Order Book Queue Hawkes Markovian Modeling: Supplementary Material

Philip Protter *
Department of Statistics, Columbia University
and
Qianfan Wu
Kelley School of Business, Indiana University Bloomington
and
Shihao Yang†
H. Milton Stewart School of Industrial and Systems Engineering.

December 16, 2023

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

* *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.*

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 $p_{ref} = p_{mid} = (Q_{\text{best-bid}} + Q_{\text{best-ask}})/2$; When $(Q_{\text{best-ask}} - Q_{\text{best-bid}}) = 2n, n \in \mathbb{Z}$, we have $p_{ref} = (Q_{best-bid} + Q_{best-ask})/2 + \alpha/2$ or $p_{ref} = (Q_{best-bid} + Q_{best-ask})/2 - \alpha/2$, 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 $(\forall i = 1, 2, \ldots, d)$ 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:*

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

, 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 $[L_1, L_2, \ldots, L_{10}]$ and $t_k^{(\Delta)}$ can be bucketized into 126 categories $[T_1, T_2, \ldots, T_{126}]$ 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:

$$\left(\hat{\Phi}_1^{(\Delta,s)}, \hat{\Phi}_2^{(\Delta,s)}, ..., \hat{\Phi}_p^{(\Delta,s)}, \hat{\mu}_1^{(\Delta,s)}, \ldots, \hat{\mu}_{10}^{(\Delta,s)}, \hat{\theta}_1^{(\Delta,s)}, \ldots, \hat{\theta}_{126}^{(\Delta,s)}\right) := \hat{\Phi}^{(\Delta,s)} \in \mathbb{R}^{d\times(dp+10+126)},$$
$$\textit{with } p := \lfloor s/\Delta \rfloor \tag{1}$$

Specifically,

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

, where the matrix element

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

are weakly consistent estimators for the Multivaraite Hawkes excitement function for event j stimulating event i at the r-th function discretizated 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 $T \to \infty, \Delta \to 0$ and $s = \Delta p \to \infty$) for $\phi_{ji}(t)$ as shown in Eq.(3).

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

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:*

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

The CLS estimator is defined as

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

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

,where

$$Z\left(B_1^{(\Delta)},\ldots,B_n^{(\Delta)};l_1^{(\Delta)},\ldots,l_n^{(\Delta)};t_1^{(\Delta)},\ldots,t_n^{(\Delta)}\right) :=$$

$$\begin{bmatrix}
B_p^{(\Delta)} & B_{p+1}^{(\Delta)} & \ldots & B_{n-1}^{(\Delta)}\\
B_{p-1}^{(\Delta)} & B_p^{(\Delta)} & \ldots & B_{n-2}^{(\Delta)}\\
\ldots & & & \\
B_1^{(\Delta)} & B_2^{(\Delta)} & \ldots & B_{n-p}^{(\Delta)}\\
\mathbb{1}_{l_{p+1}^{(\Delta)}\in L_1} & \mathbb{1}_{l_{p+2}^{(\Delta)}\in L_1} & \ldots & \mathbb{1}_{l_n^{(\Delta)}\in L_1}\\
\mathbb{1}_{l_{p+1}^{(\Delta)}\in L_2} & \mathbb{1}_{l_{p+2}^{(\Delta)}\in L_2} & \ldots & \mathbb{1}_{l_n^{(\Delta)}\in L_2}\\
\ldots & & & \\
\mathbb{1}_{l_{p+1}^{(\Delta)}\in L_{10}} & \mathbb{1}_{l_{p+2}^{(\Delta)}\in L_{10}} & \ldots & \mathbb{1}_{l_n^{(\Delta)}\in L_{10}}\\
[0]_{d\times 1} & [0]_{d\times 1} & \ldots & [0]_{d\times 1}\\
\mathbb{1}_{t_{p+1}^{(\Delta)}\in T_2} & \mathbb{1}_{t_{p+2}^{(\Delta)}\in T_2} & \ldots & \mathbb{1}_{t_n^{(\Delta)}\in T_2}\\
\mathbb{1}_{t_{p+1}^{(\Delta)}\in T_3} & \mathbb{1}_{t_{p+2}^{(\Delta)}\in T_3} & \ldots & \mathbb{1}_{t_n^{(\Delta)}\in T_3}\\
\ldots & & & \\
\mathbb{1}_{t_{p+1}^{(\Delta)}\in T_{126}} & \mathbb{1}_{t_{p+2}^{(\Delta)}\in T_{126}} & \ldots & \mathbb{1}_{t_n^{(\Delta)}\in T_{126}}
\end{bmatrix} \in \mathbb{R}^{(dp+10+126)\times(n-p)}$$

is the design matrix and

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

is the response.

Within the design matrix Z, $\left(\mathbb{1}_{l_k^{(\Delta)}\in L_1},\ldots,\mathbb{1}_{l_k^{(\Delta)}\in L_{10}}\right)$ *and* $\left(\mathbb{1}_{t_k^{(\Delta)}\in L_1},\ldots,\mathbb{1}_{l_k^{(\Delta)}\in L_{126}}\right)$ *with* $\forall k = p+1,\ldots,n$ *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:*

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

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

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

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

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

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

In model implementation, we prefer conducting the OLS estimations based on $y_i = Z^\top \phi_i^{(\Delta,s)} + u_i$ over all dimensions $1, 2, \ldots, d$, 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 $\phi_i^{(\Delta,s),excitements} := \left[\left(\hat{\phi}_{i1,1}^{(\Delta,s)}, \ldots, \hat{\phi}_{id,1}^{(\Delta,s)}\right), \ldots, \left(\hat{\phi}_{i1,p}^{(\Delta,s)}, \ldots, \hat{\phi}_{id,p}^{(\Delta,s)}\right) \right]^\top \in \mathbb{R}^{dp \times 1}$ 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:

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

, 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.

