2024

A Limit Order Book Model for High Frequency Trading with Rough Volatility

Yun S. Chen-Shue
University of Central Florida

A LIMIT ORDER BOOK MODEL FOR HIGH FREQUENCY TRADING WITH ROUGH VOLATILITY

by

YUN SU CHEN-SHUE
M.S. University of Central Florida, 2019
B.S. University of Central Florida, 2015

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
in the College of Science
at the University of Central Florida
Orlando, Florida

Spring Term
2024

Major Professor: Jiongmin Yong

ABSTRACT

We introduce a financial model for limit order book with two main features: First, the limit orders and market orders for the given asset both appear and interact with each other. Second, the high frequency trading (HFT, for short) activities are allowed and described by the scaling limit of nearly-unstable multi-dimensional Hawkes processes with power law decay. The model eventually becomes a stochastic partial differential equation (SPDE, for short) with the diffusion coefficient determined by a Volterra integral equation governed by a Hawkes process, whose Hurst exponent is less than 1/2, which makes the volatility path of the stochastic PDE rougher than that driven by a Brownian motion. We have further established the well-posedness of such a system so that a foundation is laid down for further studies in this direction.

This paper is dedicated to my enduring husband and loving parents.

ACKNOWLEDGMENTS

I would like to express my appreciation for the contributions of Dr. Jiongmin Yong for mentoring me and guiding me with endless insight and wisdom. I would also like to thank Dr. Liqiang Ni, Dr. Jason Swanson, and Dr. Christian Keller for their patience and sharing their invaluable advice throughout the course of this dissertation's development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

A limit order book (LOB, for short), a list of prices and volumes for a traded asset, can be used as a mechanism to facilitate trades in the financial market: traders can place limit orders in the order book with pre-determined prices and volumes waiting for execution as well as submit market orders that are executed immediately against the existing limit orders by the best available prices. For each time t, the LOB provides a snapshot of the market by presenting the volumes of outstanding limit orders at each price level. The price level increments by the minimum price change is called the tick size. In the LOB example below, the tick size is 1 cent. The green columns visualize the volumes of the bid orders (or, buy orders) and are negative by convention. The red columns show the volumes of the ask orders (or, sell orders) and are positive by convention also. The highest bid offer, \$100.00 in the example, is called the bid price, while the lowest ask offer (\$100.01) is called the ask price. The mid-price of a LOB is often calculated as the average of the bid and ask prices, which is \$100.005 in the example below.

Since the LOB dynamics shows the supply and demand of a certain asset in a fundamental way and forms the price dynamics of this asset, there has been an increasing interest in modeling the LOB dynamics. However, most modeling attempts are hard to be analytically or computationally tractable [40] [18] [46] [8].

Cont and Müller [13] proposed a model in which the dynamics of the centered order book density is described by a stochastic partial differential equation (SPDE, for short) with multiplicative Gaussian noise. We will refer to this model as the Cont-Müller model (C-M model, for short) in the rest of this paper. The centered order book density, $u(t, x)$, is the volume per unit price (tick size) of the limit order at time t and the position x is the distance away from the mid-price, with $x \in [-L, L]$ for some $L > 0$. It is easy to see that rational investors will not submit limit orders

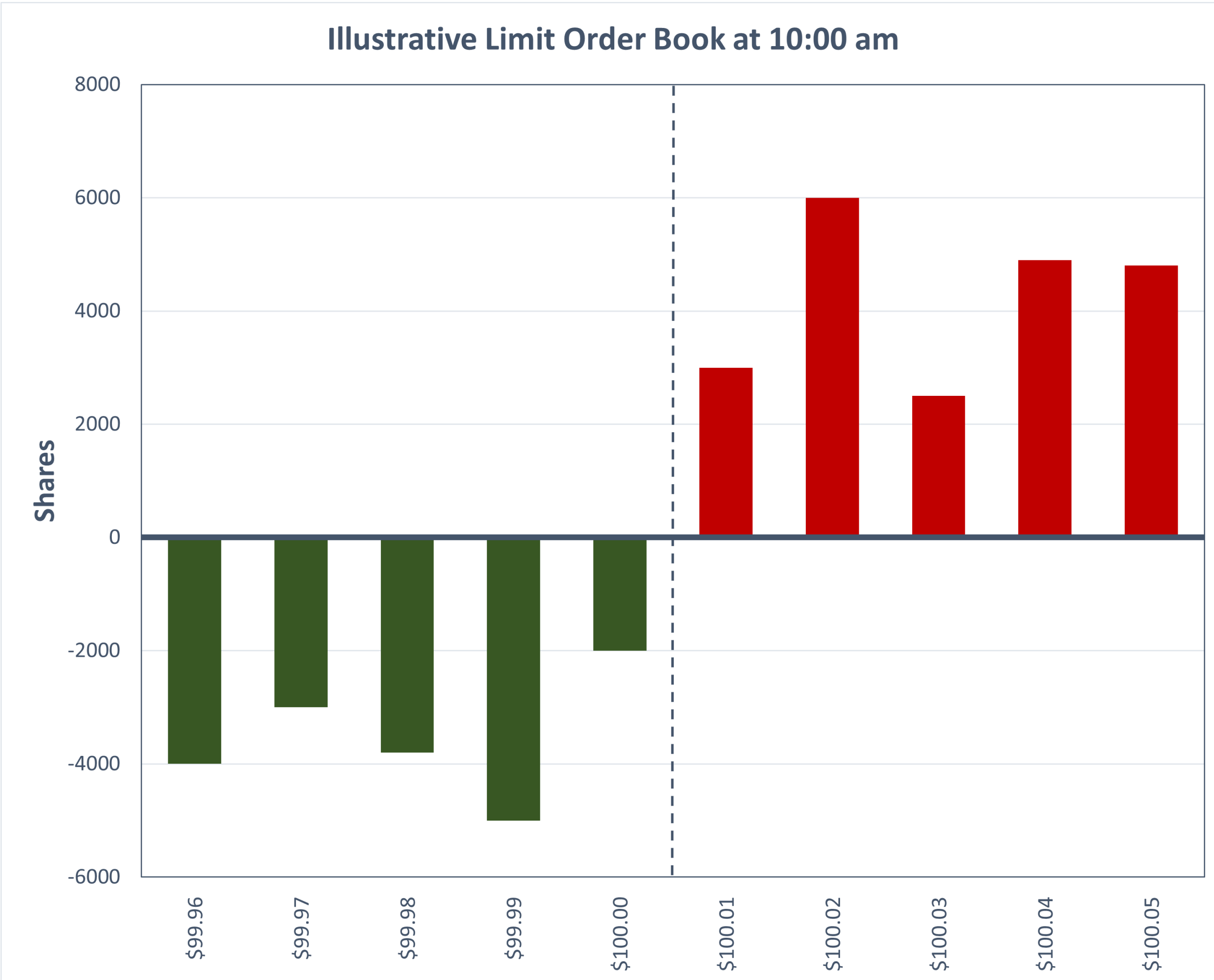


Figure 1.1: Illustrative LOB at 10:00 am

far away from the mid-price, and that all the previously-submitted orders were cancelled as soon as their price levels became too far away from the mid-price. This assumption is reflected by the setting that $u(t, x) = 0$ when $x \notin (-L, L)$ (See [13]).

The C-M model that presents the dynamics of centered book density $u(t, x)$ is as follows ([13], with small modifications):

$$du(t,x) = \Big[\eta_a \Delta u(t,x) + \beta_a \nabla u(t,x) - \alpha_a u(t,x) + f^a(x)\Big]dt + \sigma_a u(t,x) dW^a(t), \qquad x \in (0,L)$$

$$du(t,x) = \Big[\eta_b \Delta u(t,x) - \beta_b \nabla u(t,x) - \alpha_b u(t,x) - f^b(x)\Big]dt + \sigma_b u(t,x) dW^b(t), \qquad x \in (-L,0)$$

$$u(t,x) \leq 0, \quad x < 0, \qquad u(t,x) \geq 0, \quad x > 0$$

$$u(t,0+) = u(t,0-) = 0, \qquad u(t,-L) = u(t,L) = 0,$$

where $\eta_a, \eta_b, \beta_a, \beta_b, \sigma_a, \sigma_b, \alpha_a, \alpha_b > 0$ are some constants, $f^a, f^b : [-L, L] \to [0, \infty)$ are given functions, and (W^a, W^b) is a two-dimensional Brownian motion (with possibly correlated components). In these equations, non-high frequency trading (non-HFT, for short) order submissions are modeled by $f^a(x)$ and $f^b(x)$, all kinds of non-HFT order cancellations by

$$\Big[\eta_a \Delta u(t,x) + \beta_a \nabla u(t,x) - \alpha_a u(t,x)], \qquad \Big[\eta_b \Delta u(t,x) - \beta_b \nabla u(t,x) - \alpha_b u(t,x)\Big],$$

and high frequency trading (HFT, for short) order dynamics by $\sigma_a u(t,x) dW^a(t)$ and $\sigma_b u(t,x) dW^b(t)$, on the ask and bid sides respectively. We will provide detailed explanations of the relevant terms when introducing our model in Section 3.

The C-M model [13] has both the analytical and computational tractability for applications, and the price dynamics was naturally derived from the model. However, there are two main limitations in that model.

First, the C-M model did not reflect the effect to the centered order book density from market ask/bid orders. Indeed, the only terms regarding order submissions are $f^a(x)$ and $f^b(x)$, which only increase the volumes on the ask and bid sides, whereas the market order submissions affect the LOB in a different way since they decrease the LOB volumes. Thus, the market orders should be taken into account.

Second, the C-M model used multiplicative Gaussian noise terms to model the order dynamics from HFT at coarse-grained time scale of the average (non-HFT) market participants. This implies that each increment of the HFT is independent of the previous HFT incremental changes. However, many evidence shows that HFT markets are highly endogenous, meaning HFT orders tend to generate other HFT orders. Furthermore, many HFT orders are part of a larger order (or metaorder) that takes a relatively long time to fully execute, which causes a given HFT order to have a relatively long-term influence on other HFT orders. Thus, it is better to use self-exciting and long term dependency process to model HFT, rather than Brownian motions (as in the C-M model) [16].

In this paper, we propose a new model. First, we include the effect from market orders so that the limit orders and market orders interact with each other, which looks more realistic. Second, we have used the scaling limit of a sequence of nearly-unstable multivariate Hawkes process (which is self-exciting) with power-law tails to model the HFT dynamics at a coarse-grained time scale, reflecting the dependencies among HFT orders.

The remaining of this paper is arranged as follows. Chapter 2 provides a brief overview on the Hawkes process. Chapter 3 presents our new model with both the non-HFT and HFT components. Chapter 4 gives the proof for the scaling limit of the Hawkes processes, while Chapter 5 presents the SPDE of the market model and its well-posedness In Chapter 6, we derive the price dynamics based on the order book dynamics. We also provide analyses for the parameters in the price model. Some simulation results will be collected in Chapter 7. Finally, some lengthy and technical results will be put in the appendices.

CHAPTER 2: MATHEMATICAL PRELIMINARIES

In this chapter, we provide an overview of the d-dimensional Hawkes process (with $d \geq 1$). Most of the following definitions and propositions are from [34].

2.1 One-Dimensional Hawkes Process

Definition 2.1. A discrete random variable X is said to have a *Poisson distribution* with parameter $\lambda^* > 0$, if it has a discrete probability distribution:

$$f(k; \lambda^*) = \mathbb{P}(X = k) = \frac{(\lambda^*)^k e^{-\lambda^*}}{k!}, \qquad \forall k = 0, 1, 2, ...$$

We denote it as $X \sim \text{Poi}(\lambda^*)$.

Definition 2.2. A *counting process* is a stochastic process $(N(t) : t \geq 0)$ taking values in the set $\{0, 1, 2, ...\}$ that satisfies $N(0) = 0$, almost surely finite, and is a right-continuous non-decreasing step function with increments of size $+1$.

Further, denote by $(\mathcal{H}(u) : u \geq 0)$ a right continuous *filtration*, that is, an increasing sequence of σ-algebras, such that $\mathcal{H}(u) = \bigcap_{\epsilon>0} \mathcal{H}(u + \epsilon)$. The filtration $\mathcal{H}(u)$ represents the history of the counting process $N(\cdot)$, namely, it is generated by $N(\cdot)$.

Definition 2.3. Consider a counting process $N(\cdot)$ with associated histories $\mathcal{H}(\cdot)$. If a (non-negative) function $\lambda(t)$ exists such that

$$\lambda(t) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(t+h) - N(t) | \mathcal{H}(t)]}{h}, \qquad t \geq 0, \tag{2.1}$$

then it is called the *conditional intensity function* of $N(\cdot)$.

Definition 2.4. A counting process $(N(t) : t \geq 0)$ is called an *(inhomogeneous) Poisson process* with rate function $\lambda(t) > 0$ if

1. For any interval $I = (a, b]$, $N(I)$ has a Poisson distribution with parameter $\int_a^b \lambda(s)ds$, i.e.,

$$N(I) \sim \text{Poi}\Big(\int_a^b \lambda(s)ds\Big), \quad \text{or} \quad \mathbb{P}(N(I) = k) = \frac{(\int_a^b \lambda(s)ds)^k e^{-(\int_a^b \lambda(s)ds)}}{k!}, \quad \forall k = 0, 1, 2, ...$$

2. For any n disjoint interval $I_1, I_2, ..., I_n$, the random variables $N(I_1), N(I_2), ..., N(I_n)$ are independent.

If the rate function is a constant $\lambda > 0$, then $N(\cdot)$ is called a *homogeneous* Poisson process.

Definition 2.5. A counting process $(N(t) : t \geq 0)$ is called a *Hawkes process* if the following conditions hold:

(i). The conditional increment against its history $(\mathcal{H}(t) : t \geq 0)$ satisfies

$$\mathbb{P}(N(t+h) - N(t) = m|\mathcal{H}(t)) = \begin{cases} 1 - \lambda(t)h + o(h), & m = 0 \\ \lambda(t)h + o(h), & m = 1 \\ o(h), & m > 1 \end{cases} \tag{2.2}$$

for some conditional intensity function $\lambda(\cdot)$.

(ii). The conditional intensity function $\lambda(\cdot)$ is of the form

$$\lambda(t) = \mu(t) + \int_0^t \phi(t-s)dN(s), \tag{2.3}$$

where $\mu(t)$, called the *background intensity*, is a deterministic function of t that is integrable over any finite intervals and has a finite limit $\mu(\infty) > 0$ as $t \to \infty$, and $\phi : (0, \infty) \to$

$[0, \infty)$, called the *excitation function*, is assumed to be a positive function. This means that the exogenous events arrive according to an inhomogeneous Poisson process with the rate function $\mu(t)$, and the direct offspring of any event arrives according to an inhomogeneous Poisson process with the rate function $\phi(t)$.

When $\phi(\cdot) = 0$, the Hawkes process $N(t)$ becomes an inhomogeneous Poisson process. Thus the former is an extension of the latter.

According to [24] and [51], Hawkes process exists as long as $\int_0^\infty \phi(t)dt < 1$, $\mu(t)$ is positive, integrable over any finite intervals, and has a finite limit $\mu(\infty) > 0$ as $t \to \infty$.

For any $f(\cdot)$ which is integrable on $\mathbb{R}$, we define

$$f^{*1}(\cdot) = f(\cdot), \qquad f^{*(n+1)}(\cdot) = \int_{-\infty}^{\infty} f^{*n}(\cdot - s)f(s)ds = \int_{-\infty}^{\infty} f(\cdot - s)f^{*n}(s)ds,$$

which are called convolution powers of $f(\cdot)$. The following lemma will be useful below.

Lemma 2.1. For any integrable function $f : \mathbb{R} \to \mathbb{R}$, the following holds for all $n \in \mathbb{Z}^+$:

$$\int_{-\infty}^{\infty} f^{*n}(t)dt = \Big(\int_{-\infty}^{\infty} f(t)dt\Big)^n$$

Proof. We use the principle of mathematical induction to prove this. When $n = 1$, we have

$$\int_{-\infty}^{\infty} f^{*1}(t)dt = \int_{-\infty}^{\infty} f(t)dt = \Big(\int_{-\infty}^{\infty} f(t)dt\Big)^1$$

Suppose that the equation holds for n, then using the definition of convolution and Fubini's theorem, we have

$$\int_{-\infty}^{\infty} f^{*(n+1)}(t)dt = \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} f(t-s)f^{*n}(s)ds\Big]dt$$

$$= \int_{-\infty}^{\infty} f^{*n}(s)\Big[\int_{-\infty}^{\infty} f(t-s)dt\Big]ds$$
$$= \Big(\int_{-\infty}^{\infty} f(s)ds\Big)^n\Big[\int_{-\infty}^{\infty} f(t)dt\Big] = \Big(\int_{-\infty}^{\infty} f(t)dt\Big)^{n+1}$$

By the principle of mathematical induction, we have finished this proof. □

Proposition 2.1. Let N be a one-dimensional Hawkes process, and $\lambda(t)$ be its conditional intensity process of form (2.3) with $\mu(\cdot)$ and $\phi(\cdot)$ given as in Definition 2.5. In addition, if $\int_0^\infty \phi(s)ds < 1$, then

$$\lim_{t\to\infty} \mathbb{E}[\lambda(t)] = \frac{\mu(\infty)}{1-\int_0^\infty \phi(s)ds}$$

Proof. Note that from (2.1), we have

$$\mathbb{E}[dN(t)] = \mathbb{E}[\lambda(t)]dt.$$

Denote $\bar{\lambda}(t) = \mathbb{E}[\lambda(t)]$, then

$$\bar{\lambda}(t) = \mu(t) + \int_0^t \phi(t-s)\bar{\lambda}(s)ds = \Big[\mu + \phi * \bar{\lambda}\Big](t)$$

$$= \Big[\mu + \phi * (\mu + \phi * \bar{\lambda})\Big](t) = \cdots = \Big[\Big(\sum_{n=0}^{\infty} \phi^{*n}\Big) * \mu\Big](t).$$

As t goes to infinity, by the dominated convergence and monotone convergence theorems, along with Lemma 2.1, we have

$$\lim_{t\to\infty} \mathbb{E}[\lambda(t)] = \lim_{t\to\infty} \int_0^t \mu(t-s)\sum_{n=0}^{\infty} \phi^{*n}(s)ds = \int_0^\infty \mu(\infty)\sum_{n=0}^{\infty} \phi^{*n}(s)ds$$

$$= \mu(\infty)\sum_{n=0}^{\infty}\int_0^\infty \phi^{*n}(s)ds = \mu(\infty)\sum_{n=0}^{\infty}\Big(\int_0^\infty \phi(s)ds\Big)^n = \frac{\mu(\infty)}{1-\int_0^\infty \phi(s)ds}$$

□

When $\int_0^\infty \phi(s)ds \geq 1$, $\lambda(t) \to \infty$ as $t \to t^*$ for some $t^* \leq \infty$, and hence the Hawkes process in this case explodes almost surely.

2.2 Multi-Dimensional Hawkes Process

An extension of the one-dimensional Hawkes process is a multi-space valued Hawkes process. In [34], the term multi-dimensional Hawkes process was reserved only for the multi-dimensional space valued process where the components are decoupled, and hence the components are not mutually exciting. Meanwhile, in [16], [7], [4], [36], [2], it was assumed that the components are coupled so that they are mutually exciting. In our paper, by multi-dimensional Hawkes process, we mean the process is not only multidimensional but also mutually exciting. More precisely, we have the following definition:

Definition 2.6. Consider $(\{\mathbf{N}(t)\} : t \geq 0)$ a vector process formed by m counting processes $\{N_1(t), ...N_m(t)\}$. If for each $i = 1, ..., m$, $N_i(t)$ has a conditional intensity of the form

$$\lambda_i(t) = \mu_i(t) + \sum_{j=1}^{m} \int_0^t \phi_{i,j}(t-s)dN_j(s)$$

for some positive function $\mu_i(t)$ with $\lim_{t\to\infty} \mu_i(t) = \mu_i(\infty) > 0$, $\phi_{i,j} : (0,\infty) \to [0,\infty)$, and $\phi_{i,j}(\cdot) \in L^1(0,\infty)$, then $\mathbf{N}$ is called a *multi-dimensional* Hawkes process.

We can also write conditional intensity of the multi-dimensional Hawkes process in vector form as

$$\boldsymbol{\lambda}(t) = \boldsymbol{\mu}(t) + \int_0^t \mathbf{\Phi}(t-s)d\mathbf{N}(s).$$

Note that $\boldsymbol{\mu}(\cdot)$ is an m-dimensional vector-valued functions, $\boldsymbol{\lambda}(\cdot)$ and $\mathbf{N}(\cdot)$ are m-dimensional processes, and $\mathbf{\Phi}(\cdot)$ is an $m \times m$ square matrix-valued function with the entries $\phi_{i,j}(\cdot)$.

Remark 2.1. In our definition of the multi-dimensional Hawkes process, the background intensity $\boldsymbol{\mu}(t)$ is allowed to be a vector-valued function that converges to a constant vector with positive components as $t \to \infty$. This is different from [34], where the background intensity can only be a constant vector with positive components.

CHAPTER 3: THE MODEL

We now propose a model that describes the LOB dynamics of orders from both HFT and non-HFT investors. For the non-HFTs, we use the centered order book density model similar to that in [13]. To model the dynamics of HFT orders, we use the multi-dimensional Hawkes process.

Let the volume of orders awaiting execution at time t and price p be $U(t,p)$. By convention, $U(t,p) \geq 0$ for ask orders, and $U(t,p) \leq 0$ for bid orders. We define the ask price (the lowest ask offer) $s^a(t)$ and bid price (the highest bid offer) $s^b(t)$ as follows:

$$s^a(t) := \inf\{p > 0, U(t,p) > 0\}, \qquad s^b(t) := \sup\{p > 0, U(t,p) < 0\}$$

We assume that all the investors are rational. Thus, they could not offer a lower price to sell than any ask price, or a higher price to buy than any bid price. Therefore,

$$s^b(t) < s^a(t), \qquad \Big\{U(t,p) \,\Big|\, s^b(t) < p < s^a(t)\Big\} = \varnothing.$$

With the above $s^a(t)$ and $s^b(t)$, we define the mid-price to be

$$S(t) = \frac{s^a(t) + s^b(t)}{2}.$$

We can see that $p < S(t) < s^a(t)$ implies $U(t,p) \leq 0$, and $p > S(t) > s^b(t)$ implies $U(t,p) \geq 0$. Let the tick size of the market be $\delta > 0$, and let $v(t,p) \approx U(t,p)/\delta$ be the volume density. We define

$$u(t,x) = \begin{cases} v(t, S(t) + x), & \text{for } x \in [-L, L] \\ 0, & \text{otherwise} \end{cases}$$

where $L > 0$, and x represents a distance from the mid-price. When $x < 0$, $S(t) + x < S(t)$, and hence $u(t,x) = v(t, S(t) + x) \leq 0$. Similarly, when $x > 0$, $u(t,x) \geq 0$. We call $u(t,x)$ the centered order book density at (t,x).

3.1 Non-HFT Orders

In this subsection, we are modeling non-HFT orders. We observe the following different LOB events with each corresponding term appeared on the right-hand side of the equation:

1. **Outright cancellation of orders without replacement**: Let $\zeta_a, \zeta_b > 0$. When $x > 0$, then $u(t,x) \geq 0$, and we use a term $-\zeta_a u(t,x)$ to model the decrease of $u(t,x)$ from the outright proportional cancellation of limit ask orders at the price level $S(t) + x$. When $x < 0$, then $u(t,x) \leq 0$, and we use a term $-\zeta_b u(t,x) = \zeta_b |u(t,x)|$ to model the decrease of the *absolute* value of $u(t,x)$ from the outright cancellation of limit bid orders at the price level $S(t) + x$. The C-M Model [13] also contained these two terms.

2. **Symmetric changes**:

   (1) $x > 0$: $\eta_a u_{xx}(t,x)$ with $\eta_a > 0$:

   This term models the symmetric changes of limit ask orders at a distance x from the mid-price. For example, in the illustrative LOB (1.1), the volume at the price level \$100.03 is lower than all the neighboring price levels, \$100.02, \$100.04, and \$100.05, which acts roughly like a local minimum and makes $u_{xx}(t,x) > 0$, assuming everything is smooth. Some of the limit ask orders at the neighboring price levels will be cancelled and added to the price level \$100.03. So at the price level \$100.03, $u(t,x)$ goes up with a possible change $\eta_a u_{xx}(t,x) > 0$. On the other hand, the volume at the price level \$100.02 is higher than the neighboring price levels, \$100.01, \$100.03, \$100.04 and \$100.05, which acts roughly like a local maximum and it makes $u_{xx}(t,x) < 0$, assuming again everything is smooth. Some of the limit ask orders at the price level \$100.02 will be cancelled and added to the neighboring price levels. So at the price level \$100.02, $u(t,x)$ goes down with a possible change $\eta_a u_{xx}(t,x) < 0$.

(2) $x < 0$: $\eta_b u_{xx}(t,x)$ with $\eta_b > 0$:

This term models the symmetric changes of limit bid orders at a distance $|x|$ from the mid-price. It is similar to the ask case but applied in the opposite way since $u(t,x) < 0$ by convention. For example, in the illustrative LOB (1.1), the volume at the price level \$99.97 is lower than all the neighboring price levels, \$99.96, \$99.98, and \$99.99. Since $u(t,x) < 0$, it acts roughly like a local maximum and leads to $u_{xx}(t,x) < 0$, assuming everything is smooth. Some of the limit bid orders at the neighboring price levels will be cancelled and added to the price level \$99.97. So $u(t,x)$ at the price level \$99.97 should go down with a possible change $\eta_b u_{xx}(t,x) < 0$, which makes $u(t,x)$ smaller or the *absolute* value $|u(t,x)|$ larger. On the other hand, the volume at the price level \$99.99 is higher than the neighboring price levels, \$99.96, \$99.97, \$99.98 and \$100. Since $u(t,x) < 0$, it acts roughly like a local minimum and leads to $u_{xx}(t,x) > 0$, assuming again everything is smooth. Some of the limit bid orders at the price level \$99.99 will be cancelled and added to the neighboring price levels. So $u(t,x)$ at the price level \$99.99 should go up with a possible change $\eta_b u_{xx}(t,x) > 0$, which makes $u(t,x)$ larger or the *absolute* value $|u(t,x)|$ smaller.

The C-M Model [13] also contained these two terms. We slightly modify the notation: for example, instead of $\Delta u(t,x)$, we use $u_{xx}(t,x)$ to clarify that x is one-dimensional.

