel](#) and the [power-law kernel](#), we implemented a regression-based non-parametric method (Kirchner 2017). As argued by (Kirchner 2017), assuming parametric form of Hawkes excitement kernels, such as exponential distribution, gives less flexibility for estimating kernels that exhibit more complex shapes other the assumed parametric form. In addition, the non-parametric method can reduce numerical issues compared to the typically used maximum likelihood method to estimate parametric kernels, since the likelihood optimization can be very complicated as the number of parameters increases. Moreover, Wu et al. (2019) discusses that the maximum likelihood method (MLE) adopted to estimate kernels under the parametric assumption can lead to biases, since in general the MLE does not admit negative kernel at certain time points. To support these views, we illustrate in Figure 7 that some Hawkes kernels do not exhibit monotonic decreasing shapes. For example, the stimulation effect of trading on best bid ($-1(t)$) to trading on third ask ($+3(t)$) may not be monotonically diminishing, and some delayed-stimulation effect might exist.

There are several reasons to explain the noise in estimated Hawkes functions in Figure 7. First, as we consider order size in the estimation (an order of size 10 is equivalent to 10 individual orders with size 1), some extremely large order arrivals can generate outliers. Second, the estimation can be noisy if the sample size of a specific type of order is small. Given our order classification method, the orders types that move the mid-price $\{p^+(i), p^+(c), p^+(t), p^-(i), p^-(c), p^-(t)\}$ overall have a smaller sample size, because these orders must deplete all the liquidity of a queue or be inserted to an extreme price level. Specifically, for the first trading day within our estimation, the average fraction

of order number among all the 24 types is 4.17% while the average fraction is only 1.27% for those orders that move the reference price.

One may concern that the model might be over-parametrized due to the discretization. The number of parameters in our model is indeed very large, with 49,314 parameters for the model with $s = 20s$ and $\Delta = 0.25s$. We mitigate any potential over-parameterization issue by (1) using a L_1 penalty on all Hawkes process parameters, and (2) fit our model with an even larger volume of high-frequency transaction data. In fact, with such large volume of data (the current data contains around 20 million events for all 20 trading days), the representation power enabled by large number of parameters help to reveal patterns in the order book data that will not be able to discover using a limited-flexibility parametric model.

6.2 choice of bin-size

During the estimation process, we indeed use discretization under choices of bin-size Δ and piece-wise constant to approximate the non-parametric Hawkes kernel, following the approach of Kirchner (2017). For the choice of bin-size Δ , Kirchner (2017) has discussed that the choice of Δ is a bias/variance trade-off as well as a bias/computational-issue trade-off when Δ is extremely small. Our estimation demonstrates that for all else being equal, smaller AICs are achieved when we reduce Δ from 0.5 seconds to 0.25 seconds for all 20 days discussed, taking the AIC difference between Figure 11 and Figure S26 as an example. In addition, when $\Delta = 0.25$ seconds, **Model ⑤** (Hawkes+LASSO) achieves smaller AIC than **Model ③** (liquidity state+time factor) for all 20 days discussed (see Table 2) while the number of days reduces to 19 when $\Delta = 0.5$ seconds with all else equal (see Table S2). This result implies that choosing $\Delta = 0.25s$ is always better than choosing $\Delta = 0.5s$ throughout our estimation. We further enhance this implication by showing that the model with a smaller $\Delta = 0.125s$ achieves better goodness-of-fit than the model with $\Delta = 0.5s$,

as indicated by the quantile-to-quantile (Q-Q) plot in Supplementary Material section L.

It is a limitation that our model doesn't achieve perfect goodness-of-fit, as indicated in the Q-Q plot of Figure S27. We anticipate better model performance will be achieved if Δ is reduced to a further smaller value until the model reaches a threshold when the increasing number of parameters brought by the decreasing Δ generates too much penalty.

However, the computational budget required for model estimation increases dramatically as Δ decreases. Due to this limitation in computation, any experiment with Δ smaller than 0.125s is left for future research.

6.3 existence and stability

Our proposed model is new to the literature since the state variables in the baseline intensity is a function of past order arrival histories. Here we discuss three considerations about the existence of this type of counting process and its stability properties.

First, the Hawkes kernel norm must be below 1 to ensure stability. Our empirical estimates confirm that such criterion is met, and thus the Hawkes part of Eq.(3) is not explosive. Second, although our non-parametric kernel function ϕ is allowed to take negative value, the final form of Eq.(3) is always positive in practice, which ensures the existence (Bacry et al. 2016). Third, the state-dependent baseline intensities coupled with the self-exciting Hawkes component could lead Eq.(3) to be ill-posed even if the kernel norm is below 1. To study the identification and stationary properties, some regularity conditions have been proposed for this “spillover” effect in generalised Hawkes models where the baseline intensity is influenced by past events (Bowsher 2007). For example, Wu et al. (2019) argues a simple case that the invariant distribution exists when state-dependent baseline intensities are bounded.

Our considerations here are not comprehensive. It is a future research direction to rigorously analyze the theoretical properties of Hawkes process coupled with Markovian

state-dependent baseline intensity.

6.4 conclusion

Though we have demonstrated our choice of maximum support s and bin-size Δ is appropriate enough to derive expected results, research on the choice of (s, Δ) and their dynamics is a natural direction of further study. Besides, as an extension to Kirchner (2017), Kirchner & Bercher (2018) has introduced the comparison between the Hawkes model non-parametric estimation and maximum likelihood estimation (MLE), and concluded that the non-parametric method outperforms MLE under some circumstances. Whether this result holds under our proposed framework is also worth investigating. Another possible direction for future research would be to explore more LOB state variables that could reduce the estimation AIC. For example, it is worthwhile to test queue imbalance as proposed by Morariu-Patrichi & Pakkanen (2018). Moreover, based on Huang & Rosenbaum (2017)’s work on the mathematical properties for LOB Markov models, further research on the stationary and ergodicity properties for our proposed model is also a direction of investigation.

To conclude, in this work we have proposed a comprehensive method in high-frequency limit order book data modeling, integrated Markovian state factors into plain-vanilla Hawkes process for limit order book modeling, and applied a flexible non-parametric methods for high-dimensional parameter estimation. Our model provides more careful classification rules for LOB event types and doesn’t require strict parametric assumptions on event stimulating kernels. The mathematical property of our model enables us to implement the estimation in large scale under a parallel and distributed computing framework. We believe our proposed model will bring valuable insights for researchers, financial institutions, and policymakers who attempt to understand the distribution of the order book, the stimulating effects between orders, and more topics related to market microstructure.

References

- Abergel, F. & Jedidi, A. (2015), ‘Long-time behavior of a hawkes process–based limit order book’, *SIAM Journal on Financial Mathematics* **6**(1), 1026–1043.
- Bacry, E., Delattre, S., Hoffmann, M. & Muzy, J.-F. (2013a), ‘Modelling microstructure noise with mutually exciting point processes’, *Quantitative finance* **13**(1), 65–77.
- Bacry, E., Delattre, S., Hoffmann, M. & Muzy, J.-F. (2013b), ‘Some limit theorems for hawkes processes and application to financial statistics’, *Stochastic Processes and their Applications* **123**(7), 2475–2499.
- Bacry, E., Jaisson, T. & Muzy, J.-F. (2016), ‘Estimation of slowly decreasing hawkes kernels: application to high-frequency order book dynamics’, *Quantitative Finance* **16**(8), 1179–1201.
- Bacry, E., Mastromatteo, I. & Muzy, J.-F. (2015), ‘Hawkes processes in finance’, *Market Microstructure and Liquidity* **1**(01), 1550005.
- Bacry, E. & Muzy, J.-F. (2014a), ‘Hawkes model for price and trades high-frequency dynamics’, *Quantitative Finance* **14**(7), 1147–1166.
- Bacry, E. & Muzy, J.-F. (2014b), ‘Second order statistics characterization of hawkes processes and non-parametric estimation’, *arXiv preprint arXiv:1401.0903* .
- Blanc, P., Donier, J. & Bouchaud, J.-P. (2017), ‘Quadratic hawkes processes for financial prices’, *Quantitative Finance* **17**(2), 171–188.
- Bowsher, C. G. (2007), ‘Modelling security market events in continuous time: Intensity based, multivariate point process models’, *Journal of Econometrics* **141**(2), 876–912.

- Chavez-Demoulin, V. & McGill, J. (2012), ‘High-frequency financial data modeling using hawkes processes’, *Journal of Banking & Finance* **36**(12), 3415–3426.
- Clements, A., Herrera, R. & Hurn, A. (2015), ‘Modelling interregional links in electricity price spikes’, *Energy Economics* **51**, 383–393.
- Daley, D. J., Vere-Jones, D. et al. (2003), *An introduction to the theory of point processes: volume I: elementary theory and methods*, Springer.
- Ferriani, F. & Zoi, P. (2020), ‘The dynamics of price jumps in the stock market: an empirical study on europe and us’, *The European Journal of Finance* pp. 1–25.
- Filimonov, V. & Sornette, D. (2015), ‘Apparent criticality and calibration issues in the hawkes self-excited point process model: application to high-frequency financial data’, *Quantitative Finance* **15**(8), 1293–1314.
- Hawkes, A. G. (1971), ‘Spectra of some self-exciting and mutually exciting point processes’, *Biometrika* **58**(1), 83–90.
- Hawkes, A. G. (2018), ‘Hawkes processes and their applications to finance: a review’, *Quantitative Finance* **18**(2), 193–198.
- Hawkes, A. G. & Oakes, D. (1974), ‘A cluster process representation of a self-exciting process’, *Journal of Applied Probability* pp. 493–503.
- Huang, W., Lehalle, C.-A. & Rosenbaum, M. (2015), ‘Simulating and analyzing order book data: The queue-reactive model’, *Journal of the American Statistical Association* **110**(509), 107–122.
- Huang, W. & Rosenbaum, M. (2017), ‘Ergodicity and diffusivity of markovian order book models: a general framework’, *SIAM Journal on Financial Mathematics* **8**(1), 874–900.

- Jaisson, T. & Rosenbaum, M. (2015), ‘Limit theorems for nearly unstable hawkes processes’.
- Kirchner, M. (2017), ‘An estimation procedure for the hawkes process’, *Quantitative Finance* **17**(4), 571–595.
- Kirchner, M. & Bercher, A. (2018), ‘A nonparametric estimation procedure for the hawkes process: comparison with maximum likelihood estimation’, *Journal of Statistical Computation and Simulation* **88**(6), 1106–1116.
- Lallouache, M. & Challet, D. (2016), ‘The limits of statistical significance of hawkes processes fitted to financial data’, *Quantitative Finance* **16**(1), 1–11.
- Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer Science & Business Media.
- Maloney, T. & Almeida, R. J. (2019), ‘Lengthening the investment time horizon’, *MFS White Paper Series November* .
- Morariu-Patrichi, M. & Pakkanen, M. S. (2018), ‘State-dependent hawkes processes and their application to limit order book modelling’, *arXiv preprint arXiv:1809.08060* .
- Muni Toke, I. & Pomponio, F. (2012), ‘Modelling trades-through in a limited order book using hawkes processes’, *Economics: The Open-Access, Open-Assessment E-Journal* **6**(2012-22).
- Omi, T., Hirata, Y. & Aihara, K. (2017), ‘Hawkes process model with a time-dependent background rate and its application to high-frequency financial data’, *Physical Review E* **96**(1), 012303.
- The World Bank (2019), ‘World stocks traded, total value in \$US’. data retrieved from The World Bank and the World Federation of Exchanges database, <https://data.worldbank.org/indicator/CM.MKT.TRAD.CD>.

Tibshirani, R. (1996), ‘Regression shrinkage and selection via the lasso’, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Wu, P., Rambaldi, M., Muzy, J.-F. & Bacry, E. (2019), ‘Queue-reactive hawkes models for the order flow’, *arXiv preprint arXiv:1901.08938* .

Supplementary Material

This Supplementary Material is organized as following: section A illustrates the construction of the reference price in the order book representation mentioned in section 2; section B illustrates the mathematical representation details of the estimation procedure following section 4.3; section C demonstrates the mathematical representation for the LASSO regularization discussed in section 4.4; section D provides estimation results on a fully simulated level-1 order book data; section E demonstrates the estimated Hawkes excitement functions of insertion event at the bid side following section 5.2; section F presents more examples on the liquidity state estimation following section 5.4; section G demonstrates the estimation results when the size of order is ignored as mentioned in section 5.7; section H demonstrates the estimation results when the LASSO regularization is removed as mentioned in section 5.7; section I demonstrates the estimation results when the bin-size is enlarged as mentioned in section 5.7; section J presents results when the maximum support is enlarged as mentioned in section 5.7; section K presents additional model selection results as mentioned in section 5.8; section L presents the goodness-of-fit evaluations of the model using quantile-to-quantile plot.

A Specifications on reference price

Following section 2, the supporting section demonstrates the construction of the reference price p_{ref} . The construction method presented here is mainly adopted from Huang et al. (2015).

When the bid-ask spread is odd in tick unit, it is intuitive to use the mid price p_{mid} to approximate the reference price p_{ref} . Though one can still use p_{mid} as a proxy of p_{ref} when the spread is even in tick unit, it is no longer appropriate enough since p_{mid} itself can be a

position for order arrivals. To be more strict, when $(Q_{\text{best-ask}} - Q_{\text{best-bid}}) = 2n + 1, n \in \mathbb{Z}$, we have $p_{\text{ref}} = p_{\text{mid}} = (Q_{\text{best-bid}} + Q_{\text{best-ask}})/2$; When $(Q_{\text{best-ask}} - Q_{\text{best-bid}}) = 2n, n \in \mathbb{Z}$, we have $p_{\text{ref}} = (Q_{\text{best-bid}} + Q_{\text{best-ask}})/2 + \alpha/2$ or $p_{\text{ref}} = (Q_{\text{best-bid}} + Q_{\text{best-ask}})/2 - \alpha/2$, whichever is closer to the previous value of p_{ref} .