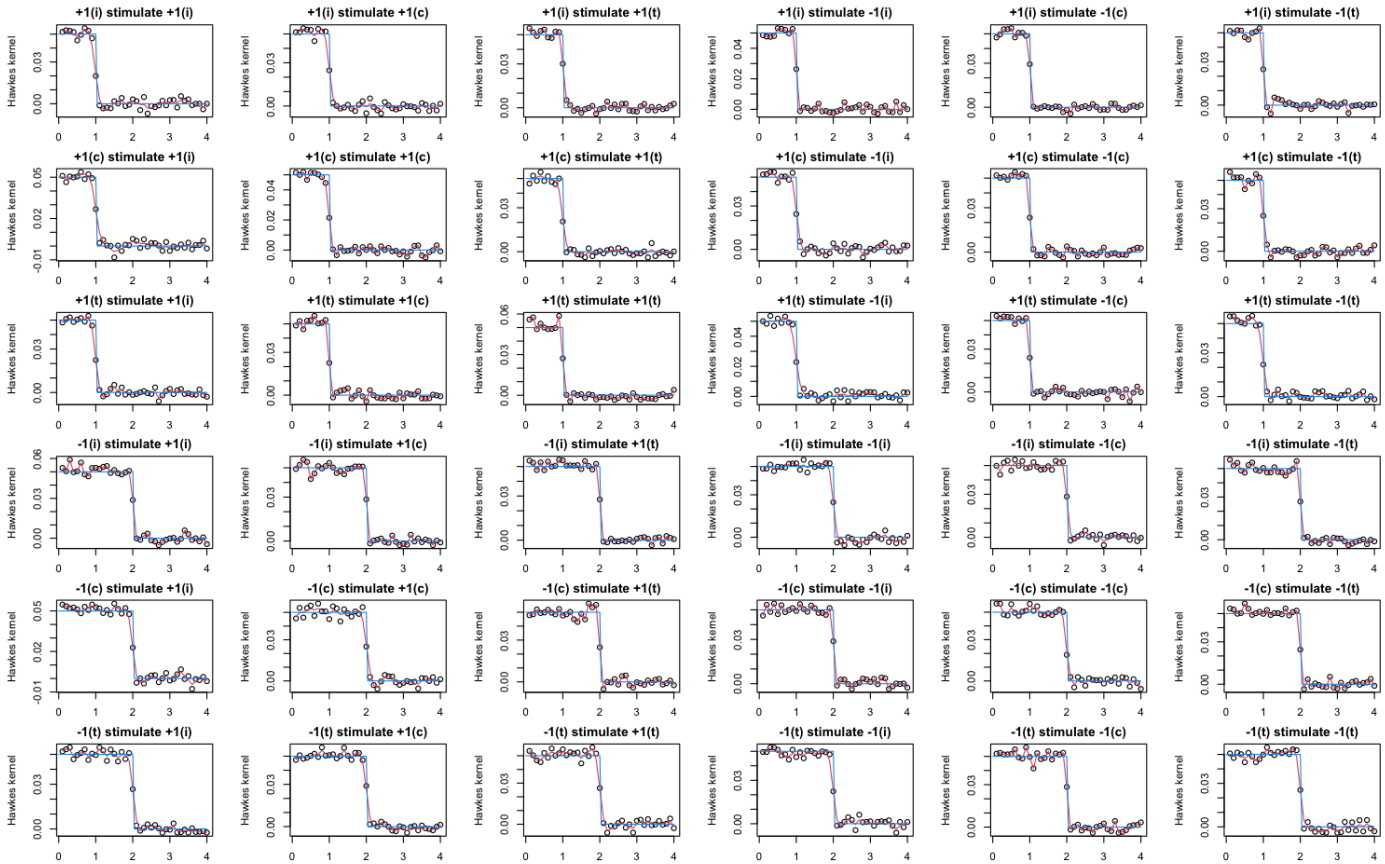


Figure S1: Simulation and estimation results of Hawkes kernel based on simulated level-1 order book with 6 events. The points illustrate the discrete estimator of the Hawkes stimulating function. The red line illustrates the smoothing spline for the points. The blue line illustrates the true stimulating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

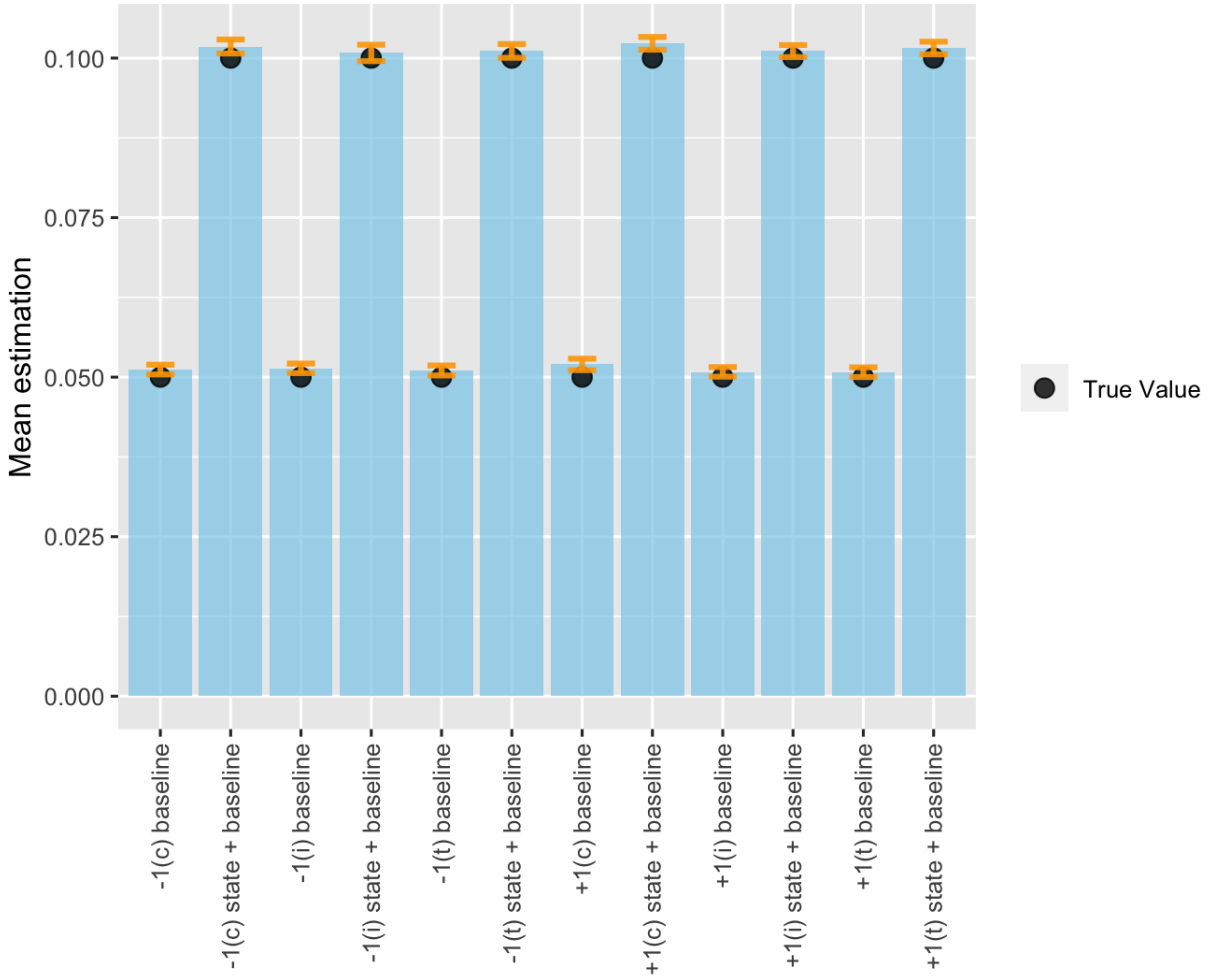


Figure S2: Simulation and estimation results of Hawkes process baseline intensity state based on simulated level-1 order book with 6 events. The blue bar indicates the average estimated value of the state parameters. The orange bar indicates the 95% confidence interval of the estimated values. The black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

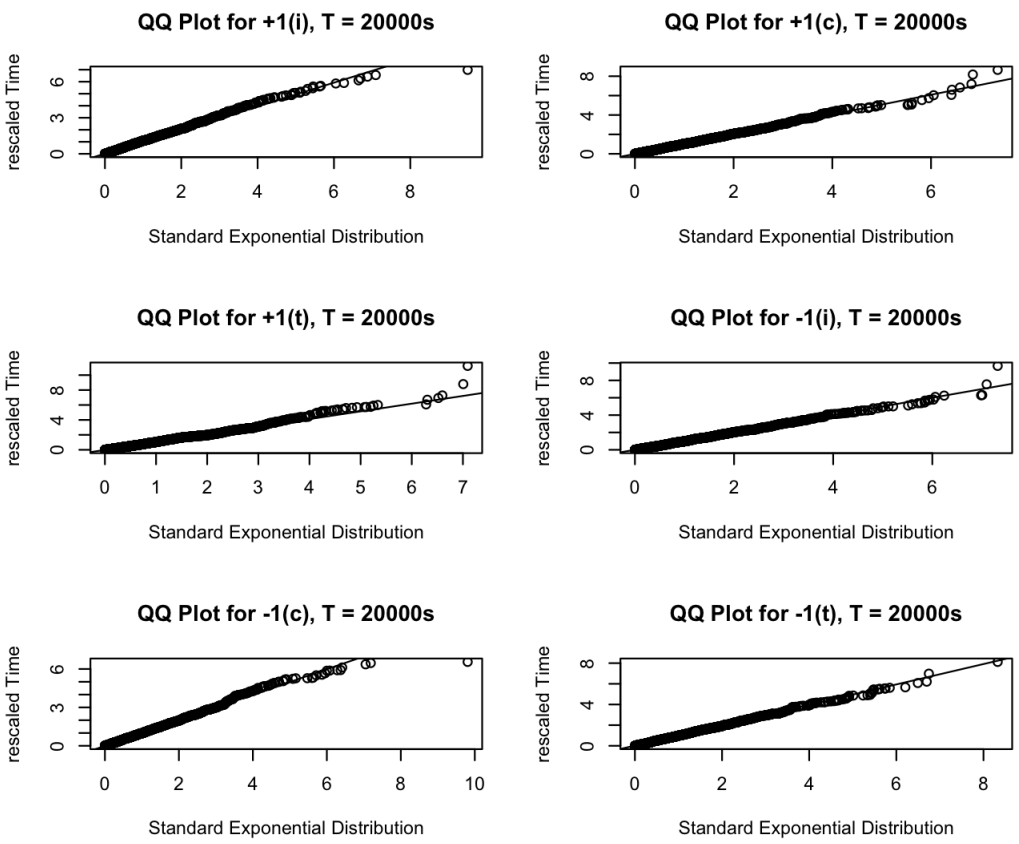
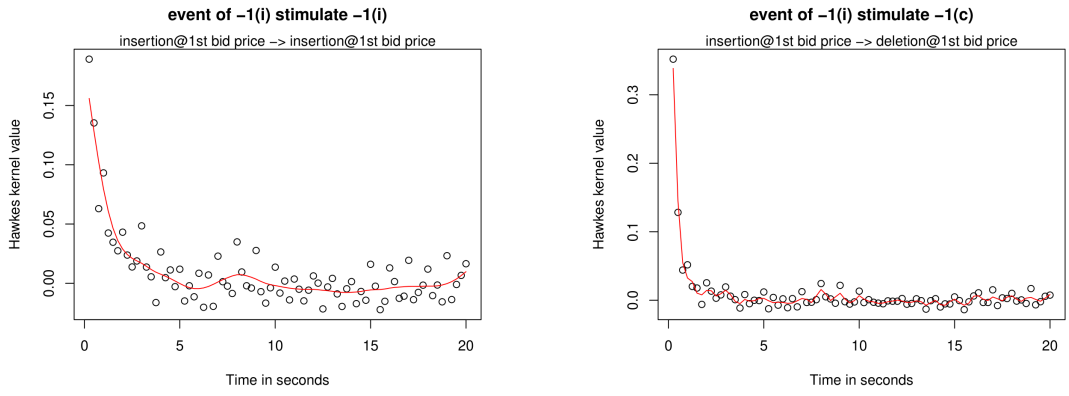