3. **Cancellation of orders with asymmetric replacement**:

(1) $x > 0$: $-\beta_a [u_x(t,x)]^-$ with $\beta_a > 0$:

This term models the cancellation of ask orders at a distance x from the mid-price and replacement of these orders closer to the mid-price. When $u_x(t, x) < 0$, it roughly means that there are more ask orders at lower prices than $S(t) + x$. Therefore, in order to sell the shares at the price level $S(t) + x$ faster, some investors will likely cancel their limit ask orders and resubmit them as limit ask orders at a price level closer to the mid-price, or even market ask orders. Thus, at price level $S(t) + x$, a certain portion of the volume will be decreased. This amount is assumed to be $-\beta_a[u_x(t, x)]^-$. When $u_x(t, x) > 0$, it roughly means that there are more ask orders at higher prices than $S(t) + x$. Therefore, most rational investors will not cancel the orders, as their ask orders are already better than most other orders. Hence, these orders will most likely be unchanged or the change will be $-\beta_a[u_x(t, x)]^- = 0$.

(2) $x < 0$: $\beta_b[u_x(t, x)]^-$ with $\beta_b > 0$:

This term models the cancellation of bid orders at a distance $|x|$ from the mid-price and replacement of these orders closer to the mid-price. When $u_x(t, x) < 0$, it roughly means that there are more bid orders at higher prices than $S(t) - |x| = S(t) + x$. Therefore, in order to buy the shares at the price level $S(t) - |x|$ faster, some investors will likely cancel their limit bid orders and resubmit them as limit bid orders at a price level closer to the mid-price, or even submit market bid orders. Thus, at price level $S(t) - |x|$, a certain portion of the volume will be decreased. This amount is assumed to be $\beta_b[u_x(t, x)]^- > 0$. When $u_x(t, x) > 0$, it roughly means that there are more bid orders at lower prices than $S(t) - |x| = S(t) + x$. Therefore, most rational investors will not cancel the orders as their bid orders are already better than most other orders. Hence, these orders will most likely be unchanged or the change will be $\beta_b[u_x(t, x)]^- = 0$.

This treatment is different from the C-M model [13]. We zero out the term $u_x(t, x)$ when $u_x(t, x) > 0$ so that the dynamics of limit order resubmission only goes towards the middle price. This is significantly different from the usual convection in the heat transfer situation.

4. **Cancellation of orders with market order replacement**:

   When the queues are long around the mid-price, some investors will likely cancel their limit orders in these queues, and resubmit the orders as market orders so that their orders can get executed immediately.

   For example, in the illustrative LOB (1.1), an investor originally placed a limit ask order of 70 shares at the price level \$100.01 at 10:00 am. She wants to sell her shares relatively quickly, but she has to wait until the 3000 shares, placed before 10:00 am at the same or lower prices, to be sold first. If she wants to sell her 70 shares by 10:15 am, and she does not think the 3000 shares will be sold by that time, she might cancel her order and resubmit it as a market ask order. She would take a total loss of \$0.70, but the 70 shares can be sold immediately, executed against the existing limit bid queue at the price level \$100.00. In this case, the limit ask queue at the price level \$100.01 is decreased by 70 shares due to the cancellation, and the limit bid queue at \$100.00 is also decreased by 70 shares due to the resubmitted market ask order.

   On the opposite side, another investor originally placed a limit bid order of 80 shares at the price level \$100.00 at 10:00 am. He wants to buy 80 shares relatively quickly, but he has to wait until the 2000 offers, placed before 10:00 am at same or higher prices, to be executed first. If he wants to buy 80 shares by 10:10 am, and he does not think the 2000 offers will be executed by that time, he might cancel his order and resubmit it as a market bid order. He would have to pay \$0.80 more than his previous offer, but he would get the 80 shares immediately from the existing limit ask queue at the price level \$100.01. In this case, the limit bid queue at the price level \$100.00 is decreased by 80 offers due to the cancellation, and the limit ask queue at \$100.01 is also decreased by 80 shares due to the resubmitted market bid order.

To model the impact from this LOB event, we first set a threshold $u_0 > 0$ such that a queue $u(t,x)$ is "too long" if $|u(t,x)| \geq u_0$. In other words, this LOB event will not happen when $\big|u(t,|x|)\big| < u_0$.

For $x > 0$, when $u(t,x) \geq u_0$, it means that the limit ask queue is too long for the investors. Therefore, the investors that want to sell their shares of the stock quickly will likely cancel their limit ask orders and resubmit them as market ask orders. The cancellation will cause the limit ask volume density to decrease, and we model this impact by $-j(x)\big(u(t,x)-u_0\big)^+$, with $j(x)$ a positive function decreasing in $x > 0$, meaning that the positively farther away a price level is from the mid-price, the less likely the investors will cancel the limit ask orders at the price level, as otherwise the loss would be too large. Assuming all the cancelled limit ask orders become market orders, these orders will cause the *absolute value* of the bid volume density to decrease, and we model this impact by $j(|x|)\big(u(t,|x|) - u_0\big)^+$. In summary, we model this scenario by

$$-j(x)\big(u(t,x) - u_0\big)^+\mathbf{1}_{\{x>0\}} + j(|x|)\big(u(t,|x|) - u_0\big)^+\mathbf{1}_{\{x<0\}}.$$

For $x < 0$, it is symmetric. When $u(t,x) \leq -u_0$, it means that the limit bid queue is too long for the investors. Therefore, the investors that want to buy the stock quickly will likely cancel their limit bid orders and resubmit them as market bid orders. The cancellation will cause the *absolute value* of the limit bid volume density to decrease, and we model this impact by $j(x)\big(u(t,x) + u_0\big)^-$, with $j(x)$ a positive function increasing in $x < 0$. The meaning is similar to the case of $x > 0$. Also, assuming that all cancelled limit bid orders become market bid orders, then these orders will cause the limit ask volume density to decrease, and we model this impact by $-j(-|x|)\big(u(t,-|x|) + u_0\big)^-$. In summary, we model this scenario by

$$j(x)\big(u(t,x) + u_0\big)^-\mathbf{1}_{\{x<0\}} - j(-|x|)\big(u(t,-|x|) + u_0\big)^-\mathbf{1}_{\{x>0\}}.$$

Therefore, the rate of cancellation with market order replacement at time t and price level $S(t)+x$ can be modeled as

$$\begin{aligned} J(x,u(t,x)) = \mathbf{1}_{\{x>0\}}\big[-j(x)\big(u(t,x)-u_0\big)^+ - j(-|x|)\big(u(t,-|x|)+u_0\big)^-\big] \\ + \mathbf{1}_{\{x<0\}}\big[j(|x|)\big(u(t,|x|)-u_0\big)^+ + j(x)\big(u(t,x)+u_0\big)^-\big], \end{aligned}$$

with $j(x)$ a positive function decreasing in $x > 0$ and increasing in $x < 0$. Since the decreased density should not exceed the existing volume density above the threshold, we set $j(\cdot) \leq 1$.

5. **Submission of Orders**: The submission of limit orders and market orders are both largely influenced by the price, which in turn is largely influenced by the difference between the volume of the ask and bid orders around the mid-price. We introduce

$$\ell(t) = \int_{-\iota}^{\iota} u(t,y)dy,$$

with $\delta \leq \iota \ll L$. If $\ell(t) > 0$, then there are more limit ask orders than limit bid orders around the mid-price. If $\ell(t) < 0$, then there are more limit bid orders than limit ask orders around the mid-price.

For $x > 0$, when $\ell(t) > 0$, there are already too many ask orders. Therefore, rational investors are less likely to submit limit ask orders and maybe some investors will cancel their limit ask orders and wait until the ask and bid queues are balanced again. This will lead to the decreasing tendency of the limit ask orders. Clearly, it is acceptable that the larger the $\ell(t)$, the larger the decreasing tendency. Hence, we model this by $G(x,\ell(t))$, with a function $G(x,\ell)$ strictly decreasing in ℓ and $G(x,0)=0$.

For $x > 0$ and $\ell(t) < 0$, there are already too many bid orders, which might signal a large demand for the stock. Therefore, rational investors are more likely to submit limit ask orders than to rush the sale with market ask orders, for a potential increase in the mid-price. This will lead to the increasing tendency of the limit ask orders. Clearly, it is acceptable that the smaller the $\ell(t)$, the larger the increasing tendency. Hence, we still model this by $G(x, \ell(t))$ strictly decreasing in ℓ and $G(x, 0) = 0$.

For $x < 0$, it is symmetric: If $\ell(t) < 0$, then there are already too many bid orders. Thus, rational investors are less likely to submit limit bid orders and maybe some investors will cancel their limit bid orders and wait until the ask and bid queues are balanced again. This will lead to the decreasing tendency of the limit bid orders. Clearly, it is acceptable that the smaller the $\ell(t)$, the larger the decreasing tendency to the *absolute value* of limit bid orders, which means the larger the increasing tendency to the bid volume density. Hence, we still model this by $G(x, \ell(t))$ strictly decreasing in ℓ and $G(x, 0) = 0$.

For $x < 0$ and $\ell(t) > 0$, there are already too many ask orders, which might signal a large supply for the stock. Therefore, rational investors are more likely to submit limit bid orders that to rush the purchase with market bid orders, for a potential decrease in the mid-price. This will lead to the increasing tendency of the limit bid orders. Clearly, it is acceptable that the larger the $\ell(t)$, the larger the increasing tendency to the *absolute value* of limit bid orders, which means the larger the decreasing tendency to the bid volume density. Hence, we still model this by $G(x, \ell(t))$ strictly decreasing in ℓ and $G(x, 0) = 0$.

The impact of the non-HFT order flows can be summarized by the following differential equation for the centered order book density u:

$$du(t,x) = \Big[\eta(x)u_{xx}(t,x) - \beta(x)\,\mathrm{sgn}(x)[u_x(t,x)]^- - \zeta(x)u(t,x) + J(x,u(t,x)) + G(x,\ell(t))\Big]dt,$$

where

$$\eta(x) = \begin{cases} \eta_a & x \in (0, L] \\ \eta_b & x \in [-L, 0) \end{cases}, \qquad \beta(x) = \begin{cases} \beta_a & x \in (0, L] \\ \beta_b & x \in [-L, 0) \end{cases}, \qquad \zeta(x) = \begin{cases} \zeta_a & x \in (0, L] \\ \zeta_b & x \in [-L, 0) \end{cases}$$

with $\eta_a, \eta_b, \beta_a, \beta_b, \zeta_a, \zeta_b$ positive constants, and

$$\begin{aligned} J(x, u(t,x)) = \mathbf{1}_{\{x>0\}} \big[- j(x)\big(u(t,x) - u_0\big)^+ - j(-|x|)\big(u(t,-|x|) + u_0\big)^- \big] \\ + \mathbf{1}_{\{x<0\}} \big[j(|x|)\big(u(t,|x|) - u_0\big)^+ + j(x)\big(u(t,x) + u_0\big)^- \big], \end{aligned}$$

with $u_0 > 0$, and $j(x) \leq 1$ a positive function decreasing in $x > 0$ and increasing in $x < 0$. The function $G(x, \ell(t))$ is strictly decreasing in $\ell(t)$ and $G(x, 0) = 0$, with

$$\ell(t) = \int_{-\iota}^{\iota} u(t, y) dy.$$

3.2 HFT Orders

In this subsection, we are modeling the HFT orders. We assume that the HFT orders mainly occur near the mid-price and on average provide zero or very small net contribution in volume to the limit order book. Thus, roughly speaking, the HFT dynamics are alsmost like a zero mean noise process.

3.2.1 A microscopic volume model

In order to model the volume of HFT orders on both sides of the market, we consider the following six types of market events: Submission of limit ask/bid orders, cancellation of limit ask/bid orders, and submission of market ask/bid orders. To simplify the HFT microscopic model, we reduce the dimension of our model by combining the cancellation of limit ask (bid) orders with the submission of market bid (ask) orders since their impacts on the order dynamics are the same: decrease the volume of limit ask (bid) orders. For example, by cancelling an limit ask order, it is equivalent to putting a same size market bid order since both orders are executed against the existing limit ask orders.

Viewing HFT macroscopically, it is just like a noise, and viewing it microscopically, it is mutually self-exciting. Assume the average trading speed of HFT is n times per millisecond. Then, during the time interval $[0, t]$ (with t being measured by second), there would be $1000nt$ tradings. Thus, the number of HFTs is roughly the same as that of non-HFTs during $[0, 1000t]$. Now, in general if the ratio of the fast and slow times is T (instead of 1000), then within the (normal) time interval $[0, t]$, the average number of HFTs is roughly the same as those non-HFTs during $[0, tT]$. Hence, it is a suitable approach to investigate the HFT as follows: For a very large $T > 0$, consider a multi-dimensional Hawkes process (so that it is mutually exciting) on $[0, tT]$. Then letting $T \to \infty$ with a proper scaling (normalization), the limit will be a good approximation of a model for the HFT.

Our microscopic volume model is based on one 4-dimensional Hawkes process on $[0, tT]$, defined as the following:

$$\{\mathbf{N}(tT)\}_{t\geq 0} = \begin{pmatrix} N^{a,+}(tT) \\ N^{b,+}(tT) \\ N^{a,-}(tT) \\ N^{b,-}(tT) \end{pmatrix} \qquad ,$$

where $N^{a,+}(tT)$ ($N^{b,+}(tT)$) corresponds to the accumulative number of limit ask (bid) orders submitted in the time interval $[0, tT]$, and $N^{a,-}(tT)$ ($N^{b,-}(tT)$) to the accumulative number of combined market ask (bid) orders and cancelled bid (ask) orders in the time interval $[0, tT]$. See right below for details.

3.2.2 The HFT density

How do we use the limit of $\mathbf{N}_T(tT)$ to model the market macroscopically? First, we let the volume of HFT around the mid price $S_T(tT)$ at time tT be $V_T(tT)$. Then $V_T(tT)$ can be written as

$$V_T(tT) = N_T^{a,+}(tT) + N_T^{b,+}(tT) - N_T^{a,-}(tT) - N_T^{b,-}(tT).$$

Since $V_T(tT)$ is not density like $u(t,x)$, we cannot simply add $V_T(tT)$ to the centered order book density equation. However, we can write $u(t,x)$ into the following generic equation:

$$u(t,x) = \text{non-HFT density} + \text{HFT density}.$$

Since the HFT density is a part of $u(t,x)$, we let

$$\text{HFT density} = f(t) \cdot u(t,x),$$

with some function $f(t)$ valued in $(0,1)$, which serves as a ratio function in the model, so that $f(t) \cdot u(t,x)$ preserves the same macroscopic properties of a normalized $V_T(tT)$. Such a normalization is necessary because the amplitude of $V_T(tT)$ is divergent as $T \to \infty$, and only the limit of the normalized HFT volume can eventually capture the nature of the mean zero noise of HFT. Hence, we have

$$f(t) = \lim_{T\to\infty} \frac{V_T(tT)}{h(T)},$$

for some scalar factor $h(T)$. Therefore, we can model the change of HFT density as

$$df(t) \cdot u(t,x) + f(t) \cdot du(t,x).$$

Since the change of $u(t,x)$, observed in normal time like seconds, is significantly slower than the change of $V_T(tT)$, the impact from du to the change of HFT density is almost negligible. So we set $f(t)\cdot du(t,x)\approx 0$. Combining with the non-HFT density model, we have the following centered order book density equation:

$$
\begin{aligned}
du(t,x) = \Big[&\eta u_{xx}(t,x) - \beta\,\mathrm{sgn}(x)[u_x(t,x)]^- - \zeta u(t,x) + J(x,u(t,x)) + G(x,u(t,\cdot))\Big]dt \\
&+ u(t,x)df(t),
\end{aligned}
$$

where

$$
\begin{cases}
f(t) = \lim_{T\to\infty} \dfrac{V_T(tT)}{h(T)}, \\
V_T(tT) = N_T^{a,+}(tT) + N_T^{b,+}(tT) - N_T^{a,-}(tT) - N_T^{b,-}(tT).
\end{cases}
$$

3.2.3 *Settings of the Hawkes conditional intensity process*

In this subsection, we provide settings of $\boldsymbol{\lambda}(tT)$, the conditional intensity process associated with the Hawkes process $\{\mathbf{N}(tT)\}_{tT\geq 0}$, to encode properties of the HFT market. We define

$$
\boldsymbol{\lambda}(tT) := \begin{pmatrix} \lambda^{a,+} \\ \lambda^{b,+} \\ \lambda^{a,-} \\ \lambda^{b,-} \end{pmatrix}(tT),
$$

and it is of the form

$$
\boldsymbol{\lambda}(tT) = \boldsymbol{\mu}(tT) + \int_0^{tT} \boldsymbol{\Phi}(tT-s)d\mathbf{N}(s) \equiv \mu(tT) + \Big[\boldsymbol{\Phi} * d\mathbf{N}\Big](tT),
$$

where

$$\boldsymbol{\mu}(\cdot) = \begin{pmatrix} \mu^{a,+} \\ \mu^{b,+} \\ \mu^{a,-} \\ \mu^{b,-} \end{pmatrix}(\cdot), \qquad \boldsymbol{\Phi}(\cdot) = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{14} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{24} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi_{34} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} \end{pmatrix}(\cdot).$$

For the subscripts of each entry of $\boldsymbol{\Phi}(\cdot)$, 1 stands for limit ask orders, 2 for limit bid orders, 3 for market ask orders, and 4 for market bid orders.

In the conditional intensity process, $\boldsymbol{\mu}$ models the conditional intensity that a new HFT order event is induced exogenously. For example, $\mu^{a,+}(\cdot)$ models the conditional intensity that a new HFT limit ask order is submitted due to some exogenous reason. The kernel matrix $\boldsymbol{\Phi}(\cdot)$ models the endogenous induction power from past events. For example, $\int_0^{\cdot} \varphi_{42}(\cdot - s)dN^{b,+}(s)$ models the conditional intensity of market bid order submission induced by past limit bid order submissions. We summarize in the following table the endogenous induction power from all the entries in $\boldsymbol{\Phi}(\cdot)$:

Table 3.1: Endogeneous Induction Power of Conditional Intensity Parameters

Conditional intensity of current order sub- -mission induced by past order submission		Current Order Submission			
		Limit Ask	Limit Bid	Market Ask	Market Bid
Past Order Submission	Limit Ask	$\varphi_{11} * dN^{a,+}$	$\varphi_{21} * dN^{a,+}$	$\varphi_{31} * dN^{a,+}$	$\varphi_{41} * dN^{a,+}$
	Limit Bid	$\varphi_{12} * dN^{b,+}$	$\varphi_{22} * dN^{b,+}$	$\varphi_{32} * dN^{b,+}$	$\varphi_{42} * dN^{b,+}$
	Market Ask	$\varphi_{13} * dN^{a,-}$	$\varphi_{23} * dN^{a,-}$	$\varphi_{33} * dN^{a,-}$	$\varphi_{43} * dN^{a,-}$
	Market Bid	$\varphi_{14} * dN^{b,-}$	$\varphi_{24} * dN^{b,-}$	$\varphi_{34} * dN^{b,-}$	$\varphi_{44} * dN^{b,-}$

Now we examine the relationships among the functions $\varphi_{ij}(\cdot)$'s to simplify the matrix $\boldsymbol{\Phi}(\cdot)$:

1. Institutional investors normally split large orders (called parent orders) into smaller orders (called children orders) and execute these smaller orders in an extended time period [1] [35]. Therefore, we can assume that the conditional intensity of one type of HFT limit (market) order induced by the same type of HFT limit (market) orders in the past should be the same.

In other words, the conditional intensity of the submission of one limit ask order induced by past submissions of limit ask orders can be assumed to be the same with the conditional intensity of the submission of one limit bid order induced by past submissions of limit bid orders. We let this inducing effect be

$$\varphi_{11}(\cdot) = \varphi_{22}(\cdot) = \varphi(\cdot), \text{ where } \varphi(\cdot) \text{ is a positive bounded function with } \int_0^\infty \varphi(s)ds = 1. \tag{3.1}$$

As for the market orders, we also assume that the conditional intensity of the submission of one market ask order induced by past submissions of market ask orders is the same with the conditional intensity of the submission of one market bid order induced by past submissions of market bid orders. However, institutional investors tend to use limit parent orders over market parent orders, due to the lack of price control of the market parent orders [47]. Therefore, we let the inducing effect between market orders be $\beta_1\varphi(\cdot)$ with $\beta_1 < 1$.

We also assume momentum effect in market orders because some individual investors want to follow the market move and they usually want to execute their orders immediately. However, these individual investors usually do not have orders nearly as large as the ones from institutional investors, and hence this inducing power is less than that from the parent orders. We use $\beta_2\varphi(\cdot)$ with $0 < \beta_2 < \beta_1$ to model this momentum effect. Combining the momentum effect with the inducing effect on market orders, we let

$$\varphi_{33}(\cdot) = \varphi_{44}(\cdot) = (\beta_1 + \beta_2)\varphi(\cdot). \tag{3.2}$$

Note that there will be no restriction of the size on the positive number $\beta_1 + \beta_2$, so it is possible that $\beta_1 + \beta_2 > 1$, or namely, it is possible that

$$\varphi_{33}(\cdot) = \varphi_{44}(\cdot) > \varphi_{11}(\cdot) = \varphi_{22}(\cdot)$$

2. Market orders near the mid-price can potentially deplete the queues near the mid-price, which could lead to price changes, and the price changes in turn could lead to the submission of limit orders on the same side. For example, in the illustrative LOB (1.1), if an investor places a market ask order for 2000 shares at 10:01 am, the market ask order will be executed at the price level \$100.00 against the bid queue at that price level. Since the bid queue at the price level \$100.00 only has 2000 shares, it will be depleted and the best bid price will decrease by 1 tick to \$99.99. Meanwhile, the best ask price will also decrease by 1 tick to \$100.00. This will likely induce the submission of limit ask orders at the new best ask price by market makers, who place limit orders at the best bid and ask prices to earn the spread. We assume that this inducing effect from market orders to limit orders on the same side is the same momentum effect between market orders, since they are both driven by price changes, so we also use β_2 to model this effect, namely,

$$\varphi_{13}(\cdot) = \varphi_{24}(\cdot) = \beta_2\varphi(\cdot)$$

On the other hand, the high frequency limit orders signal a demand on one side, which could induce market order on the same side because speculating investors might want to act before large limit orders. For example, if an investor observes a stable flow of incoming limit bid orders from the same institution, this could signal a parent limit bid order, which will typically take hours or even days to complete and will potentially raise the price due to the increased demand. The investor might want to submit market bid orders so that she can buy shares of the stock at \$100.01, the current best ask price, before the potential price increase caused by the completion of this parent limit bid order. After the entire parent limit bid order is placed, she could place a market ask order to sell these shares back to the institution at a price higher than \$100.01. For the first step of this strategy, we assume that this inducing effect from limit orders to market orders on the same side is the same momentum effect between market orders, since they are both driven by price changes, so we still use β_2 to model this effect, namely,

$$\varphi_{31}(\cdot) = \varphi_{42}(\cdot) = \beta_2 \varphi(\cdot)$$

As for the second step of this strategy, the investor in our example might have the wrong speculation: The flow of limit bid orders might not end up being a parent limit bid order, or the price might not increase from the sequence of limit bid orders. In this case, the investor might not submit the market ask order since it would not profit her. Therefore, we assume that the inducing effect from limit orders to market orders on the opposite side is less than that to the market orders on the same side, namely,

$$\varphi_{41}(\cdot) < \varphi_{31}(\cdot), \qquad \varphi_{32}(\cdot) < \varphi_{42}(\cdot).$$

We use $\beta_3 < 1$ to model the inequalities and have

$$\varphi_{41}(\cdot) = \beta_3 \varphi_{31}(\cdot) = \beta_3 \beta_2 \varphi(\cdot), \qquad \varphi_{32}(\cdot) = \beta_3 \varphi_{42}(\cdot) = \beta_3 \beta_2 \varphi(\cdot).$$

3. We assume that the same event on opposite sides induce each other in the same way but very close to 0. For example, we assume that the submissions of limit ask orders barely induce the submissions of limit bid order, which is observed by the numerical results from [2] and [7]. So we have

$$\varphi_{12}(\cdot) = \varphi_{21}(\cdot) = \varphi_{34}(\cdot) = \varphi_{43}(\cdot) = 0 \tag{3.3}$$

4. We assume that the inducing power between child orders of the same parent order is much larger than the inducing power between different types of orders. For example, an institutional investor wants to buy 50,000 shares of a stock and he uses an HFT algorithm to submit the limit bid orders sequentially. Some individual speculators might want to submit market

bid orders to buy some shares before the parent order and then submit market ask orders to sell these shares back to the institutional investor to make a profit. A child limit bid order almost guarantees the submission of another child limit bid order since they are both a part of the same parent order, while the market bid and ask orders might not be induced by a child limit bid order, since the speculators might not foresee the parent order or believe the price will increase. Adding (3.3), we can assume that the past limit ask order submissions are more likely to induce limit ask order submissions than they induce limit bid order submission, market ask order submission, and market bid order submission combined. This example roughly translates to

$$\varphi_{11}(\cdot) > \varphi_{21}(\cdot) + \varphi_{31}(\cdot) + \varphi_{41}(\cdot).$$

After we apply this assumption to all the order events, we have

$$\varphi_{22}(\cdot) > \varphi_{12}(\cdot) + \varphi_{32}(\cdot) + \varphi_{42}(\cdot),$$

$$\varphi_{33}(\cdot) > \varphi_{13}(\cdot) + \varphi_{23}(\cdot) + \varphi_{43}(\cdot),$$

$$\varphi_{44}(\cdot) > \varphi_{14}(\cdot) + \varphi_{24}(\cdot) + \varphi_{34}(\cdot),$$

These inequalities lead to:

$$1 - \beta_2\beta_3 - \beta_2 > 0 \tag{3.4}$$

5. Since we assume that the HFT orders provide zero net contribution in volume to the limit order book on average, we have

$$\mathbb{E}[dV(tT)] = \mathbb{E}[dN^{a,+}(tT) + dN^{b,+}(tT) - dN^{a,-}(tT) - dN^{b,-}(tT)] = 0$$

