Analysis of order book dynamics in the Japanese stock market using the Queue-Reactive Hawkes process

Makoto Nohara$^1$ and Hidetoshi Nakagawa$^2$

$^1$The Mitsubishi UFJ Trust Investment Technology Institute, 4-2-6 Akasaka, Minato-ku, Tokyo 107-0052, Japan
$^2$Hitotsubashi University, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8439, Japan

*Corresponding author: hnakagawa@hub.hit-u.ac.jp

Received October 26, 2020, Accepted December 12, 2020

Abstract

We examine the effects of various types of orders and order book states on stock price formation in the Japanese stock market. For the purpose, we use the Queue-Reactive Hawkes (QRH) process to model the order book dynamics since the QRH process can reflect the influence of order book states as well as self-excitation and/or mutual excitation of past orders on the arrival intensities of next orders. As a result, we observe whether the mid price moves or not strongly depends on the order book state.

Keywords Queue-Reactive Hawkes process, high-frequency trading, limit order book

Research Activity Group Mathematical Finance

1. Introduction

The purpose of this study is to analyze the effects of not only various types of order (e.g. limit orders, market orders, and cancellations) but also order book states, namely the best sell/buy quotes and the number of limit orders for each quote, on stock price formation in the Japanese stock market.

The literature on high-frequency trading in the financial market and so on has been increasing rapidly. Not a few researches, from a viewpoint of order reaction analysis, regard the occurrences of orders in the market as events and use the Hawkes process as the model for analysis to examine if there exist self-exciting and/or mutually exciting properties between the orders. For one example, [1] applied the Hawkes process to modeling the dynamics of orders at the futures market of DAX (German stock index) and BUND (German government bonds denominated in euros issued), and showed that the mid-price is likely to fall immediately after the mid-price rises, and vice versa.

On the other hand, [2] introduced the Queue-Reactive Hawkes (QRH) process, whose arrival intensities of orders in the market can depend on the state of the order book as well as self-excitation and/or mutual excitation of past orders, and insisted that the feature of mid price changes observed in [1] can be largely influenced by the order book state.

In this study, we model the occurrences of some types of order in stock trading using the QRH process to see not only if there exist some self-exciting and/or mutually exciting properties among such orders but also how the order book state has influence on the arrival intensity of the orders. Then, we tentatively introduce a specific model for demonstration to estimate the parameters of the QRH process using the high-frequency trading data of stocks issued by a Japanese financial group for several hours in total.

As a result of analysis using a high-frequency trading data of a Japanese stock in the middle of trading hours for several days, we observe whether or not the mid-price is liable to change depends largely on the order book state: the mid-price is likely to rise successively after the mid-price rises, unless the limit order volume of the best ask quote is relatively much thicker. The same feature can be also seen when the mid-price falls down.

2. Queue-Reactive Hawkes (QRH) process

The Hawkes process is a counting process and is often used to model the occurrence times of a particular event in various fields. In the research field of finance, usually, the arrivals of some types of orders (e.g. limit orders, market orders, and cancellations) at the financial market like the stock market are regarded as events and the Hawkes process is utilized as the model for analysis to examine if there exist self-exciting and/or mutually exciting properties between the orders.

In general, the Hawkes process is specified on a probability space by the intensity process that has self-exciting and/or mutually exciting property as follows. Let $E$ be the set of event types. For each event type $e \in E$, the occurrence intensity $\lambda_e(t)$ of $e$-type event at time $t$ is defined by

$$
\lambda_e(t) = \mu_e + \int_0^t \sum_{e' \in \mathcal{E}} h(e' \rightarrow e)(t - s) dN_{s}^{e'},
$$

where $\{N_{s}^{e'}\}$ denotes the Hawkes process for $e'$-type event, that is, the cumulative number up to time $t$ of occurrences of $e'$-type event, $\mu_e$ is the constant inten-
sity component that is exogenously given, $\phi^{e,e'}(t)$ is so-called the kernel function that stands for the remaining impact on the intensity for $e$-type event after the period $t(\geq 0)$ has passed after an $e'$-type event occurred.