B Mathematical details on non-parametric estimation

Followed section 4.3, this supporting section illustrates the mathematical representations of the non-parametric estimation over the $(6K + 6)$ -dimensional LOB data.

Precisely, based on Definition 3.3, Theorem 3.5, and Definition 3.6 from Kirchner (2017), as well as Lütkepohl (2005)(page70-75), the mathematical details of the estimation procedure is given in Definition 1 and Definition 2.

According to our model specification discussed in section 3, a total number of $6K + 6$ event types is considered for a level- K order book and the number of event types serves as the dimensions of the multivariate Hawkes process. For notation simplicity, we denote $6K + 6$ as d in the following discussions (Definition 1, Definition 2, Remark 1, and Definition 3 in Supplementary Materials C).

Definition 1 *Let X_i ($\forall i = 1, 2, \dots, d$) be a d -variate Hawkes process derived from LOB data with varying baseline intensities controlled by liquidity state sequence $l_i(t)$ and time factor t . Let $T > 0$ and consider the time interval $(0, T]$. For some bin-size $\Delta > 0$, construct the following bin-count sequences according to section 4.1:*

$$\left(B_k^{(\Delta)}, l_k^{(\Delta)}, t_k^{(\Delta)} \right), \forall k = 1, 2, \dots, n := \lfloor T/\Delta \rfloor$$

, where $B_k^{(\Delta)}$, $l_k^{(\Delta)}$ and $t_k^{(\Delta)}$ are $d \times 1$ column vectors defined on $\mathbb{R}^d$.

Then assume $l_k^{(\Delta)}$ can be bucketized into 10 categories $[L_1, L_2, \dots, L_{10}]$ and $t_k^{(\Delta)}$ can be bucketized into 126 categories $[T_1, T_2, \dots, T_{126}]$ according to section 4.2. Given some maximum

support s such that $\Delta < s < T$, The d -variate estimator for the proposed model is defined as:

$$\left(\hat{\Phi}_1^{(\Delta,s)}, \hat{\Phi}_2^{(\Delta,s)}, \dots, \hat{\Phi}_p^{(\Delta,s)}, \hat{\mu}_1^{(\Delta,s)}, \dots, \hat{\mu}_{10}^{(\Delta,s)}, \hat{\theta}_1^{(\Delta,s)}, \dots, \hat{\theta}_{126}^{(\Delta,s)} \right) := \hat{\Phi}^{(\Delta,s)} \in \mathbb{R}^{d \times (dp+10+126)},$$

with $p := \lfloor s/\Delta \rfloor$ (4)

Specifically,

$$\hat{\Phi}_r^{(\Delta,s)} := \begin{bmatrix} \hat{\phi}_{11,r}^{(\Delta,s)} & \hat{\phi}_{12,r}^{(\Delta,s)} & \hat{\phi}_{1d,r}^{(\Delta,s)} \\ \hat{\phi}_{21,r}^{(\Delta,s)} & \hat{\phi}_{22,r}^{(\Delta,s)} & \hat{\phi}_{2d,r}^{(\Delta,s)} \\ \dots & & \\ \hat{\phi}_{d1,r}^{(\Delta,s)} & \hat{\phi}_{d2,r}^{(\Delta,s)} & \hat{\phi}_{dd,r}^{(\Delta,s)} \end{bmatrix} \in \mathbb{R}^{d \times d}, \text{ with } \forall r = 1, 2, \dots, p$$

, where the matrix element

$$\hat{\phi}_{ji,r}^{(\Delta,s)}, \forall i, j = 1, 2, \dots, d; \forall r = 1, 2, \dots, p$$

are weakly consistent estimators for the Multivariate Hawkes excitement function for event j stimulating event i at the r -th function discretized short period. Substitute r with $\lfloor t/\Delta \rfloor$ yields that $\hat{\phi}_{ji, \lfloor t/\Delta \rfloor}^{(\Delta,s)}$ are weakly consistent estimator (for $T \rightarrow \infty, \Delta \rightarrow 0$ and $s = \Delta p \rightarrow \infty$) for $\phi_{ji}(t)$ as shown in Eq.(3).

Also, $\left(\hat{\mu}_1^{(\Delta,s)}, \dots, \hat{\mu}_{10}^{(\Delta,s)} \right) \in \mathbb{R}^{d \times 10}$ and $\left(\hat{\theta}_1^{(\Delta,s)}, \dots, \hat{\theta}_{126}^{(\Delta,s)} \right) \in \mathbb{R}^{d \times 126}$ are weakly consistent estimators for function $M(\cdot)$ and $\Theta(\cdot)$ (for $T \rightarrow \infty, \Delta \rightarrow 0$ and $s = \Delta p \rightarrow \infty$).

Definition 1 gives the detailed description on the structure of the estimator. Then we elucidate the estimation formulas for estimator $\hat{\Phi}^{(\Delta,s)}$ in Definition 2:

Definition 2 Followed from Definition 1, $\hat{\Phi}^{(\Delta,s)}$ can be obtained by applying the following multivariate conditional least-squares (CLS) estimator:

$$\hat{\Phi}^{(\Delta,s)} := \frac{1}{\Delta} \hat{\theta}_{CLS}^{(p,n)} \left(B_k^{(\Delta)}, l_k^{(\Delta)}, t_k^{(\Delta)} \right)_{k=1, \dots, n}$$

The CLS estimator is defined as

$$\hat{\theta}_{CLS}^{(p,n)} : \mathbb{R}^{d \times (n-p)} \rightarrow \mathbb{R}^{d \times (dp+10+126)}$$

$$\begin{aligned} \left(B_1^{(\Delta)}, \dots, B_n^{(\Delta)}; l_1^{(\Delta)}, \dots, l_n^{(\Delta)}; t_1^{(\Delta)}, \dots, t_n^{(\Delta)} \right) &\rightarrow \hat{\theta}_{CLS}^{(p,n)} \left(B_1^{(\Delta)}, \dots, B_n^{(\Delta)}; l_1^{(\Delta)}, \dots, l_n^{(\Delta)}; t_1^{(\Delta)}, \dots, t_n^{(\Delta)} \right) \\ &:= Y Z^\top (Z Z^\top)^{-1} \end{aligned}$$

, where

$$Z \left(B_1^{(\Delta)}, \dots, B_n^{(\Delta)}; l_1^{(\Delta)}, \dots, l_n^{(\Delta)}; t_1^{(\Delta)}, \dots, t_n^{(\Delta)} \right) := \begin{bmatrix} B_p^{(\Delta)} & B_{p+1}^{(\Delta)} & \dots & B_{n-1}^{(\Delta)} \\ B_{p-1}^{(\Delta)} & B_p^{(\Delta)} & \dots & B_{n-2}^{(\Delta)} \\ \dots & & & \\ B_1^{(\Delta)} & B_2^{(\Delta)} & \dots & B_{n-p}^{(\Delta)} \\ \mathbb{1}_{l_{p+1}^{(\Delta)} \in L_1} & \mathbb{1}_{l_{p+2}^{(\Delta)} \in L_1} & \dots & \mathbb{1}_{l_n^{(\Delta)} \in L_1} \\ \mathbb{1}_{l_{p+1}^{(\Delta)} \in L_2} & \mathbb{1}_{l_{p+2}^{(\Delta)} \in L_2} & \dots & \mathbb{1}_{l_n^{(\Delta)} \in L_2} \\ \dots & & & \\ \mathbb{1}_{l_{p+1}^{(\Delta)} \in L_{10}} & \mathbb{1}_{l_{p+2}^{(\Delta)} \in L_{10}} & \dots & \mathbb{1}_{l_n^{(\Delta)} \in L_{10}} \\ [0]_{d \times 1} & [0]_{d \times 1} & \dots & [0]_{d \times 1} \\ \mathbb{1}_{t_{p+1}^{(\Delta)} \in T_2} & \mathbb{1}_{t_{p+2}^{(\Delta)} \in T_2} & \dots & \mathbb{1}_{t_n^{(\Delta)} \in T_2} \\ \mathbb{1}_{t_{p+1}^{(\Delta)} \in T_3} & \mathbb{1}_{t_{p+2}^{(\Delta)} \in T_3} & \dots & \mathbb{1}_{t_n^{(\Delta)} \in T_3} \\ \dots & & & \\ \mathbb{1}_{t_{p+1}^{(\Delta)} \in T_{126}} & \mathbb{1}_{t_{p+2}^{(\Delta)} \in T_{126}} & \dots & \mathbb{1}_{t_n^{(\Delta)} \in T_{126}} \end{bmatrix} \in \mathbb{R}^{(dp+10+126) \times (n-p)}$$

is the design matrix and

$$Y \left(B_1^{(\Delta)}, \dots, B_n^{(\Delta)} \right) := \left(B_{p+1}^{(\Delta)}, B_{p+2}^{(\Delta)}, \dots, B_n^{(\Delta)} \right) \in \mathbb{R}^{d \times (n-p)}$$

is the response.

Within the design matrix Z , $\left(\mathbb{1}_{l_k^{(\Delta)} \in L_1}, \dots, \mathbb{1}_{l_k^{(\Delta)} \in L_{10}}\right)$ and $\left(\mathbb{1}_{t_k^{(\Delta)} \in L_1}, \dots, \mathbb{1}_{t_k^{(\Delta)} \in L_{126}}\right)$ with $\forall k = p + 1, \dots, n$ are $d \times 1$ column indicator functions that returns 1 (returns 0 otherwise) when the i -th dimension element falling into the corresponding category; $[0]_{d \times 1}$ is d -dimensional column vector consisting of zeros.

The design matrix Z in Definition 2 contains a $(n - p)$ dimensional row consisting of zeros. This row references the part of design matrix such that the time factor sequence $t_k^{(\Delta)}$ belongs to the T_1 category. The T_1 category is treated as the “reference group” in the present of the two categorical variables $[l_k^{(\Delta)}, t_k^{(\Delta)}]$ and the regular (non-categorical) variables derived from $B_k^{(\Delta)}$. The R programming language we use is built with its default “contrast coding” system that requires the existence of at least one “reference group” when implementing linear regression models with more than one categorical variables. The CLS estimators for the “reference group” are automatically set to zeros according to the “contrast coding” rule. Therefore, the CLS estimator in Definition 2 returns a zero vector of dimensional d as the estimator for category T_1 . The “contrast coding” system may be different for other programming languages and their statistical packages. The choice of categorical variable coding rules and the reference group can be considered as adding/subtracting a constant on the estimators for one categorical variable and subtracting/adding it back on another, leading to no changes on model goodness-of-fit.

Followed Definition 2, the following definition gives an equivalent but easier way for model implementation:

Remark 1 *Given Definition 1 and 2, Lütkepohl (2005)(page72) illustrates that the multivariate CLS-estimation is equivalent to d individual Ordinary-Least-Squared(OLS) estimations, in which d is the dimension of the CLS-estimation. Using design matrix Z , response Y , and estimator $\hat{\Phi}^{(\Delta, s)}$ given in Definition 1 and 2, and let y_i be the transpose of the i -th*

row vector of response Y :

$$y_i = \left(B_{i,p+1}^{(\Delta)}, B_{i,p+2}^{(\Delta)}, \dots, B_{i,n}^{(\Delta)} \right)^\top \in \mathbb{R}^{(n-p) \times 1}, \forall i = 1, 2, \dots, d$$

; Let $\hat{\phi}_i^{(\Delta,s)}$ be the transpose of the i -th row vector of $\hat{\Phi}^{(\Delta,s)}$:

$$\begin{aligned} \hat{\phi}_i^{(\Delta,s)} = & \left[\left(\hat{\phi}_{i1,1}^{(\Delta,s)}, \dots, \hat{\phi}_{id,1}^{(\Delta,s)} \right), \dots, \left(\hat{\phi}_{i1,p}^{(\Delta,s)}, \dots, \hat{\phi}_{id,p}^{(\Delta,s)} \right), \right. \\ & \left. \left(\hat{\mu}_{i,1}^{(\Delta,s)}, \dots, \hat{\mu}_{i,10}^{(\Delta,s)} \right), \left(\hat{\theta}_{i,1}^{(\Delta,s)}, \dots, \hat{\theta}_{i,126}^{(\Delta,s)} \right) \right]^\top \in \mathbb{R}^{(dp+10+126) \times 1} \end{aligned}$$

We have that $\hat{\phi}_i^{(\Delta,s)} := (ZZ^\top)^{-1} Zy_i$ is the OLS estimator for the model:

$$y_i = Z^\top \phi_i^{(\Delta,s)} + u_i, \forall i = 1, 2, \dots, d$$

, where u_i is $(n-p) \times 1$ white-noise column vector $(u_{i,p+1}, u_{i,p+2}, \dots, u_{i,n})^\top$ with $i = 1, \dots, d$.

In model implementation, we prefer conducting the OLS estimations based on $y_i = Z^\top \phi_i^{(\Delta,s)} + u_i$ over all dimensions $1, 2, \dots, d$, over the one single CLS-estimation shown in Definition 2, since the OLS estimations involve less dimensions and thereby more computationally efficient under a parallel computing setting.

C Non-parametric estimation with LASSO

The following definition illustrates LASSO regularization for our proposed model discussed in section 4.4. Consistent with Supplementary Materials B, we denote the dimension of our estimation (the number of event types considered) $6 \times (K+1)$ as d for simplicity. Note K represents the level of the order book we consider.