Figure S3: Q-Q (quantile-quantile) plot of the estimation results based on simulated level-1 order book with 6 events. The points indicate the empirical distribution quantiles of the rescaled time derived from the estimated model parameters. The black line indicates the quantiles of the standard exponential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



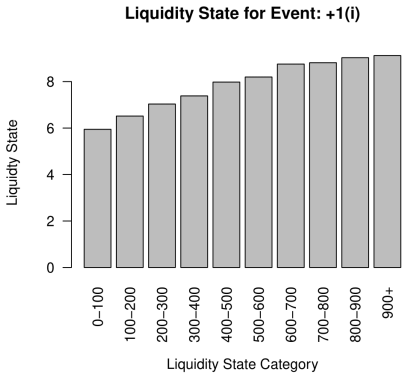
(a) `-1(i)` stimulate `-1(i)`

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

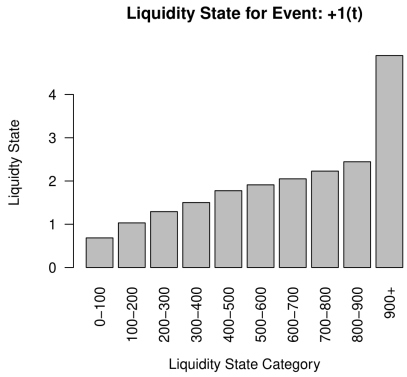
Figure S4: Aggregated Hawkes excitement function estimation result under (s = 20 seconds, Δ = 0.25 seconds) with LASSO regularization. The points illustrate the discrete function valued estimator. The red line illustrates the cubic smoothing spline for the points.

F More examples on liquidity state

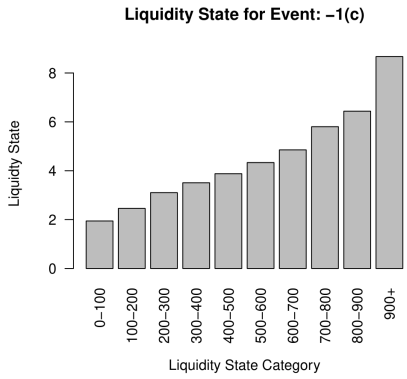
Following section 5.4 and Figure 8, this section presents more examples of the estimated result for liquidity state. The examples are presented in Figure S5 and Figure S6. In general, the arrival intensities increase as the liquidity state increases for trade/cancellation events and insertion events 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)`

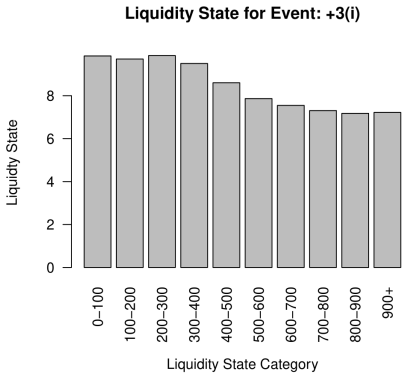


(b) Liquidity state for `+1(t)`

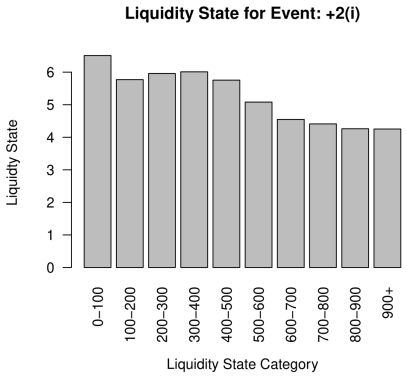


(c) Liquidity state for `-1(c)`

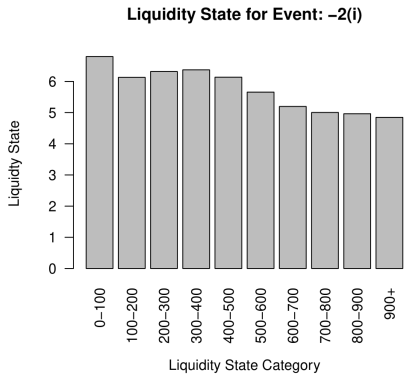
Figure S5: Aggregated estimation results for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds).



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

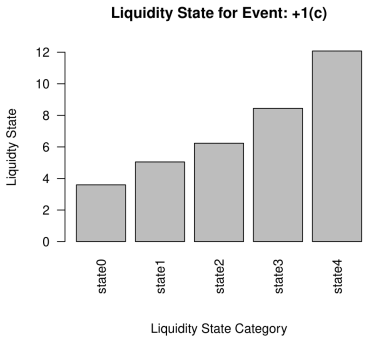


(b) Liquidity state for `+2(i)`

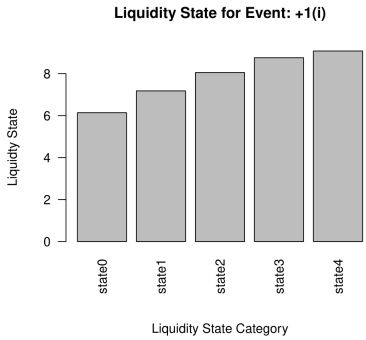


(c) Liquidity state for `-2(i)`

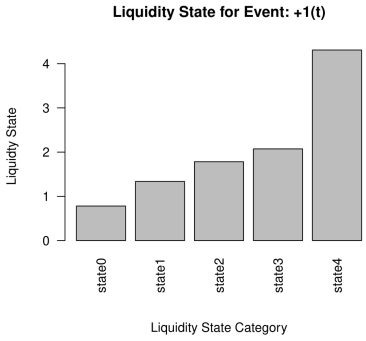
Figure S6: Aggregated estimation results for liquidity state for selected events under ($s =$ 20 seconds, $\Delta = 0.25$ seconds).