We remark that the kernel function $\phi^{e,e'}$ for the same type $e$ identifies the self-exciting property of $e$-type event, while $\phi^{e,e'}$ for different types $e \neq e'$ does the mutually exciting property from $e'$-type to $e$-type.

The Hawkes process is useful since it is possible to estimate the intensity only with history data on event occurrence, but it is impossible to make the intensity dependent on other state variables.

In this study, we use the Queue-Reactive Hawkes (QRH) process, that is almost the same as the intensity model called the QRH-II in [2], instead of the naive Hawkes process, so that we can make the jump size of the intensity process at the time when some event happens depend on the order book state as follows. Specifically, the QRH process is characterized by the occurrence intensity $\lambda^e_t$ for $e$-type event at time $t$ that is given by the product of a positive-valued function $f^e$ for $e$-event of some common variable $X_t$ related to the order book state at the time and the intensity of having the same formula given in (1) as follows:

$$X_t = f^e(X_t) \left( \mu^e + \int_0^t \sum_{e' \in \mathcal{E}} \phi^{e,e'}(t-s)dN^e_s \right), \quad (2)$$

where $\{N^e_s\}$ denotes the QRH process for $e'$-type event, and the exogenous intensity $\mu^e$ and the kernels $\phi^{e,e'}(t)$ can be interpreted in the same way as the naive Hawkes model. Hereafter we call the function $f^e$ the state effect function, as it determines the magnitude of the effect of the state variable on the intensity.

The remaining problem for empirical studies using the QRH process is how to specify the state variable $X_t$. We suppose that the state variable $X_t$ is specified in terms of the imbalance between the limit order quantity of the best bid (resp. ask) price relatively thick. In short, $X_t$ is closer to $0$ if the order book state is closer to $t$ bid (resp. ask) price. Thus, we call $X_t$ the imbalance indicator.

Finally, we assume that the state variable $X_t$ at time $t$ takes value in a finite set of states that simply represent the direction and the degree of limit order imbalance in the best quotes, according to the value of the imbalance indicator $QI_t$. The specific formulation of for our empirical analysis will be described in the next section.

3. Data

In the empirical analysis seen later, we use the FLEX Historical data provided by JPX Data Cloud: the security code is 8411 (Mizuho Financial Group, Inc.) in the Tokyo Stock Exchange for the six business days (November 9-11 and 14-16, 2016), and the target period is a total of two hours (one hour from 10 am and one hour from 1 pm) for each day.

In general, transactions are more active just after the market opens and before it closes. It does not seem appropriate to analyze the data in such special time zones with our model, so we limit the samples to the middle of trading hours.

The target stock is arbitrarily selected, but we should mention that it is a constituent of the stock index TOPIX100, a low-priced stock, and actively traded. In addition, we remark that the sample period is just after Donald Trump was elected in the U.S. presidential election, when trading of this stock became active, trading volume increased sharply, and the stock price rose, owing to expectations for his deregulation policies for banks.

The original FLEX Historical data contains the following items: Time (when the order book changed), Tag ID (identifying the information on trading), Price (at the change in order book or at the contract time), Trading volume, Turnover value, and Order quantity (after the change in order book). By sequentially combining the information in these records, it is possible to reproduce the dynamics of order book state, concretely the best sell/buy quotes and the number of limit orders, so that we can obtain time series data of the imbalance indicator $\{QI_t\}$.

Moreover, similar to [2], we classify the orders in the data into any one element in the set $\mathcal{E} = \{P_{ap}, P_{+}, P_{-}, P_{dum}\}$, depending on whether the order changes the mid-price and whether it raises or lowers the mid-price. Specifically, $P_{ap}$ (resp. $P_{dum}$) stands for the set of orders that raise (resp. lower) the mid-price, while $P_{+}$ (resp. $P_{-}$) does the set of orders that do not change the mid-price but make the limit order volume of the best bid (resp. ask) price relatively thick. In short, the set $P_{+}$ contains limit buy orders to best buy quote, market buy orders, and cancel orders at best sell quote and the set $P_{-}$ contains limit sell orders to the best sell price, market sell orders, cancel orders at the best buy quote. Hence, we can focus our interest only on the dynamics of the best quotes and the order quantity waited at best quotes.