Note that

$$\mathbb{E}[dN^{a,+}(tT)] + \mathbb{E}[dN^{b,+}(tT)] = \mathbb{E}[\lambda^{a,+}(tT)]d(tT) + \mathbb{E}[\lambda^{b,+}(tT)]d(tT),$$

$$\mathbb{E}[dN^{a,-}(tT)] + \mathbb{E}[dN^{b,-}(tT)] = \mathbb{E}[\lambda^{a,-}(tT)]d(tT) + \mathbb{E}[\lambda^{b,-}(tT)]d(tT)$$

and

$$\begin{aligned}
\mathbb{E}[\lambda^{a,+}(tT)] &= \mu^{a,+}(tT) + \int_0^{tT} \varphi_{11}(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds + \int_0^{tT} \varphi_{12}(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds \\
&+ \int_0^{tT} \varphi_{13}(tT-s)\mathbb{E}[\lambda^{a,-}(s)]ds + \int_0^{tT} \varphi_{14}(tT-s)\mathbb{E}[\lambda^{b,-}(s)]ds \\
&= \mu^{a,+}(tT) + \int_0^{tT} \varphi(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds \\
&+ \int_0^{tT} \beta_2\varphi(tT-s)\mathbb{E}[\lambda^{a,-}(s)]ds + \int_0^{tT} \varphi_{14}(tT-s)\mathbb{E}[\lambda^{b,-}(s)]ds,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\lambda^{b,+}(tT)] &= \mu^{b,+}(tT) + \int_0^{tT} \varphi_{21}(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds + \int_0^{tT} \varphi_{22}(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds \\
&+ \int_0^{tT} \varphi_{23}(tT-s)\mathbb{E}[\lambda^{a,-}(s)]ds + \int_0^{tT} \varphi_{24}(tT-s)\mathbb{E}[\lambda^{b,-}(s)]ds \\
&= \mu^{b,+}(tT) + \int_0^{tT} \varphi(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds \\
&+ \int_0^{tT} \varphi_{23}(tT-s)\mathbb{E}[\lambda^{a,-}(s)]ds + \int_0^{tT} \beta_2\varphi(tT-s)\mathbb{E}[\lambda^{b,-}(s)]ds,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\lambda^{a,-}(tT)] &= \mu^{a,-}(tT) + \int_0^{tT} \varphi_{31}(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds + \int_0^{tT} \varphi_{32}(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds \\
&+ \int_0^{tT} \varphi_{33}(tT-s)\mathbb{E}[\lambda^{a,-}(s)]ds + \int_0^{tT} \varphi_{34}(tT-s)\mathbb{E}[\lambda^{b,-}(s)]ds \\
&= \mu^{a,-}(tT) + \int_0^{tT} \beta_2\varphi(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds
\end{aligned}$$

$$+\int_0^{tT}\beta_2\beta_3\varphi(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds+\int_0^{tT}[(\beta_1+\beta_2)\varphi(tT-s)]\mathbb{E}[\lambda^{a,-}(s)]ds$$

$$\begin{aligned}\mathbb{E}[\lambda^{b,-}(tT)] &= \mu^{b,-}(tT)+\int_0^{tT}\varphi_{41}(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds+\int_0^{tT}\varphi_{42}(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds\\ &+\int_0^{tT}\varphi_{43}(tT-s)\mathbb{E}[\lambda^{a,-}(s)]ds+\int_0^{tT}\varphi_{44}(tT-s)\mathbb{E}[\lambda^{b,-}(s)]ds\\ &=\mu^{b,-}(tT)+\int_0^{tT}\beta_2\beta_3\varphi(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds\\ &+\int_0^{tT}\beta_2\varphi(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds+\int_0^{tT}[(\beta_1+\beta_2)\varphi(tT-s)]\mathbb{E}[\lambda^{b,-}(s)].ds\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathbb{E}[\lambda^{a,+}(tT)]+\mathbb{E}[\lambda^{b,+}(tT)] &= \mu^{a,+}(tT)+\mu^{b,+}(tT)\\ &+\int_0^{tT}\varphi(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds\\ &+\int_0^{tT}\varphi(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds\\ &+\int_0^{tT}[\beta_2\varphi(tT-s)+\varphi_{23}(tT-s)]\mathbb{E}[\lambda^{a,-}(s)]ds\\ &+\int_0^{tT}[\varphi_{14}(tT-s)+\beta_2\varphi(tT-s)]\mathbb{E}[\lambda^{b,-}(s)]ds,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\lambda^{a,-}(tT)]+\mathbb{E}[\lambda^{b,-}(tT)] &= \mu^{a,-}(tT)+\mu^{b,-}(tT)\\ &+\int_0^{tT}(\beta_2+\beta_2\beta_3)\varphi(tT-s)\mathbb{E}[\lambda^{a,+}(s)]ds\\ &+\int_0^{tT}(\beta_2+\beta_2\beta_3)\varphi(tT-s)\mathbb{E}[\lambda^{b,+}(s)]ds\\ &+\int_0^{tT}[(\beta_1+\beta_2)\varphi(tT-s)]\mathbb{E}[\lambda^{a,-}(s)]ds\end{aligned}$$

$$+\int_0^{tT}[(\beta_1+\beta_2)\varphi(tT-s)]\mathbb{E}[\lambda^{b,-}(s)]ds$$

Inspired by the simplification of a similar equation in [16], we assume that

$$\mu^{a,+}(tT)+\mu^{b,+}(tT)=\mu^{a,-}(tT)+\mu^{b,-}(tT), \qquad \varphi_{23}(tT)=\varphi_{14}(tT),$$

$$\mathbb{E}[\lambda^{a,+}(tT)]+\mathbb{E}[\lambda^{b,+}(tT)]=\mathbb{E}[\lambda^{a,-}(tT)]+\mathbb{E}[\lambda^{b,-}(tT)].$$

and get

$$\varphi(\cdot)+\beta_2\varphi(\cdot)+\varphi_{14}(\cdot)=\beta_2\beta_3\varphi(\cdot)+\beta_2\varphi(\cdot)+(\beta_1+\beta_2)\varphi(\cdot),$$

$$\varphi(\cdot)+\beta_2\varphi(\cdot)+\varphi_{23}(\cdot)=\beta_2\beta_3\varphi(\cdot)+\beta_2\varphi(\cdot)+(\beta_1+\beta_2)\varphi(\cdot),$$

which gives us

$$\varphi_{14}(\cdot)=(\beta_1+\beta_2+\beta_2\beta_3-1)\varphi(\cdot), \quad \varphi_{23}(\cdot)=(\beta_1+\beta_2+\beta_2\beta_3-1)\varphi(\cdot). \tag{3.5}$$

6. We assume that the same-side limit-market order induction power is greater than the opposite-side limit-market induction power, which is observed by the numerical results from [2] and [7]. For example, limit ask order submissions are more likely induced by past market ask order submissions than past market bid order submissions. Therefore, we have $\beta_2 > \beta_2\beta_3$, which is consistent with the setting that $\beta_3 < 1$. From this assumption, we also have

$$\beta_2 > \beta_1+\beta_2+\beta_2\beta_3-1 \tag{3.6}$$

Having the above, we have $\mathbf{\Phi}(\cdot) = \mathbf{\Phi}_0\varphi(\cdot)$, where

$$\mathbf{\Phi}_0 = \begin{pmatrix} 1 & 0 & \beta_2 & (\beta_1+\beta_2+\beta_2\beta_3-1) \\ 0 & 1 & (\beta_1+\beta_2+\beta_2\beta_3-1) & \beta_2 \\ \beta_2 & \beta_2\beta_3 & (\beta_1+\beta_2) & 0 \\ \beta_2\beta_3 & \beta_2 & 0 & (\beta_1+\beta_2) \end{pmatrix}, \tag{3.7}$$

with $\varphi(\cdot)$ a positive bounded function that satisfies $\int_0^\infty \varphi(s)ds = 1$, and

$$0 < \beta_2 < \beta_1 < 1, \qquad 0 < \beta_3 < 1, \qquad \beta_1 + \beta_2\beta_3 < 1 < \beta_1 + \beta_2 + \beta_2\beta_3.$$

Proposition 3.1.

1. The eigenvalues of $\mathbf{\Phi}_0$ are given by the following:

$$\lambda_1 = \beta_1 + \beta_2\beta_3 + 2\beta_2, \qquad \lambda_2 = -\beta_2\beta_3 + \beta_2 + 1,$$

$$\lambda_3 = \beta_1 + \beta_2\beta_3, \qquad \lambda_4 = -\beta_2\beta_3 - \beta_2 + 1$$

   Moreover, it holds that $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$. This also means that $\mathbf{\Phi}_0$ is diagonalizable.

2. Define:

$$v_1 = \begin{pmatrix} \frac{\beta_2(\beta_3+1)}{\beta_1+\beta_2\beta_3+2\beta_2-1} \\ \frac{\beta_2(\beta_3+1)}{\beta_1+\beta_2\beta_3+2\beta_2-1} \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} \frac{\beta_2(\beta_3-1)}{\beta_1+\beta_2\beta_3-1} \\ -\frac{\beta_2(\beta_3-1)}{\beta_1+\beta_2\beta_3-1} \\ -1 \\ 1 \end{pmatrix}, \quad v_4 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix},$$

   They are eigenvectors of $\mathbf{\Phi}_0^\top$ corresponding to λ_1, λ_2, λ_3, λ_4, respectively.

Proof. We provide the calculation of the eigenvalues in the appendices at A. We show here the inequalities among the eigenvalues can be easily shown using the assumptions. First, since $\beta_2 > 0$, $\lambda_2 > \lambda_4$. Second, since

$$\beta_1 + \beta_2 + \beta_2\beta_3 - 1 > 0 > -\beta_2\beta_3,$$

we have

$$\beta_1 + \beta_2\beta_3 > -\beta_2\beta_3 - \beta_2 + 1,$$

which leads to $\lambda_3 > \lambda_4$. As for λ_2 and λ_3, using (3.6), we have

$$\beta_1 + \beta_2\beta_3 = \Big[(\beta_1 + \beta_2 + \beta_2\beta_3 - 1) + \beta_2\beta_3\Big] - \beta_2 + 1 - \beta_2\beta_3$$

$$< 2\beta_2 - \beta_2 + 1 - \beta_2\beta_3 = -\beta_2\beta_3 + \beta_2 + 1$$

We have shown that $\lambda_2 > \lambda_3$. Lastly, we are left to show $\lambda_1 > \lambda_2$. Indeed, since

$$\beta_1 + 2\beta_2\beta_3 + \beta_2 - 1 > \beta_1 + \beta_2\beta_3 + \beta_2 - 1 > 0,$$

$$\beta_1 + \beta_2\beta_3 + 2\beta_2 > -\beta_2\beta_3 + \beta_2 + 1,$$

and hence $\lambda_3 < \lambda_1$. $\square$

We define

$$\mathbf{\Phi}_T(\cdot) = \frac{a_T}{\lambda_1}\varphi(\cdot)\mathbf{\Phi}_0, \tag{3.8}$$

with $\{a_T\}_{T\in\mathbb{Z}^+}$ an undetermined sequence of positive constants that are less than 1 but converge to 1. Then the largest eigenvalue of each $\int_0^\infty \mathbf{\Phi}_T(\cdot)$ forms the sequence $\{a_T\}_{T\in\mathbb{Z}^+}$. Based on the cluster representation of Hawkes process, illustrated and discussed in the appendices at B, we use the largest eigenvalue of $\int_0^\infty \mathbf{\Phi}_T(s)ds$ to model the percentage of endogenous orders in the HFT market, and thus in our model, the HFT market gets more and more endogenous over time.

CHAPTER 4: SCALING LIMIT OF THE MICROSCOPIC VOLUME MODEL

In this section, we find $f(t)$, the scaling limit of $V_T(tT)$ as well as the ratio function of the HFT density. We also find the proper expression of the normalizing factor $h(T)$ in the process.

4.1 An asymptotic framework

We first set up an asymptotic framework for our Microscopic volume mode. Let $\{\mathbf{N}_T(tT)\}_{t\geq 0}$ be a sequence of 4-dimensional Hawkes processes defined on $[0,T]$ in a sequence of probability spaces $(\Omega_T, \mathscr{F}_T, \mathbb{P}_T)$, indexed by $T \in \mathbb{Z}^+$, where T goes to infinity. For each T, $\mathbf{N}_T(0) = 0$, $\mathscr{F}_T$ is the σ-algebra generated by $\mathbf{N}_T(tT)$, and the conditional intensity process $\{\boldsymbol{\lambda}_T(tT)\}_{t\geq 0}$ is

$$\boldsymbol{\lambda}_T(tT) = \boldsymbol{\mu}_T(tT) + \int_0^{tT} \mathbf{\Phi}_T(tT-s)d\mathbf{N}_T(s), \tag{4.1}$$

with $\mathbf{\Phi}_T(\cdot)$ defined in (3.8). This setting allows $\{\mathbf{N}_T(tT)\}_{t\geq 0}$ to tend to an unstable Hawkes process, while maintaining stability for each $\mathbf{N}_T(tT)$. For each $\mathbf{\Phi}_T(\cdot)$ and each $i = 1, 2, 3, 4$, the eigenvalues can be calculated as

$$\lambda_{T,i}(\cdot) = \frac{a_T \lambda_i \varphi(\cdot)}{\lambda_1}$$

Furthermore, the eigenvectors of each $(\mathbf{\Phi}_T(\cdot))^\top$ can be taken as the ones of $\mathbf{\Phi}_0^\top$, which are calculated in Proposition 3.1, so we still denote them as v_1, v_2, v_3, v_4.

We let

$$\mathbf{M}_T(t) = \mathbf{N}_T(t) - \int_0^t \boldsymbol{\lambda}_T(s)ds, \tag{4.2}$$

which is the martingale associated with $\mathbf{N}_T(t)$, see [34]. Then

$$d\mathbf{N}_T(t) = d\mathbf{M}_T(t) + \boldsymbol{\lambda}_T(s)ds,$$

and we have

$$\begin{aligned}
\boldsymbol{\lambda}_T(t) &= \boldsymbol{\mu}_T(t) + \int_0^t \boldsymbol{\Phi}_T(t-s)d\mathbf{N}_T(s) \\
&= \boldsymbol{\mu}_T(t) + \int_0^t \boldsymbol{\Phi}_T(t-s)d\mathbf{M}_T(s) + \int_0^t \boldsymbol{\Phi}_T(t-s)\boldsymbol{\lambda}_T(s) \\
&= [\boldsymbol{\mu}_T + \boldsymbol{\Phi}_T * \mathbf{M}_T + \boldsymbol{\Phi}_T * \boldsymbol{\lambda}_T](t) \\
&= [\boldsymbol{\mu}_T + \boldsymbol{\Phi}_T * \mathbf{M}_T + \boldsymbol{\Phi}_T * (\boldsymbol{\mu}_T + \boldsymbol{\Phi}_T * \mathbf{M}_T + \boldsymbol{\Phi}_T * \boldsymbol{\lambda}_T)](t) \\
&= [\boldsymbol{\mu}_T + \boldsymbol{\Phi}_T * \boldsymbol{\mu}_T + \boldsymbol{\Phi}_T * \mathbf{M}_T + \boldsymbol{\Phi}_T * \boldsymbol{\Phi}_T * \mathbf{M}_T + \boldsymbol{\Phi}_T * \boldsymbol{\Phi}_T * \lambda_T](t) \\
&= ... \\
&= \big[\big(\sum_{k=0}^{m} \boldsymbol{\Phi}_T^{*k}\big) * \boldsymbol{\mu}_T + \big(\sum_{k=1}^{m+1} \boldsymbol{\Phi}_T^{*k}\big) * \mathbf{M}_T + \boldsymbol{\Phi}^{*(m+1)} * \boldsymbol{\lambda}_T\big)\big](t)
\end{aligned}$$

Note that

$$\boldsymbol{\Phi}_T^{*2}(t) = (\boldsymbol{\Phi}_T * \boldsymbol{\Phi}_T)(t) = \int_0^t \boldsymbol{\Phi}_T(t-s)\boldsymbol{\Phi}_T(s)ds = \frac{a_T}{\lambda_1^2}\big(\int_0^t \varphi(t-s)\varphi(s)ds\big)\boldsymbol{\Phi}_0^2 = \frac{a_T^2}{\lambda_1^2}\varphi^{*2}(t)\boldsymbol{\Phi}_0^2$$

By induction,

$$\boldsymbol{\Phi}_T^{*k}(t) = \frac{a_T^k}{\lambda_1^k}\varphi^{*k}(t)\boldsymbol{\Phi}_0^k, \qquad \forall k \geq 0.$$

Since

$$\sup_{t\geq 0} \|\varphi^{*k}(t)\| \equiv \|\varphi^{*k}\|_\infty \leq \|\varphi^{*(k+1)}\|_\infty \|\varphi\|_1 \leq ... \leq \|\varphi\|_\infty \|\varphi\|_1^{k-1} \leq \|\varphi\|_\infty < \infty,$$

and since $\mathbf{\Phi}_0$ is diagonalizable, there is an invertible P so that

$$\mathbf{\Phi}_0 = PDP^{-1}, \qquad D = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}$$

Hence,

$$\mathbf{\Phi}_T^{*k}(t) = a_T^k \varphi^{*k}(t) P \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \left(\frac{\lambda_2}{\lambda_1}\right)^k & 0 & 0 \\ 0 & 0 & \left(\frac{\lambda_3}{\lambda_1}\right)^k & 0 \\ 0 & 0 & 0 & \left(\frac{\lambda_4}{\lambda_1}\right)^k \end{pmatrix} P^{-1}$$

Consequently,

$$|\mathbf{\Phi}_T^{*k}(t)| \leq C a_T^k, \qquad \forall k \geq 0,$$

with the constant $C > 0$ independent of $k \geq 0$. Hence, by $0 < a_T < 1$, we see that

$$\boldsymbol{\lambda}_T(t) = \sum_{k=0}^{\infty} \left(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\right)^k [\varphi^{*k} * \boldsymbol{\mu}_T](t) + \sum_{k=1}^{\infty} \left(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\right)^k [\varphi^{*k} * \mathbf{M}_T](t) \tag{4.3}$$

Equation (4.3) also allows us to calculate the expected value of $\boldsymbol{\lambda_T}(tT)$. Indeed, we have

$$\mathbb{E}[\boldsymbol{\lambda}_T(t)] = \sum_{k=0}^{\infty} \left(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\right)^k [\varphi^{*k} * \boldsymbol{\mu}_T](t) \tag{4.4}$$

Note that

$$\begin{aligned} [\varphi^{*k} * \boldsymbol{\mu}_T](tT) &= \int_0^{tT} \varphi^{*k}(tT - s)\boldsymbol{\mu}_T(s)ds \\ &= \int_0^{t} \varphi^{*k}(tT - sT)\boldsymbol{\mu}_T(sT)d(sT) \end{aligned}$$

$$= T \int_0^t \varphi^{*k}\big((t-s)T\big)\boldsymbol{\mu}_T(sT)ds$$

and similarly,

$$\begin{aligned}[\varphi^{*k} * \mathbf{M}_T](tT) &= \int_0^{tT} \varphi^{*k}(tT-s)d\mathbf{M}_T(s) \\ &= \int_0^t \varphi^{*k}\big(tT-sT\big)d\mathbf{M}_T(sT)\end{aligned}$$

4.2 Assumptions and Intuitions of the Result

We first introduce an assumption on $\varphi(\cdot)$. As we mentioned in the introduction, many HFT orders are part of a larger parent order that typically takes hours or even days to fully execute, which can be observed in an HFT market by child orders exciting each other during this relatively long execution window. To model this relatively long-term influence, we need to choose an excitation function $\varphi(\cdot)$ such that the conditional intensity has a relatively slow decay. The Dirac delta function is obviously not a good choice for the excitation function, since it would mean that an order no longer excites other orders right after it arrives at the HFT market. Although the exponential function, i.e. ae^{-bt} with $a, b > 0$, is a common choice for the Hawkes excitation function, it yields an exponentially decaying conditional intensity, which is faster than the power-law decay. Therefore, we model this long-term influence by giving each $\varphi(\cdot)$ a power-law tail. This leads to the following assumption:

Assumption 1: The function $\varphi(\cdot)$ is positive, bounded, integrable, with $\int_0^\infty \varphi(s)ds = 1$, and

$$\lim_{t\to\infty} t^\alpha \Big(1 - \int_0^t \varphi(s)ds\Big) = K,$$

with some $\alpha \in (0,1)$ and positive constant K.

The above gives the speed of convergence of $\int_0^t \varphi(s) \to 1$ as $t \to \infty$.

Remark 4.1. One example of such a function $\varphi(\cdot)$ is $f^{\alpha,1}(\cdot)$, the *Mittag-Leffler density function* with $\lambda = 1$. For $(\alpha, \lambda) \in (0,1) \times \mathbb{R}_+$, namely,

$$f^{\alpha,\lambda}(t) = \lambda t^{\alpha-1} E_{\alpha,\alpha}(-\lambda t^{\alpha}), \qquad t > 0,$$

where $E_{\alpha,\alpha}$, called the *Mittag-Leffler function*, is defined for $z \in \mathbb{C}$ as

$$E_{\alpha,\alpha}(z) = \sum_{n \geq 0} \frac{z^n}{\Gamma(\alpha n + \alpha)},$$

with $\Gamma(\cdot)$ the Gamma function. See Definition F.1.

Since the conditional intensity process essentially defines $\mathbf{N}_T(tT)$, we next try to find the asymptotic behavior, along with the proper normalizing factor, of $\boldsymbol{\lambda}_T(tT)$. Our first step is to find the limit of $\mathbb{E}[\boldsymbol{\lambda}_T(tT)]$ as $t \to \infty$.

Proposition 4.1. The expectation of the conditional intensity $\boldsymbol{\lambda}_T(tT)$, calculated in (4.4), is as follows:

$$\lim_{t\to\infty} \mathbb{E}[\boldsymbol{\lambda}_T(tT)] = \mathbf{P} \begin{pmatrix} \frac{1}{1-a_T} & 0 & 0 & 0 \\ 0 & \frac{1}{1-a_T(\lambda_2/\lambda_1)} & 0 & 0 \\ 0 & 0 & \frac{1}{1-a_T(\lambda_3/\lambda_1)} & 0 \\ 0 & 0 & 0 & \frac{1}{1-a_T(\lambda_4/\lambda_1)} \end{pmatrix} \mathbf{P}^{-1} \boldsymbol{\mu}_T(\infty),$$

where

(1). The constants $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are eigenvalues of $\boldsymbol{\Phi}_0$, which are calculated in Proposition 3.1.

(2). The matrix $\mathbf{P}$ is invertible and is consisting of the eigenvectors of $\mathbf{\Phi}_0$ correspondent to $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.

(3). The vector $\boldsymbol{\mu}_T(\infty)$, defined as

$$\boldsymbol{\mu}_T(\infty) := \lim_{t \to \infty} \boldsymbol{\mu}_T(tT),$$

is a four dimensional vector with positive constant entries.

Proof. Denote $\bar{\boldsymbol{\lambda}}_T(tT) = \mathbb{E}[\boldsymbol{\lambda_T}(tT)]$, then by (4.4), we have

$$\bar{\boldsymbol{\lambda}}_T(tT) = \boldsymbol{\mu}_T(tT) + \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n [\varphi^{*n} * \boldsymbol{\mu}_T](tT)$$

As t goes to infinity, by the dominated convergence theorem and monotone convergence theorem, along with Lemma 2.1, we have

$$\begin{aligned}
\lim_{t \to \infty} \bar{\boldsymbol{\lambda}}_T(tT) &= \lim_{T \to \infty} \boldsymbol{\mu}_T(tT) + \lim_{T \to \infty} \int_0^{tT} \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \varphi^{*n}(s) \boldsymbol{\mu}_T(tT - s) ds \\
&= \boldsymbol{\mu}_T(\infty) + \int_0^{\infty} \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \varphi^{*n}(s) \boldsymbol{\mu}_T(\infty) ds \\
&= \boldsymbol{\mu}_T(\infty) + \Big[\sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \int_0^{\infty} \varphi^{*n}(s) ds\Big] \boldsymbol{\mu}_T(\infty) \\
&= \boldsymbol{\mu}_T(\infty) + \Big[\sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \Big(\int_0^{\infty} \varphi(s) ds\Big)^n\Big] \boldsymbol{\mu}_T(\infty) \\
&= \Big[\sum_{n=0}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n\Big] \boldsymbol{\mu}_T(\infty)
\end{aligned}$$

Since $\mathbf{\Phi}_0$ has distinct eigenvalues, it is diagonalizable. Thus, there exists a diagonal matrix $\mathbf{D}$ whose diagonal entries are the eigenvalues of $\mathbf{\Phi}_0$, and an invertible matrix $\mathbf{P}$ consisting of the correspondent eigenvectors such that

$$\mathbf{\Phi}_0 = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

with

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}.$$

Therefore, we have

$$\lim_{t\to\infty} \bar{\boldsymbol{\lambda}}_T(tT) = \Big[\sum_{n=0}^{\infty}\Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n\Big]\boldsymbol{\mu}_T(\infty) = \mathbf{P}\Big(\sum_{n=0}^{\infty}\Big(\frac{a_T}{\lambda_1}\mathbf{D}\Big)^n\Big)\mathbf{P}^{-1}\boldsymbol{\mu}_T(\infty).$$

Since $a_T < 1$ and $\lambda_1 > \max\{\lambda_2, \lambda_3, \lambda_4\}$, all the diagonal entries of $\mathbf{D}$ are less than 1, and hence we have

$$\lim_{t\to\infty} \mathbb{E}[\boldsymbol{\lambda}_T(tT)] = \mathbf{P}\begin{pmatrix} \frac{1}{1-a_T} & 0 & 0 & 0 \\ 0 & \frac{1}{1-a_T(\lambda_2/\lambda_1)} & 0 & 0 \\ 0 & 0 & \frac{1}{1-a_T(\lambda_3/\lambda_1)} & 0 \\ 0 & 0 & 0 & \frac{1}{1-a_T(\lambda_4/\lambda_1)} \end{pmatrix}\mathbf{P}^{-1}\boldsymbol{\mu}_T(\infty).$$

$\square$

For each T, we define μ_T as the sum of the entries of $\boldsymbol{\mu}_T(\infty)$. Note that μ_T and a_T have different financial meanings as well as functions in our model, and thus they are not necessarily related.