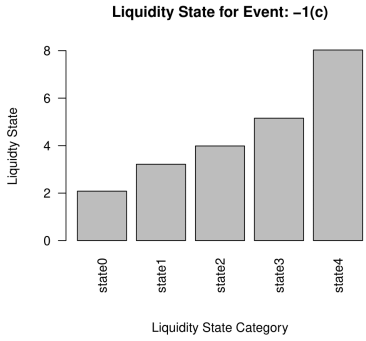
(a) Liquidity state for `+1(c)`



(b) Liquidity state for `+1(i)`

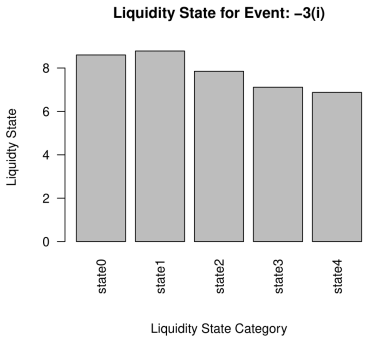


(c) Liquidity state for `+1(t)`

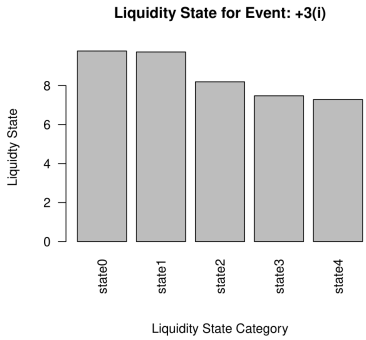


(d) Liquidity state for `-1(c)`

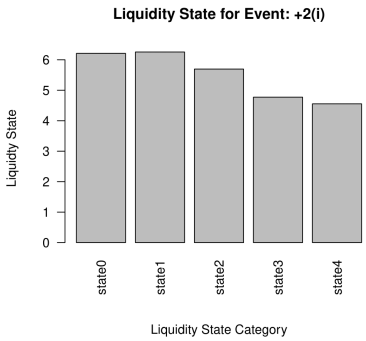
Figure S7: Aggregated estimation results for liquidity state for selected events under ($s = 20$ seconds, $\Delta = 0.25$ seconds). Compared to Figure S5, the number of liquidity state is set to be 5 for this estimation.



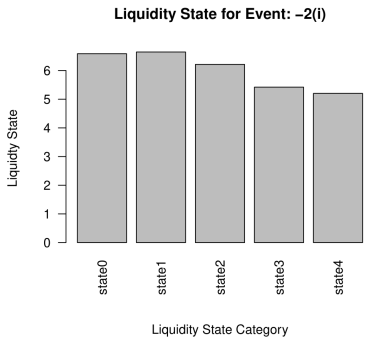
(a) Liquidity state for `-3(i)`



(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.

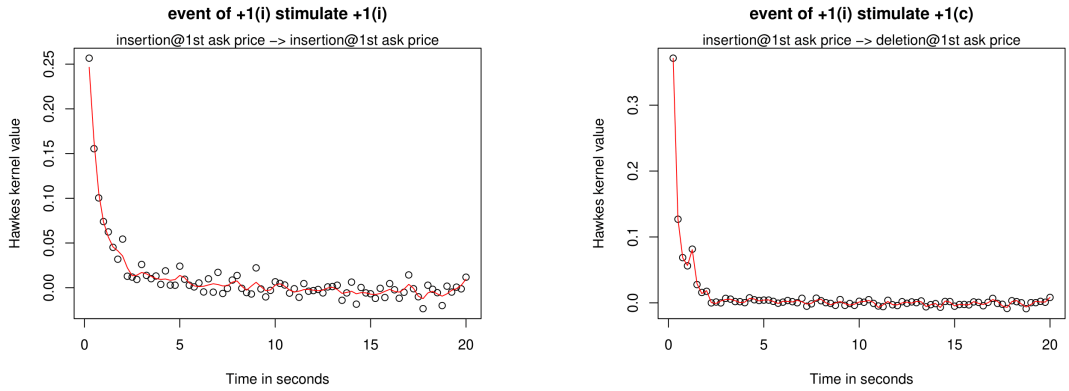
G Empirical results when order size is ignored

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

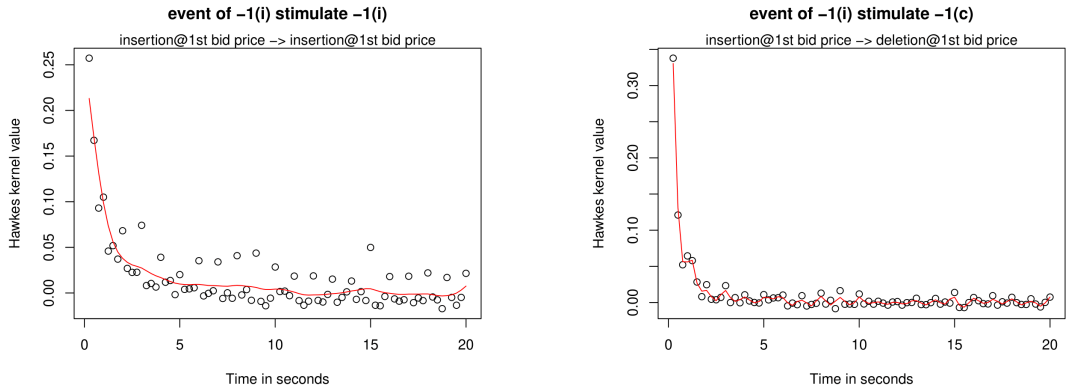
Based on Figure 6 and Figure S4, the following Figure S9 and Figure S10 demonstrate the estimated Hawkes excitement functions when the order size is ignored.



(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)`

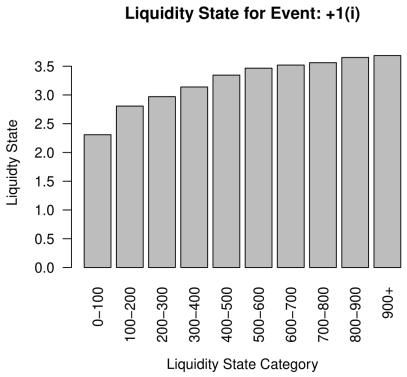
(b) `-1(i)` stimulate `-1(c)`

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.

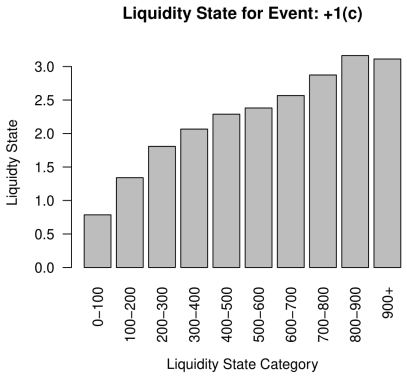
As we can observe, the above estimated functions are consistent with the 1st-ask and 1st-bid similarity patterns discussed in section 5.2 and section E.

G.2 liquidity state

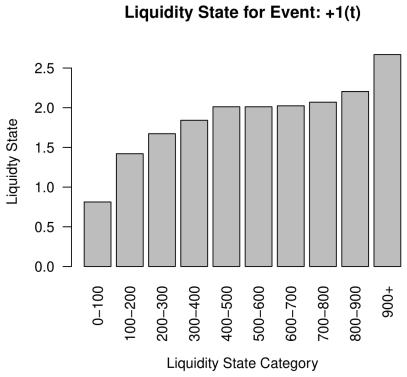
Based on Figure S5 and Figure S6 in section F, the following Figure S11 and Figure S12 demonstrate the liquidity state estimations of the model with LASSO regularization.



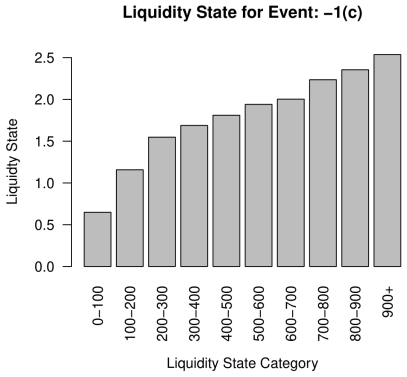
(a) Liquidity state for `+1(i)`



(b) Liquidity state for `+1(c)`

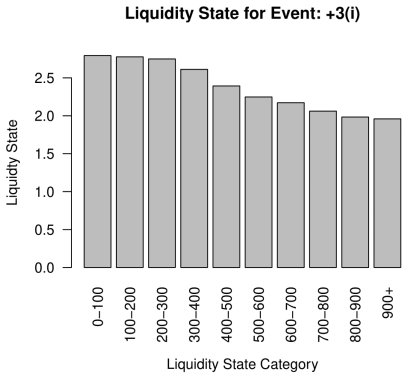


(c) Liquidity state for `+1(t)`

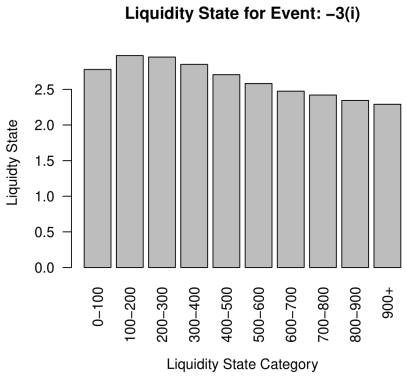


(d) Liquidity state for `-1(c)`

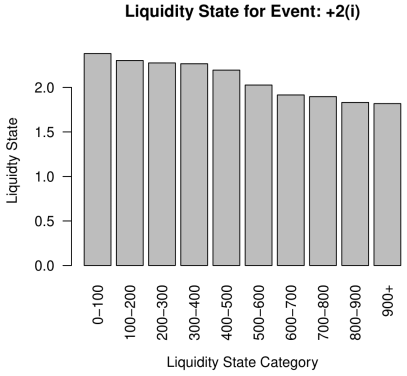
Figure S11: Aggregated estimation result for liquidity state for event `+1(i)`, `+1(c)`, `+1(t)`, `-1(c)` under ($s = 20$ seconds, $\Delta = 0.25$ seconds). All orders are considered to have size 1. For these events the event arrival intensity increases as liquidity state increases.