Definition 3 Consider the design matrix Z , the response Y , and the OLS-estimators proposed in Definition 2 and Remark 1. The OLS-estimators minimize the loss function:

$$\left(y_i - Z^\top \phi_i^{(\Delta,s)} \right)^\top \left(y_i - Z^\top \phi_i^{(\Delta,s)} \right)$$

Then consider adding a LASSO regularization term that only applies to the Hawkes excitement function $\phi_i^{(\Delta,s),excitements} := \left[\left(\hat{\phi}_{i1,1}^{(\Delta,s)}, \dots, \hat{\phi}_{id,1}^{(\Delta,s)} \right), \dots, \left(\hat{\phi}_{i1,p}^{(\Delta,s)}, \dots, \hat{\phi}_{id,p}^{(\Delta,s)} \right) \right]^\top \in \mathbb{R}^{d_p \times 1}$ to the loss function, the LASSO loss function becomes:

$$\left(y_i - Z^\top \phi_i^{(\Delta,s)} \right)^\top \left(y_i - Z^\top \phi_i^{(\Delta,s)} \right) + \lambda_i \|\phi_i^{(\Delta,s),excitements}\|_1$$

where λ_i denotes the regularization penalty and $\|\cdot\|_1$ denotes the ℓ_1 -norm for estimators.

D Simulation

In this section, we evaluate the proposed estimation method on a fully simulated order book dataset. For computational simplicity in the simulation, we set the order book level $K = 1$, and there are no events changing the reference price. Therefore, we consider 6 types of events $\{-1(i), -1(c), -1(t), +1(i), +1(c), +1(t)\}$ in the simulation environment.

To obtain a clear demonstration of the varying baseline intensity estimation, we specify the Hawkes intensity as:

$$\lambda_i(t) = \Theta_i(t) + \sum_{j=1}^6 \int \phi_{j,i}(t-s) dX_j(s), \quad \forall i, j = 1, 2, \dots, 6 \quad (5)$$

, in which i, j are indexes for the 6 events described above. Specifically, the Hawkes kernel is set as step functions in the following form:

$$\phi_{j,i} = \begin{cases} 0.05 \cdot \mathbb{1}_{\{t \leq 2\}}, j = 1, 2, 3 \\ 0.05 \cdot \mathbb{1}_{\{t \leq 1\}}, j = 4, 5, 6 \end{cases}$$

The Hawkes intensity also depends on the state-dependent baseline intensity $\Theta_i(t)$, which is defined as:

for $i = 1, 2, 3$:

$$\Theta_i(t) = \begin{cases} 0.1, & \text{if } \lfloor t \rfloor = 3n, n \in \mathbb{N}^+ \\ 0.05, & \text{else} \end{cases}$$

for $i = 4, 5, 6$:

$$\Theta_i(t) = \begin{cases} 0.1, & \text{if } \lfloor t \rfloor = 2n, n \in \mathbb{N}^+ \\ 0.05, & \text{else} \end{cases}$$

The varying baseline intensity depends on the event arrival time t . i.e. for the first three types of events, when the timestamp (in seconds) of event arrival falls into $[0s - 1s)$, $[3s - 4s)$, or $[6s - 7s)$..., the baseline intensity becomes 0.1 but otherwise 0.05. Similarly, for the last three types of events, the baseline intensity elevates by 0.05 whenever the timestamp of arrival falls into $[0s - 1s)$, $[2s - 3s)$, or $[4s - 5s)$, etc. This baseline intensity setup is analogous to the “liquidity state” and “time factor” effects in estimating real order book data, in which the baseline intensity depends on both the event arrival time and the current state of the order book. The difference is that the baseline intensity in the simulation is a simpler version, which can provide a more intuitive demonstration and reduces the computational cost.

We simulate the specified 6-dimensional Hawkes process over a period of $[0, T = 20000s)$ for 100 times and then estimate the model parameters under the proposed non-parametric procedure. The estimated Hawkes kernels are demonstrated in Figure S1, in which we can observe very accurate and stable Hawkes kernel estimations for each type of event, with the red line (estimated value) matching very closely to the blue line (true value). Furthermore, we also observe in Figure S2 that the estimation method can recover the values of the state parameters $\Theta_i(t)$ very closely, as the true values fall into the 95% confidence intervals of the estimated values for all 6 events. The Q-Q plots demonstrated in Figure S3 enhance the validity of the estimation procedure by showing decent goodness-of-fit. It is observed that the empirical distribution of the rescaled time based on our estimation ensembles very closely to the standard exponential distribution. Overall, the estimation result based on the simulated order book data demonstrates the validity of the non-parametric estimation procedure, and therefore we are confident enough to implement this method to the more complex real order book data on higher levels.

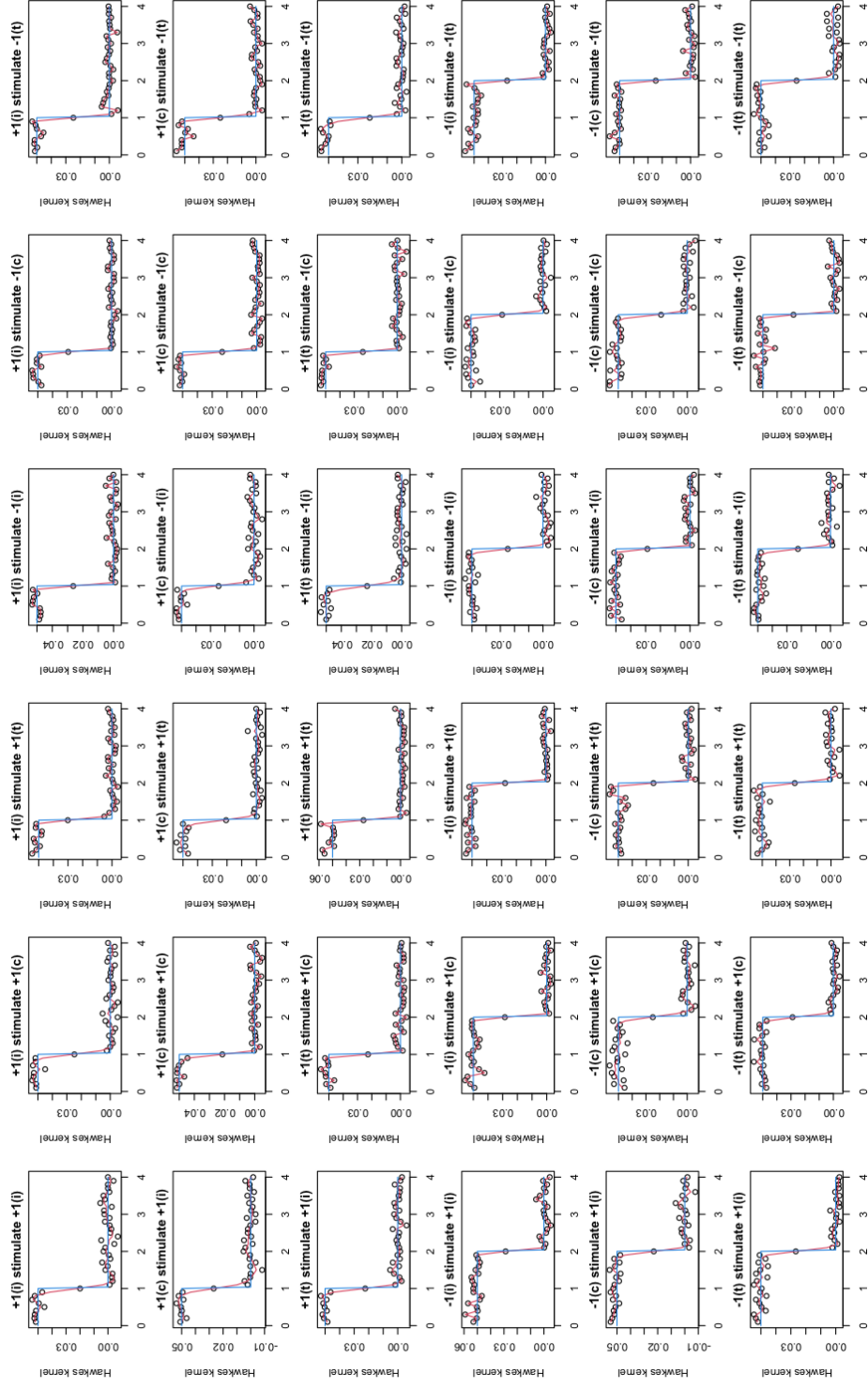


Figure S1: Simulation and estimation results of Hawkes kernel based on simulated level-1 order book with 6 events. The points illustrate the discrete estimator of the Hawkes stimulating function. The red line illustrates the smoothing spline for the points. The blue line illustrates the true stimulating function. The reported values are the average of 100 independent simulation and estimation runs.

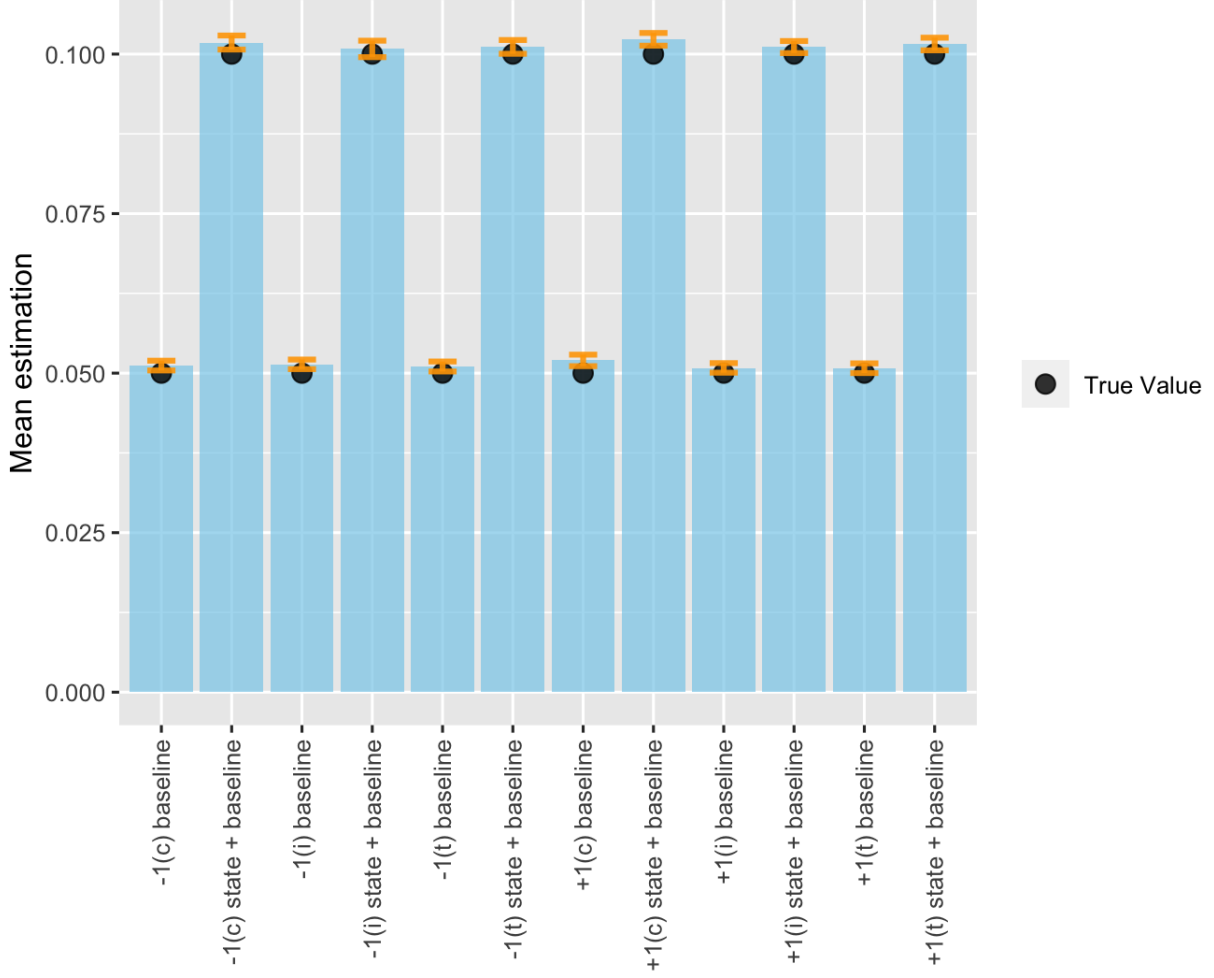


Figure S2: Simulation and estimation results of Hawkes process baseline intensity state based on simulated level-1 order book with 6 events. The blue bar indicates the average estimated value of the state parameters. The orange bar indicates the 95% confidence interval of the estimated values. The black dot indicates the true value of the state parameters. The reported values are the average of 100 independent simulation and estimation runs.