4. Estimation method

For our empirical study, we tentatively define the state variable process $\{X_t\}$ by taking value in a set of five states $\{Ask_{++}, Ask_+, Eqv, Bid_+, Bid_{++}\}$ depending on the value of $QI_t$ given in (3) as follows:

$$X_t := \begin{cases} Ask_{++} & \text{if } QI_t \in [-1, -0.6) \\ Ask_+ & \text{if } QI_t \in [-0.6, -0.2) \\ Eqv & \text{if } QI_t \in [-0.2, 0.2) \\ Bid_+ & \text{if } QI_t \in [0.2, 0.6) \\ Bid_{++} & \text{if } QI_t \in [0.6, 1] \end{cases}$$
We suppose to divide the range $[-1, 1]$ equally into the five subintervals for the order book state, while we try dividing the range $[-1, 1]$ of the imbalance indicator with the first four quintiles on the actual data as the threshold values for subintervals. As a consequence, we observe that the results of parameter estimation are almost the same for both cases, so we assume the above specification for the order book state.

We note that the state $\text{Ask}_{++}$ (resp. $\text{Bid}_{++}$) shows the state in which the limit order volume for the best ask (resp. bid) quote is relatively much larger than that of the opposite side.

Then, we set the range of the state effect function $f'$ appeared in (2) as the parameter set $\varphi' := \{\varphi_{\text{Ask}_{++}}', \varphi_{\text{Ask}_{-}}', \varphi_{\text{Bid}_{++}}', \varphi_{\text{Bid}_{-}}'\}$ to be estimated. As such, we suppose that the state effect function $f'$ maps the state $X_t$ to the corresponding value $\varphi'_{X_t}$; for example, $f(\text{Ask}_{++}) = \varphi_{\text{Ask}_{++}}'$. Hereafter, we assume that $\varphi_{\text{Ask}_{++}}' = 1$ for normalization and the other four values are to be estimated.

Moreover, to simply represent the size of jump impact introduced in the next section, we suppose that the kernels $\varphi'^{-\epsilon}$ are given by an exponential decay function as follows:

$$\varphi'^{-\epsilon}(\tau) := \alpha^{e^{-\epsilon}} \beta^e \exp (-\beta^e \tau) \cdot 1_{\{\tau \geq 0\}},$$  \hspace{1cm} (4)

where $\alpha^{e}$ and $\beta^e$ are parameters called the order impact parameter and the decay parameter, respectively. In fact, it was supposed in [2] that the kernels are given by the sum of several exponential functions. However, we apply the above single exponential function since the impact on the intensity is almost the same for both the kernels given by the sum of two exponential functions tried in a preliminary analysis and the single exponential kernels.

Hence, we have to estimate the parameters $\theta := \{\varphi', \epsilon, \beta^e, \alpha^e\}$ for $e$-type event using the data presented in the previous section. However, we set the decay parameter $\beta^e$ in (4) as the hyper-parameter, similar to the approach of [2]. We assume that the value for $\beta^e$ is chosen from the set $\{100, 200, 500, 1000, 2000, 5000\}$.

The rest of the parameters are estimated by using the least squares method proposed in [2]. (We can refer to [3] for mathematical argument of this method.) Specifically, we achieve the estimates by minimizing the objective function $C^e(\theta)$ for the data period $[0, T]$ defined by

$$C^e(\theta) = \int_0^T \left( \lambda_s^e(\theta)^2 \right) ds - 2 \sum_{k=1}^{N^e_k} \lambda_s^e(\theta).$$

Indeed, the estimation procedure of the 36 parameters (four values of the state effect function, one exogenous intensity, and four order impact parameters for each of the four event types) is executed 72 (= 12 × 6) times for each one-hour period (one hour from 10 am and one hour from 1 pm for the six business days) as well as for each value of the hyper-parameter $\beta^e$.

| $e$ | mean | median | stdev | min | max |
|-----|------|--------|-------|-----|-----|
| $P_{up}$ | 0.01 | 0.01 | 0.02 | 0.00 | 0.07 |
| $P_{down}$ | 0.03 | 0.03 | 0.09 | -0.18 | 0.14 |

Table 1. The basic statistics (mean, median, standard deviation, minimum, and maximum) of 36 estimated parameters. We fix $\varphi_{\text{Ask}_{++}}' = 1$.