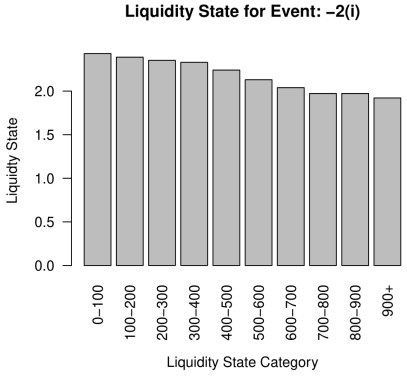
(a) Liquidity state for `+3(i)`



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



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



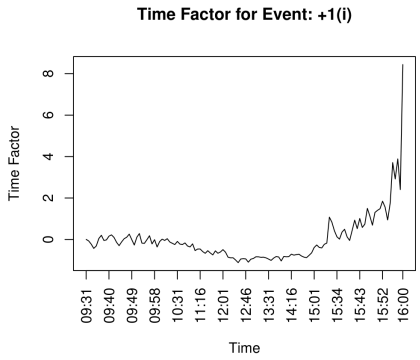
(d) Liquidity state for `-2(i)`

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.

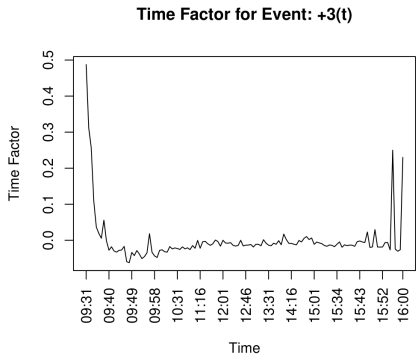
The demonstrated liquidity state estimation results of the model with LASSO regularization is consistent with the results discussed in section 5.4.

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.



(a) Time factor for event `+1(i)`



(b) Time factor for event `+3(t)`

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.

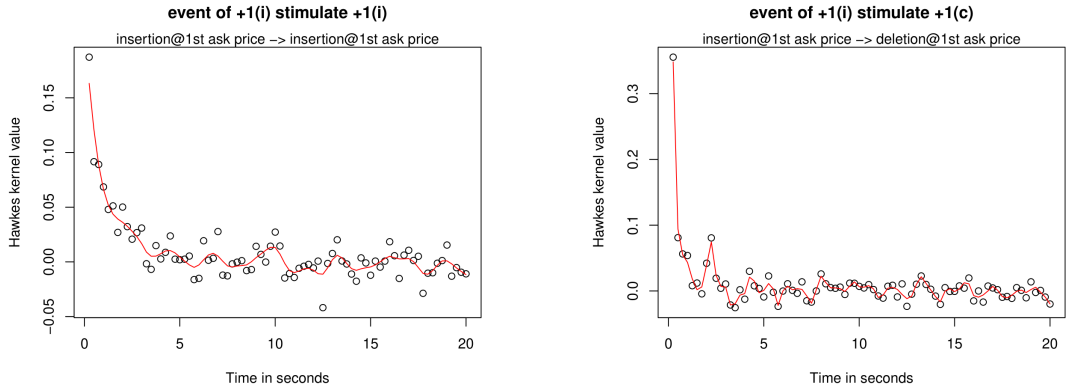
H Empirical results without LASSO

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

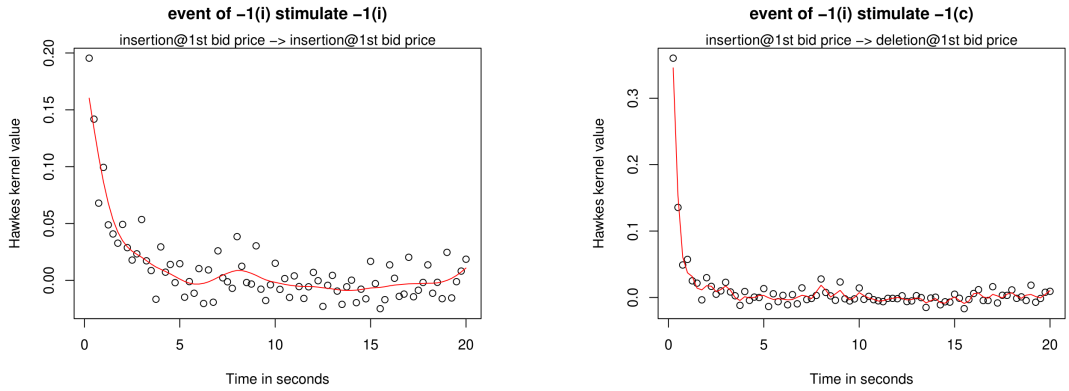
Based on Figure 6 and Figure S4, the following Figure S14 and Figure S15 demonstrate the estimated Hawkes excitement functions when the LASSO regularization is removed.



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

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

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)`

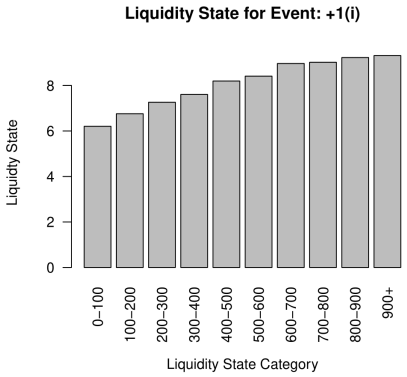
(b) `-1(i)` stimulate `-1(c)`

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.

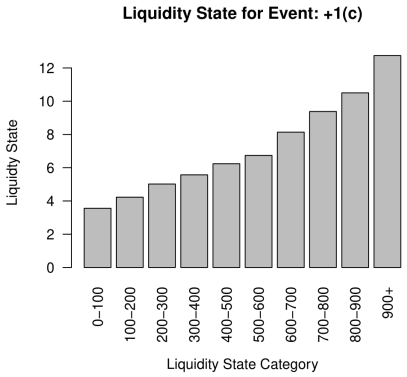
As we can observe, the above estimated functions are consistent with the 1st-ask and 1st-bid similarity patterns discussed in section 5.2 and section E, when the LASSO regularization is removed.

H.2 liquidity state

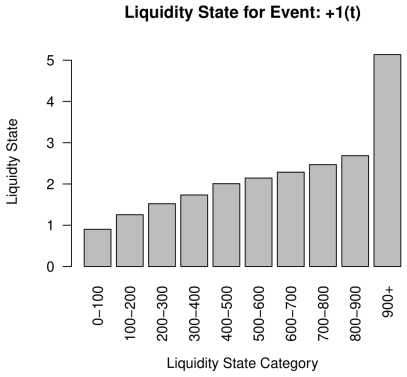
Based on Figure S5 and Figure S6 in section F, the following Figure S16 and Figure S17 demonstrate the liquidity state estimations of the model when the LASSO regularization is removed.