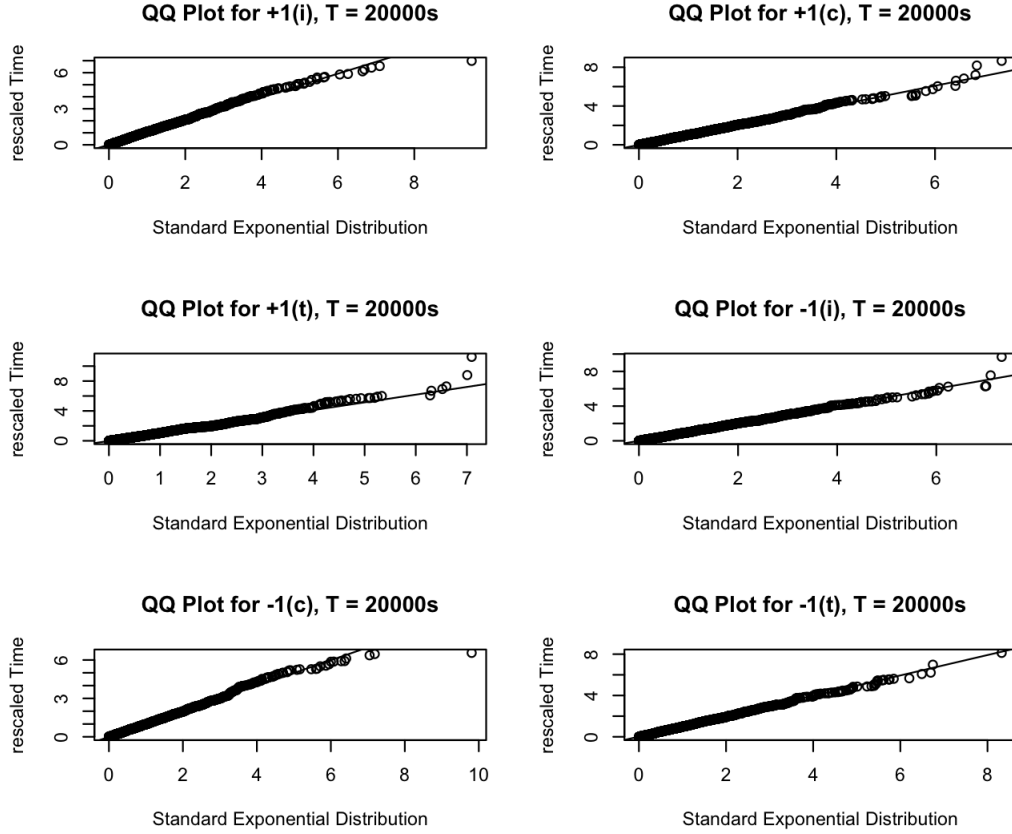


Figure S3: Q-Q (quantile-quantile) plot of the estimation results based on simulated level-1 order book with 6 events. The points indicate the empirical distribution quantiles of the rescaled time derived from the estimated model parameters. The black line indicates the quantiles of the standard exponential distribution. The reported values are the average of 100 independent simulation and estimation runs.

E Estimated excitement functions: bid orders

Following section 5.2, this section illustrates excitement functions of insertion event at 1st bid (event $-1(i)$) stimulating insertion and cancellation at the 1st bid (event $-1(i)$ and

$-1(c)$). The shape of the estimated excitement functions exhibit similar time-decaying patterns as shown in section 5.2 for ask orders.

Additionally, we observed similarities between Fig.S4(a) and Fig.S4(b), which show the stimulation of $-1(i)$ to $-1(i)$ and $-1(c)$. Both excitement functions spike at around 8 seconds and 20 seconds. This observation suggests that the estimated Hawkes excitement functions are similar for the effect towards the insertion and cancellation at the 1st bid (effect towards $-1(i)$ and $-1(c)$).

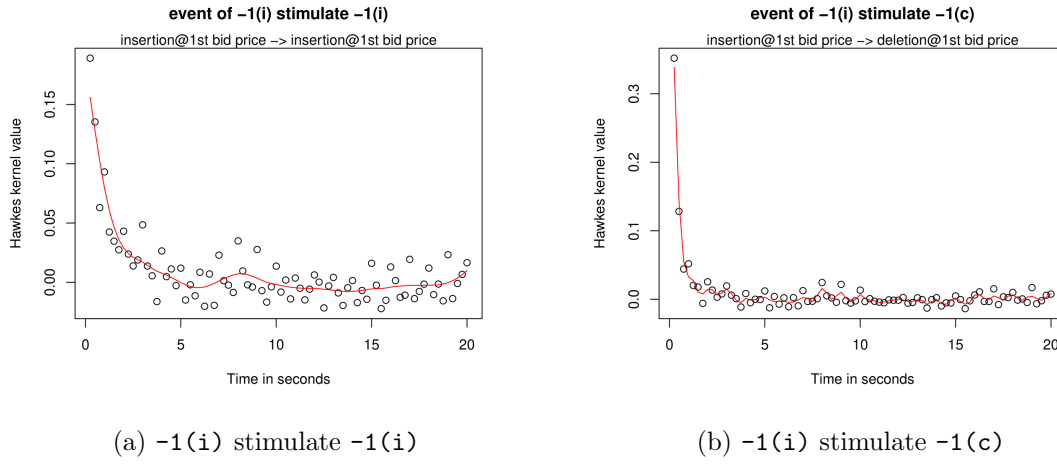
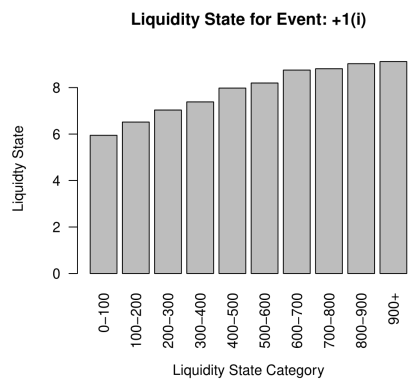


Figure S4: Aggregated Hawkes excitement function estimation result under ($s = 20$ seconds, $\Delta = 0.25$ seconds) with LASSO regularization. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

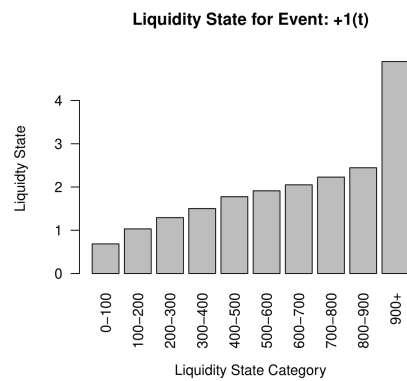
F More examples on liquidity state

Following section 5.4 and Figure 8, this section presents more examples of the estimated result for liquidity state. The examples are presented in Figure S5 and Figure S6. In

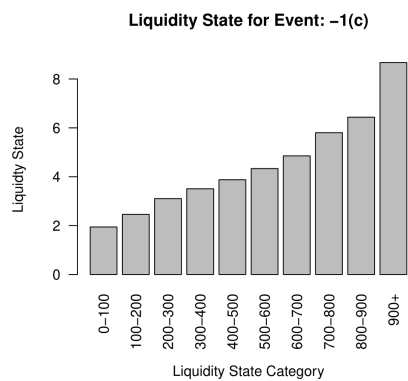
general, the arrival intensities increase as the liquidity state increases for trade/cancellation events and insertion events on the 1st level (i.e., $-3(c)$, $-3(t)$, $-2(c)$, $-2(t)$, $-1(i)$, $-1(c)$, $-1(t)$, $+1(i)$, $+1(c)$, $+1(t)$, $+2(c)$, $+2(t)$, $+3(c)$, $+3(t)$); the liquidity state increases for the insertion events on the 2nd and 3rd level (i.e., $+3(i)$, $-3(i)$, $+2(i)$, $-2(i)$). Moreover, we also perform a sensitivity analysis on the number of liquidity state in the estimation. Specifically, compared to our baseline estimation demonstrated in Figure S5 and S6 that uses 10 liquidity states, we alternatively estimate the full model under 5 liquidity states, with the results demonstrated in Figure S7 and S8.



(a) Liquidity state for +1(i)

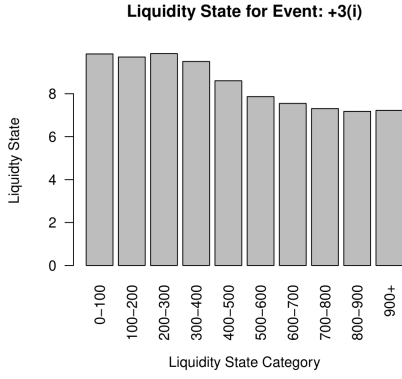


(b) Liquidity state for +1(t)

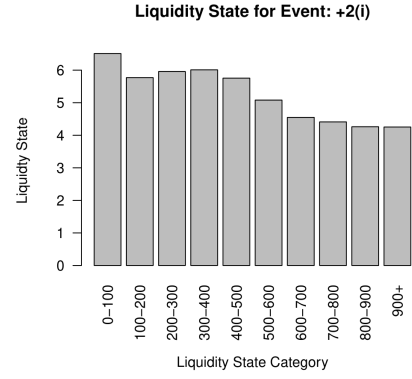


(c) Liquidity state for -1(c)

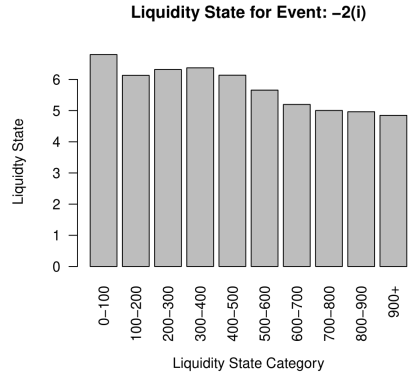
Figure S5: Aggregated estimation results for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds).



(a) Liquidity state for +3(i)

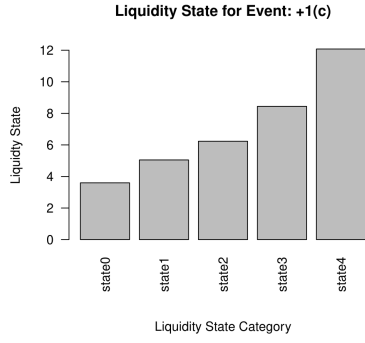


(b) Liquidity state for +2(i)

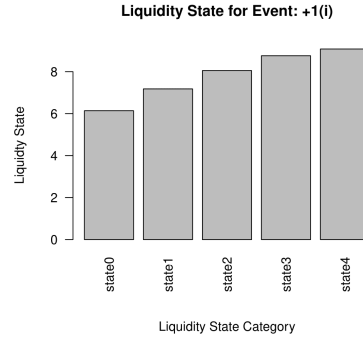


(c) Liquidity state for -2(i)

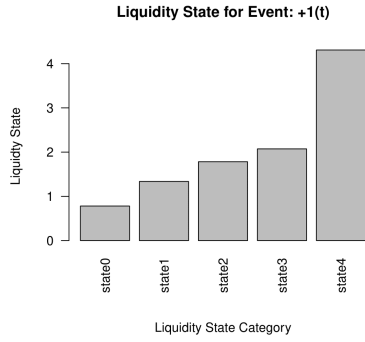
Figure S6: Aggregated estimation results for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds).



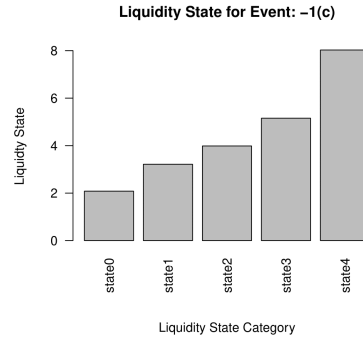
(a) Liquidity state for +1(c)



(b) Liquidity state for +1(i)

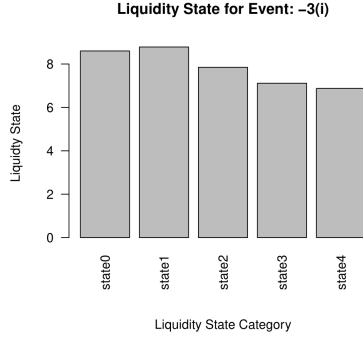


(c) Liquidity state for +1(t)

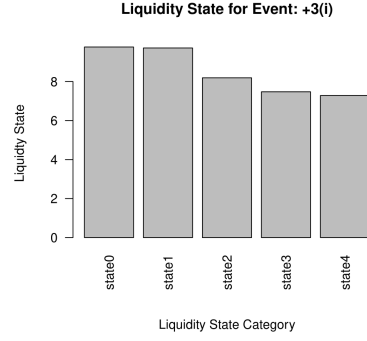


(d) Liquidity state for -1(c)

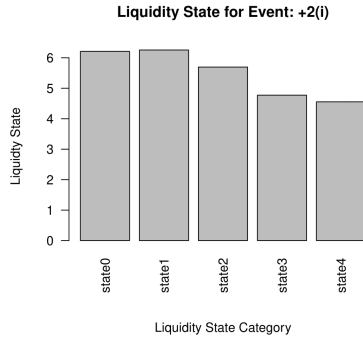
Figure S7: Aggregated estimation results for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds). Compared to Figure S5, the number of liquidity state is set to be 5 for this estimation.



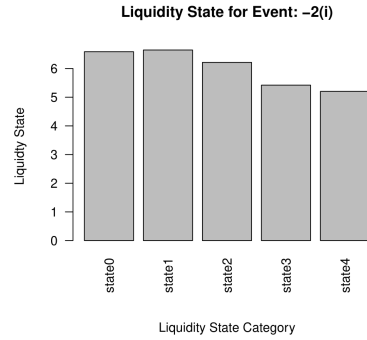
(a) Liquidity state for $-3(i)$



(b) Liquidity state for $+3(i)$



(c) Liquidity state for $+2(i)$



(d) Liquidity state for $-2(i)$

Figure S8: Aggregated estimation results for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds). Compared to Figure S6, the number of liquidity state is set to be 5 for this estimation.

G Empirical results when order size is ignored

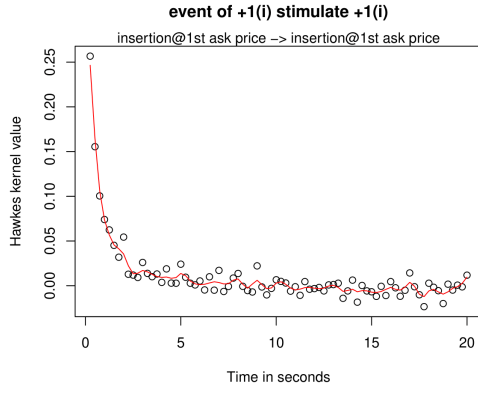
This supporting section demonstrates the empirical estimation results when the order size is ignored, as mentioned in section 5.7. The demonstrations will be presented in a similar format as the demonstrations from section 5.2 to section 5.5. In general, the results on

excitement function, liquidity state, and time factor still hold qualitatively in the sense that most estimated function have similar shapes.