Then we have

$$
\begin{aligned}
\lim_{t\to\infty} \frac{1-a_T}{\mu_T}\mathbb{E}[\boldsymbol{\lambda}_T(tT)] &= \frac{1-a_T}{\mu_T}\lim_{t\to\infty}\mathbb{E}[\boldsymbol{\lambda}_T(tT)] \\
&= \mathbf{P}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1-a_T}{1-a_T(\lambda_2/\lambda_1)} & 0 & 0 \\ 0 & 0 & \frac{1-a_T}{1-a_T(\lambda_3/\lambda_1)} & 0 \\ 0 & 0 & 0 & \frac{1-a_T}{1-a_T(\lambda_4/\lambda_1)} \end{pmatrix}\mathbf{P}^{-1}\frac{1}{\mu_T}\boldsymbol{\mu}_T(\infty).
\end{aligned}
$$

Since $\lambda_i < \lambda_1$ for all $i = 2, 3, 4$, this is a nontrivial limit as $t \to \infty$ for each T. Therefore, as $T \to \infty$, with proper assumptions on $\frac{1-a_T}{\mu_T}$, the normalized $\boldsymbol{\lambda}_T(tT)$ has a nontrivial limit, meaning that it is neither 0 nor ∞.

After rescaling $\boldsymbol{\lambda}_T(tT)$ with $\dfrac{1-a_T}{\mu_T}$, and by (4.3), we have

$$
\begin{aligned}
\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) = (1-a_T)\frac{\boldsymbol{\mu}_T(tT)}{\mu_T} + (1-a_T)\sum_{n=1}^{\infty}\left(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\right)^n[\varphi^{*n} * \frac{\boldsymbol{\mu}_T}{\mu_T}](tT) \\
+ (1-a_T)\sum_{n=1}^{\infty}\left(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\right)^n[\varphi^{*n} * \frac{\mathbf{M}_T}{\mu_T}](tT)
\end{aligned}
\tag{4.5}
$$

Like in [16], we use the orthogonal decomposition of $\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ to find its limit. Let (e_1, e_2, e_3, e_4) be an orthonormal basis of $\mathbb{R}^4$, such that $e_1 \cdot v_1 > 0$ and

$$
\text{span}(e_2, e_3, e_4) = \text{span}(v_2, v_3, v_4).
$$

Decomposing $\boldsymbol{\lambda}_T(tT)$ in the basis $\{e_1, e_2, e_3, e_4\}$, we have

$$\begin{aligned}\boldsymbol{\lambda}_T(tT) &= \sum_{i=1}^{4} \Big(e_i^\top \boldsymbol{\lambda}_T(tT)\Big) e_i \\ &= \frac{1}{e_1^\top v_1}\Big(v_1^\top \boldsymbol{\lambda}_T(tT)\Big) e_1 + \Big((v')^\top \boldsymbol{\lambda}_T(tT)\Big) e_1 + \sum_{i=2}^{4} \Big(e_i^\top \boldsymbol{\lambda}_T(tT)\Big) e_i,\end{aligned} \tag{4.6}$$

with

$$v' = e_1 - \frac{1}{e_1^\top v_1} v_1 \in \text{span}(v_2, v_3, v_4).$$

We can see that the re-scaled asymptotic behavior of $\boldsymbol{\lambda}_T(tT)$ depends on the re-scaled asymptotic behaviors of $v_i^\top \boldsymbol{\lambda}_T(tT)$. In particular, under proper settings, we can show that for $i = 2, 3, 4$, we have $v_i^\top \boldsymbol{\lambda}_T(tT) = 0$, which leads to

$$(v')^\top \boldsymbol{\lambda}_T(tT) = e_i^\top \frac{1 - a_T}{\mu_T} \boldsymbol{\lambda}_T(tT) = 0,$$

This means that the re-scaled asymptotic behavior of $\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ depends on that of $v_1^\top \boldsymbol{\lambda}_T(tT)$.

Remark 4.2. One example of such an orthonormal basis is

$$\left\{ \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \right\}$$

Meanwhile, we can see for each $i = 1, 2, 3, 4$ and for any $k \geq 1$,

$$v_i^\top \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1} \boldsymbol{\Phi}_0\Big)^n \varphi^{*n}(t) = \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\Big)^n v_i^\top \boldsymbol{\Phi}_0^k \varphi^{*n}(t)$$

$$= \Big[\sum_{n=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^n \varphi^{*n}(t)\Big] v_i^\top$$
$$= \psi_{T,i}(t) v_i^\top,$$

where

$$\psi_{T,i}(t) := \sum_{n=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^n \varphi^{*n}(t) \tag{4.7}$$

Therefore, along with (4.3), for each $i = 1, 2, 3, 4$, we have

$$\begin{aligned} v_i^\top \boldsymbol{\lambda}_T(tT) &= v_i^\top \boldsymbol{\mu}_T(tT) + v_i^\top \sum_{n=1}^{\infty}\Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n [\varphi^{*n} * \boldsymbol{\mu}_T](tT) + v_i^\top \sum_{n=1}^{\infty}\Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n [\varphi^{*n} * \mathbf{M}_T](tT) \\ &= v_i^\top \boldsymbol{\mu}_T(tT) + \int_0^t T\psi_{T,i}(T(t-s))(v_i^\top \boldsymbol{\mu}_T(sT))ds + \int_0^t \psi_{T,i}(T(t-s))(v_i^\top d\mathbf{M}_T(sT)). \end{aligned} \tag{4.8}$$

For each i, the Laplace transform of $\psi_{T,i}(Tt)$ is:

$$\begin{aligned} \hat{\psi}_{T,i}(T\cdot)(z) &= \int_0^\infty \Psi_{T,i}(Tx)e^{-xz}dx \\ &= \int_0^\infty \sum_{k=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^k \varphi^{*k}(Tx)e^{-xz}dx \\ &= \sum_{k=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^k \int_0^\infty \varphi^{*k}(Tx)e^{-xz}dx \\ &= \sum_{k=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^k \widehat{\varphi^{*k}}(T\cdot)(z) \\ &= \sum_{k=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^k \frac{1}{T}\widehat{\varphi^{*k}}\Big(\frac{z}{T}\Big) \\ &= \frac{1}{T}\sum_{k=1}^{\infty}\Big(\frac{a_T\lambda_i}{\lambda_1}\Big)^k \Big(\hat{\varphi}\Big(\frac{z}{T}\Big)\Big)^k \\ &= \frac{1}{T}\sum_{k=1}^{\infty}\Big[\frac{a_T\lambda_i}{\lambda_1}\hat{\varphi}\Big(\frac{z}{T}\Big)\Big]^k \end{aligned}$$

$$= \frac{\frac{a_T\lambda_i}{\lambda_1}\hat{\varphi}\Big(\dfrac{z}{T}\Big)}{T\Big(1 - \frac{a_T\lambda_i}{\lambda_1}\hat{\varphi}\Big(\dfrac{z}{T}\Big)\Big)}$$

Since

$$\hat{\varphi}\Big(\frac{z}{T}\Big) = \int_0^\infty \varphi(x)e^{-\frac{zx}{T}}dx,$$

for $i = 2, 3, 4$, as $T \to \infty$, we have

$$\frac{a_T\lambda_i}{\lambda_1}\hat{\varphi}\Big(\frac{z}{T}\Big) \to \frac{\lambda_i}{\lambda_1} < 1.$$

Hence, $\psi_{T,i}(T\cdot)$ goes to 0 as $T \to \infty$. Plugging this back in (4.8), we have

$$v_i^\top \boldsymbol{\lambda}_T(tT) \sim v_i^\top \boldsymbol{\mu}_T(tT)$$

This means that if we assume all the entries of $\boldsymbol{\mu}_T(tT)$ are identical, as T goes to infinity, $v_i^\top \boldsymbol{\lambda}_T(tT)$ goes to 0. Therefore, we have the following assumption:

Assumption 2': For each T, the background intensity $\boldsymbol{\mu}_T(tT)$ has identical entries.

Under this assumption, the re-scaled asymptotic behavior of $\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ depends on that of $v_1^\top \boldsymbol{\lambda}_T(tT)$. Indeed, we have

$$\begin{aligned}
\frac{1-a_T}{\mu_T}v_1^\top \boldsymbol{\lambda}_T(tT) &= \frac{1-a_T}{\mu_T}(v_1^\top \boldsymbol{\mu}_T(tT)) + \int_0^t \frac{T(1-a_T)}{\mu_T}\psi_{T,1}(T(t-s))(v_1^\top \boldsymbol{\mu}_T(sT))ds \\
&+ \int_0^t \frac{1-a_T}{\mu_T}\psi_{T,1}(T(t-s))(v_1^\top d\mathbf{M}_T(sT)) \\
&= \frac{1-a_T}{\mu_T}(v_1^\top \boldsymbol{\mu}_T(tT)) + \int_0^t \frac{T(1-a_T)}{\mu_T}\psi_{T,1}(T(t-s))(v_1^\top \boldsymbol{\mu}_T(sT))ds \\
&+ \sqrt{\frac{T(1-a_T)}{\mu_T}}\int_0^t \psi_{T,1}(T(t-s))\sqrt{\frac{1-a_T}{\mu_T}(v_1^2)^\top \boldsymbol{\lambda}_T(sT)}dB_{T,1}(s).
\end{aligned}$$

where

$$B_{T,1}(t) = \frac{1}{\sqrt{T}} \int_0^{tT} \frac{v_1^\top d\mathbf{M}_T(s)}{\sqrt{(v_1^2)^\top \boldsymbol{\lambda}_T(s)}}, \tag{4.9}$$

with $v_1^2 = (v_{1,1}^2, v_{1,2}^2, v_{1,3}^2, v_{1,4}^2)$. Note that

$$(v_1^2)^\top \boldsymbol{\lambda}_T(t) = (v_1)^\top \text{diag}[\boldsymbol{\lambda}_T(t)] v_1.$$

Thus,

$$\mathbb{E}[B_{T,1}(t)^2] = \mathbb{E}\Big[\Big| \int_0^{tT} \frac{d(v_1^\top \mathbf{M}_T)(s)}{\sqrt{T(v_1^2)^\top \boldsymbol{\lambda}_T(s)}} \Big|^2\Big] = \frac{1}{T} \int_0^{tT} ds = t.$$

Therefore, the limit of $B_{T,1}(\cdot)$ as $T \to \infty$ is a Brownian motion (see [16], p.254). Next, we examine the asymptotic behavior of

$$\rho_{T,1}(x) := T(1 - a_T)\psi_{T,1}(Tx)$$

Indeed, the Laplace transform of ρ is

$$\hat{\rho}_{T,1}(z) = \int_0^\infty \rho_{T,1}(x) e^{-zx} dx = (1 - a_T)\hat{\psi}_{T,1}\Big(\frac{z}{T}\Big) = (1 - a_T) \frac{\hat{\lambda}_{T,1}\Big(\frac{z}{T}\Big)}{\Big(1 - \hat{\lambda}_{T,1}\Big(\frac{z}{T}\Big)\Big)}$$

Since

$$\begin{aligned}
\hat{\lambda}_{T,1}(z) &= \int_0^\infty \lambda_{T,1}(x) e^{-zx} dx \\
&= a_T\Big(1 - z \int_0^\infty \int_x^\infty \varphi(s) ds e^{-zx} dx\Big) \\
&= a_T\Big(1 - z^\alpha \int_0^\infty \Big(\frac{x}{z}\Big)^\alpha \int_{\frac{x}{z}}^\infty \varphi(u) du x^{-\alpha} e^{-x} dx\Big),
\end{aligned}$$

using Assumption 1 and the dominated convergence theorem,

$$\hat{\lambda}_{T,1}(z) = a_T\Big(1 - K\Gamma(1-\alpha)z^{\alpha} + o(z)\Big)$$

Then we have

$$\begin{aligned}\hat{\rho}_{T,1}(z) &= \frac{(1-a_T)a_T\Big(1 - K\Gamma(1-\alpha)\frac{z^{\alpha}}{T^{\alpha}} + o(\frac{z}{T})\Big)}{1 - a_T\Big(1 - K\Gamma(1-\alpha)\frac{z^{\alpha}}{T^{\alpha}} + o(\frac{z}{T})\Big)} \\ &= \frac{\frac{T^{\alpha}(1-a_T)a_T}{K\Gamma(1-\alpha)} - (1-a_T)a_T z^{\alpha} + \frac{(1-a_T)a_T}{K\Gamma(1-\alpha)}T^{\alpha}o(\frac{z}{T})}{\frac{(1-a_T)T^{\alpha}}{K\Gamma(1-\alpha)} + a_T z^{\alpha} - \frac{a_T}{K\Gamma(1-\alpha)}T^{\alpha}o(\frac{z}{T})}.\end{aligned}$$

Let $\nu_T = \dfrac{T^{\alpha}(1-a_T)}{K\Gamma(1-\alpha)}$, then as $T \to \infty$,

$$\hat{\rho}_{T,1} = \frac{\nu_T}{\nu_T + z^{\alpha}},$$

which is equal to the Laplace transformation of

$$\nu_T x^{\alpha-1}E_{\alpha,\alpha}(-\nu_T x^{\alpha}), \text{ where } E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty}\frac{z^n}{\Gamma(\alpha n + \beta)}$$

Therefore, we have

$$T(1-a_T)\psi_{T,1}(Tt) = \nu_T t^{\alpha-1}E_{\alpha,\alpha}(-\nu_T t^{\alpha})$$

Plugging this back in the equation (4.8), we can expect (for $\alpha > \frac{1}{2}$)

$$v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) \sim \frac{1-a_T}{\mu_T}(v_1 \cdot \boldsymbol{\mu}_T(tT)) + \nu_T\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\nu_T(t-s)^{\alpha})\frac{1}{\mu_T}(v_1 \cdot \boldsymbol{\mu}_T(Ts))ds$$

$$+ \frac{\nu_T}{\sqrt{T(1-a_T)\mu_T}}\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\nu_T(t-s)^{\alpha})\sqrt{\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(sT)\cdot v_1^2}dB_{T,1}(s)$$

with $v_1^2 = (v_{1,i}^2)$ where $i \in 1,2,3,4$ and $B_{T,1}$ defined at (4.9). Decomposing v_1^2 in the basis

(e_1, e_2, e_3, e_4), we get

$$v_1^2 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) = \frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}\Big(v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big) + (e_1 \cdot v_1^2)\Big(v' \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big) + \sum_{2\le i\le 4}(e_i \cdot v_1^2)\Big(e_i \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big)$$

with $v' = e_1 - \dfrac{1}{e_1 \cdot v_1}v_1 \in \operatorname{span}(v_2, v_3, v_4)$. We can simplify this equation along with (4.6) by finding a $\boldsymbol{\mu}_T$ such that

$$v_i \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) = 0$$

for all $i = 2, 3, 4$. Indeed, if that is true, (4.6) can be simplified to

$$\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) = \frac{1}{e_1 \cdot v_1}\Big(v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big)e_1,$$

and since for any vector $v \in \operatorname{span}(v_2, v_3, v_4)$, $v \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ converges to zero, $v_1^2 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ has the same asymptotic behavior as

$$\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}\Big(v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big)$$

How should we choose such a $\boldsymbol{\mu}_T$? First, per Assumption 2', we need all the entries of $\boldsymbol{\mu}_T$ to be identical. However, if we set $\boldsymbol{\mu}_T(tT) := \mu_T(1,1,1,1)^\top$ as in [16] and [30], $v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(0)$ will disappear as well. This means that as $T \to \infty$, $\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ converges to a process with zero initial value.

Consider a $\boldsymbol{\mu}_T$ that can solve the following equation

$$\boldsymbol{\mu}_T(t) + \int_0^t \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n \varphi^{*n}(t-s)(\boldsymbol{\mu}_T(s))ds = \frac{\mu_T}{1-a_T}\mathbf{1} + \mu_T \int_0^t \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n \varphi^{*n}(s)\mathbf{1}ds \tag{4.10}$$

Recall that

$$v_i^\top \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n \varphi^{*n}(t) = \psi_{T,i}(t)v_i^\top$$

Plugging it back in (4.10), we have

$$v_i \cdot \boldsymbol{\mu}_T(t) + \int_0^t \psi_{T,i}(t-s)(v_i \cdot \boldsymbol{\mu}_T(s))ds = \frac{\mu_T}{1-a_T} v_i \cdot \mathbf{1} + \int_0^t \mu_T \psi_{T,i}(s) v_i \cdot \mathbf{1} ds$$

Then we have

$$\begin{aligned} v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) &= v_1 \cdot \mathbf{1} + \int_0^t T(1-a_T)\psi_{T,1}(Ts) v_1 \cdot \mathbf{1} ds \\ &\quad + \sqrt{\frac{T(1-a_T)}{\mu_T}} \int_0^t \psi_{T,1}(T(t-s)) \sqrt{\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(sT) \cdot v_1^2} dB_{T,1}(s) \\ &\sim v_1 \cdot \mathbf{1} + (v_1 \cdot \mathbf{1})\nu_T \int_0^t s^{\alpha-1} E_{\alpha,\alpha}(-\nu_T s^\alpha) ds \\ &\quad + \frac{\nu_T}{\sqrt{T(1-a_T)\mu_T}} \int_0^t (t-s)^{\alpha-1} E_{\alpha,\alpha}(-\nu_T (t-s)^\alpha) \sqrt{\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(sT) \cdot v_1^2} dB_{T,1}(s), \end{aligned}$$

which has a nonzero initial value. This gives us the following assumption:

Assumption 2: For each T, the background intensity $\boldsymbol{\mu}_T(tT)$ is given by

$$\boldsymbol{\mu}_T(tT) = \frac{\mu_T}{1-a_T}\Big[\mathbf{I} - \Big(a_T^2 \int_0^{tT} \varphi(s)ds\Big)\frac{\boldsymbol{\Phi}_0}{\lambda_1}\Big]\mathbf{1} = \frac{\mu_T}{1-a_T}\Big(1 - a_T^2 \int_0^{tT} \varphi(s)ds\Big)\mathbf{1} \tag{4.11}$$

with **1** the unit vector.

Remark: Note that with this assumption, all the entries of $\boldsymbol{\mu}_T$ are identical, $v_i \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ will vanish for $i = 2, 3, 4$. Therefore, Assumption 2 includes Assumption 2'. We will also show in the proof of the following proposition, provided at 4.4.2, that the $\boldsymbol{\mu}_T$ defined in Assumption 2 is indeed the solution to (4.10).

Proposition 4.2. The solution to the following integral equation

$$\boldsymbol{\mu}_T(t) + \int_0^t \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n \varphi^{*n}(t-s)(\boldsymbol{\mu}_T(s))ds = \frac{\mu_T}{1-a_T}\mathbf{1} + \mu_T \int_0^t \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\boldsymbol{\Phi}_0\Big)^n \varphi^{*n}(s)\mathbf{1}ds$$

is $\boldsymbol{\mu}_T$ defined in Equation (4.11).

With Assumption 2, we have

$$\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) = \frac{1}{e_1 \cdot v_1}\Big(v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big)e_1,$$

where

$$\begin{aligned} v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) &\sim v_1 \cdot \mathbf{1} + (v_1 \cdot \mathbf{1})\nu_T \int_0^t s^{\alpha-1}E_{\alpha,\alpha}(-\nu_T s^\alpha)ds \\ &+ \frac{\nu_T}{\sqrt{T(1-a_T)\mu_T}} \int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\nu_T(t-s)^\alpha)\sqrt{\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(sT)\cdot v_1^2}dB_{T,1}(s), \end{aligned}$$

with $v_1^2 = (v_{1,i}^2)$ where $i \in 1, 2, 3, 4$ and $B_{T,1}(t)$ defined in (4.9).

Note that $v_1^2 \cdot \dfrac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ having the same asymptotic behavior as

$$\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}\Big(v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)\Big)$$

We can see that the asymptotic behavior of $\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ depends on the asymptotic settings of

$$\nu_T \qquad \text{and} \qquad \frac{1}{\sqrt{T(1-a_T)\mu_T}}$$

As discussed in [16] and [30], to get a non-deterministic limit of $\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$, we need both ν_T and $\frac{1}{\sqrt{T(1-a_T)\mu_T}}$ to be positive constants and hence have the following assumption:

Assumption 3: There are two positive constants θ and μ such that

$$\lim_{T\to\infty} T^{\alpha}(1-a_T) = \theta K\Gamma(1-\alpha) \qquad \text{and} \qquad \lim_{T\to\infty} T^{1-\alpha}\mu_T = \frac{\mu}{K\Gamma(1-\alpha)}$$

With this assumption, as $T \to \infty$,

$$\nu_T \to \theta, \text{ and } \frac{\nu_T}{\sqrt{T(1-a_T)\mu_T}} \to \sqrt{\frac{\theta}{\mu}}.$$

Then we have

$$\begin{aligned} v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT) &\sim v_1 \cdot \mathbf{1} + (v_1 \cdot \mathbf{1}) \int_0^t f^{\alpha,\theta}(s)ds \\ &\quad + \frac{1}{\sqrt{\theta\mu}}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}} \int_0^t f^{\alpha,\theta}(t-s)\sqrt{v_1 \cdot \frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(sT)}dB_{T,1}(s) \end{aligned}$$

Thus, if we introduce

$$Y_T(t) := v_1 \cdot \frac{1-a_T}{\mu_T},$$

then as $T \to \infty$, we can see that the limit $Y(\cdot)$ of $Y_T(\cdot)$ satisfies

$$Y(t) = v_1 \cdot \mathbf{1} + (v_1 \cdot \mathbf{1}) \int_0^t f^{\alpha,\theta}(s)ds + \frac{1}{\sqrt{\theta\mu}}\sqrt{\frac{e_1 \cdot v_1^2}{e_1 \cdot v_1}} \int_0^t f^{\alpha,\theta}(t-s)\sqrt{Y(s)}dB(s),$$

where B is the limit of $B_{T,1}$ as $T \to \infty$. Since we have shown that B is a Brownian motion, we can use the following proposition, whose proof is provided in [17], to link the limit of Y_T with a stochastic Volterra integral equation:

Proposition 4.3. Let c_1, c_2, c_3, and θ be positive constants, $\alpha \in (\frac{1}{2}, 1)$ and B a Brownian motion. Let $F^{\alpha,\theta}(t) = \int_0^t f^{\alpha,\theta}(s)ds$. The process V is the solution of the following stochastic integral equation

$$V_1(t) = c_1 + c_2 F^{\alpha,\theta}(t) + c_3 \int_0^t f^{\alpha,\theta}(t-s)\sqrt{V_1(s)}dB(s)$$

if and only if it is the solution of

$$V_2(t) = c_1 + \frac{\theta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(c_1 + c_2 - V_2(s))ds + \frac{c_3\theta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}\sqrt{V_2(s)}dB(s)$$

By this proposition, we can see that $Y(t)$ is also the solution of

$$Y(t) = v_1 \cdot \mathbf{1} + \frac{\theta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(2v_1 \cdot \mathbf{1} - Y(s))ds + \frac{1}{\Gamma(\alpha)} \frac{\theta(e_1 \cdot v_1^2)}{\mu(e_1 \cdot v_1)} \int_0^t (t-s)^{\alpha-1}\sqrt{Y(s)}dB(s).$$

Therefore, as $T \to \infty$, we can derive that $\dfrac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ converges to

$$\frac{1}{e_1 \cdot v_1} Y(t) e_1$$

The derivation in this subsection is very useful in finding the necessary assumptions for the sequence of Hawkes processes to converge. Next, we work directly on the Hawkes processes.

4.3 Main Results

Similar to how we rescale the conditional intensity process, We first rescale the Hawkes process with the space normalization factor:

$$\mathbf{X}_T(t) := \frac{1-a_T}{T\mu_T}\mathbf{N}_T(tT), \qquad \mathbf{\Lambda}_T(t) := \frac{1-a_T}{T\mu_T}\int_0^{tT}\boldsymbol{\lambda}_T(s)ds,$$

$$\mathbf{Z}_T(t) := \sqrt{\frac{T\mu_T}{1-a_T}}(\mathbf{X}_T(t)-\mathbf{\Lambda}_T(t)) = \sqrt{\frac{1-a_T}{T\mu_T}}\mathbf{M}_T(Tt)$$

The following result, proved in 4.4.1, shows that we can work with $\mathbf{\Lambda}_T$ rather than $\mathbf{X}_T(t)$:

Proposition 4.4. $\sup_{t\in[0,1]}\|\mathbf{\Lambda}_T(t)-\mathbf{X}_T(t)\|$ goes to 0 in probability as $T\to\infty$.

Since we have shown that as $T\to\infty$, the rescaled conditional intensity $\dfrac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(tT)$ converges to

$$\frac{1}{e_1\cdot v_1}Y(t)e_1,$$

where

$$Y(t) = v_1\cdot\mathbf{1}+\frac{\theta}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}(2v_1\cdot\mathbf{1}-Y(s))ds+\frac{1}{\Gamma(\alpha)}\frac{\theta(e_1\cdot v_1^2)}{\mu(e_1\cdot v_1)}\int_0^t(t-s)^{\alpha-1}\sqrt{Y(s)}dB(s).$$

Therefore, as $T\to\infty$, we can derive that the rescaled integral

$$\mathbf{\Lambda}_T(t) = \int_0^t\frac{1-a_T}{\mu_T}\boldsymbol{\lambda}_T(sT)ds$$

converges to $\int_0^t Y(s)ds$. By Proposition 4.4, $\mathbf{\Lambda}_T$ and $\mathbf{X}_T$ share the same limit as $T\to\infty$. Therefore, as $T \to \infty$, $\mathbf{X}_T(t)$ converges also to

$$\frac{1}{e_1 \cdot v_1}(\int_0^t Y(s)ds)e_1.$$

Theorem 4.1. Under the assumptions above, as T tends to infinity, the process $(\mathbf{\Lambda}_T(t), \mathbf{X}_T(t), \mathbf{Z}_T(t))$ converges in law for the topology of the convergence in measure to $(\mathbf{\Lambda}(t), \mathbf{X}(t), \mathbf{Z}(t))_{t\in[0,1]}$, where

$$\mathbf{\Lambda}(t) = \mathbf{X}(t) = \frac{1}{e_1 \cdot v_1}(\int_0^t Y(s)ds)e_1,$$

and for $1 \le i \le 4$

$$\mathbf{Z}^i(t) = \int_0^t \sqrt{\frac{e_{1,i}}{e_1 \cdot v_1} Y(s)} dB^i(s)$$

where (B^1, B^2, B^3, B^4) is a 4-dimensional Brownian motion and Y is the unique solution of

$$Y(t) = v_1 \cdot \mathbf{1} + \frac{\theta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(2v_1 \cdot \mathbf{1} - Y(s))ds + \frac{1}{\Gamma(\alpha)} \sqrt{\frac{\theta(e_1 \cdot v_1^2)}{\mu(e_1 \cdot v_1)}} \int_0^t (t-s)^{\alpha-1} \sqrt{Y(s)} dB(s) \tag{4.12}$$

with

$$B = \frac{1}{\sqrt{e_1 \cdot v_1^2}} \sum_{i=1}^{4} \sqrt{e_{1,i}} v_{1,i} B^i$$

and that Y has Hölder regularity $\alpha - 1/2 - \epsilon$ for any $\epsilon > 0$. Furthermore, Y has Hölder regularity $\alpha - \frac{1}{2} - \epsilon$ for any $\epsilon > 0$.