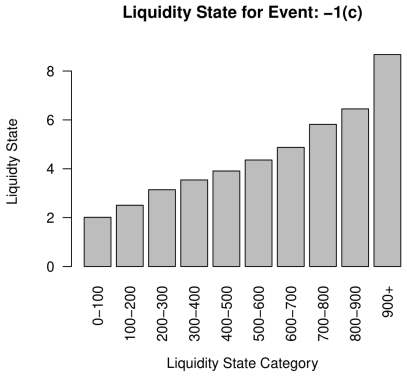
(a) Liquidity state for `+1(i)`



(b) Liquidity state for `+1(c)`

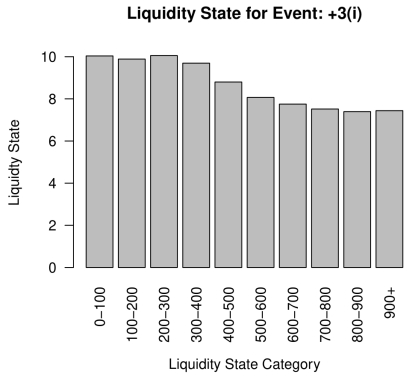


(c) Liquidity state for `+1(t)`

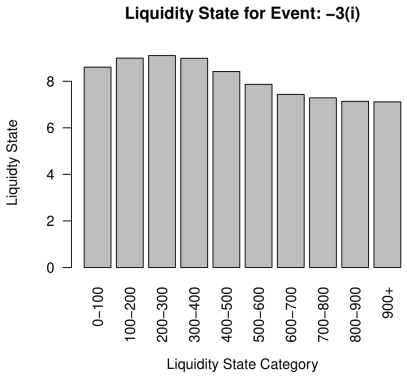


(d) Liquidity state for `-1(c)`

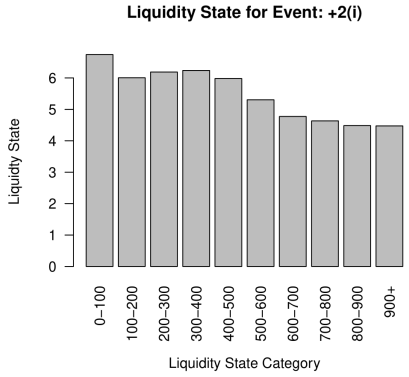
Figure S16: Aggregated estimation result for liquidity state for event `+1(i)`, `+1(c)`, `+1(t)`, `-1(c)` under ($s = 20$ seconds, $\Delta = 0.25$ seconds) when LASSO regularization is removed. For these events the event arrival intensity increases as liquidity state increases.



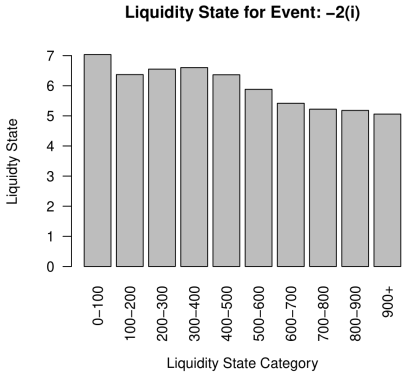
(a) Liquidity state for `+3(i)`



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



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



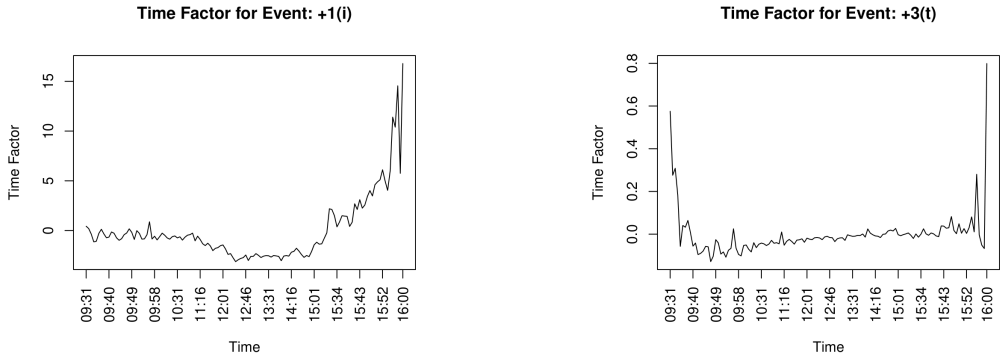
(d) Liquidity state for `-2(i)`

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 are 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.



(a) Time factor for event `+1(i)`

(b) Time factor for event `+3(t)`

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 are consistent with the results discussed in section 5.5.

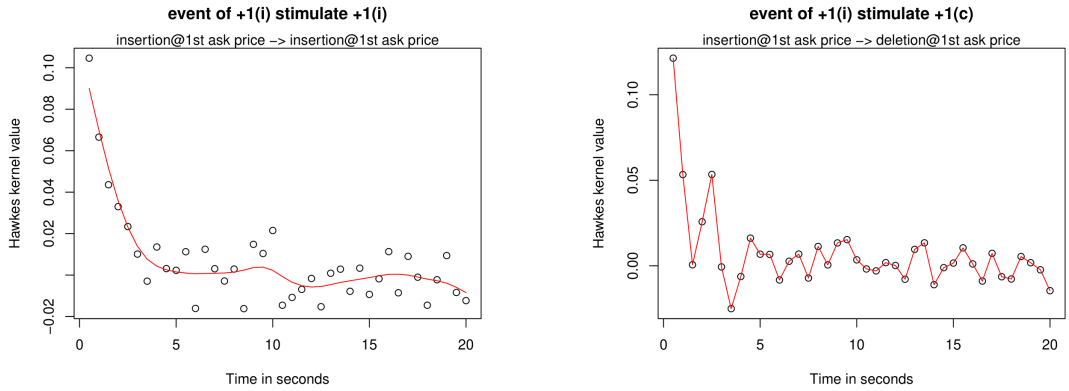
I Empirical results with enlarged bin-size

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

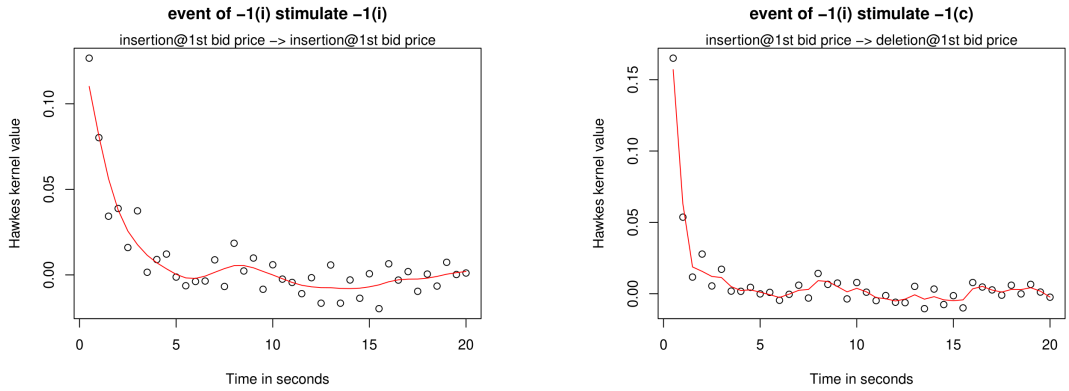
Based on Figure 6 and Figure S4, the following Figure S19 and Figure S20 demonstrate the estimated Hawkes excitement functions when the bin-size Δ is enlarged from 0.25 seconds to 0.5 seconds.



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

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

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)`

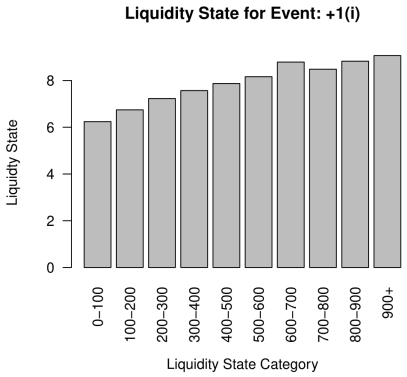
(b) `-1(i)` stimulate `-1(c)`

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.

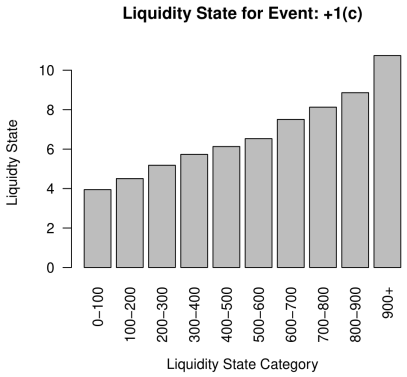
As we can observe, the above estimated functions are consistent with the 1st-ask and 1st-bid similarity patterns discussed in section 5.2 and section E, when the bin-size is enlarged.