5. Results

First, we present the basic statistics of 36 estimated parameters in Table 1.

Then, we illustrate some estimation results on $q^e \times \alpha^{e^{-\epsilon}}$. That is, the product of the value of state effect function and the order impact parameter. This quantity $q^e \times \alpha^{e^{-\epsilon}}$ can be viewed as the size of jump impact of $e$-type event occurrence on the arrival intensity for $e$-type event when the order book state is given as $e$.

Fig. 1 displays, for each $e \in \{P_{up}, P_{+}, P_{-}, P_{down}\}$, the heat map of the average impact size over all the sample periods for each pair of the order book state and the type of event that just occurred. Although negative jump impacts are estimated in some cases, their absolute values are so small that we can view them as zero impact. We remark that the darker the green color is, the larger the jump impact of the happened event on the
arrival intensity for $e$-type event.

For example, the upper-left heat map shows the distribution of average impact size $q_{\text{up}}^e \times \alpha_{\text{up}-e'}$ over the twelve sample periods.

It immediately follows from the upper-left heat map for $P_{\text{up}}$-type event that the darkest green cell in the map corresponds to the impact of occurrence of the same $P_{\text{up}}$-type event when the order book state is Bid$++$. This implies that the mid-price is likely to rise successively after the mid-price rises when the limit order volume of the best bid quote is relatively thicker than that of the best ask quote. Such a consequence is quite natural since the limit sell order at the best ask quote, if relatively less, can be easily offset by arrival of some market buy order.

Similarly, the same feature can be also seen in the lower-right heat map for $P_{\text{dwn}}$-type. In fact, the darkest green cell in the map corresponds to the impact of occurrence of the same $P_{\text{dwn}}$-type event when the order book state is Ask$++$. As the limit buy order at the best bid quote is relatively less in this case, the order can be easily offset by some market sell order, thus, it seems that the mid-price is likely to fall down successively after the mid-price falls.

Even in the other cases of $P_+$-type and $P_-$-type, we can observe some degree of self-excitation from their heat maps, but it seems that the impact size hardly depends on the book order state, unlike the cases of $P_{\text{up}}$-type and $P_{\text{dwn}}$-type.

6. Concluding remarks

We use the Queue-Reactive Hawkes (QRH) process presented in [2] to model the occurrences of some types of orders in stock trading. To examine the impact of the order book state on the arrival intensity of the orders as well as the existence of self-exciting and/or mutually exciting properties among the orders, we tentatively propose a parametric model for empirical study.

As a result of the model estimation using the the high-frequency trading data of stocks issued by a Japanese company for a short time, we observe that the mid price strongly depends on the order book state, and that in particular, the mid-price is likely to rise (resp. fall) successively after the mid-price rises (resp. falls), unless the limit order volume of the best ask (resp. bid) quote is relatively much thicker than the opposite.

There are some remaining issues: robustness check by increasing of the number of stocks for analysis, revision of the assumptions on the model, statistical testing of the parameter estimates, and so on.

Acknowledgments

This study is supported by KAKENHI Nos. 17K01248 and 20K04960.

Disclaimer

The views expressed in the article are the authors’ own and do not represent the official views of the institutions to which the authors belong.

References

[1] E. Bacry, T. Jaisson and J.-F. Muzy, Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics, Quant. Finance, 16 (2016), 1179–1201.
[2] P. Wu, M. Rambaldi, J.-F. Muzy and E. Bacry, Queue-reactive Hawkes models for the order flow, arXiv:1901.08938 [q-fin.TR].
[3] P. Reynaud-Bouret and S. Schbath, Adaptive estimation for Hawkes processes; application to genome analysis, Ann. Statist., 38 (2010), 2781–2822.