Remark 4.3. The fractional Brownian motion $B^H(t)$ can be expressed as

$$B^H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \Big(\int_0^t (t-s)^{H-\frac{1}{2}} dW(s) + \int_{-\infty}^0 (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} dW(s) \Big)$$

where $W(t)$ is a Brownian motion, and H is the Hurst parameter associated with $B^H(t)$ [38]. Therefore, we can interpret (4.12) as a Volterra integral equation with Hurst parameter $\alpha - \frac{1}{2}$. Since $\alpha \in (\frac{1}{2}, 1)$, the Hurst parameter is less than $\frac{1}{2}$.

Applying this result to the microscopic model in 3.2.1, we have the following corollary. We provide the proof in 4.4.3.

Corollary 4.1. We let the normalizing factor of $V_T(tT)$ be

$$h(T) = \sqrt{\frac{T\mu_T}{1-a_T}}.$$

Then under the assumptions above, as T tends to infinity, $\dfrac{V_T(tT)}{h(T)}$ converges in certain sense to the following rough Heston model

$$f(t) = \Big(\frac{2\beta_1 + 5\beta_2 + 3\beta_2\beta_3 - 1}{\beta_1 + 3\beta_2 + 2\beta_2\beta_3 - 1}\Big)\frac{1}{\sqrt{2\gamma+2}}\int_0^t \sqrt{Y(s)}dW(s)$$

Y is the unique solution of the following rough stochastic differential equation

$$Y(t) = 2(\gamma+1)+\frac{\theta}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(4(\gamma+1)-Y(s))ds+\frac{1}{\Gamma(\alpha)}\sqrt{\frac{\theta}{2\mu(\gamma+1)}}\int_0^t (t-s)^{\alpha-1}\sqrt{Y(s)}dB(s)$$

where $\gamma = \frac{\beta_2(\beta_3+1)}{\beta_1+\beta_2\beta_3+2\beta_2-1}$,

$$W = B^1 + B^2 - B^3 - B^4, \qquad B = \gamma B^1 + \gamma B^2 + B^3 + B^4$$

are two 1-dimensional Brownian motions, with B^1, B^2, B^3, B^4 four 1-dimensional independent Brownian motions. Furthermore, Y has Hölder regularity $\alpha - \frac{1}{2} - \epsilon$ for any $\epsilon > 0$.

4.4 Proofs

We provide the proofs of Proposition 4.4, Proposition 4.2, Proposition 4.3, and Corollary 4.1 in this subsection. In the following proofs, Assumption 1, Assumption 2, and Assumption 3 hold, and we use c to denote a generic positive constant.

4.4.1 Proof of Proposition 4.4

Proof. We show that $\mathbf{X}_T - \mathbf{\Lambda}_T$ converges uniformly to zero in probability. Since

$$\mathbf{X}_T - \mathbf{\Lambda}_T = K\Gamma(1-\alpha)\frac{1-a_T}{T^\alpha \mu}\mathbf{M}_T(tT),$$

by Doob's inequality and the fact that $[\mathbf{M}_T, \mathbf{M}_T] = \mathbf{N}_T$, we have

$$\mathbb{E}[\sup_{t\in[0,1]} |\mathbf{X}_T - \mathbf{\Lambda}_T|^2] \leq cT^{-4\alpha}\mathbb{E}[\mathbf{M}_T(T)]^2 = cT^{-4\alpha}\mathbb{E}[N_T(T)] \leq cT^{-2\alpha}$$

This shows that $\mathbf{X}_T - \mathbf{\Lambda}_T$ converges uniformly to zero in probability.

$\square$

4.4.2 Proof of Proposition 4.2

We show that the solution to the following equation

$$\boldsymbol{\mu}_T(t) + \int_0^t \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \varphi^{*n}(t-s)(\boldsymbol{\mu}_T(s))ds = \frac{\mu_T}{1-a_T}\mathbf{1} + \mu_T \int_0^t \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \varphi^{*n}(s)\mathbf{1}ds$$

is

$$\boldsymbol{\mu}_T(t) = \frac{\mu_T}{1-a_T}\Big[\mathbf{I} - \Big(a_T^2 \int_0^t \varphi(s)ds\Big)\frac{\mathbf{\Phi}_0}{\lambda_1}\Big]\mathbf{1} = \frac{\mu_T}{1-a_T}\Big(1 - a_T^2 \int_0^t \varphi(s)ds\Big)\mathbf{1}$$

Proof. We use again the helpful expression

$$\mathbf{\Psi}_T = \sum_{n=1}^{\infty} (\mathbf{\Phi}_T)^{*n}$$

with the property

$$\mathbf{\Psi}_T * \mathbf{\Phi}_T = \mathbf{\Psi}_T - \mathbf{\Phi}_T.$$

Then from the left hand side, we have:

$$\begin{aligned}
&\int_0^t \Big(\boldsymbol{\mu}_T(s) + \int_0^s \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \varphi^{*n}(s-u)(\boldsymbol{\mu}_T(u))du\Big)\mathbf{\Phi}_T(t-s)ds \\
&= \int_0^t \Big(\boldsymbol{\mu}_T(s) + \int_0^s \mathbf{\Psi}_T(s-u)(\boldsymbol{\mu}_T(u))du\Big)\mathbf{\Phi}_T(t-s)ds \\
&= \int_0^t \boldsymbol{\mu}_T(s)\mathbf{\Phi}_T(t-s)ds + \int_0^t \int_0^{t-u} \mathbf{\Psi}_T(s)\mathbf{\Phi}_T(t-u-s)ds\boldsymbol{\mu}_T(u)du \\
&= \int_0^t \boldsymbol{\mu}_T(s)\mathbf{\Phi}_T(t-s)ds + \int_0^t (\mathbf{\Psi}_T(s-u) - \mathbf{\Phi}_T(s-u))\boldsymbol{\mu}_T(u)du \\
&= \int_0^t \mathbf{\Psi}_T(t-s)\boldsymbol{\mu}_T(s)ds
\end{aligned}$$

From the right hand side, we have

$$\begin{aligned}
&\int_0^t \Big(\frac{\mu_T}{1-a_T}\mathbf{1} + \mu_T \int_0^s \sum_{n=1}^{\infty} \Big(\frac{a_T}{\lambda_1}\mathbf{\Phi}_0\Big)^n \varphi^{*n}(u)\mathbf{1}du\Big)\mathbf{\Phi}_T(t-s)ds \\
&= \int_0^t \Big(\frac{\mu_T}{1-a_T}\mathbf{1} + \mu_T \int_0^s \mathbf{\Psi}_T(u)du\mathbf{1}\Big)\mathbf{\Phi}_T(t-s)ds \\
&= \int_0^t \frac{\mu_T}{1-a_T}\mathbf{\Phi}_T(t-s)\mathbf{1}ds + \int_0^t \int_0^s \mu_T \mathbf{\Psi}_T(s-u)\mathbf{\Phi}_T(t-s)\mathbf{1}duds \\
&= \frac{\mu_T}{1-a_T}\int_0^t \mathbf{\Phi}_T(t-s)\mathbf{1}ds + \mu_T \int_0^t \int_0^{t-u} \mathbf{\Psi}_T(s)\mathbf{\Phi}_T(t-u-s)duds\mathbf{1} \\
&= \frac{\mu_T}{1-a_T}\int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1} + \mu_T \int_0^t \Big(\mathbf{\Psi}_T(t-u) - \mathbf{\Phi}_T(t-u)\Big)ds\mathbf{1} \\
&= \frac{\mu_T}{1-a_T}\int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1} + \int_0^t \mu_T\mathbf{\Psi}_T(t-s)ds\mathbf{1} - \int_0^t \mu_T\mathbf{\Phi}_T(t-s)ds\mathbf{1}
\end{aligned}$$

This means that

$$\int_0^t \mathbf{\Psi}_T(t-s)\boldsymbol{\mu}_T(s)ds = \frac{\mu_T}{1-a_T}\int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1} + \int_0^t \mu_T\mathbf{\Psi}_T(t-s)ds\mathbf{1} - \int_0^t \mu_T\mathbf{\Phi}_T(t-s)ds\mathbf{1}$$

Combining with (4.10), we have

$$\begin{aligned}
&\boldsymbol{\mu}_T(t) + \frac{\mu_T}{1-a_T}\int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1} + \int_0^t \mu_T\mathbf{\Psi}_T(t-s)ds\mathbf{1} - \int_0^t \mu_T\mathbf{\Phi}_T(t-s)ds\mathbf{1} \\
&= \frac{\mu_T}{1-a_T}\mathbf{1} + \mu_T\int_0^t \mathbf{\Psi}_T(t-s)ds\mathbf{1}
\end{aligned}$$

This leads to

$$\boldsymbol{\mu}_T(t) + \frac{\mu_T}{1-a_T}\int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1} - \int_0^t \mu_T\mathbf{\Phi}_T(t-s)ds\mathbf{1} = \frac{\mu_T}{1-a_T}\mathbf{1}$$

Therefore, we have

$$\begin{aligned}
\boldsymbol{\mu}_T(t) &= \frac{\mu_T}{1-a_T}\Big(\mathbf{1} - \int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1}\Big) + \mu_T\int_0^t \mathbf{\Phi}_T(t-s)ds\mathbf{1} \\
&= \frac{\mu_T}{1-a_T}\Big(\mathbf{I} - \frac{a_T^2}{\lambda_1}\int_0^t \varphi(s)ds\mathbf{\Phi}_0\Big)\mathbf{1} \\
&= \frac{\mu_T}{1-a_T}\Big(1 - a_T^2\int_0^t \varphi(s)ds\Big)\mathbf{1}
\end{aligned}$$

$\square$

4.4.3 Proof of Corollary 4.1

We apply Theorem 4.1 to the Microscopic Model:

Proof. We first write $\dfrac{V_T(tT)}{h(T)}$ in a more friendly format:

$$\begin{aligned}\sqrt{\frac{1-a_T}{T\mu_T}}V_T(tT) &= \sqrt{\frac{1-a_T}{T\mu_T}}(N_T^{a,+}(tT)+N_T^{b,+}(tT)-N_T^{a,-}(tT)-N_T^{b,-}(tT))\\ &= \int_0^{tT}\sqrt{\frac{1-a_T}{T\mu_T}}(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s))\\ &+\int_0^{tT}\sqrt{\frac{1-a_T}{T\mu_T}}(\lambda_T^{a,+}(s)+\lambda_T^{b,+}(s)-\lambda_T^{a,-}(s)-\lambda_T^{b,-}(s))ds.\end{aligned}$$

Furthermore,

$$\begin{aligned}&\lambda_T^{a,+}(t)+\lambda_T^{b,+}(t)-\lambda_T^{a,-}(t)-\lambda_T^{b,-}(t)\\ &= a_T(\int_0^t\frac{1-\beta_2-\beta_2\beta_3}{\beta_1+\beta_2\beta_3+2\beta_2}\varphi(t-s)(dN_T^{a,+}(s)+dN_T^{b,+}(s)-dN_T^{a,-}(s)-dN_T^{b,-}(s)))\\ &= \int_0^t a_T\frac{1-\beta_2-\beta_2\beta_3}{\beta_1+\beta_2\beta_3+2\beta_2}\varphi(t-s)(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s))\\ &+\int_0^t a_T\frac{1-\beta_2-\beta_2\beta_3}{\beta_1+\beta_2\beta_3+2\beta_2}\varphi(t-s)(\lambda_T^{a,+}(s)+\lambda_T^{b,+}(s)-\lambda_T^{a,-}(s)-\lambda_T^{b,-}(s))ds\end{aligned}$$

By the Lemma C.1, we have

$$\begin{aligned}&\lambda_T^{a,+}(t)+\lambda_T^{b,+}(t)-\lambda_T^{a,m}(t)-\lambda_T^{b,m}(t)\\ &= \int_0^t\psi_{T,4}(t-s)(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s))\end{aligned}$$

Then using Fubini theorem, we get

$$\begin{aligned}&\int_0^x\lambda_T^{a,+}(s)+\lambda_T^{b,+}(s)-\lambda_T^{a,-}(s)-\lambda_T^{b,-}(s)ds\\ &= \int_0^x(\int_0^{x-s}\psi_{T,4}(u)du)(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s))\end{aligned}$$

Hence the rescaled price process $\sqrt{\frac{1-a_T}{T\mu_T}}V_T(tT)$ can be written as

$$\begin{aligned}
\sqrt{\frac{1-a_T}{T\mu_T}}V_T(tT) &= \sqrt{\frac{1-a_T}{T\mu_T}}(N_T^{a,+}(tT)+N_T^{b,+}(tT)-N_T^{a,-}(tT)-N_T^{b,-}(tT)) \\
&= \int_0^{tT}\sqrt{\frac{1-a_T}{T\mu_T}}(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s)) \\
&+ \int_0^{tT}\sqrt{\frac{1-a_T}{T\mu_T}}(\lambda_T^{a,+}(s)+\lambda_T^{b,+}(s)-\lambda_T^{a,-}(s)-\lambda_T^{b,-}(s))ds \\
&= \int_0^{tT} dZ_T^{a,+}(s)+dZ_T^{b,+}(s)-dZ_T^{a,-}(s)-dZ_T^{b,-}(s) \\
&+ \sqrt{\frac{1-a_T}{T\mu_T}}\int_0^{tT}\int_0^{tT-sT}\psi_{T,4}(u)du(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s)) \\
&= (Z_T^{a,+}(t)+Z_T^{b,+}(t)-Z_T^{a,-}(t)-Z_T^{b,-}(t)) \\
&- \sqrt{\frac{1-a_T}{T\mu_T}}\int_0^{t}\int_{T(t-s)}^{\infty}\psi_{T,4}(u)du(dM_T^{a,+}(s)+dM_T^{b,+}(s)-dM_T^{a,-}(s)-dM_T^{b,-}(s)) \\
&+ \sqrt{\frac{1-a^T}{T\mu_T}}\int_0^{\infty}\psi_{T,4}(u)du(M_T^{a,+}(t)+M_T^{b,+}(t)-M_T^{a,-}(t)-M_T^{b,-}(t)) \\
&= (Z_T^{a,+}(t)+Z_T^{b,+}(t)-Z_T^{a,-}(t)-Z_T^{b,-}(t)) \\
&- \int_0^{t}\int_{T(t-s)}^{\infty}\psi_{T,4}(u)du(dZ_T^{a,+}(s)+dZ_T^{b,+}(s)-dZ_T^{a,-}(s)-dZ_T^{b,-}(s)) \\
&+ \int_0^{\infty}\psi_{T,4}(u)du(Z_T^{a,+}(t)+Z_T^{b,+}(t)-Z_T^{a,-}(t)-Z_T^{b,-}(t)) \\
&= (1+\int_0^{\infty}\psi_{T,4}(u)du)(Z_T^{a,+}(t)+Z_T^{b,+}(t)-Z_T^{a,-}(t)-Z_T^{b,-}(t))-R_T(t)
\end{aligned}$$

with

$$R_T(t) = \int_0^{t}\int_{T(t-s)}^{\infty}\psi_{T,4}(u)du(dZ_T^{a,+}(s)+dZ_T^{b,+}(s)-dZ_T^{a,-}(s)-dZ_T^{b,-}(s)).$$

Since

$$\int_0^{\infty}\psi_{T,4}(u)du = \int_0^{\infty}\sum_{n=1}^{\infty}(a_T)^n\lambda_4^{*n}(u)du$$

Since $a_T < 1$ and $\mathscr{S}(\int_0^\infty \mathbf{\Phi}_T(s)ds) < 1$,

$$= \sum_{n=1}^{\infty} \int_0^\infty \lambda_{T,4}^{*n}(u)du = \frac{1}{1 - \int_0^\infty \lambda_{T,4}(s)ds} = \frac{\beta_1 + \beta_2\beta_3 + 2\beta_2}{\beta_1 + \beta_2\beta_3 + 2\beta_2 - a^T(1 - \beta_2 - \beta_2\beta_3)\|\varphi\|_1}$$

Therefore, we have

$$\sqrt{\frac{1-a_T}{T\mu_T}} V_T(tT) = (1 + \frac{\beta_1 + \beta_2\beta_3 + 2\beta_2}{\beta_1 + \beta_2\beta_3 + 2\beta_2 - a_T(1 - \beta_2 - \beta_2\beta_3)\|f^{\alpha,1}\|_1})$$
$$(Z_T^{a,+}(t) + Z_T^{b,+}(t) - Z_T^{a,-}(t) - Z_T^{b,-}(t)) - R_T(t),$$

where

$$R_T(t) = \int_0^t \int_{T(t-s)}^\infty \psi_{T,4}(u)du(dZ_T^{a,+}(s) + dZ_T^{b,+}(s) - dZ_T^{a,-}(s) - dZ_T^{b,-}(s))$$

Next, we let $e_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, then by Theorem 4.1, the process $(\boldsymbol{\lambda}_T(t), \mathbf{X}_T(t), \mathbf{Z}_T(t))$ converges in law for the Skorokhod topology to $(\boldsymbol{\lambda}, \mathbf{X}, \mathbf{Z})$ where

$$\boldsymbol{\lambda}(t) = \mathbf{X}(t) = \frac{1}{2\gamma+2}\Big(\int_0^t Y(s)\Big)\mathbf{1}$$

and

$$\mathbf{Z}(t) = \int_0^t \sqrt{\frac{1}{2(\gamma+1)}Y(s)}ds \begin{pmatrix} dB^1(s) \\ dB^2(s) \\ dB^3(s) \\ dB^4(s) \end{pmatrix}$$

where (B^1, B^2, B^3, B^4) is a 4-dimensional Brownian motion and Y is the unique solution of the following rough stochastic differential equation

$$\begin{aligned}Y(t) &= 2(\gamma+1) + \frac{\theta}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(4(\gamma+1) - Y(s))ds \\ &+ \frac{1}{\Gamma(\alpha)}\sqrt{\frac{\theta(\gamma^2+1)}{\mu(\gamma+1)}}\int_0^t (t-s)^{\alpha-1}\sqrt{Y(s)}dB(s)\end{aligned} \tag{4.13}$$

Furthermore, Y has Holder regularity $\alpha - \frac{1}{2} - \epsilon$ for any $\epsilon > 0$. Therefore,

$$\Big(1 + \frac{\beta_1 + \beta_2\beta_3 + 2\beta_2}{\beta_1 + \beta_2\beta_3 + 2\beta_2 - a_T(1 - \beta_2 - \beta_2\beta_3)\|\varphi\|_1}\Big)\Big(Z_T^{a,+}(t) + Z_T^{b,+}(t) - Z_T^{a,-}(t) - Z_T^{b,-}(t)\Big)$$

converges in law to

$$\Big(\frac{2\beta_1 + 5\beta_2 + 3\beta_2\beta_3 - 1}{\beta_1 + 3\beta_2 + 2\beta_2\beta_3 - 1}\Big)\frac{1}{\sqrt{2\gamma+2}}\int_0^t \sqrt{Y(s)}dW(s)$$

where $W = B^1 + B^2 - B^3 - B^4$, and Y is defined by (4.13).

Now we are left to show that R_t^T converges to 0. Since

$$R_T(t) = \int_0^t \Big(\int_{T(t-s)}^{\infty} \psi_{T,4}(u)du)(dZ_T^{a,+}(t) + dZ_T^{b,+}(t) - dZ_T^{a,-}(t) - dZ_T^{b,-}(t)\Big),$$

there exists $c > 0$ such that

$$\mathbb{E}[(R_T(t))^2] \le c\int_0^t (\int_{Ts}^{\infty} \psi_{T,4}(u)du)^2 ds$$

Let $G = \sum_{n=1}^{\infty} |(\frac{-\beta_2\beta_3 - \beta_2 + 1}{\beta_1 + \beta_2\beta_3 + 2\beta_2})\varphi|^{*n}$, then we have

$$|\psi_{T,4}| = \int_0^{\infty} \psi_{T,4}(u)du = \int_0^{\infty} \sum_{n=1}^{\infty} \lambda_{T,4}^{*n}(u)du \le \sum_{n=1}^{\infty} |\frac{-\beta_2\beta_3 - \beta_2 + 1}{\beta_1 + \beta_2\beta_3 + 2\beta_2}\varphi|^{*n} = G$$

G is integrable since $\int_0^\infty |\frac{-\beta_2\beta_3-\beta_2+1}{\beta_1+\beta_2\beta_3+2\beta_2}\varphi| < 1$. Therefore, we have

$$\begin{aligned}
\mathbb{E}[(R_T(t))^2] &\leq c\int_0^t (\int_{Ts}^\infty G(u)du)^2 ds \\
&\leq c\int_0^1 (\int_{Ts}^\infty G(u)du)^2 ds \\
&\leq c(\int_0^{T^{-1/2}} (\int_{Ts}^\infty G(u)du)^2 ds + \int_{T^{-1/2}}^1 (\int_{Ts}^\infty G(u)du)^2 ds) \\
&= c(T^{-1/2}(\int_0^\infty G)^2 + (\int_{T^{1/2}}^\infty G)^2)
\end{aligned}$$

Thus, $\mathbb{E}[(R_T(t))^2] \to 0$ as $T \to \infty$, and hence R_t^T converges to 0. $\square$

CHAPTER 5: DYNAMICS OF THE LOB WITH HAWKES PROCESSES

5.1 Combining the HFT and Non-HFT Volumes

With the scaling limit of the HFT order volumes $V(t)$, and suppose the limit is on some filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ at the coarse-grained time scale of the average (non-HF) market participants, we model the HFT orders as a multiplicative noise term of the form:

$$C_\sigma(x)u(t,x)dV(t) = C_\sigma(x)u(t,x)\sqrt{Y(t)}dW(t),$$

where

$$C_\sigma(x) = \begin{cases} C_\sigma^a & x \in (0, L] \\ C_\sigma^b & x \in [-L, 0) \end{cases}$$

with $C_\sigma^a, C_\sigma^b > 0$. Combining with the non-HFT orders, and let the relative price level be $x \in [-L, L]$ for some positive constant L, we have the following SPDE for the centered order book density u in a real, separable Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$:

$$\begin{cases} du(t,x) = [Au(t,x) + F(t,x,u_x(t,x),u(t,\cdot))]dt + C_\sigma(x)u(t,x)\sqrt{Y(t)}dW(t) \\ u(0,x) = u_0(x), \\ Y(t) = 2(\gamma+1) + \dfrac{\theta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(4(\gamma+1) - Y(s))ds \\ +\dfrac{1}{\Gamma(\alpha)}\sqrt{\dfrac{\theta}{2\mu(\gamma+1)}} \int_0^t (t-s)^{\alpha-1}\sqrt{Y(s)}dB(s) \end{cases} \tag{5.1}$$

where $u : [0,T] \times [-L,L] \times \Omega \to \mathbb{R}$, $u_0 : [-L,L] \to \mathbb{R}$ for some $L > 0$, and $T \in [0, \infty)$. $A : \text{dom}(A) \subset H \to H$ is a linear operator on H defined as

$$Au = \eta(x)u_{xx} - \zeta(x)u,$$

where

$$\eta(x) = \begin{cases} \eta_a & x \in (0, L] \\ \eta_b & x \in [-L, 0) \end{cases}, \quad \zeta(x) = \begin{cases} \zeta_a & x \in (0, L] \\ \zeta_b & x \in [-L, 0) \end{cases}, \quad C_\sigma(x) = \begin{cases} C_\sigma^a & x \in (0, L] \\ C_\sigma^b & x \in [-L, 0) \end{cases}$$

with $\eta_a, \eta_b, \zeta_a, \zeta_b, C_\sigma^a, C_\sigma^b$ positive constants. Also,

$$F(t, x, u_x(t,x), u(t, \cdot)) = -\beta(x)\,\mathrm{sgn}(x)[u_x(t,x)]^- + J(x, u(t,x)) + G(x, \ell(t)),$$

where

$$[u_x(t,x)]^- = \begin{cases} -u_x(t,x) & u_x(t,x) < 0 \\ 0 & \text{otherwise} \end{cases}, \quad \text{and} \quad \beta(x) = \begin{cases} \beta_a & x \in (0, L] \\ \beta_b & x \in [-L, 0) \end{cases}$$

with β_a, β_b are positive constants. The function J is defined as

$$\begin{aligned} J(x, u(t,x)) = \mathbf{1}_{\{x>0\}}\big[-j(x)\big(u(t,x) - u_0\big)^+ - j(-|x|)\big(u(t,-|x|) + u_0\big)^-\big] \\ + \mathbf{1}_{\{x<0\}}\big[j(|x|)\big(u(t,|x|) - u_0\big)^+ + j(x)\big(u(t,x) + u_0\big)^-\big], \end{aligned}$$

with $u_0 > 0$, and $j(x) \leq 1$ a positive function decreasing in $x > 0$ and increasing in $x < 0$. The function $G(x, \ell(t))$ is strictly decreasing in $\ell(t)$ and $G(x, 0) = 0$, with

$$\ell(t) = \int_{-\iota}^{\iota} u(t,y)dy.$$

In the equation of $Y(t)$, $\alpha \in (\frac{1}{2}, 1)$, and $\gamma = \dfrac{\beta_2(\beta_3 + 1)}{\beta_1 + \beta_2\beta_3 + 2\beta_2 - 1}$. For the diffusion term for both of the equations,

$$W = B^1 + B^2 - B^3 - B^4, \qquad B = \gamma B^1 + \gamma B^2 + B^3 + B^4$$

are two one-dimensional Brownian motions, with B^1, B^2, B^3, B^4 four independent one-dimensional Brownian motions.

5.2 The Existence of the Unique Solution to the SPDE

In this section, we first show that there exists a unique solution to (5.1).

Theorem 5.1. Assume that the following conditions hold:

(i) For any $\mathscr{T} \in [0, \infty)$, there exists $C_{\mathscr{T}} > 0$ such that for all $x, y \in \mathbb{R}$ and $t \in [0, \mathscr{T}]$, $|G(x, f_1(t)) - G(x, f_2(t))| \leq C_{\mathscr{T}} |f_1(t) - f_2(t)|$.