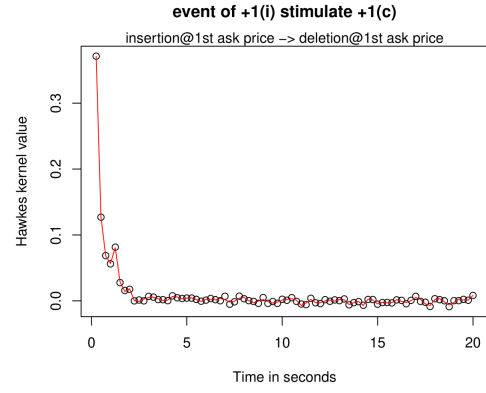
However, different order size considerations tend to give the estimated functions in different scale, where the estimations in general have lower intensity levels for many events when the order size is ignored. Also, the estimated Hawkes excitement functions tend to be less volatile if we ignore the order size. This behavior is expected since we aggregate order sizes in the original model setting while all orders are considered to have size 1 if we ignore order size. Therefore, when order size is ignored, it is natural for the estimates to have relatively lower intensity and volatility, especially during peak trading hours.

G.1 estimated excitement functions

Based on Figure 6 and Figure S4, the following Figure S9 and Figure S10 demonstrate the estimated Hawkes excitement functions when the order size is ignored.

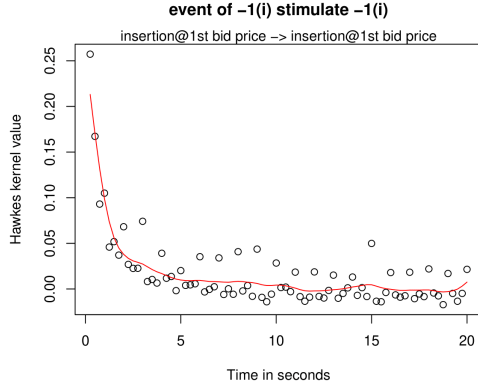


(a) +1(i) stimulate +1(i)

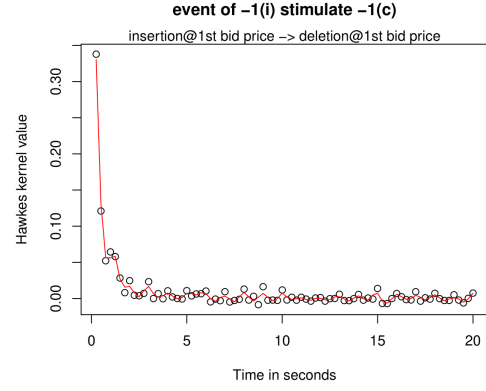


(b) +1(i) stimulate +1(c)

Figure S9: Aggregated Hawkes excitement function estimation under ($s = 20$ seconds, $\Delta = 0.25$ seconds) with LASSO regularization. All orders are considered to have size 1. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.



(a) -1(i) stimulate -1(i)



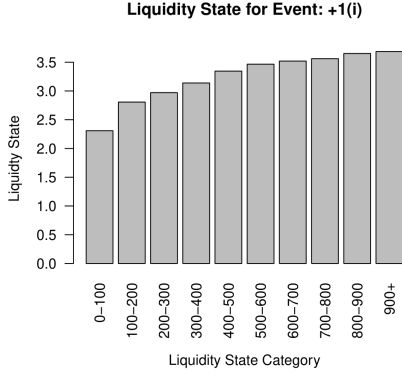
(b) -1(i) stimulate -1(c)

Figure S10: Aggregated Hawkes excitement function estimation under ($s = 20$ seconds, $\Delta = 0.25$ seconds) with LASSO regularization. All orders are considered to have size 1. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

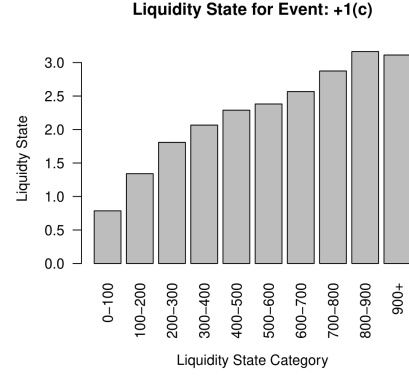
As we can observe, the above estimated functions are consistent with the 1st-ask and 1st-bid similarity patterns discussed in section 5.2 and section E.

G.2 liquidity state

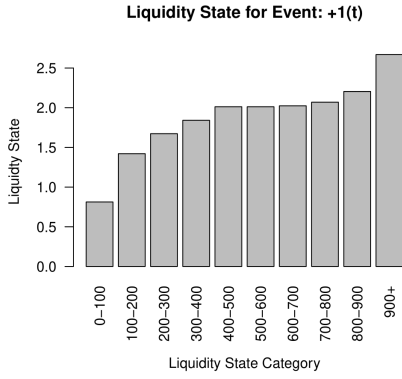
Based on Figure S5 and Figure S6 in section 5.4, the following Figure S11 and Figure S12 demonstrate the liquidity state estimations of the model with LASSO regularization.



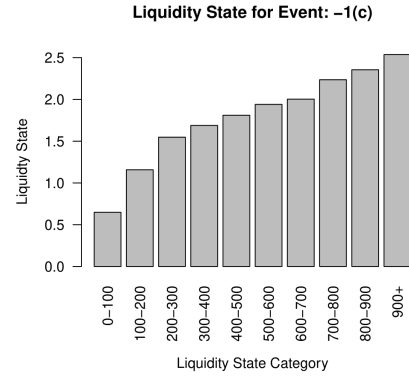
(a) Liquidity state for +1(i)



(b) Liquidity state for +1(c)

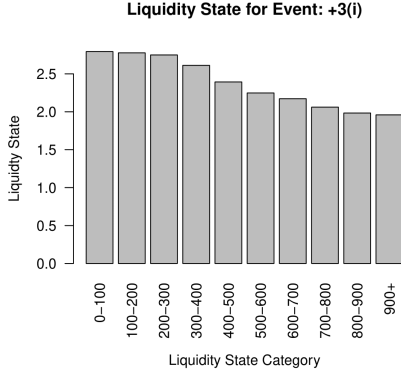


(c) Liquidity state for +1(t)

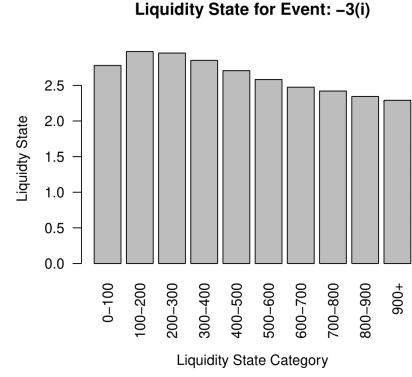


(d) Liquidity state for -1(c)

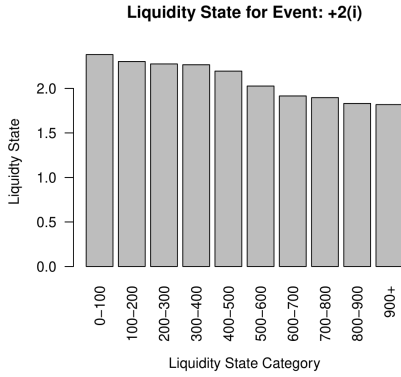
Figure S11: Aggregated estimation result for liquidity state for event +1(i), +1(c), +1(t), -1(c) under ($s = 20$ seconds, $\Delta = 0.25$ seconds). All orders are considered to have size 1. For these events the event arrival intensity increases as liquidity state increases.



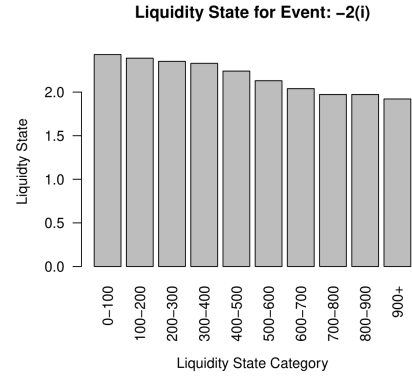
(a) Liquidity state for $+3(i)$



(b) Liquidity state for $-3(i)$



(c) Liquidity state for $+2(i)$



(d) Liquidity state for $-2(i)$

Figure S12: Aggregated estimation result for liquidity state for event $+3(i)$, $-3(i)$, $+2(i)$, $-2(i)$ under ($s = 20$ seconds, $\Delta = 0.25$ seconds). All orders are considered to have size 1. For these events the event arrival intensity decreases as liquidity state increases.

The demonstrated liquidity state estimation results of the model with LASSO regularization is consistent with the results discussed in section 5.4.

G.3 time factor

Based on Figure 9 in section 5.5, the following Figure S13 demonstrates the time factor estimations of the model with LASSO regularization.

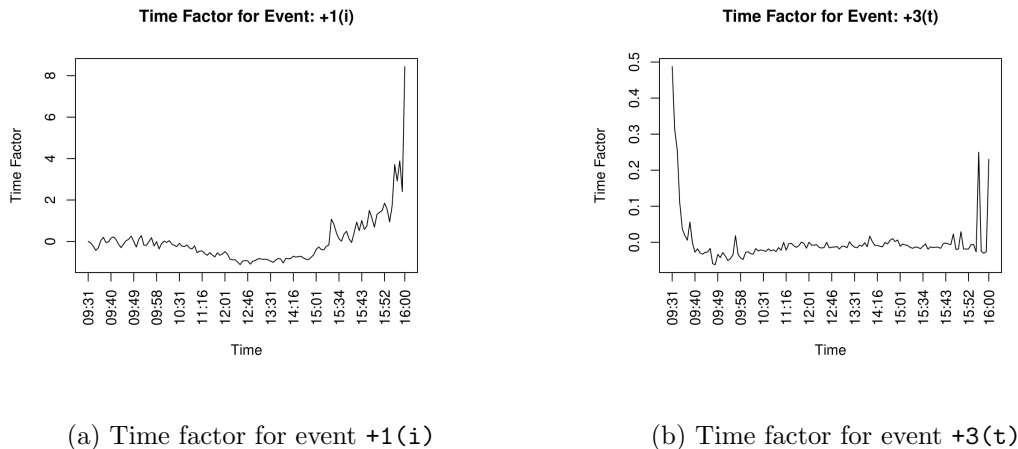


Figure S13: Aggregated estimation result for time factor between 9:30 am and 4:00 pm under ($s = 20$ seconds, $\Delta = 0.25$ seconds). All orders are considered to have size 1.

The demonstrated time factor estimation results of the model with LASSO regularization is consistent with the results discussed in section 5.5.

H Empirical results without LASSO

This supporting section demonstrates the empirical estimation results when the LASSO regularization is removed, as mentioned in section 5.7. The demonstrations will be presented in a similar format as the demonstrations from section 5.2 to section 5.5. As a whole, the results on excitement function, liquidity state, and time factor still hold qualitatively. The estimation result with or without LASSO (small regularization $\lambda_i = 0.0005$) are very

similar visually. Furthermore, the cubic smoothing spline for the LASSO model is smoother than the model without LASSO since the estimator distribution is more concentrated to zero after adding LASSO.

H.1 estimated excitement functions

Based on Figure 6 and Figure S4, the following Figure S14 and Figure S15 demonstrate the estimated Hawkes excitement functions when the LASSO regularization is removed.

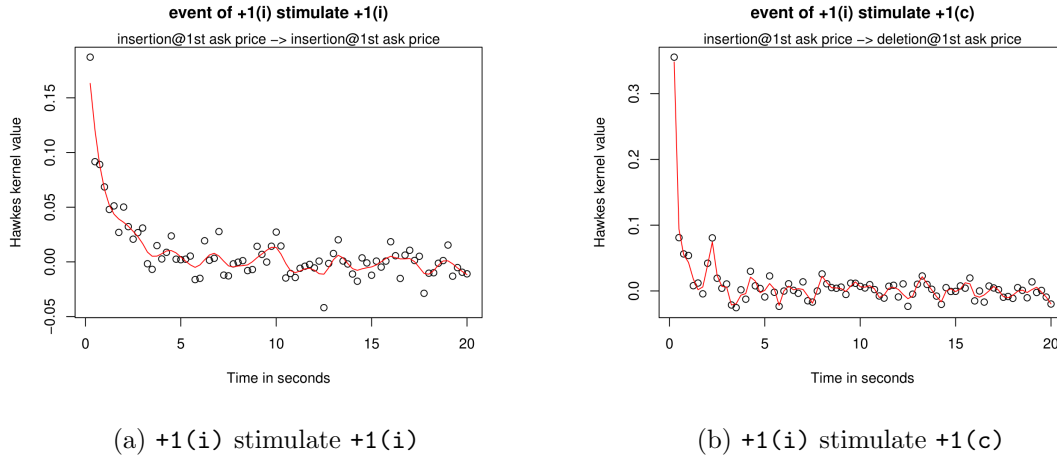
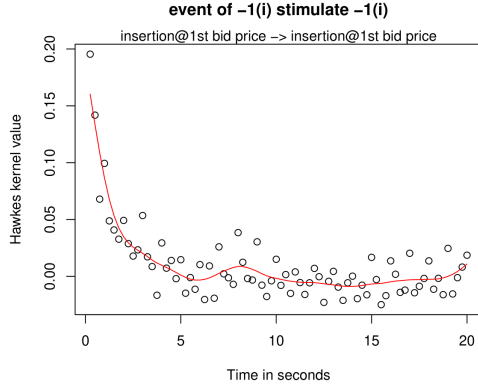
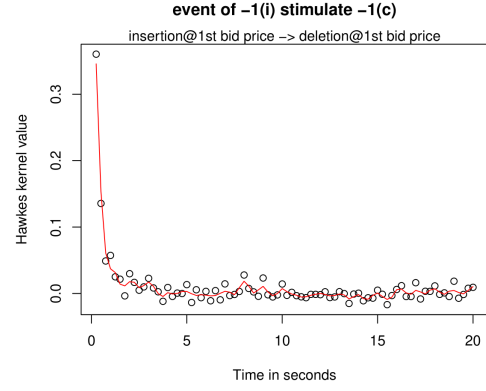


Figure S14: Aggregated Hawkes excitement function estimation under ($s = 20$ seconds, $\Delta = 0.25$ seconds) when LASSO regularization is removed. the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.