I.2 liquidity state

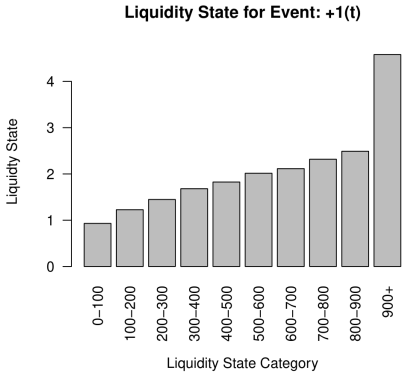
Based on Figure S5 and Figure S6 in section F, the following Figure S21 and Figure S22 demonstrate the liquidity state estimations of the model when the bin-size Δ is enlarged from 0.25 seconds to 0.5 seconds.



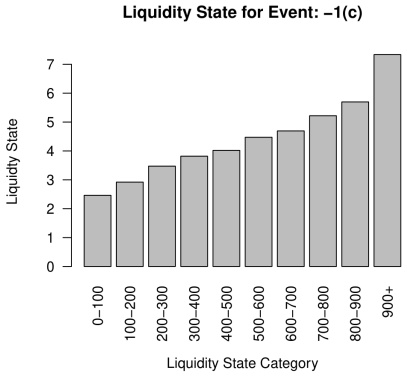
(a) Liquidity state for `+1(i)`



(b) Liquidity state for `+1(c)`

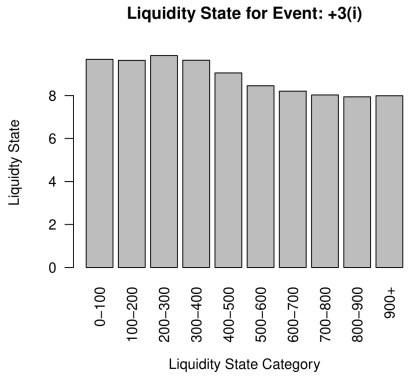


(c) Liquidity state for `+1(t)`

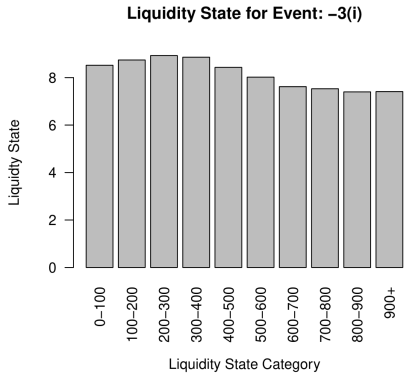


(d) Liquidity state for `-1(c)`

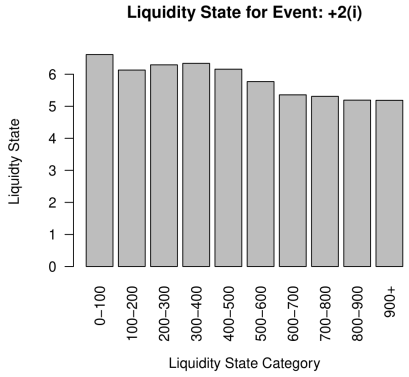
Figure S21: Aggregated estimation result for liquidity state for event `+1(i)`, `+1(c)`, `+1(t)`, `-1(c)` under ($s = 20$ seconds, $\Delta = 0.5$ seconds). For these events the event arrival intensity increases as liquidity state increases.



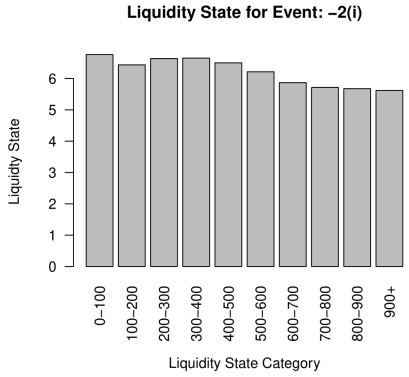
(a) Liquidity state for `+3(i)`



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



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



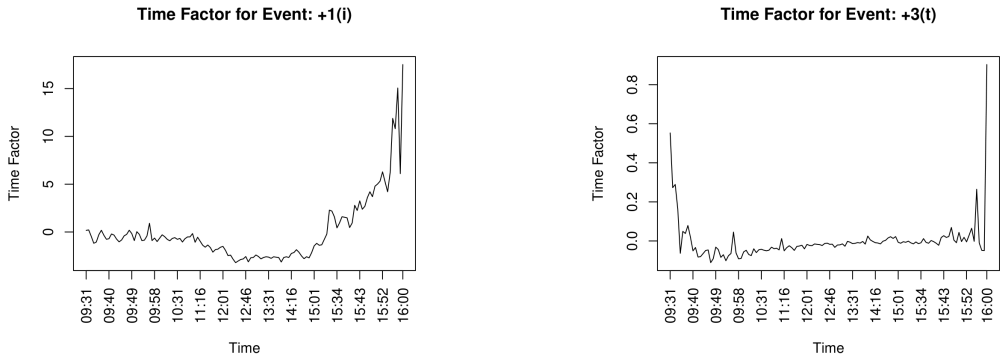
(d) Liquidity state for `-2(i)`

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.



(a) Time factor for event `+1(i)`

(b) Time factor for event `+3(t)`

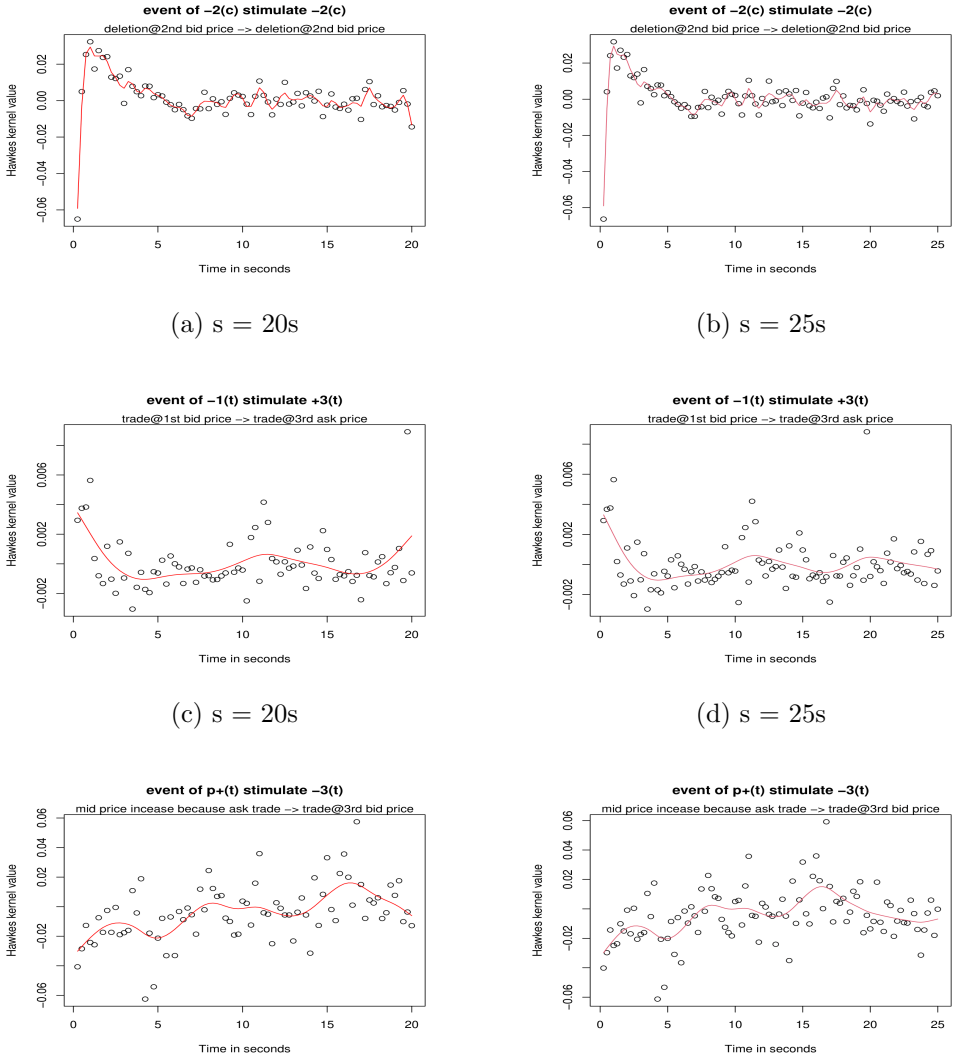
Figure S23: Aggregated estimation result for time factor between 9:30 am and 4:00 pm under ($s = 20$ seconds, $\Delta = 0.5$ seconds).