(ii) The linear operator A generates a C_0 semigroup of bounded linear operators $S(t)$ with $\|S(t)\| \leq Me^{\omega t}$ with constants $M \geq 1$ and $\omega \geq 0$,

Then there exists a unique mild solution $u(t, x)$ to (5.1) for $t \in [0, \mathscr{T}]$.

5.2.1 Weak Existence and Uniqueness of $Y(t)$

We first show that there is a unique in law continuous weak solution $Y(t)$ of the equation

$$
\begin{aligned}
Y(t) = 2(\gamma+1) &+ \frac{\theta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(4(\gamma+1) - Y(s))ds \\
&+ \frac{1}{\Gamma(\alpha)} \sqrt{\frac{\theta}{2\mu(\gamma+1)}} \int_0^t (t-s)^{\alpha-1} \sqrt{Y(s)} dB(s)
\end{aligned}
\tag{5.2}
$$

Most of the technical results in Section G needs the kernel function to meet requirements (G.2) and (G.3). We show in this section that (5.2) is an affine Volterra equation per (5.2) that meets these 2 requirements.

Proposition 5.1. (5.2) is an affine Volterra equation that meets the requirements (G.2) and (G.3).

Proof. (1). The equation (5.2) is an affine Volterra equation:

We first show that (5.2) is an affine Volterra equation. Indeed, comparing it to the equation of $\mathscr{Y}(t)$ (G.1), we can see that $d = 1$, $m = 1$, and $\mathscr{Y}_0 = Y(0) = 2(\gamma + 1) \in \mathbb{R}$. The kernel $\mathscr{K}$ corresponds to

$$\mathscr{K}(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)} \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}).$$

As for the coefficients, $b : \mathbb{R} \to \mathbb{R}$ corresponds to

$$b(y) = -\theta y + 4\theta(\gamma + 1),$$

and $\sigma : \mathbb{R} \to \mathbb{R}$ corresponds to

$$\sigma(y) = \sqrt{\frac{\theta}{2\mu(\gamma + 1)}}\sqrt{y}.$$

Therefore, $b(y)$ and

$$a(y) = \sigma(y)\sigma(y)^\top = \frac{\theta}{2\mu(\gamma + 1)}y$$

are affine of the form

$$a(y) = A^0 + yA^1, \text{ with } A^0 = 0 \text{ and } A^1 = \frac{\theta}{2\mu(\gamma + 1)}$$

as well as

$$b(y) = b^0 + yb^1, \text{ with } b^0 = 4\theta(\gamma + 1) \text{ and } b^1 = -\theta$$

We have shown that (5.2) is an affine Volterra equation. We can also find the corresponding expressions of the following important ingredients to use the results in Section G:

$$B = (b^1) = -\theta, \qquad A(u) := uA^1u^\top = \frac{\theta}{2\mu(\gamma+1)}u^2$$

any row vector $u \in (\mathbb{C})^*$.

(2). $\mathscr{K}$ satisfies the requirement (G.2):

Next, we show that the kernel $\mathscr{K}$ satisfies the requirement (G.2). Indeed, let $\gamma = 2\alpha - 1$, we have

$$\int_0^h \mathscr{K}^2 dt = \frac{h^\gamma}{(\Gamma(\alpha))^2\gamma}$$

and

$$\begin{aligned}\int_0^T (\mathscr{K}(t+h) - \mathscr{K}(t))^2 dt &\leq \frac{h^\gamma}{(\Gamma(\alpha))^2}\int_0^\infty ((t+1)^{\alpha-1} - t^{\alpha-1})^2 dt \\ &\leq \frac{h^\gamma}{(\Gamma(\alpha))^2}(\frac{1}{\gamma} + \frac{1}{2-2\gamma})\end{aligned}$$

Since $\gamma \in (0, 2]$, $\int_0^h \mathscr{K}(t)^2 dt = \mathcal{O}(h^\gamma)$, and $\int_0^T (\mathscr{K}(t+h) - \mathscr{K}(t))^2 dt = \mathcal{O}(h^\gamma)$ for every $T < \infty$, (G.2) is fulfilled.

(3). $\mathscr{K}$ satisfies the requirement (G.3):

Lastly, we show that the kernel $\mathscr{K}$ satisfies the requirement (G.3). If $\Delta_h\mathscr{K}(t) = \mathscr{K}(t + h) := \mathscr{F}(t)$ is completely monotone on $(0, \infty)$ and not identically zero, then (G.3) is fulfilled by Lemma G.1 [20]. Recall that a function $\mathscr{F}$ is completely monotone on $(0, \infty)$ if it is infinitely differentiable with $(-1)^n\mathscr{F}^{(n)}(t) \geq 0$ for all $t > 0$ and $n = 0, 1, ...$

$$\mathscr{F}^{(0)}(t) = f(t) = \mathscr{K}(t+h) = \frac{1}{\Gamma(\alpha)}(t+h)^{\alpha-1} > 0$$

We show that for $n \geq 1$.

$$\mathscr{F}^{(n)}(t) = \frac{1}{\Gamma(\alpha)} \prod_{i=1}^{n} (\alpha - i) t^{\alpha-(n+1)}$$

When $n = 1$, $\mathscr{F}'(t) = \dfrac{(\alpha-1)}{\Gamma(\alpha)}(t+h)^{\alpha-2}$. Suppose $\mathscr{F}^{(n)}(t)$ holds, then

$$f^{(n+1)}(t) = \frac{1}{\Gamma(\alpha)}(\alpha - (n+1)) \prod_{i=1}^{n} (\alpha - i) t^{\alpha-(n+2)} = \frac{1}{\Gamma(\alpha)} \prod_{i=1}^{n+1} (\alpha - i) t^{\alpha-((n+1)+1)}$$

By the Principle of Mathematical Induction, for all $n \geq 1$,

$$\mathscr{F}^{(n)}(t) = \frac{1}{\Gamma(\alpha)} \prod_{i=1}^{n} (\alpha - i) t^{\alpha-(n+1)}$$

Therefore, for all $n \geq 1$, $(-1)^n f^{(n)}(t) \geq 0$. The second condition is fulfilled.

□

Therefore, by Lemma G.2, there is a unique in law $\mathbb{R}_+$-valued continuous weak solution Y of the equation (5.2).

5.2.2 *Existence of a mild solution at stopping time*

In this section, we prove the existence of a solution $u(t,x)$ to the following SPDE:

$$\begin{cases} du(t,x) = [Au(t,x) + F(t,x,u_x(t,x)u(t,\cdot))]dt + C_\sigma(x)u(t,x)\sqrt{Y(t)}dW(t) \\ u(0,x) = u_0(x) \end{cases} \tag{5.3}$$

We write out the proof for the case where $x > 0$. The case where $x < 0$ can be proved using the same method.

Since A generates a C_0 semigroup of bounded linear operators $S(t)$, we have

$$u(t,x) = S(t)u_0(x) + \int_0^t S(t-s)F(s,x,u_x(t,x),u(s,\cdot))ds + C_\sigma^a \int_0^t S(t-s)u(s,x)\sqrt{Y(s)}dW(s)$$

We first show that the L^2 norm of $S(t)$ decays exponentially. Let $\phi \in \mathrm{Dom}(A)$, and

$$\begin{cases} -A\phi(x) = \nu\phi(x), & x \in (-L, L) \\ \phi(x) = 0, & x = -L \text{ or } x = L \end{cases}$$

We find the eigenvalue ν_n and eigenfunction ϕ of the operator $-A$ as

$$\begin{cases} \nu_n = \dfrac{n^2\pi^2\eta_a^2}{L^2} + \zeta_a \\ \phi_n(x) = \sin(\frac{n\pi}{L}x) \end{cases}$$

Then for any $\psi \in \mathrm{Dom}(A)$,

$$\langle A\psi, \psi\rangle \le -\nu_1 \|\psi\|^2$$

and so

$$\frac{d}{dt}\|S(t)\psi\|^2 = \langle \frac{d}{dt}S(t)\psi, S(t)\psi\rangle + \langle S(t)\psi, \frac{dt}{d}S(t)\psi\rangle$$

$$= \langle AS(t)\psi, S(t)\psi\rangle + \langle S(t)\psi, AS(t)\psi\rangle$$

$$\le -2\nu_1\|S(t)\psi\|^2$$

Hence, we have

$$\|S(t)\psi\|^2 \le e^{-2\nu_1 t}\|S(0)\psi\|^2,$$

which means

$$\|S(t)\psi\| \leq e^{-\nu_1 t}\|\psi\|$$

Since $\nu_1 = \dfrac{\pi^2\eta_a^2}{L^2} + \zeta_a > 0$, we have shown that $\|S(t)\|$ decays exponentially.

Let $\tau_k = \inf\{t \in [0,T] : Y(t) > k\} \wedge T$, we use the Picard's Iteration to prove the existence of the mild solution:

$$u^{(n+1)}(\tau_k \wedge t, x) = S(\tau_k \wedge t)u_0(x) + \int_0^{\tau_k\wedge t} S(\tau_k \wedge t - s)F(s, x, u^{(n)}(s,\cdot))ds$$

$$+C_\sigma^a \int_0^{\tau_k\wedge t} S(\tau_k \wedge t - s)u^{(n)}(s,x)\sqrt{Y(s)}dW(s)$$

We write out the proof for the case where $x > 0$. The case where $x < 0$ can be proved using the same method.

$$\begin{aligned}
&\mathbb{E}\Big[\Big|u^{(n+1)}(\tau_k \wedge t, x) - u^{(n)}(\tau_k \wedge t, x)\Big|^2\Big] \\
&= \mathbb{E}\Big[\Big|\int_0^{\tau_k\wedge t} -\beta_a S(\tau_k \wedge t - s)\Big([u_x^{(n)}(s,x)]^- - [u_x^{(n-1)}(s,x)]^-\Big)ds \\
&+ \int_0^{\tau_k\wedge t} -j(x)S(\tau_k \wedge t - s)\Big([u^{(n)}(s,x) - u_0]^+ - [u^{(n-1)}(s,x) - u_0]^+\Big)ds \\
&+ \int_0^{\tau_k\wedge t} -j(-x)S(\tau_k \wedge t - s)\Big([u^{(n)}(s,-x) + u_0]^- - [u^{(n-1)}(s,-x) + u_0]^-\Big)ds \\
&+ \int_0^{\tau_k\wedge t} S(\tau_k \wedge t - s)\Big(G(x, \int_{-\iota}^{\iota} u^{(n)}(s,y)dy) - G(x, \int_{-\iota}^{\iota} u^{(n-1)}(s,y)dy)\Big)ds \\
&+ C_\sigma^a \int_0^{\tau_k\wedge t} \sqrt{Y(s)}S(\tau_k \wedge t - s)(u^{(n)}(s,x) - u^{(n-1)}(s,x))dW(s)\Big|^2\Big] \\
&\leq 5\mathbb{E}\Big[\Big|\int_0^{\tau_k\wedge t} \beta_a S(\tau_k \wedge t - s)\Big([u_x^{(n)}(s,x)]^- - [u_x^{(n-1)}(s,x)]^-\Big)ds\Big|^2\Big] \\
&+ 5\mathbb{E}\Big[\Big|\int_0^{\tau_k\wedge t} j(x)S(\tau_k \wedge t - s)\Big([u^{(n)}(s,x) - u_0]^+ - [u^{(n-1)}(s,x) - u_0]^+\Big)ds\Big|^2\Big] \\
&+ 5\mathbb{E}\Big[\Big|\int_0^{\tau_k\wedge t} j(-x)S(\tau_k \wedge t - s)\Big([u^{(n)}(s,-x) + u_0]^- - [u^{(n-1)}(s,-x) + u_0]^-\Big)ds\Big|^2\Big]
\end{aligned}$$

$$
+5\mathbb{E}\Big[\Big|\int_0^{\tau_k\wedge t} S(\tau_k\wedge t-s)\Big(G(x,\int_{-\iota}^{\iota} u^{(n)}(s,y)dy)-G(x,\int_{-\iota}^{\iota} u^{(n-1)}(s,y)dy)\Big)ds\Big|^2\Big]
$$
$$
+5\mathbb{E}\Big[\Big|C_\sigma^a\int_0^{\tau_k\wedge t}\sqrt{Y(s)}S(\tau_k\wedge t-s)[(u^{(n)}(s,x)-u^{(n-1)}(s,x))]dW(s)\Big|^2\Big]
$$
$$
:= I+II+III+IV+V
$$

Since $S(\cdot)$ is analytical, and $[u_x]^-$ is roughly $[A^{\frac{1}{2}}u]^-$. Therefore, there exists a positive constant C_1 such that

$$
\begin{aligned}
I &\le 5C_1\beta_a^2(\tau_k\wedge t-s)^{-\frac{1}{2}}\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,x)-u^{(n-1)}(s,x)\Big|^2 ds\\
&\le 5C_1\beta_a^2\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,x)-u^{(n-1)}(s,x)\Big|^2 ds
\end{aligned}
$$

As for II, since

$$
\begin{aligned}
&\Big|[u^{(n)}(t,x)-u_0]^+-[u^{(n-1)}(t,x)-u_0]^+\Big|\\
&=\Big|\frac{[u^{(n)}(t,x)-u_0]+|u^{(n)}(t,x)-u_0|}{2}-\frac{[u^{(n-1)}(t,x)-u_0]+|u^{(n-1)}(t,x)-u_0|}{2}\Big|\\
&=\frac{1}{2}\Big|[u^{(n)}(t,x)-u_0]-[u^{(n-1)}(t,x)-u_0]+|u^{(n)}(t,x)-u_0|-|u^{(n-1)}(t,x)-u_0|\Big|\\
&\le\frac{1}{2}\Big[\Big|[u^{(n)}(t,x)-u_0]-[u^{(n-1)}(t,x)-u_0]\Big|+\Big||u^{(n)}(t,x)-u_0|-|u^{(n-1)}(t,x)-u_0|\Big|\Big]\\
&\le\frac{1}{2}\Big[\Big|[u^{(n)}(t,x)-u_0]-[u^{(n-1)}(t,x)-u_0]\Big|+\Big|[u^{(n)}(t,x)-u_0]-[u^{(n-1)}(t,x)-u_0]\Big|\Big]\\
&\le\Big|u^{(n)}(t,x)-u^{(n-1)}(t,x)\Big|,
\end{aligned}
$$

we have

$$
II\le 5\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,x)-u^{(n-1)}(s,x)\Big|^2 ds
$$

Similarly, we have

$$III \leq 5\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,-x)-u^{(n-1)}(s,-x)\Big|^2 ds$$
$$\leq 5C_2\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,x)-u^{(n-1)}(s,x)\Big|^2 ds$$

for some $C_2 > 0$, assuming that the volume density on the opposite side with the same distance away from the mid-price are not too different from each other.

Using the first condition of G in Theorem 5.1, we have

$$IV \leq 5\iota C_{\mathscr{T}}\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,x)-u^{(n-1)}(s,x)\Big|^2$$

As for the last term, we have

$$V \leq 5(C_\sigma^a)^2 C_p\int_0^{\tau_k\wedge t}\mathbb{E}\Big|Y(s)\Big|\Big|S(\tau_k\wedge t-s)(u^{(n)}(s,x)-u^{(n-1)}(s,x))\Big|^2 ds$$

$$\leq 5(C_\sigma^a)^2 C_p k e^{-2\nu_1(\tau_k\wedge t)}\int_0^{\tau_k\wedge t}\mathbb{E}\Big|u^{(n)}(s,x)-u^{(n-1)}(s,x)\Big|^2 ds.$$

Gathering I, II, III, IV, and V, there exits some constant C that depends on $T, \iota, C_\sigma^a, C_{\mathscr{T}}, C_1, C_2, \nu_1, k$ such that

$$\mathbb{E}\Big[\Big|u^{(n+1)}(\tau_k\wedge t,x)-u^{(n)}(\tau_k\wedge t,x)\Big|^2\Big] \leq C\int_0^{\tau_k\wedge t}\mathbb{E}\Big[|u^{(n)}(s,x)-u^{(n-1)}(s,x)|^2\Big]ds,$$

Since $u^{(0)}(\tau_k\wedge t,x)=u_0(x)$, we also have

$$\mathbb{E}\Big[\Big|u^{(1)}(\tau_k\wedge t,x)-u^{(0)}(\tau_k\wedge t,x)\Big|^2\Big] = \mathbb{E}\Big[\Big|S(\tau_k\wedge t)u_0(x)-u_0(x)$$
$$+\int_0^{\tau_k\wedge t}-\beta_a S(\tau_k\wedge t-s)[u_0(x)]^- ds$$

$$
\begin{aligned}
&+ \int_0^{\tau_k \wedge t} -j(x) S(\tau_k \wedge t - s)[u_0(x) - u_0]^+ ds \\
&+ \int_0^{\tau_k \wedge t} -j(-x) S(\tau_k \wedge t - s)[u_0(-x) + u_0]^- ds \\
&+ \int_0^{\tau_k \wedge t} S(\tau_k \wedge t - s) G(x, \int_{-\iota}^{\iota} u_0(y) dy) ds \\
&+ C_\sigma^a \int_0^{\tau_k \wedge t} \sqrt{Y(s)} S(\tau_k \wedge t - s) u_0(x) dW(s) \Big|^2 \Big] \\
&\leq \mathbb{E}\Big[6 \Big| S(\tau_k \wedge t) u_0(x) - u_0(x) \Big|^2 \\
&+ 6 \Big| \int_0^{\tau_k \wedge t} -\beta_a S(\tau_k \wedge t - s)[u_0(x)]^- ds \Big|^2 \\
&+ 6 \Big| \int_0^{\tau_k \wedge t} -j(x) S(\tau_k \wedge t - s)[u_0(x) - u_0]^+ ds \Big|^2 \\
&+ 6 \Big| \int_0^{\tau_k \wedge t} -j(-x) S(\tau_k \wedge t - s)[u_0(-x) + u_0]^- ds \Big|^2 \\
&+ 6 \Big| \int_0^{\tau_k \wedge t} S(\tau_k \wedge t - s) G(x, \int_{-\iota}^{\iota} u_0(y) dy) ds \Big|^2 \\
&+ 6 C_\sigma^a \Big| \int_0^{\tau_k \wedge t} \sqrt{Y(s)} S(\tau_k \wedge t - s) u_0(x) dW(s) \Big|^2 \Big]
\end{aligned}
$$

Since $j(x) \in (0, 1]$ and $S(t)$ is a bounded linear operator for all $t \geq 0$, we have

$$
\begin{aligned}
\mathbb{E}\Big[\Big| u^{(1)}(\tau_k \wedge t, x) - u^{(0)}(\tau_k \wedge t, x) \Big|^2 \Big] &\leq 6C |u_0(x)|^2 + 6\beta_a T |u_0(x)|^2 \\
&+ 6CT |u_0(x)|^2 + 6CT |u_0(x)|^2 \\
&+ 6\iota C_{\mathscr{T}} T |u_0(x)|^2 + 6(C_\sigma^a)^2 C_p k e^{-2\nu_1 t} T |u_0(x)|^2 := C'
\end{aligned}
$$

for some constant $C' > 0$ that depends on $|u_0(x)|, C, T, \iota, C_\sigma^a, \nu_1, k, C_p, C_{\mathscr{T}}$. Therefore, we have

$$
\mathbb{E}\Big[|u^{(n+1)}(\tau_k \wedge t, x) - u^{(n)}(\tau_k \wedge t, x)|^2 \Big] \leq C^n \int_0^{\tau_k \wedge t} \ldots \int_0^{\tau_k \wedge t} C' ds \ldots ds = C^n \frac{C'}{n!} \tau_k \wedge t^n \leq \frac{C_4}{n!}
$$

Therefore,

$$\|u^{(m)}(\tau_k \wedge t, x) - u^{(n)}(\tau_k \wedge t, x)\| = \|\sum_{k=n}^{m-1} u^{(k+1)}(\tau_k \wedge t, x) - u^{(k)}(\tau_k \wedge t, x)\|$$

$$\leq \sum_{k=n}^{m-1} \|u^{(k+1)}(\tau_k \wedge t, x) - u^{(k)}(\tau_k \wedge t, x)\|$$

$$= \sum_{k=n}^{m-1} \Big(\mathbb{E}\Big[\int_0^1 |u^{(k+1)}(\tau_k \wedge t, x) - u^{(k)}(\tau_k \wedge t, x)|^2 dt\Big]\Big)^{1/2}$$

$$\leq \sum_{k=n}^{m-1} \Big(\int_0^t \frac{C_4}{n!}\Big)^{1/2} = \sum_{k=n}^{m-1} \Big(\frac{C_4}{n!}\Big)^{1/2} \to 0$$

as $m, n \to \infty$. Therefore, $\{u^{(n)}(\tau_k \wedge t, x)\}_{n=0}^{\infty}$ is a Cauchy sequence in $(H, \langle \cdot, \cdot \rangle_H)$. Hence $\{u^{(n)}(\tau_k \wedge t, x)\}_{n=0}^{\infty}$ converges in $(H, \langle \cdot, \cdot \rangle_H)$. Define

$$u(\tau_k \wedge t, x) := \lim_{n \to \infty} u^n(\tau_k \wedge t, x)$$

Now we prove that $u(\tau_k \wedge t, x)$ satisfies (5.3): For all n and all $t \in [0, T]$, we have

$$u^{(n+1)}(\tau_k \wedge t, x) = A \int_0^{\tau_k \wedge t} u^{(n)}(s, x) ds + \int_0^{\tau_k \wedge t} F(s, x, u^{(n)}(t, \cdot)) ds + C_\sigma^a \int_0^{\tau_k \wedge t} u^{(n)}(s, x) \sqrt{Y(s)} dW(s)$$

Now let $n \to \infty$. Then by the Hölder inequality we get that

$$\int_0^{\tau_k \wedge t} u^{(n)}(s, x) ds \to \int_0^t u(s, x) ds$$

$$\int_0^{\tau_k \wedge t} F(s, x, u^{(n)}(s, \cdot)) ds \to \int_0^{\tau_k \wedge t} F(s, x, u(s, \cdot)) ds$$

by the Itô isometry, it follows that

$$\int_0^{\tau_k \wedge t} u^{(n)}(s,x)\sqrt{Y(s)}dW(s) \to \int_0^{\tau_k \wedge t} u(s,x)\sqrt{Y(s)}dW(s)$$

We conclude that for all $t > 0$, $u(\tau_k \wedge t, x)$ is a solution to

$$u(\tau_k \wedge t, x) = \int_0^{\tau_k \wedge t} Au(s,x)ds + \int_0^{\tau_k \wedge t} F(s,x,u(s,\cdot))ds + C_\sigma(x)\int_0^{\tau_k \wedge t} u(s,x)\sqrt{Y(s)}dW(s)$$

5.2.3 Uniqueness of $u(\tau_k \wedge t, x)$

In this section, we check uniqueness the solution $u(t,x)$. Suppose u and $\tilde{u}$ are both solutions to the equation above. We let $s_k = \tau_k \wedge \tilde{\tau}_k$ Then we have

$$\begin{aligned}
&\mathbb{E}\Big|u(s_k \wedge t, x) - \tilde{u}(s_k \wedge t, x)\Big|^2 \\
&= \mathbb{E}\Big|\int_0^{\tau_k \wedge t} S(\tau_k \wedge t, x)\Big(F(s,x,u(s,\cdot)) - F(s,x,\tilde{u}(s,\cdot))\Big)ds \\
&+ C_\sigma^a \int_0^{\tau_k \wedge t} S(\tau_k \wedge t, x)\Big(u(s,x) - \tilde{u}(s,x)\Big)\sqrt{Y(s)}dW(s)\Big|^2 \\
&\le C\int_0^{\tau_k \wedge t} \mathbb{E}\Big[|u(s,x) - \tilde{u}(s,x)|^2\Big]ds
\end{aligned}$$

for some constant C that depends on $T, \iota, C_\sigma^a, C_{\mathscr{T}}, C_1, C_2, \nu_1, k$. Using Gronwall inequality, we can conclude that for $t \in [0, \tau_k]$, $u(s_k \wedge t, x)$ and $\tilde{u}(s_k \wedge t, x)$ are modifications of each other, and thus indistinguishable. Therefore, we have shown that for all $t \in [0, \tau_k]$, $u(\tau_k \wedge t, x)$ is the unique mild solution to

$$u(\tau_k \wedge t, x) = \int_0^{\tau_k \wedge t} Au(s,x)ds + \int_0^{\tau_k \wedge t} F(s,x,u_x(s,x),u(s,\cdot))ds + C_\sigma^a \int_0^{\tau_k \wedge t} u(s,x)\sqrt{Y(s)}dW(s)$$

5.2.4 $\tau_k \to T$ *Almost Surely*

Proposition 5.2. $\tau_k \to T$ almost surely. This means that there exists $N \in \mathcal{F}$ such that $\mathbb{P}(N) = 0$ and $\tau_k(\omega) \to T$ for all $\omega \in N^c$.