(a) -1(i) stimulate -1(i)



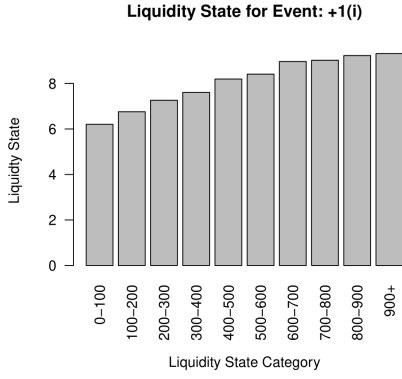
(b) -1(i) stimulate -1(c)

Figure S15: Aggregated Hawkes excitement function estimation under ($s = 20$ seconds, $\Delta = 0.25$ seconds) when LASSO regularization is removed. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

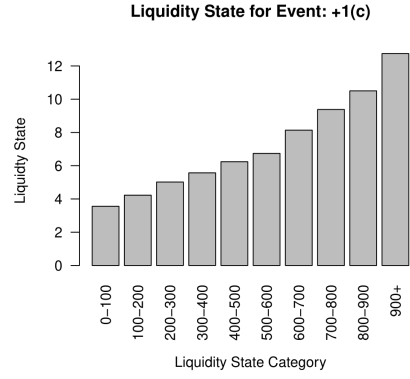
As we can observe, the above estimated functions are consistent with the 1st-ask and 1st-bid similarity patterns discussed in section 5.2 and section E, when the LASSO regularization is removed.

H.2 liquidity state

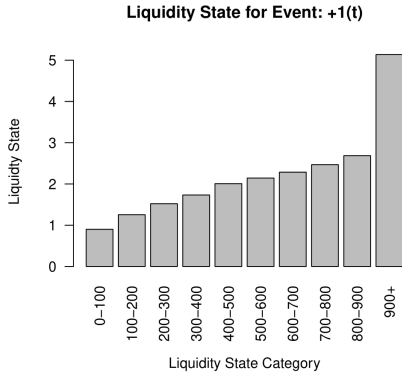
Based on Figure S5 and Figure S6 in section 5.4, the following Figure S16 and Figure S17 demonstrate the liquidity state estimations of the model when the LASSO regularization is removed.



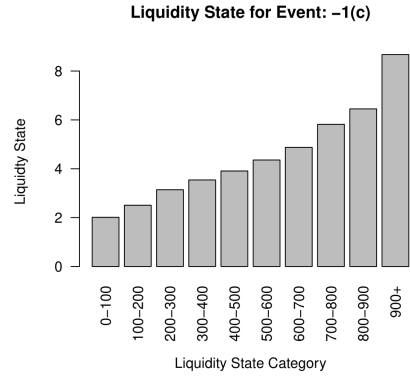
(a) Liquidity state for +1(i)



(b) Liquidity state for +1(c)

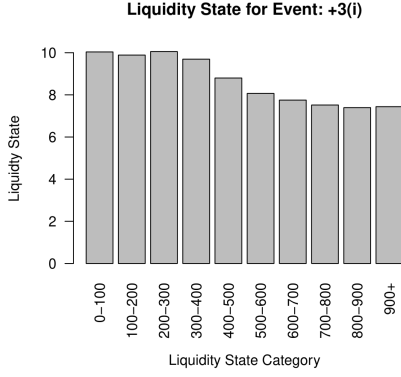


(c) Liquidity state for +1(t)

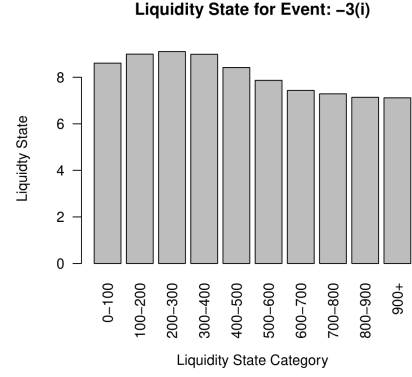


(d) Liquidity state for -1(c)

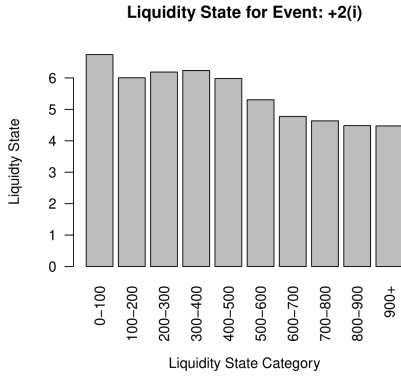
Figure S16: Aggregated estimation result for liquidity state for event +1(i), +1(c), +1(t), -1(c) under ($s = 20$ seconds, $\Delta = 0.25$ seconds) when LASSO regularization is removed. For these events the event arrival intensity increases as liquidity state increases.



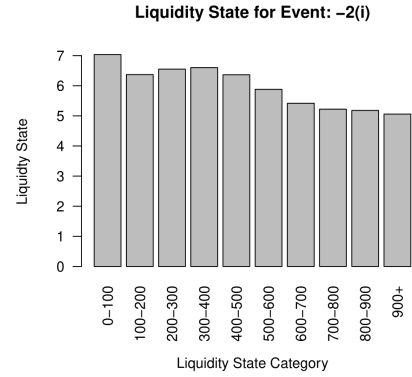
(a) Liquidity state for $+3(i)$



(b) Liquidity state for $-3(i)$



(c) Liquidity state for $+2(i)$



(d) Liquidity state for $-2(i)$

Figure S17: Aggregated estimation result for liquidity state for event $+3(i)$, $-3(i)$, $+2(i)$, $-2(i)$ under ($s = 20$ seconds, $\Delta = 0.25$ seconds) when LASSO regularization is removed. For these events the event arrival intensity decreases as liquidity state increases.

The demonstrated liquidity state estimation results of the model without LASSO regularization is consistent with the results discussed in section 5.4.

H.3 time factor

Based on Figure 9 in section 5.5, the following Figure S18 demonstrates the time factor estimations of the model when the LASSO regularization is removed.

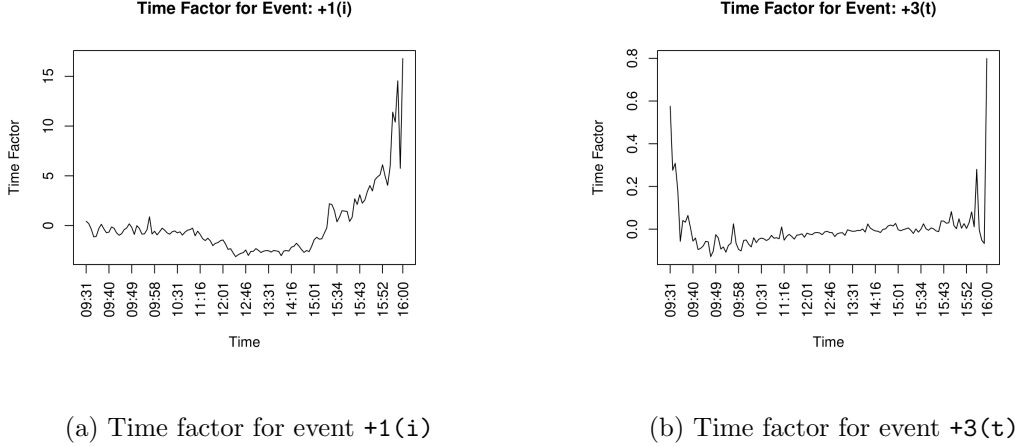


Figure S18: Aggregated estimation result for time factor between 9:30 am and 4:00 pm under ($s = 20$ seconds, $\Delta = 0.25$ seconds) when LASSO regularization is removed. All orders are considered to have size 1.

The demonstrated time factor estimation results of the model without LASSO regularization is consistent with the results discussed in section 5.5.

I Empirical results with enlarged bin-size

This supporting section demonstrates the empirical estimation results when the bin-size Δ is enlarged from 0.25 seconds to 0.5 seconds, as mentioned in section 5.7. The demonstrations will be presented in a similar format as the demonstrations from section 5.2 to section 5.5. As a whole, the results on excitement function, liquidity state, and time factor still

hold qualitatively. This meets our expectation that enlarging the bin-size won't change the estimated result significantly as estimations are obtained from the same dataset and the $\Delta = 0.5s$ estimation result is just a coarse version of the $\Delta = 0.25s$ result.

I.1 estimated excitement functions

Based on Figure 6 and Figure S4, the following Figure S19 and Figure S20 demonstrate the estimated Hawkes excitement functions when the bin-size Δ is enlarged from 0.25 seconds to 0.5 seconds.

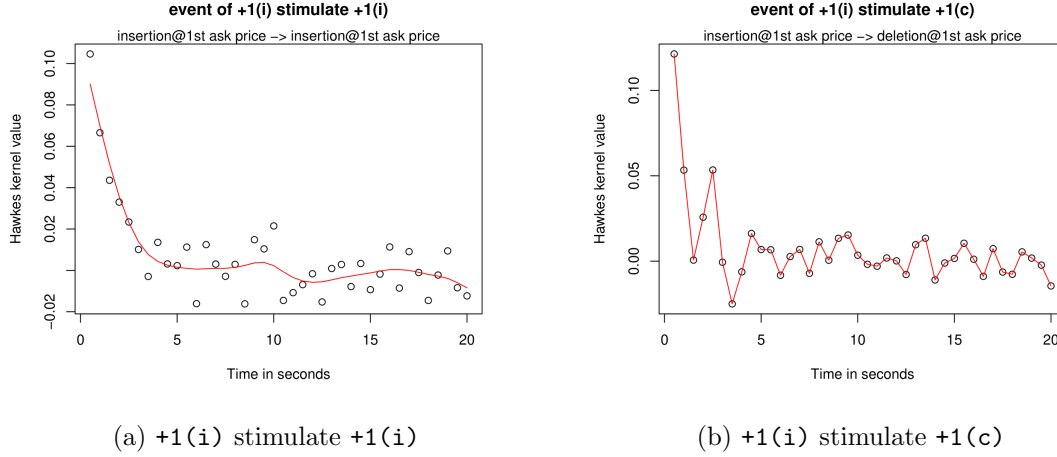
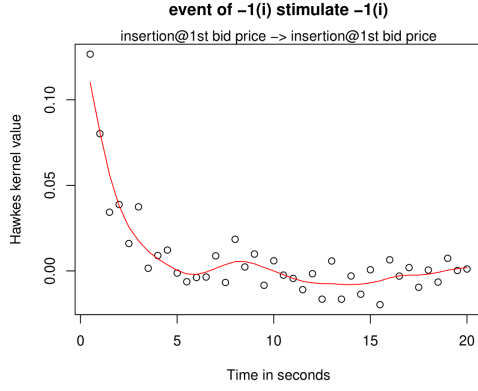
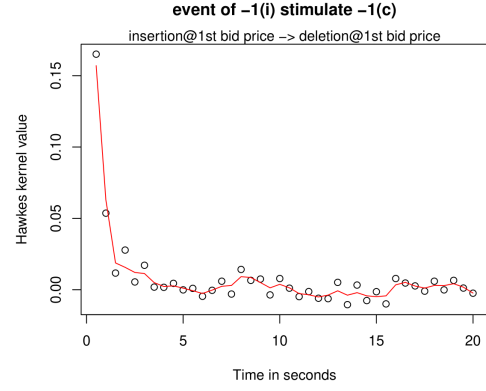


Figure S19: Aggregated Hawkes excitement function estimation under ($s = 20$ seconds, $\Delta = 0.5$ seconds) with LASSO regularization. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.



(a) $-1(i)$ stimulate $-1(i)$



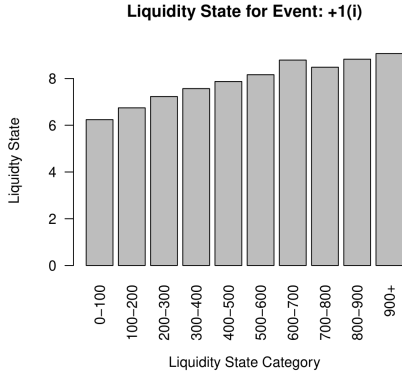
(b) $-1(i)$ stimulate $-1(c)$

Figure S20: Aggregated Hawkes excitement function estimation under ($s = 20$ seconds, $\Delta = 0.5$ seconds) with LASSO regularization. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

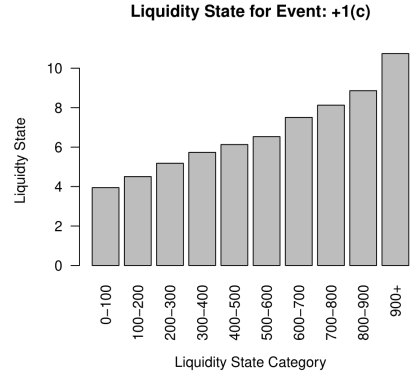
As we can observe, the above estimated functions are consistent with the 1st-ask and 1st-bid similarity patterns discussed in section 5.2 and section E, when the bin-size is enlarged.