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.

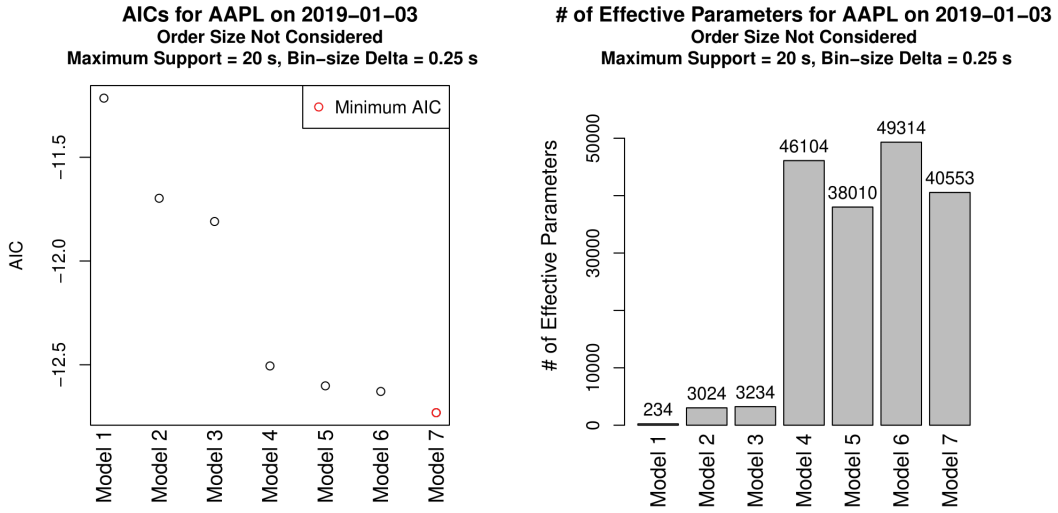


Figure S25: AICs and number of effective parameters for seven model types for Apple.Inc on 2019-01-03 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.25$ seconds. All orders are considered to have size 1. The LASSO regularization parameter is chosen as $\lambda = 0.0005$.

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.

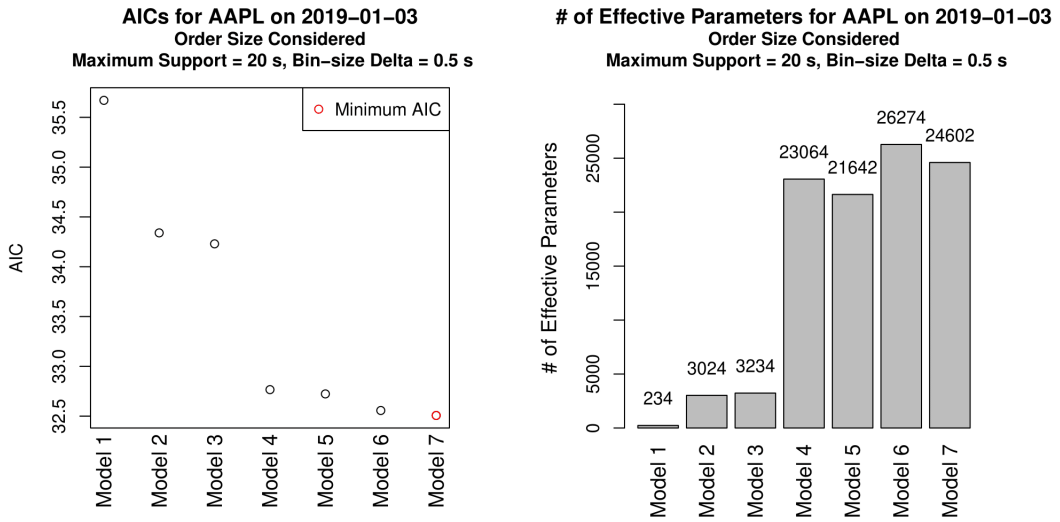


Figure S26: AICs and number of effective parameters for seven model types for Apple.Inc on 2019-01-03 with maximum support $s = 20$ seconds and bin-size $\Delta = 0.5$ seconds. Order sizes are considered in bin count sequence construction. The LASSO regularization parameter is chosen as $\lambda = 0.0005$.

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 = 20s, 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 $\Delta = 0.25$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 $\lambda = 0.00025$.

L Goodness-of-fit evaluation

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

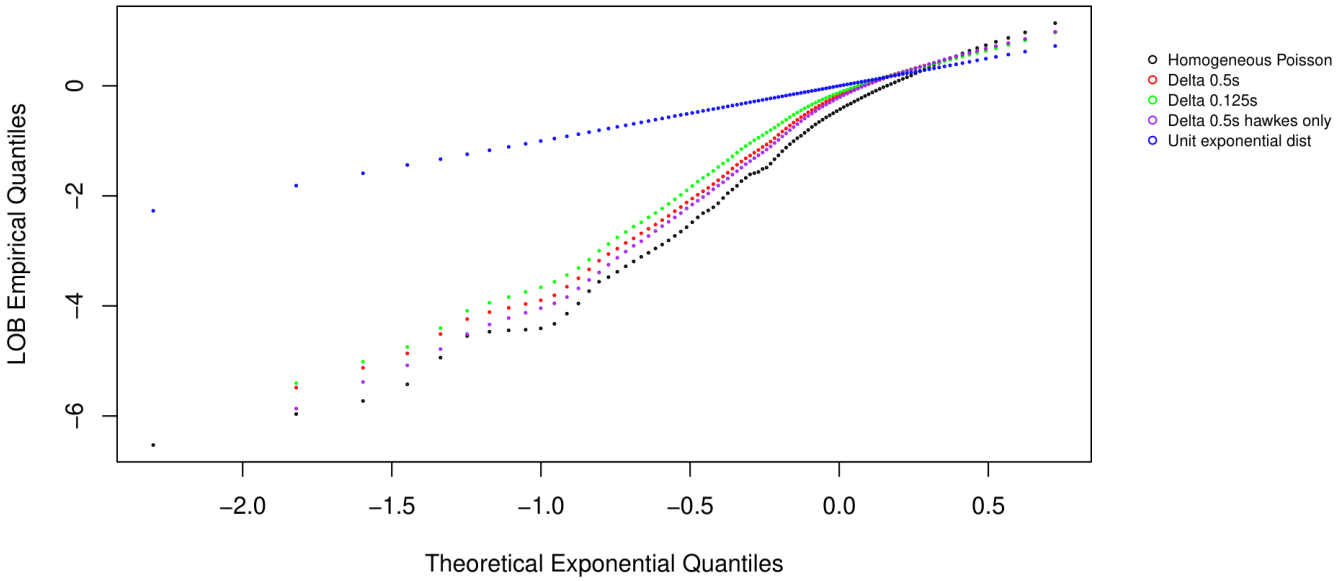


Figure S27: The figure demonstrates the Q-Q (quantile-quantile) plot of the LOB estimated based on maximum support s = 20s for selected events. All order sizes are considered to be one. The x-axis plots the log quantiles of the standard exponential distribution. The y-axis plots the log quantiles of 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 $\Delta = 0.5$s. The green dots show the distribution with bin-size $\Delta = 0.125$s. The purple dots show the distribution with bin-size $\Delta = 0.5$s 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 $\Lambda(t_1), \ldots, \Lambda(t_k)$ should follow a Poisson process with intensity 1, given a point process $t_1, \ldots, t_k$ 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 $\Lambda(t_2) - \Lambda(t_2), \ldots, \Lambda(t_k) - \Lambda(t_{k-1})$ 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 $\Delta = 0.125$s significantly outperforms the model with $\Delta = 0.5$s. 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.

References

Daley, D. J., Vere-Jones, D. et al. (2003), *An introduction to the theory of point processes: volume I: elementary theory and methods*, Springer.

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.

Kirchner, M. (2017), 'An estimation procedure for the hawkes process', *Quantitative Finance* **17**(4), 571–595.

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer Science & Business Media.