Proof. Let $\epsilon > 0$, need to show that

$$\lim_{m\to\infty} \mathbb{P}\Big\{|\tau_k - T| < \epsilon \text{ for every } k \geq m\Big\} = 1$$

Considering the complement event, we have

$$\mathbb{P}\Big\{|\tau_k - T| \geq \epsilon \text{ for some } k \geq m\Big\}$$

$$= \mathbb{P}\Big\{\tau_k \leq T - \epsilon \text{ for some } k \geq m\Big\} + \mathbb{P}\Big\{\tau_k \geq T + \epsilon \text{ for some } k \geq m\Big\}$$

$$= \mathbb{P}\Big\{\tau_k \leq T - \epsilon \text{ for some } k \geq m\Big\}$$

$$= \mathbb{P}\Big\{\sup_{t \leq T-\epsilon} Y(t) > k \text{ for some } k \geq m\Big\}$$

$$\leq \mathbb{P}\Big\{\sup_{t \leq T-\epsilon} Y(t) > m\Big\}$$

$$\leq \mathbb{P}\{Y(0) + \sup_{0 \leq s \leq t \leq T-\epsilon} |Y(t) - Y(s)| > m\}$$

$$\leq \Big(\frac{1}{m - y_0}\Big)^p \mathbb{E}[\sup_{0 \leq s \leq t \leq T-\epsilon} |Y(t) - Y(s)|^p] \qquad \text{for some } p \geq 4$$

Now we check if we can apply Lemma G.4. Indeed, we can write $Y(t)$ as $Y = \mathscr{K} * (bdt + dM)$, where

$$\mathscr{K}(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}, \qquad b(y) = -\theta y + 4\theta(\gamma + 1), \qquad a(y) = \frac{\theta}{2\mu(\gamma+1)} y$$

By Lemma G.3, for any $p \geq 2$ and $T < \infty$,

$$\sup_{t \leq T-\epsilon} \mathbb{E}[|Y(t)|^p] \leq c$$

for some constant c that only depends on

$$\gamma, \quad \int_0^T \Big[\frac{t^{\alpha-1}}{\Gamma(\alpha)}\Big]^2 dt, \quad C_{LG}, \quad p, \quad \epsilon, \quad \text{and } T.$$

Therefore, by Lemma G.4, we have

$$\mathbb{E}\Big[\Big(\sup_{0\leq s<t\leq T-\epsilon} |Y(t)-Y(s)|\Big)^p\Big] \leq c' \sup_{t\leq T-\epsilon} \mathbb{E}[|a(t)|^{p/2} + |b(t)|^p]$$

$$\leq c' \sup_{t\leq T-\epsilon} \mathbb{E}[C|Y(t)|^{p/2} + C(1+|Y(t)|^p)] \leq C'$$

for some C' depending on $\gamma, \theta, \mu, p, T, \epsilon, C_{LG}$. Therefore, we have

$$\mathbb{P}\Big\{|\tau_k - T| \geq \epsilon \text{ for some } k \geq m\Big\} \leq \Big(\frac{1}{m-y_0}\Big)^p \mathbb{E}[\sup_{0\leq s\leq t\leq T-\epsilon} |Y(t)-Y(s)|^p]$$

$$\leq \Big(\frac{1}{m-y_0}\Big)^p C'$$

When $m \to \infty$, $\mathbb{P}\Big\{|\tau_k - T| \geq \epsilon \text{ for some } k \geq m\Big\} \to 0$. Therefore, we have proved that

$$\lim_{m\to\infty} \mathbb{P}\Big\{|\tau_k - T| < \epsilon \text{ for every } k \geq m\Big\} = 1,$$

and hence $\tau_k \to T$ almost surely. This means that there exists $N \in \mathcal{F}$ such that $\mathbb{P}(N) = 0$ and $\tau_k(\omega) \to T$ for all $\omega \in N^c$.

$\square$

5.2.5 *Extension of the unique mild solution from* $[0,\tau_k]$ *to* $[0,T]$

In this section, we treat the model by truncating $Y(t)$. Let $t\in[0,T]$, for each $k\in\mathbb{Z}^+$, we have

$$u^k(t,x)=\begin{cases} u_0^k(x)+\int_0^t[Au^k(s,x)+F(s,x,u_x^k(s,x),u^k(s,\cdot))]ds \\ +\int_0^t C_\sigma^a u^k(s,x)\sqrt{k}dW(s), \quad Y(t)>k \\ u_0^k(x)+\int_0^t[Au^k(s,x)+F(s,x,u_x^k(s,x),u^k(s,\cdot))]ds \\ +\int_0^t C_\sigma^a u^k(s,x)\sqrt{Y(s)}dW(s), \quad Y(t)\le k \end{cases}$$

$$Y(t)=2(\gamma+1)+\frac{\theta}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}(4(\gamma+1)-Y(s))ds+\frac{1}{\Gamma(\alpha)}\sqrt{\frac{\theta}{2\mu(\gamma+1)}}\int_0^t(t-s)^{\alpha-1}\sqrt{Y(s)}dB(s)$$

Meanwhile, for each $k\in\mathbb{Z}^+$, and for $t\in[0,\tau_k]$, we also have

$$\begin{aligned} u_k(t,x)=u_{k,0}(x)&+\int_0^t[Au_k(s,x)+F(s,x,[u_k]_x(s,x),u_k(s,\cdot))]ds \\ &+\int_0^t C_\sigma^a u_k(s,x)\sqrt{Y(s)}dW(s), \quad Y(t)\le k \end{aligned}$$

$$Y(t)=2(\gamma+1)+\frac{\theta}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}(4(\gamma+1)-Y(s))ds+\frac{1}{\Gamma(\alpha)}\sqrt{\frac{\theta}{2\mu(\gamma+1)}}\int_0^t(t-s)^{\alpha-1}\sqrt{Y(s)}dB(s)$$

By Proposition 5.2, there exists $N\in\mathcal{F}$ such that $\mathbb{P}(N)=0$ and $\tau_k(\omega)\to T$ for all $\omega\in N^c$. Assuming the functions $u_0^k=u_{k,0}$ for all $k\in\mathbb{Z}^+$, we want to show that for each $\omega\in N^c$, each $t\in[0,T]$, there exists $\tilde{k}(\omega)$ such that $t\le\tau_{\tilde{k}(\omega)}$. Following similar steps in the proof of Proposition 5.2, we have

$$\begin{gathered} \mathbb{P}\Big\{\tau_k<T \text{ for all } k\Big\} \\ \le\mathbb{P}\Big\{\sup_{t\le T}Y(t)>k \text{ for all } k\Big\} \\ \le\mathbb{P}\Big\{2(\gamma+1)+\sup_{0\le s\le t\le T}|Y(t)-Y(s)|>k \text{ for all } k\Big\} \end{gathered}$$

$$\leq \Big(\frac{1}{k-2(\gamma+1)}\Big)^p \mathbb{E}[\sup_{0\leq s\leq t\leq T} |Y(t)-Y(s)|^p] \qquad \text{for some } p \geq 4 \text{ and all } k$$

$$\leq \Big(\frac{1}{k-2(\gamma+1)}\Big)^p C'$$

for all k and some $p > 4$, with C' depending on $\gamma, \theta, \mu, \alpha, p, T, \epsilon, C_{LG}$. Therefore, for all $k \in \mathbb{Z}^+$, we have

$$\mathbb{P}\Big\{\tau_k < T \text{ for all } k\Big\} \leq \Big(\frac{1}{k-2(\gamma+1)}\Big)^p C' \text{ for all } k$$

We want to show that $\mathbb{P}\Big\{\tau_k < T \text{ for all } k\Big\} = 0$. Suppose for contradiction that $\mathbb{P}\Big\{\tau_k < T \text{ for all } k\Big\} > 0$, and let it be ϵ, then by the Archimedean Property, there exists an $N \in \mathbb{Z}^+$ such that

$$N\epsilon^{1/p} > (C')^{1/p} \Rightarrow N^p\epsilon > C',$$

There also exists a $k \in \mathbb{Z}^+$ such that $N = \lfloor k-2(\gamma+1)\rfloor$. Therefore, there exists a $k \in \mathbb{Z}^+$ such that

$$\mathbb{P}\Big\{\tau_k < T \text{ for all } k\Big\} = \epsilon > \frac{C'}{N^p} = \frac{C'}{\lfloor k-2(\gamma+1)\rfloor^p} \geq \frac{C'}{(k-2(\gamma+1))^p}$$

We have reached a contradiction. Therefore,

$$\mathbb{P}\Big\{\tau_k < T \text{ for all } k\Big\} = 0,$$

which means

$$\mathbb{P}\Big\{\tau_k = T \text{ for some } k\Big\} = 1,$$

or equivalently, $\tau_k = T$ for some k almost surely. Therefore, for each $\omega \in N^c$ and $t \in [0,T]$, there exists $\tilde{k}(\omega)$ such that $\tau_{\tilde{k}(\omega)} = T \geq t$.

When $t \leq \tau_{\tilde{k}(\omega)}$, $Y(t) \leq \tilde{k}(\omega)$, and $u^{\tilde{k}}(t,x) = u_{\tilde{k}}(t,x)$. Therefore, we have shown that for each $\omega \in N^c$ and $t \in [0,T]$, $u^{\tilde{k}}(t,x) = u_{\tilde{k}}(t,x)$ for all $k \geq \tilde{k}(\omega)$.

CHAPTER 6: PRICE DYNAMICS

The bid and ask price dynamics are determined by the LOB dynamics. When the ask (bid) queue is depleted, the price moves up (down) to the next level of the order book. We assume that the order book contains no gaps so that the price increments are equal to one tick, which is δ as defined in section 3.1. When the bid queue is depleted, the price decreases by one tick. When the ask queue is depleted, the price increases by one tick. On the other hand, if the queue sizes increase rapidly in a short period of time, it means there are excessive amount of limit orders, which will likely be transferred to market orders and be executed towards the opposite direction. When the ask queue size increases n times, the price will move down n ticks. When the bid queue size increases n times, the price will move up n ticks.

We use a simple example to illustrate how the LOB dynamics determine the bid and ask prices. Suppose in the illustrative LOB (1.1), there is a bid order of 10,000 shares, then the first 2 queues on the ask side will be depleted, and the ask price will moving up 2 ticks, rising from \$100.01 to to \$100.03.

All 3 LOB activities affect the ask and bid queues. Submission of limit orders increase the queues, while cancellation of limit orders as well as market orders from the opposite side decrease the queues. Therefore, the price changes are determined by the volume changes, and we model the volume by the order book depth D^a and D^b [13].

6.1 Price Dynamics Model

Let $D^a(t)$ ($D^b(t)$) be the volume of limit ask (bid) orders at the top of the LOB at time t. The order book depth can be expressed as

$$D^a(t) = \int_0^{\iota} u(t,x)dx, \qquad D^b(t) = \int_{-\iota}^{0} u(t,x)dx$$

Let the change of the order book depth in the time interval $[t, t+dt]$ be $dD^a(t)$ and $dD^b(t)$. Note that since $u(t,x) > 0$ on the ask side and $u(t,x) < 0$ on the bid side, $D^a(t) > 0$ and $D^b(t) < 0$.

When $dD^a(t) < 0$, the ask queue decreases and the ask price increases by $-\dfrac{dD^a(t)}{D^a(t)}$ ticks. When $dD^a(t) > 0$, the ask price decreases by $\dfrac{dD^a(t)}{D^a(t)}$ ticks. Therefore, the price impact from the ask queue is $-\dfrac{dD^a(t)}{D^a(t)}$. On the other hand, when $dD^b(t) > 0$, $D^b(t)$ increases, but since $D^b(t) < 0$, this means that the bid queue decreases, and the bid price decreases by $-\dfrac{dD^b(t)}{D^b(t)}$ ticks. When $dD^b(t) > 0$, the bid queue increases, and the bid price increases by $\dfrac{dD^b(t)}{D^b(t)}$ ticks. Therefore, the price impact from the bid queue is $\dfrac{dD^b(t)}{D^b(t)}$. In summary, the ask and bid price changes will be:

$$ds^a(t) = -\delta\frac{dD^a(t)}{D^a(t)}, \qquad ds^b(t) = \delta\frac{dD^b(t)}{D^b(t)}$$

and the price change will be

$$dS(t) = \frac{1}{2}(ds^a(t) + ds^b(t)) = \frac{\delta}{2}\Big(\frac{dD^b(t)}{D^b(t)} - \frac{dD^a(t)}{D^a(t)}\Big)$$

We first find the dynamics of $D^a(t)$ and $D^b(t)$. Indeed, by the Leibniz integral rule, we have

$$
\begin{aligned}
dD^a(t) &= d\int_0^\iota u(t,x)dx \\
&= \int_0^\iota du(t,x)dx \\
&= \int_0^\iota \Big\{\Big[\eta_a u_{xx}(t,x) - \beta_a[u_x(t,x)]^- - \zeta_a u(t,x) \\
&- j(x)\big(u(t,x)-u_0\big)^+ - j(-x)\big(u(t,-x)+u_0\big)^- \\
&+ G(x,\ell(t))\Big]dt\Big\}dx + \int_0^\iota \Big\{C_\sigma^a u(t,x)\sqrt{Y(t)}dW(t)\Big\}dx \\
&= \int_0^\iota \Big\{\Big[\eta_a u_{xx}(t,x) - \beta_a[u_x(t,x)]^- - \zeta_a u(t,x) \\
&- j(x)\big(u(t,x)-u_0\big)^+ - j(-x)\big(u(t,-x)+u_0\big)^- \\
&+ G(x,\ell(t))\Big]dt\Big\}dx + \Big(\int_0^\iota u(t,x)dx\Big)C_\sigma^a\sqrt{Y(t)}dW(t) \\
&= \int_0^\iota \Big\{\Big[\eta_a u_{xx}(t,x) - \beta_a[u_x(t,x)]^- - \zeta_a u(t,x) \\
&- j(x)\big(u(t,x)-u_0\big)^+ - j(-x)\big(u(t,-x)+u_0\big)^- \\
&+ G(x,\ell(t))\Big]dt\Big\}dx + D^a(t)C_\sigma^a\sqrt{Y(t)}dW(t)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
dD^b(t) = \int_0^\iota \Big\{\Big[&\eta_b u_{xx}(t,x) + \beta_b[u_x(t,x)]^- - \zeta_b u(t,x) + j(-x)\big(u(t,-x)-u_0\big)^+ \\
&+ j(x)\big(u(t,x)+u_0\big)^- + G(x,\ell(t))\Big]dt\Big\}dx + D^b(t)C_\sigma^b\sqrt{Y(t)}dW(t)
\end{aligned}
$$

Therefore, we have the price dynamics model as

$$\begin{cases} dS(t) = \dfrac{\delta}{2}[\nu_b(t) - \nu_a(t)]dt + \dfrac{\delta}{2}(C^b_\sigma - C^a_\sigma)\sqrt{Y(t)}dW(t), \qquad S(0) = S_0 > 0 \\ Y(t) = 2(\gamma+1) + \dfrac{\theta}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(4(\gamma+1) - Y(s))ds \\ +\dfrac{1}{\Gamma(\alpha)}\sqrt{\dfrac{\theta}{2\mu(\gamma+1)}}\int_0^t (t-s)^{\alpha-1}\sqrt{Y(s)}dB(s) \end{cases} \tag{6.1}$$

where

$$\begin{aligned} \nu_a(t) = \frac{1}{D^a(t)}\int_0^\iota \Big\{ \Big[&\eta_a u_{xx}(t,x) - \beta_a[u_x(t,x)]^- - \zeta_a u(t,x) \\ &- j(x)\big(u(t,x) - u_0\big)^+ - j(-x)\big(u(t,-x) + u_0\big)^- + G(x,\ell(t))\Big] dt \Big\} dx \end{aligned}$$

$$\begin{aligned} \nu_b(t) = \frac{1}{D^b(t)}\int_0^\iota \Big\{ \Big[&\eta_b u_{xx}(t,x) + \beta_b[u_x(t,x)]^- - \zeta_b u(t,x) \\ &+ j(-x)\big(u(t,-x) - u_0\big)^+ + j(x)\big(u(t,x) + u_0\big)^- + G(x,\ell(t))\Big] dt \Big\} dx \end{aligned}$$

with $\eta_a, \eta_b, \beta_a, \beta_b, \zeta_a, \zeta_b, C^a_\sigma, C^b_\sigma$ all positive constants. $\delta > 0$ is the tick size of the market.

In order to analyze the parameters, we need to explicitly express the parameters in the price dynamics. Therefore, we write (6.1) as

$$\begin{cases} dS(t) = \dfrac{\delta}{2}[\nu_b(t) - \nu_a(t)]dt + \dfrac{\delta}{2}(C^b_\sigma - C^a_\sigma)\sqrt{Y(t)}dW(t) \\ S(0) = S_0, \\ Y(t) = 2(\gamma+1) + \dfrac{\theta}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(4(\gamma+1) - Y(s))ds \\ +\dfrac{1}{\Gamma(\alpha)}\sqrt{\dfrac{\theta}{2\mu(\gamma+1)}}\int_0^t (t-s)^{\alpha-1}\sqrt{Y(s)}dB(s) \end{cases}$$

6.2 Parameters Analysis

6.2.1 Financial Meaning of the Parameter

We summarize all the parameters in the following table:

Table 6.1: Parameters

Parameter	Expression	Range
α	$\varphi(x) \underset{x\to\infty}{\sim} \frac{K}{x^{1+\alpha}}$ as $x \to \infty$, with $K > 0$ a constant	$(\frac{1}{2}, 1)$
γ	$\gamma = \frac{\beta_2(\beta_3+1)}{\beta_1+\beta_2\beta_3+2\beta_2-1}$	$(0, \infty)$
θ	$\lim\limits_{T\to\infty}(1 - a_T)T^{\alpha} = \theta K\Gamma(1 - \alpha)$	$(0, \infty)$
μ	$\lim\limits_{T\to\infty} T^{1-\alpha}\mu_T = \dfrac{\mu}{K\Gamma(1 - \alpha)}$	$(0, \infty)$

The settings of θ and μ are mainly to provide the convergence order of the sequences a_T and μ_T, so we focus on the financial meaning of α and γ:

(1) α:

$\varphi(x) \underset{x\to\infty}{\sim} \frac{K}{x^{1+\alpha}}$ means that the Hawkes kernel has a power law decay. Financially, this means that the inducing power of the same type of LOB event decays slower than an exponential decay, which is mainly caused by the metaorder splitting strategy. The smaller α, the slower the decay, and the more frequently the metaorder splitting strategy is used. This will induce a larger volatility.

(2) γ:

$\gamma = \frac{\beta_2(\beta_3+1)}{\beta_1+\beta_2\beta_3+2\beta_2-1}$, and we can see from our Hawkess process:

$$\begin{pmatrix} \lambda_t^{a,+} \\ \lambda_t^{b,+} \\ \lambda_t^{a,-} \\ \lambda_t^{b,-} \end{pmatrix} = \begin{pmatrix} \mu^{a,+} \\ \mu^{b,+} \\ \mu^{a,-} \\ \mu^{b,-} \end{pmatrix} + \int_0^t \frac{\varphi(t)}{\beta_1+\beta_2\beta_3+2\beta_2}$$

$$\begin{pmatrix} 1 & 0 & \beta_2 & (\beta_1+\beta_2+\beta_2\beta_3-1) \\ 0 & 1 & (\beta_1+\beta_2+\beta_2\beta_3-1) & \beta_2 \\ \beta_2 & \beta_2\beta_3 & (\beta_1+\beta_2) & 0 \\ \beta_2\beta_3 & \beta_2 & 0 & (\beta_1+\beta_2) \end{pmatrix} \begin{pmatrix} dN_s^{a,+} \\ dN_s^{b,+} \\ dN_s^{a,-} \\ dN_s^{b,-} \end{pmatrix}$$

that γ is the ratio between the inducing power from limit orders to market orders, and the inducing power from market orders to limit orders. This is essentially the ratio of

$$\frac{\text{Power of taking away liquidity}}{\text{Power of providing liquidity}}$$

6.2.2 Price Simulation

Since we focus on the parameters and the price volatility, we zero out the drift term, and add a small positive number ϵ to treat the singularity. We also explicitly write out all the parameters:

$$\begin{cases} dS(t) = \dfrac{\delta}{2}(C_\sigma^b - C_\sigma^a)\sqrt{Y(t)}dW(t) \\ S(0) = S_0, \\ dY(t) = \dfrac{\theta\epsilon^{\alpha-1}}{\Gamma(\alpha)}(4(\gamma+1) - Y(s))dt + \dfrac{\epsilon^{\alpha-1}}{\Gamma(\alpha)}\sqrt{\dfrac{\theta}{2\mu(\gamma+1)}}\sqrt{Y(t)}dB(t) \\ Y(0) = 2(\gamma+1) \end{cases}$$

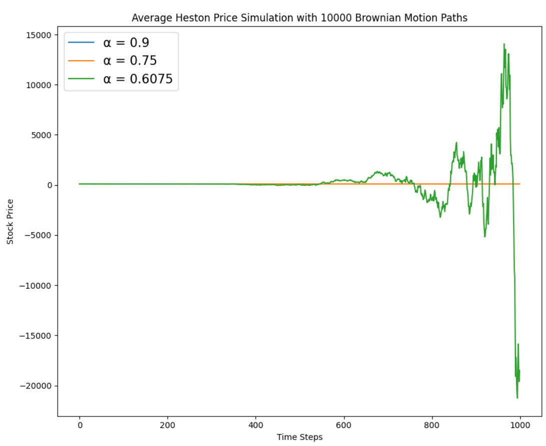
where

$$W = B^1 + B^2 - B^3 - B^4, \qquad B = \gamma B^1 + \gamma B^2 + B^3 + B^4$$

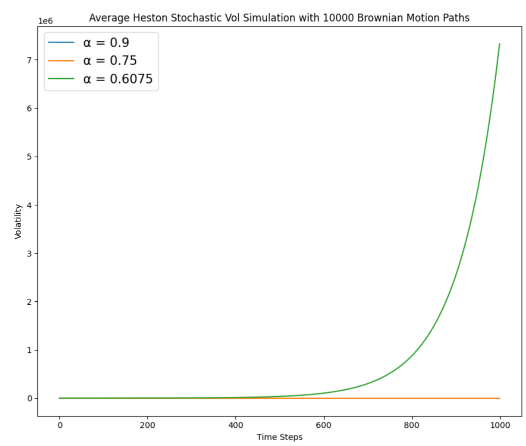
are two 1-dimensional Brownian motions, with B^1, B^2, B^3, B^4 four 1-dimensional independent Brownian motions. Let $S_0 = 100$, $\delta = 2$, $C_\sigma^b - C_\sigma^a = 0.6$, $\epsilon = 10^{-15}$, $\theta = 0.1$, $\mu = 0.8$, B^1, B^2, B^3, B^4 are standard 1-dimensional Brownian motions that are independent from each other.

(1) α:

We fix all the other parameters and check S and Y with $0.6 \leq \alpha \leq 0.95$. Let $\gamma = 1$.

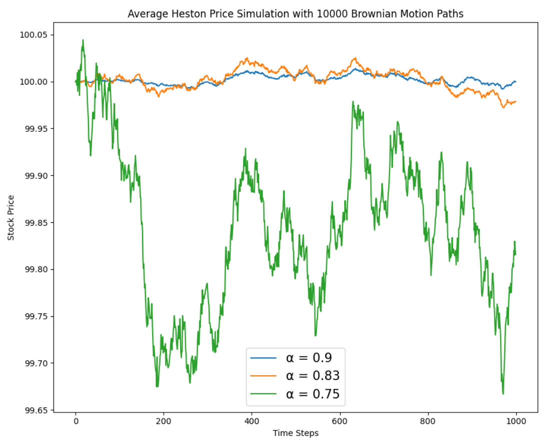


(a) $S(t)$

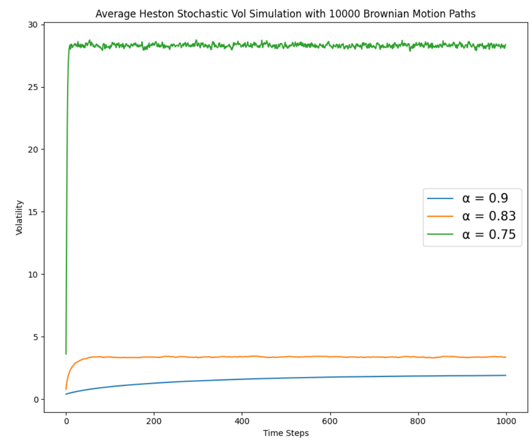


(b) $Y(t)$

Figure 6.1: Simulated Price and Volatility of $\alpha \in \{0.9, 0.75, 0.6075\}$

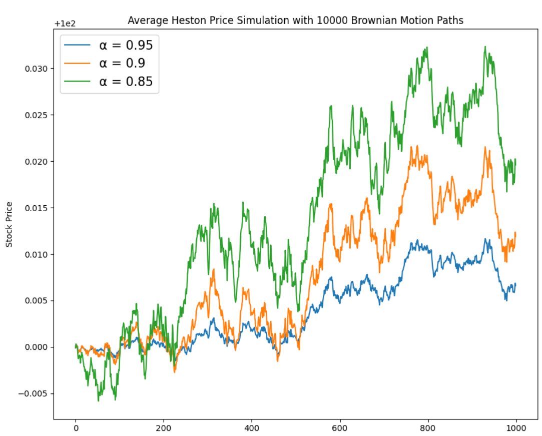


(a) $S(t)$

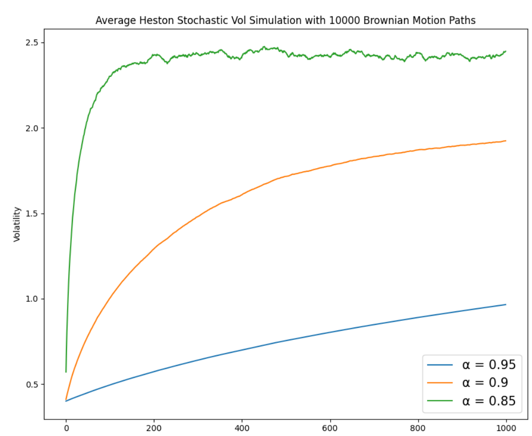


(b) $Y(t)$

Figure 6.2: Simulated Price and Volatility of $\alpha \in \{0.9, 0.83, 0.75\}$



(a) $S(t)$

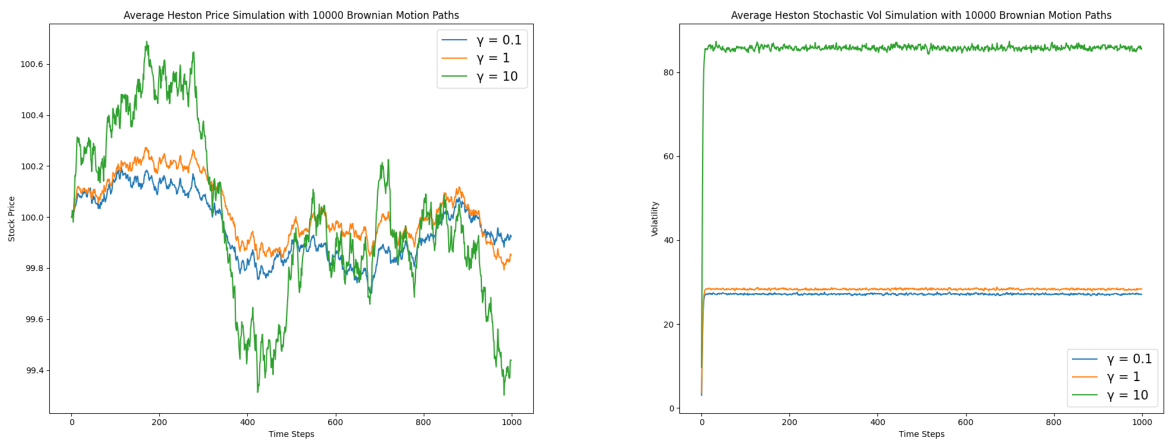


(b) $Y(t)$

Figure 6.3: Simulated Price and Volatility of $\alpha \in \{0.95, 0.9, 0.85\}$

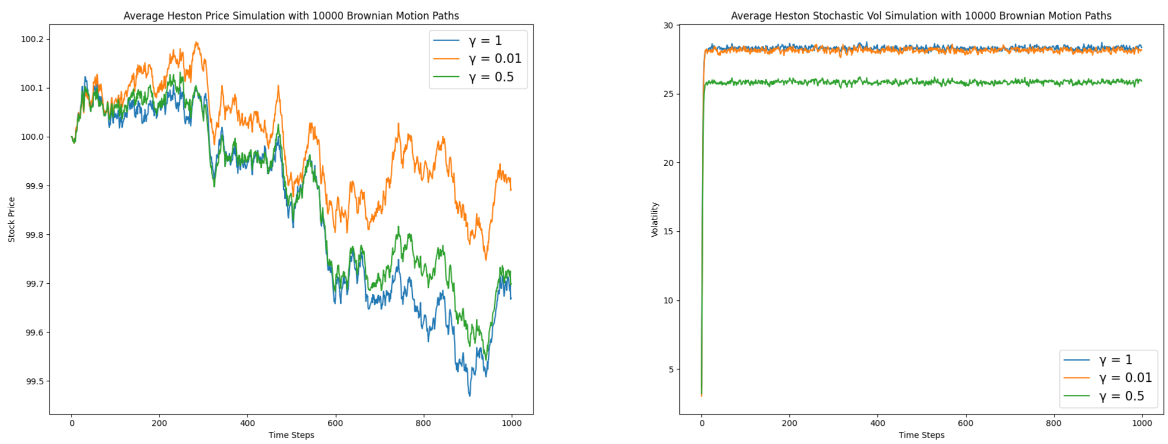
(2) γ:

We fix all the other parameters and check S and Y with varying γ. Let $\alpha = 0.75$.