I.2 liquidity state

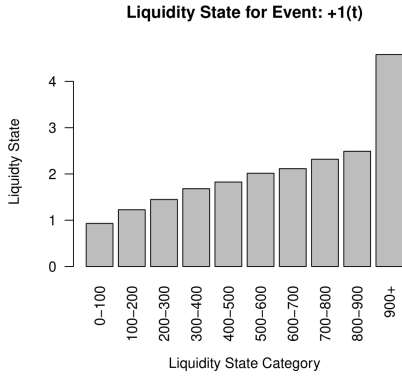
Based on Figure S5 and Figure S6 in section 5.4, the following Figure S21 and Figure S22 demonstrate the liquidity state estimations of the model when the bin-size Δ is enlarged from 0.25 seconds to 0.5 seconds.



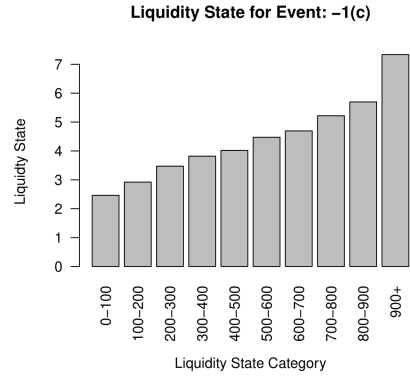
(a) Liquidity state for +1(i)



(b) Liquidity state for +1(c)

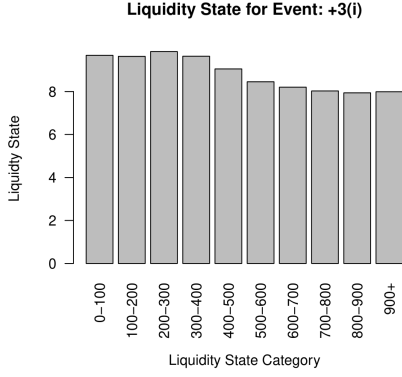


(c) Liquidity state for +1(t)

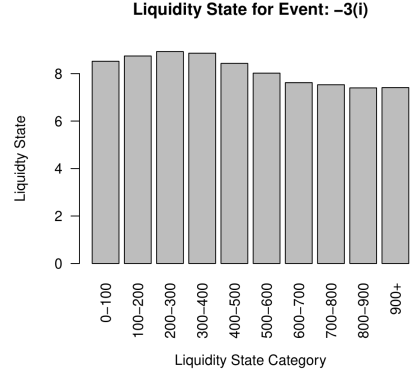


(d) Liquidity state for -1(c)

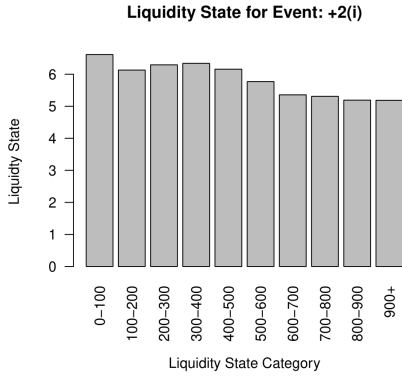
Figure S21: Aggregated estimation result for liquidity state for event +1(i), +1(c), +1(t), -1(c) under ($s = 20$ seconds, $\Delta = 0.5$ seconds). For these events the event arrival intensity increases as liquidity state increases.



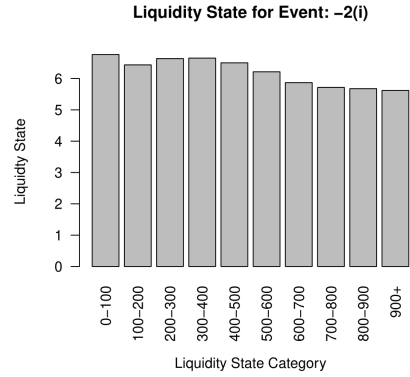
(a) Liquidity state for +3(i)



(b) Liquidity state for -3(i)



(c) Liquidity state for +2(i)



(d) Liquidity state for -2(i)

Figure S22: Aggregated estimation result for liquidity state for event +3(i), -3(i), +2(i), -2(i) under ($s = 20$ seconds, $\Delta = 0.25$ seconds). For these events the event arrival intensity decreases as liquidity state increases.

The demonstrated liquidity state estimation results of the model with LASSO regularization is consistent with the results discussed in section 5.4 when the bin-size is enlarged.

I.3 time factor

Based on Figure 9 in section 5.5, the following Figure S23 demonstrates the time factor estimations of the model with LASSO regularization.

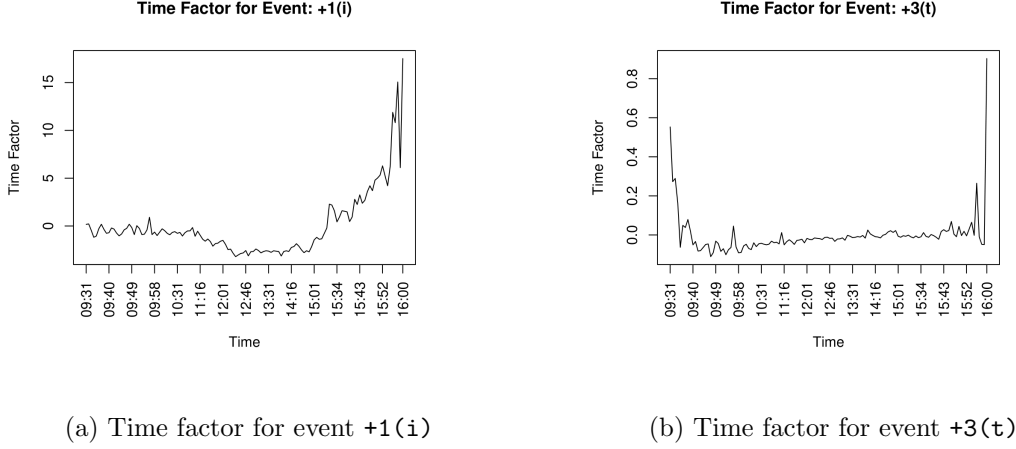


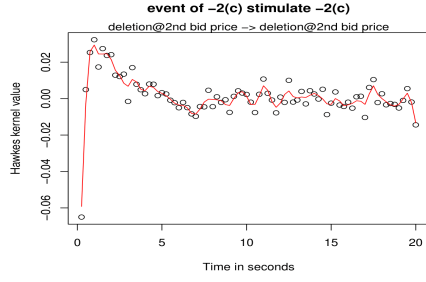
Figure S23: Aggregated estimation result for time factor between 9:30 am and 4:00 pm under ($s = 20$ seconds, $\Delta = 0.5$ seconds).

The demonstrated time factor estimation results of the model with LASSO regularization is consistent with the results discussed in section 5.5 when the bin-size is enlarged.

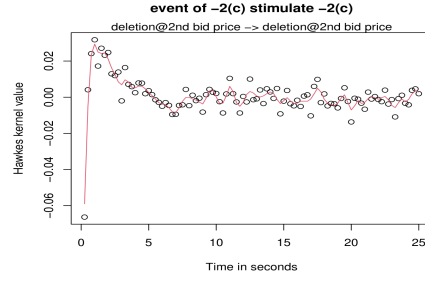
J Empirical results with extended maximum support

This supporting section demonstrates the empirical estimation results when the maximum support s is extended from 20 seconds to 25 seconds, as mentioned in section 5.7. Here we present some Hawkes kernels that do not exhibit monotonic decreasing shapes in the following Figure S24. The figure demonstrates that the estimated Hawkes kernel generally moves to zero as time elapses on the x-axis, indicating that the kernel is not explosive as we

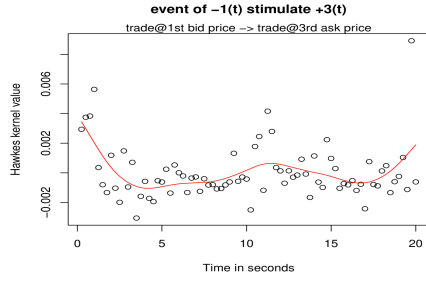
increase the maximum support. Intuitively, the figures imply that the stimulating effects gradually disappear as time goes by.



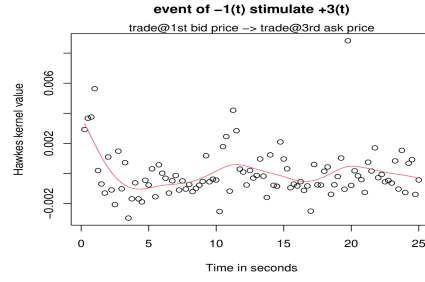
(a) $s = 20s$



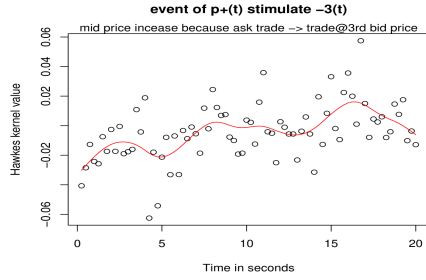
(b) $s = 25s$



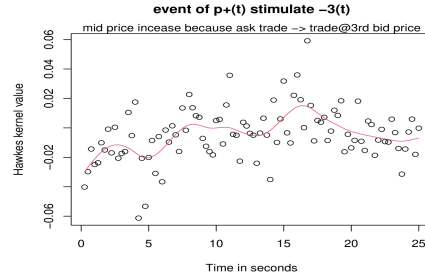
(c) $s = 20s$



(d) $s = 25s$



(e) $s = 20s$



(f) $s = 25s$

Figure S24: Aggregated Hawkes excitement function estimation under ($\Delta = 0.25$ seconds) with LASSO regularization. The subplots (a), (c), and (d) use maximum support $s = 20$ seconds. The subplots (b),(d), and (f) use maximum support $s = 25$ seconds. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

K Additional model selection results

This section presents additional model selection results based on AIC for the models when the size of order is ignored and when the bin-size is enlarged from 0.25 seconds to 0.5 seconds. These selection results are presented in the same format in Figure 11 and Table 2. Furthermore, this section also presents the model fit with different LASSO regularization parameters ($\lambda = 0.001$ and $\lambda = 0.00025$) in Table S3 and Table S4, which serves as additional sensitivity analysis of the model.

The following Figure S25 and Table S1 demonstrate the model selection result when the size of order is ignored. Model ①-⑦ are explained in section 5.8.

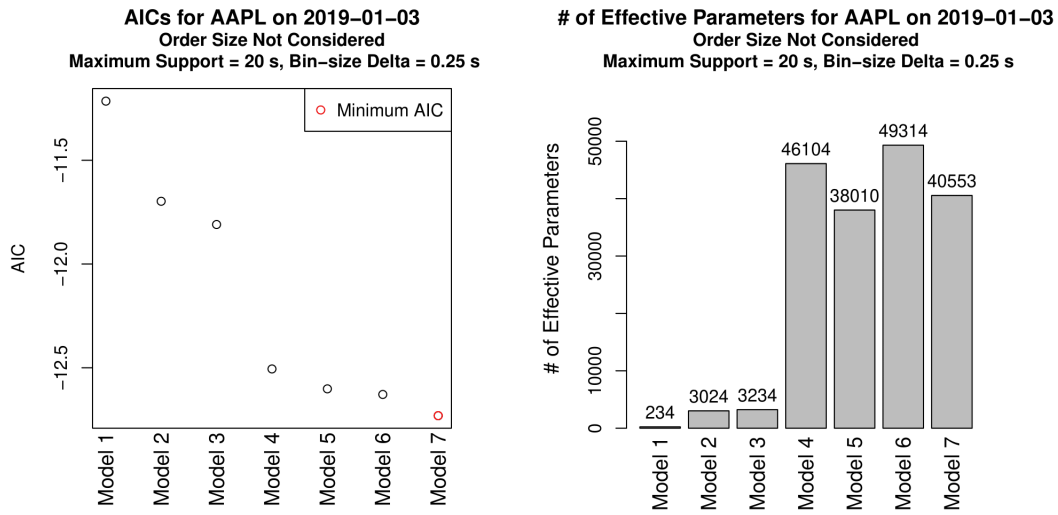


Figure S25: AICs and number of effective parameters for seven model types for Apple.Inc on 2019-01-03 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.25$ seconds. All orders are considered to have size 1. The LASSO regularization parameter is chosen as $\lambda = 0.0005$.

AIC Difference	Min	1st Quantile	Median	Mean	3rd Quantile	Max	# of days with decreased AIC
④ - ③ ¹	-1.43	-0.30	-0.15	-0.23	-0.005	0.13	15 out of 20
⑥ - ④ ²	-0.23	-0.19	-0.18	-0.18	-0.16	-0.12	20 out of 20
⑤ - ④ ³	-0.15	-0.12	-0.11	-0.10	-0.08	-0.02	20 out of 20
⑤ - ③ ⁴	-1.46	-0.38	-0.26	-0.33	-0.12	-0.016	20 out of 20
⑦ - ⑥ ⁵	-0.16	-0.12	-0.11	-0.10	-0.09	-0.02	20 out of 20

Interpretations:

¹ The Hawkes part has stronger explanation power than the liquidity state and time factor part.

² Adding the liquidity state and time factor to the Hawkes part further improves explanation power.

³ Adding LASSO (LASSO parameter 0.0005) to the Hawkes part further improves explanation power.

⁴ The Hawkes part with LASSO (LASSO parameter 0.0005) generates stronger explanation power than the liquidity state and time factor part.

⁵ Adding LASSO (LASSO parameter 0.0005) further improves the explanation power of the model with the liquidity state, time factor, and Hawkes.

Table S1: AIC difference summary statistics of Apple. Inc from 2019-01-02 to 2019-01-31 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.25$ seconds. All orders are considered to have size 1. AIC has been adjusted for sample size so that it reflects the AIC per single sample. The LASSO regularization parameter is chosen as $\lambda = 0.0005$.

The following Figure S26 and Table S2 demonstrate the model selection result when the bin size is enlarged from 0.25 seconds to 0.5 seconds. Model ①-⑦ are explained in section 5.8.

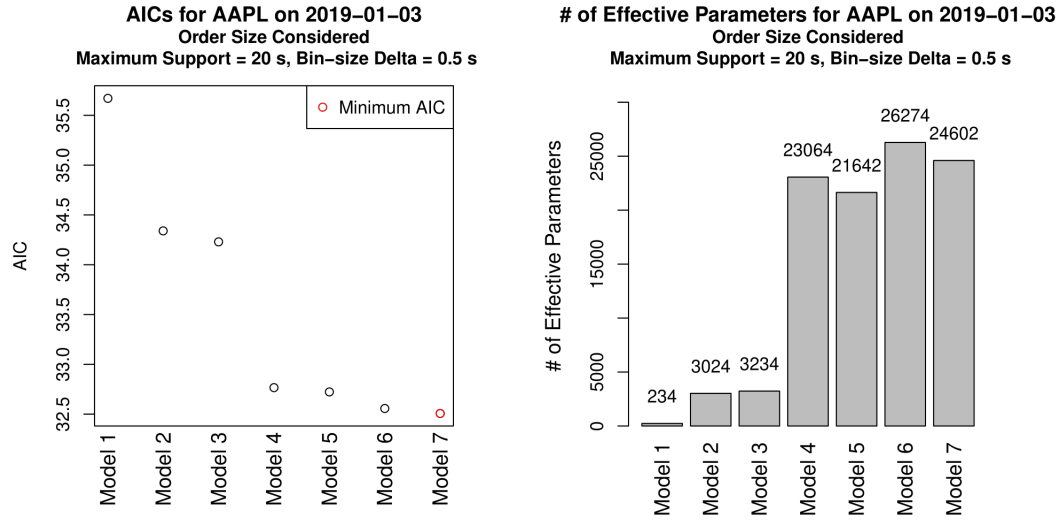


Figure S26: AICs and number of effective parameters for seven model types for Apple.Inc on 2019-01-03 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.5$ seconds. Order sizes are considered in bin count sequence construction. The LASSO regularization parameter is chosen as $\lambda = 0.0005$.

AIC Difference	Min	1st Quantile	Median	Mean	3rd Quantile	Max	# of days with decreased AIC
④ - ③ ¹	-4.25	-1.19	-0.61	-0.88	-0.18	0.06	18 out of 20
⑥ - ④ ²	-0.4	-0.31	-0.29	-0.28	-0.24	-0.18	20 out of 20
⑤ - ④ ³	-0.09	-0.06	-0.06	-0.05	-0.04	-0.03	20 out of 20
⑤ - ③ ⁴	-4.29	-1.26	-0.65	-0.94	-0.23	0.001	19 out of 20
⑦ - ⑥ ⁵	-0.09	-0.07	-0.06	-0.06	-0.05	-0.03	20 out of 20

Interpretations:

¹ The Hawkes part has stronger explanation power than the liquidity state and time factor part.

² Adding the liquidity state and time factor to the Hawkes part further improves explanation power.

³ Adding LASSO (LASSO parameter 0.0005) to the Hawkes part further improves explanation power.

⁴ The Hawkes part with LASSO (LASSO parameter 0.0005) generates stronger explanation power than the liquidity state and time factor part.

⁵ Adding LASSO (LASSO parameter 0.0005) further improves the explanation power of the model with the liquidity state, time factor, and Hawkes.

Table S2: AIC difference summary statistics of Apple. Inc from 2019-01-02 to 2019-01-31 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.5$ seconds. Order sizes are considered in bin count sequence construction. AIC has been adjusted for sample size so that it reflects the AIC per single sample. The LASSO regularization parameter is chosen as $\lambda = 0.0005$.

Besides, the following Table S3 and Table S4 demonstrate the AIC difference summary statistics for the estimated model with larger LASSO regularization parameters ($\lambda = 0.001$ and $\lambda = 0.00025$). All other parameters are the same as the ones presented in Figure 11

(support = 20s, bin size $\Delta = 0.25$ s).

AIC Difference	Min	1st Quantile	Median	Mean	3rd Quantile	Max	# of days with decreased AIC
④ - ③ ¹	-4.09	-1.19	-0.58	-0.89	-0.16	0.09	18 out of 20
⑥ - ④ ²	-0.31	-0.24	-0.23	-0.22	-0.18	-0.13	20 out of 20
⑤ - ④ ³	-0.14	-0.09	-0.08	-0.07	-0.07	-0.06	19 out of 20
⑤ - ③ ⁴	-4.16	-1.28	-0.60	-0.97	-0.24	-0.02	20 out of 20
⑦ - ⑥ ⁵	-0.15	-0.09	-0.08	-0.07	-0.06	0.04	19 out of 20

Interpretations:

¹ The Hawkes part has stronger explanation power than the liquidity state and time factor part.

² Adding the liquidity state and time factor to the Hawkes part further improves explanation power.

³ Adding LASSO (LASSO parameter 0.001) to the Hawkes part further improves explanation power.

⁴ The Hawkes part with LASSO (LASSO parameter 0.001) generates stronger explanation power than the liquidity state and time factor part.

⁵ Adding LASSO (LASSO parameter 0.001) further improves the explanation power of the model with the liquidity state, time factor, and Hawkes.

Table S3: AIC difference summary statistics of Apple. Inc from 2019-01-02 to 2019-01-31. Maximum Support $s = 20$ s, bin-size $\Delta = 0.25$ s. Order sizes are considered in bin count sequence construction. AIC has been adjusted for sample size so that it reflects the AIC per single sample. The LASSO regularization parameter is chosen as $\lambda = 0.001$.

AIC Difference	Min	1st Quantile	Median	Mean	3rd Quantile	Max	# of days with decreased AIC
④ - ③ ¹	-4.09	-1.19	-0.58	-0.89	-0.16	0.09	18 out of 20
⑥ - ④ ²	-0.31	-0.24	-0.23	-0.22	-0.18	-0.13	20 out of 20
⑤ - ④ ³	-0.09	-0.07	-0.06	-0.06	-0.05	-0.04	19 out of 20
⑤ - ③ ⁴	-4.14	-1.26	-0.64	-0.96	-0.21	0.02	19 out of 20
⑦ - ⑥ ⁵	-0.10	-0.07	-0.07	-0.07	-0.06	0.05	20 out of 20

Interpretations:

¹ The Hawkes part has stronger explanation power than the liquidity state and time factor part.

² Adding the liquidity state and time factor to the Hawkes part further improves explanation power.

³ Adding LASSO (LASSO parameter 0.00025) to the Hawkes part further improves explanation power.

⁴ The Hawkes part with LASSO (LASSO parameter 0.00025) generates stronger explanation power than the liquidity state and time factor part.

⁵ Adding LASSO (LASSO parameter 0.00025) further improves the explanation power of the model with the liquidity state, time factor, and Hawkes.

Table S4: AIC difference summary statistics of Apple. Inc from 2019-01-02 to 2019-01-31. Maximum Support $s = 20s$, bin-size $\Delta = 0.25s$. Order sizes are considered in bin count sequence construction. AIC has been adjusted for sample size so that it reflects the AIC per single sample. The LASSO regularization parameter is chosen as $\lambda = 0.00025$.

L Goodness-of-fit evaluation

The following Figure S27 shows the Q-Q (quantile-quantile) plot for the estimation to evaluate goodness-of-fit.

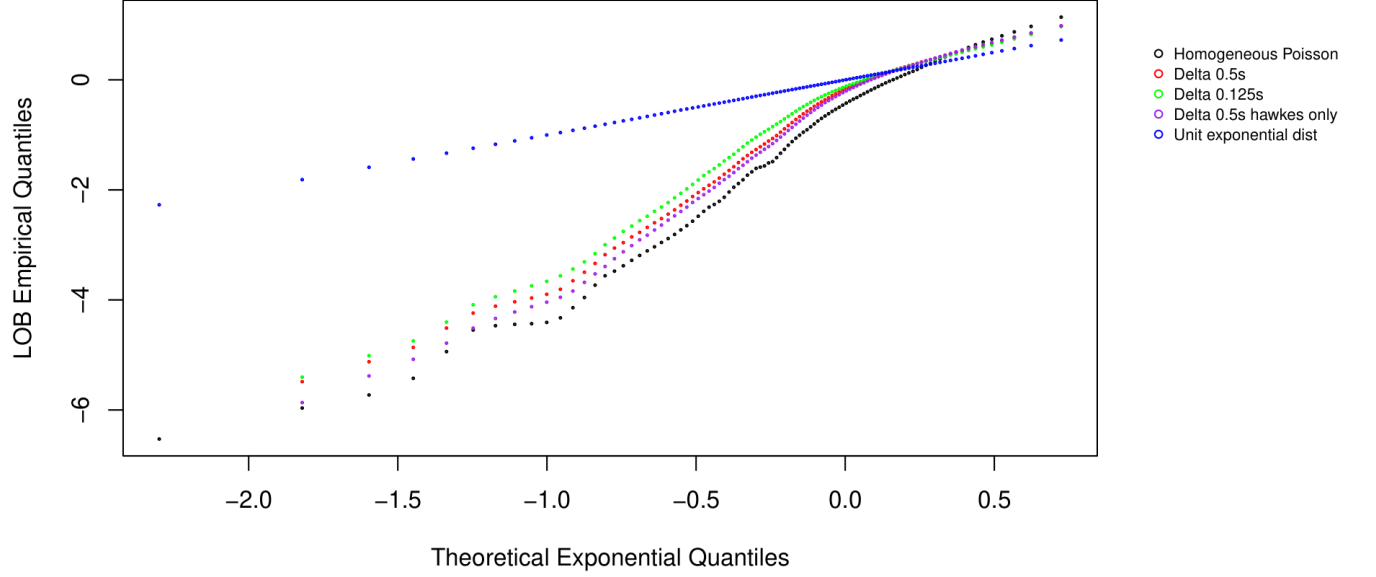


Figure S27: The figure demonstrates the Q-Q (quantile-quantile) plot of the LOB estimated based on maximum support $s = 20s$ for selected events. All order sizes are considered to be one. The x-axis plots the log quantiles of the standard exponential distribution. The y-axis plots the log quantiles of the rescaled interval time based on the estimation using real LOB data. The blue dots show the distribution of the standard exponential random variables. The red dots show the distribution with bin-size $\Delta = 0.5s$. The green dots show the distribution with bin-size $\Delta = 0.125s$. The purple dots show the distribution with bin-size $\Delta = 0.5s$ with fixed and state-independent baseline intensity (liquidity state and time factor estimation are dropped). The black dots show the distribution in which the order arrival follows a homogeneous Poisson model with the arrival rate equal to the average arrival rate of orders. There is no liquidity state, time factor, or Hawkes stimulating effect in the Poisson model.

According to the random time change theorem (Daley et al. 2003), the transformed time $\Lambda(t_1), \dots, \Lambda(t_k)$ should follow a Poisson process with intensity 1, given a point process $t_1, \dots, t_k$ with varying intensity $\lambda(\cdot)$. The change time is given by $\Lambda_i = \int_0^t \lambda_i(s) ds$, in which the index i represents the event types. This implies that the scaled interval time $\Lambda(t_2) - \Lambda(t_1), \dots, \Lambda(t_k) - \Lambda(t_{k-1})$ should follow a standard exponential distribution. Therefore, the goodness-of-fit of our model can be tested by comparing the distribution of the rescaled interval time and that of a standard exponential distribution using the Q-Q plot.

The above Q-Q plot implies several facts. First, our proposed model (high-dimensional Hawkes process with state-dependent baseline intensity) overall outperforms the simple Poisson model, as we can observe the red, purple, and green dots lie closer to the standard exponential distribution. Second, we observe that our proposed model achieves better performance compared to the specification when the estimation of liquidity state and time factor is dropped since the red dots lie closer to the standard exponential distribution compared to the purple dots. This observation indicates the explanatory power of the state-dependent baseline intensity estimated using the order book state and event arrival time. Third, we observe that our proposed model achieves better goodness-of-fit with a relatively smaller bin-size Δ , since the same model with $\Delta = 0.125s$ significantly outperforms the model with $\Delta = 0.5s$. A smaller bin-size Δ tends to increase the precision of the estimation and can better capture more grandeur-level fluctuations of the order book dynamics. Hence, we expect our model to achieve better goodness-of-fit with an even smaller bin-size Δ , which hasn't been tested yet since the limitation on computational budget.

Overall, the goodness-of-fit evaluations using the Q-Q plot meet the model-selection result using AIC presented in section 4.5. It is a limitation that our model doesn't achieve perfect goodness-of-fit, and this could be due to the unobserved features in the complex dynamics of the order book. For example, we only include the first three LOB levels

($K = 3$) in our estimation due to the constraints in computational power, and thus the effects of LOB events (insertion/cancellation/trade) higher than level-3 are not considered in the estimation. We anticipate the model fitness can be improved by reducing the bin-size Δ relative to the current model specification, and further increasing the LOB level included in the estimation.