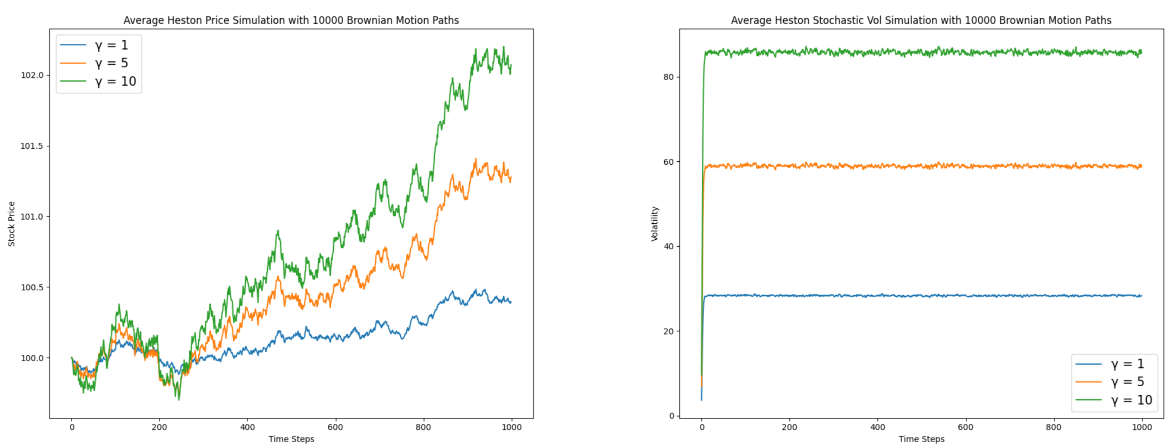
(a) $S(t)$ (b) $Y(t)$

Figure 6.4: Simulated Price and Volatility of $\gamma \in \{0.1, 1, 10\}$



(a) $S(t)$ (b) $Y(t)$

Figure 6.5: Simulated Price and Volatility of $\gamma \in \{1, 0.01, 0.5\}$



(a) $S(t)$ (b) $Y(t)$

Figure 6.6: Simulated Price and Volatility of $\gamma \in \{1, 5, 10\}$

CHAPTER 7: CONCLUSION

In this paper, we made two modifications to the C-M model: We included market orders, and instead of modeling the HFT dynamics with Brownian Motion, we used the scaling limit of a series of nearly-unstable multivariate Hawkes process with power-law tails. The second change enables our model to reflect the dependencises among HFT orders.

Based on the order book dynamics, we also created a middle price dynamics model in the same market. We analyzed parameters in the price model to determine how they impacted the price changes. We found out that among all the parameters, α, the parameter that measures how frequently the metaorder splitting strategy is used, has the most significant impact: The more frequently the strategy is used, the larger volatility there will be in the price change.

APPENDIX A: EIGENVALUE CALCULATION

In this appendix, we calculate the eigenvalues of $\mathbf{\Phi}_0$, where

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & \beta_2 & (\beta_1 + \beta_2 + \beta_2\beta_3 - 1) \\ 0 & 1 & (\beta_1 + \beta_2 + \beta_2\beta_3 - 1) & \beta_2 \\ \beta_2 & \beta_2\beta_3 & (\beta_1 + \beta_2) & 0 \\ \beta_2\beta_3 & \beta_2 & 0 & (\beta_1 + \beta_2) \end{pmatrix}$$

Let the eigenvalue be λ and the corresponding eigenvector be v, then we have

$$\mathbf{\Phi}v - \lambda v = (\mathbf{\Phi} - \lambda I)v = 0,$$

where I is the identity matrix. This equation has a nonzero solution if and only if $\det(\mathbf{\Phi} - \lambda I) = 0$. Then we have

$$
\det(\boldsymbol{\Phi} - \lambda I)
$$

$$
\begin{aligned}
&= \begin{vmatrix} 1-\lambda & 0 & \beta_2 & \beta_1+\beta_2+\beta_2\beta_3-1 \\ 0 & 1-\lambda & \beta_1+\beta_2+\beta_2\beta_3-1 & \beta_2 \\ \beta_2 & \beta_2\beta_3 & \beta_1+\beta_2-\lambda & 0 \\ \beta_2\beta_3 & \beta_2 & 0 & \beta_1+\beta_2-\lambda \end{vmatrix} \\
&= \lambda^4 - (2\beta_1+2\beta_2+2)\lambda^3 + (\beta_1^2 + 4\beta_1 - \beta_2^2 + 2\beta_1\beta_2 + 4\beta_2 - 2\beta_2^2\beta_3^2 \\
&- 2\beta_2^2\beta_3 - 2\beta_1\beta_2\beta_3 + 2\beta_2\beta_3 + 1)\lambda^2 - (2\beta_1^2 + 2\beta_1 - 2\beta_2^3 - 2\beta_1\beta_2^2 \\
&+ 4\beta_1\beta_2 + 2\beta_2 - 2\beta_2^3\beta_3^2 - 2\beta_1\beta_2^2\beta_3^2 - 2\beta_2^2\beta_3^2 - 2\beta_2^3\beta_3 - 4\beta_1\beta_2^2\beta_3 \\
&- 2\beta_1^2\beta_2\beta_3 + 2\beta_2\beta_3)\lambda + (\beta_1^2 - 2\beta_1\beta_2^3 - \beta_1^2\beta_2^2 + 2\beta_1\beta_2 + \beta_2^4\beta_3^4 \\
&+ 2\beta_2^4\beta_3^3 + 2\beta_1\beta_2^3\beta_3^3 - 2\beta_2^3\beta_3^3 - \beta_2^4\beta_3^2 + 2\beta_1\beta_2^3\beta_3^2 - 4\beta_2^3\beta_3^2 \\
&+ \beta_1^2\beta_2^2\beta_3^2 - 4\beta_1\beta_2^2\beta_3^2 + \beta_2^2\beta_3^2 - 2\beta_2^4\beta_3 - 2\beta_1\beta_2^3\beta_3 - 4\beta_1\beta_2^2\beta_3 \\
&+ 2\beta_2^2\beta_3 - 2\beta_1^2\beta_2\beta_3 + 2\beta_1\beta_2\beta_3) \\
&= \Big(\lambda - (\beta_2 - \beta_2\beta_3 + 1)\Big)\Big(\lambda^3 - (2\beta_1 + \beta_2 + \beta_2\beta_3 + 1)\lambda^2 \\
&+ (\beta_1^2 + 2\beta_1 - 2\beta_2^2 + 2\beta_2 - \beta_2^2\beta_3^2 - 2\beta_2^2\beta_3 + 2\beta_2\beta_3)\lambda \\
&- (\beta_1^2 - 2\beta_1\beta_2^2 - \beta_1^2\beta_2 + 2\beta_1\beta_2 - \beta_2^3\beta_3^3 - 3\beta_2^3\beta_3^2 - 2\beta_1\beta_2^2\beta_3^2 + \beta_2^2\beta_3^2 \\
&- 2\beta_2^3\beta_3 - 4\beta_1\beta_2^2\beta_3 + 2\beta_2^2\beta_3 - \beta_1^2\beta_2\beta_3 + 2\beta_1\beta_2\beta_3)\Big) \\
&= \Big(\lambda - (\beta_2 - \beta_2\beta_3 + 1)\Big)\Big(\lambda + (\beta_2 + \beta_2\beta_3 - 1)\Big) \\
&\Big(\lambda - (\beta_1 + \beta_2\beta_3)\Big)\Big(\lambda - (\beta_1 + 2\beta_2 + \beta_2\beta_3)\Big) = 0
\end{aligned}
$$

Therefore, we get the eigenvalues

$$
\lambda_1 = \beta_1 + \beta_2\beta_3 + 2\beta_2, \qquad \lambda_2 = -\beta_2\beta_3 + \beta_2 + 1,
$$

$$\lambda_3 = \beta_1 + \beta_2\beta_3, \qquad \lambda_4 = -\beta_2\beta_3 - \beta_2 + 1$$

APPENDIX B: THE CLUSTER REPRESENTATION OF HAWKES PROCESSES

Another way of understanding the multi-dimensional Hawkes process is through the cluster representation [24], also called "Immigration-Birth" Representation. Recall that our microscopic volume model $\mathbf{N}(\cdot)$, a four-dimensional Hawkes process, is defined as following:

$$\mathbf{N}(\cdot) := \begin{pmatrix} N^{a,+}(\cdot) \\ N^{b,+}(\cdot) \\ N^{a,-}(\cdot) \\ N^{b,-}(\cdot) \end{pmatrix}$$

with the associated conditional intensity:

$$\boldsymbol{\lambda}(\cdot) = \boldsymbol{\mu}(\cdot) + \int_0^{\cdot} \boldsymbol{\Phi}(\cdot - s) d\mathbf{N}(s),$$

where

$$\boldsymbol{\mu}(\cdot) = \begin{pmatrix} \mu^{a,+} \\ \mu^{b,+} \\ \mu^{a,-} \\ \mu^{b,-} \end{pmatrix}(\cdot), \qquad \boldsymbol{\Phi}(\cdot) = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{14} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{24} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi_{34} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} \end{pmatrix}(\cdot).$$

For the subscripts of each entry of $\boldsymbol{\Phi}(\cdot)$, 1 stands for limit ask orders, 2 for limit bid orders, 3 for market ask orders, and 4 for market bid orders. We will use the following graph to illustrate the cluster representation of $\mathbf{N}(t)$:

In this graph, each circle represents an event that happens at T_i. The arrows show the root $\rightarrow$ offspring relationship, and *Gen*$_k$ specifies the generation of the event, while $k = 0$ being an immigrant and $k > 0$ being the k-th generation from an immigrant. $Z_{i,j}$ are random variables such that $Z_{i,0} = 1$ if the event that happens at T_i is an immigrant, and $Z_{i,j} = 1$ if the event that happens at T_i is an immediate offspring of event that happens at T_j. For example, the events that hap-

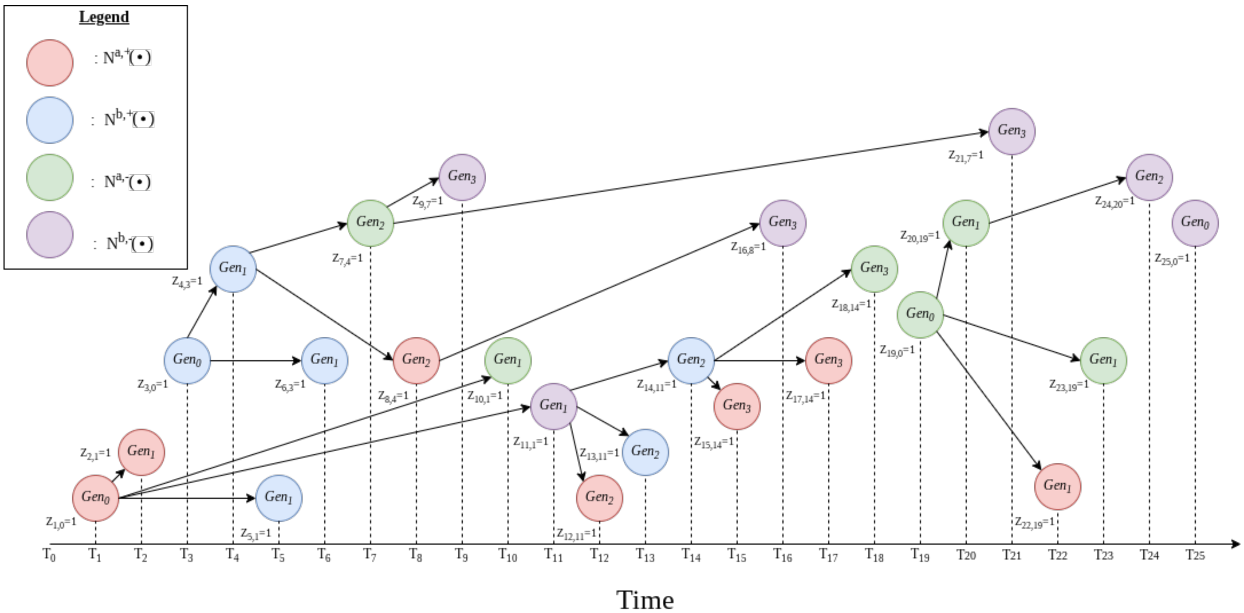


Figure B.1: Four-Dimensional Hawkes Process Cluster Representation

pen at T_2, T_5, T_{10}, and T_{11} are immediate offspring of the event that happens at T_1, and therefore $Z_{2,1} = Z_{5,1} = Z_{1,10} = Z_{1,11} = 1$. On the other hand, an example of immediate offspring of a descendent will be the events that happen at T_9 and T_{21}, which are immediate offspring of the event that happens at T_7, and therefore $Z_{9,7}$ and $Z_{21,7} = 1$.

In this representation, all the Gen_0 events occur following an inhomogeneous Poisson process with rate functions as their correspondent background intensity function. For example, the arrival of $Z_{1,0}$, $Z_{3,0}$, $Z_{19,0}$, and $Z_{25,0}$ follow respectively an inhomogeneous Poisson process with the rate function $\mu^{a,+}(\cdot)$, $\mu^{b,+}(\cdot)$, $\mu^{a,-}(\cdot)$, and $\mu^{a,-}(\cdot)$.

The immediate offspring events, $Z_{i,j}$, arrive according to an inhomogeneous Poisson process with the rate function $\varphi_{lm}(t-T_i)$ for $t > T_i$, where $m, l \in \{1, 2, 3, 4\}$ are the integers associated with the type of events, with 1 corresponding to limit ask orders, 2 to limit bid orders, 3 to market ask orders, and 4 to market bid orders. In $\varphi_{lm}(\cdot)$, m is correspondent to the type of the event that happened at

T_j, and l to the type of the event that happened at T_i. We can also write $Z_{i,j} \sim \text{Poi}(\nu_{lm})$, with

$$\nu_{lm} := \int_{T_j}^{\infty} \varphi_{lm}(t - T_j)dt = \int_0^{\infty} \varphi_{lm}(s)ds.$$

For example, the events that happen at T_7 and T_8 are immediate offspring of the event that happens at T_4. Note that the event that happens at T_4 is an $N^{b,+}$ event, the event that happens at T_7 is an $N^{a,-}$ event and the event that happens at T_8 is an $N^{a,+}$ event. Therefore, the arrivals of $Z_{7,4}$ and $Z_{8,4}$ follow respectively an inhomogeneous Poisson process with the rate function $\varphi_{32}(t - T_4)$ and $\varphi_{12}(t - T_4)$. We can also write $Z_{7,4} \sim \text{Poi}(\nu_{32})$ and $Z_{8,4} \sim \text{Poi}(\nu_{12})$, where

$$\nu_{32} = \int_0^{\infty} \varphi_{32}(s)ds, \qquad \nu_{12} = \int_0^{\infty} \varphi_{12}(s)ds.$$

All the events that are directly or indirectly connected to an immigrant form a cluster. For example, the events that happen at T_2, T_5, T_{10}, T_{11}, T_{12}, T_{13}, T_{14}, T_{15}, T_{17}, and T_{18} form a cluster as offspring of the event that happens at T_1. Similarly, the event that happes at T_{19}, T_{20}, T_{22}, T_{23}, and T_{24} form another cluster. The event happens at T_{25} is a cluster by itself.

Focusing only on the event type of the immediate offspring, we can see that the average number of immediate offspring that are $N^{a,+}$ events, regardless of their parents' event types, is

$$\int_0^{\infty} \Big[\varphi_{11}(s) + \varphi_{12}(s) + \varphi_{13}(s) + \varphi_{14}(s)\Big]ds.$$

Similarly, the average number of immediate offspring that are other events are:

$$N^{b,+}: \qquad \int_0^{\infty} \Big[\varphi_{21}(s) + \varphi_{22}(s) + \varphi_{23}(s) + \varphi_{24}(s)\Big]ds$$

$$N^{a,-}: \qquad \int_0^{\infty} \Big[\varphi_{31}(s) + \varphi_{32}(s) + \varphi_{33}(s) + \varphi_{34}(s)\Big]ds$$

$$N^{b,-}: \qquad \int_0^\infty \Big[\varphi_{41}(s) + \varphi_{42}(s) + \varphi_{43}(s) + \varphi_{44}(s)\Big] ds$$

Recall that in our model,

$$\varphi_{11}(\cdot) + \varphi_{12}(\cdot) + \varphi_{13}(\cdot) + \varphi_{14}(\cdot)$$

$$= \varphi_{21}(\cdot) + \varphi_{22}(\cdot) + \varphi_{23}(\cdot) + \varphi_{24}(\cdot)$$

$$= \varphi_{31}(\cdot) + \varphi_{32}(\cdot) + \varphi_{33}(\cdot) + \varphi_{34}(\cdot)$$

$$= \varphi_{41}(\cdot) + \varphi_{42}(\cdot) + \varphi_{43}(\cdot) + \varphi_{44}(\cdot)$$

$$= (\beta_1 + \beta_2\beta_3 + 2\beta_2)\varphi(\cdot) = \lambda_1\varphi(\cdot),$$

which is the largest eigenvalue of $\mathbf{\Phi}(\cdot)$. Therefore, the expected number of events in a cluster of our microscopic volume model can be computed as

$$\sum_{k=0}^{\infty} \Big(\int_0^\infty \lambda_1 \varphi(s) ds \Big)^k.$$

If $\int_0^\infty \lambda_1\varphi(s)ds < 1$, this infinite sum converges to

$$\frac{1}{1 - \int_0^\infty \lambda_1\varphi(s)ds}.$$

Note that $\int_0^\infty \lambda_1\varphi(s)ds$ can also be interpreted as the percentage of events in a cluster that are not immigrants. To see this, we have

$$\frac{\sum_{k=1}^\infty \Big(\int_0^\infty \lambda_1\varphi(s)ds \Big)^k}{\sum_{k=0}^\infty \Big(\int_0^\infty \lambda_1\varphi(s)ds \Big)^k} = \frac{\frac{\int_0^\infty \lambda_1\varphi(s)ds}{1-\int_0^\infty \lambda_1\varphi(s)ds}}{\frac{1}{1-\int_0^\infty \lambda_1\varphi(s)ds}} = \int_0^\infty \lambda_1\varphi(s)ds$$

APPENDIX C: WIENER-HOPF EQUATION

Lemma C.1. (Lemma 2 in [16], P277)

Let g be a measurable locally bounded function from $\mathbb{R}$ to $\mathbb{R}^d$ and $\phi : \mathbb{R}_+ \to \mathscr{M}^d(\mathbb{R})$ be a matrix-valued function with integrable components such that $\mathscr{S}(\int_0^\infty \mathbf{\Phi}(s)ds) < 1$. Then there exists a unique locally bounded function from $\mathbb{R}_+$ to $\mathbb{R}^d$ solution of

$$f(t) = g(t) + \int_0^t \mathbf{\Phi}(t-s)f(s)ds, \qquad t \geq 0$$

given by

$$f(t) = g(t) + \int_0^t \mathbf{\Psi}(t-s)g(s)ds, \qquad t \geq 0$$

where $\mathbf{\Psi} = \sum_{k=1}^{\infty}(\mathbf{\Phi})^{*k}$.

APPENDIX D: CONVERGENCE OF RANDOM PROCESSES

Definition D.1. ([28], P347)

Let E be a Polish space and $\mathscr{P}(E)$ be the space of all probability measures on $(E, \mathscr{P})$. A subset A of $\mathscr{P}(E)$ is called tight if for every $\epsilon > 0$ there exists a compact subset K in E such that $\mu(E \backslash K) \leq \epsilon$ for all $\mu \in A$.

Definition D.2. (Definition VI-3.25 [28], P351)

A sequence (X^n) of processes is called C-tight if it is tight, and if all limit points of the sequence $\{\mathscr{L}(x^n)\}$ are laws of continuous processes (i.e., if a subsequence $\{\mathscr{L}(X^{n_k})\}$ converges to a limit point P in $\mathscr{P}(\mathbb{D}(\mathbb{R}^d))$, then P charges only the set $\mathbb{C}(\mathbb{R}^d)$).

Lemma D.1. (Proposition VI-3.26 in [28], P351)

There is equivalence between

(i) The sequence (X^n) is C-tight.

(ii) The sequence (X^n) is tight, and for all $N \in \mathbb{N}^*$, $\epsilon > 0$ we have

$$\lim_n P^n\Big(\sup_{t \leq N} |\Delta X_t^n| > \epsilon\Big) = 0.$$

Lemma D.2. (Proposition VI-4.13 in [28], P358)

We suppose that $X^n - X_0^n$ is a locally square-integrable martingale on $\mathscr{B}^n$ for each n, and we set $G^n = \sum_{j \leq d} \langle X^{n,j}, X^{n,j} \rangle$. Then for the sequence (X^n) to be tight, it is sufficient that

(i) The sequence (X_0^n) is tight in $\mathbb{R}^d$.

(ii) The sequence (G^n) is C-tight (in $\mathbb{D}(\mathbb{R})$).

Lemma D.3. (Theorem VI-6.26 in [28], P384)

Assume that $X^n \xrightarrow{\mathscr{B}} X^\infty$, and that the sequence (X^n) is predictably uniformly tight. Then $(X^n, [X^n, X^n]) \xrightarrow{\mathscr{B}} (X^\infty, [X^\infty, X^\infty])$ in $\mathbb{D}(\mathbb{R}^d \times (\mathbb{R}^d \bigotimes \mathbb{R}^d))$, and in particular, $[X^n, X^n] \xrightarrow{\mathscr{B}} [X^\infty, X^\infty]$.

Lemma D.4. (Corollary IX-1.19 in [28], P527)

Let (M^n) a sequence of local martingales which converges in law to a limit process M, and assume that $|\Delta M^n| \leq b$ identically for some constant b. Then M is a local martingale with respect to the filtration it generates.

Lemma D.5. (Theorem V-3.9 in [43], P203)

Let $M = (M^1, ..., M^d)$ be a continuous vector local Martingale such that $d\langle M^i, M^i\rangle_t \ll dt$ for every i. Then there exists a d-dimensional Brownian Motion B and a $d \times d$ matrix-valued predictable process α in $L^2_{loc}(B)$ such that

$$M_t = M_0 + \int_0^t \alpha_s dB_s$$

APPENDIX E: FRACTIONAL INTEGRALS AND DERIVATIVES

Definition E.1. (A.3 in [16], P277)

The fractional integral of order $r \in (0, 1]$ of a function f is defined by

$$I^r f(t) = \frac{1}{\Gamma(r)} \int_0^t (t-s)^{r-1} f(s)ds$$

Definition E.2. (A.3 in [16], P277)

The fractional derivative of order $r \in (0, 1]$ of a function f is defined by

$$D^r f(t) = \frac{1}{\Gamma(1-r)} \frac{d}{dt} \int_0^t (t-s)^{-r} f(s)ds$$

Lemma E.1. (Corollary A.2 in [30], P2879)

Let ϕ be continuous and ψ such that $x^\mu \psi(x) \in H^\lambda$ with $u, \lambda > 0$. Then for any $\alpha < \min(1-\mu, \lambda)$, $D^\alpha \psi$ exists, belongs to L^r for some $r > 1$ and

$$\int_0^t \phi(t-s)\psi(s)ds = \int_0^t I^\alpha \phi(t-s) D^\alpha \psi(s)ds$$

Lemma E.2. (Proposition A.3 in [30], P2879)

Let f be a differentiable function on $(0, 1]$ such that for some $K > 0$, $0 < \beta < 1$ and any $x \in (0, 1]$,

$$|f(x)| \leq \frac{K}{x^\beta}, \quad \text{and} \quad |f'(x)| \leq \frac{K}{x^{\beta+1}},$$

and g a continuous function on $[0, 1]$. Then the convolution

$$f * g(t) = \int_0^t f(t-s)g(s)ds$$

has Hölder regularity $(1-\beta)$.

APPENDIX F: MITTAG-LEFFLER FUNCTIONS

Definition F.1. (A.4 in [16], P277)

Let $(\alpha, \beta) \in (R_+^*)^2$. The Mittag-Leffler function $E_{\alpha,\beta}$ is defined for $z \in \mathbb{C}$ by

$$E_{\alpha,\beta}(z) = \sum_{n \geq 0} \frac{z^n}{\Gamma(\alpha n + \beta)}$$

Definition F.2. (A.4 in [16], P277)

For $(\alpha, \lambda) \in (0, 1) \times \mathbb{R}_+$, the Mittag-Leffler density function $f^{\alpha,\lambda}$ is defined by

$$f^{\alpha,\lambda}(t) = \lambda t^{\alpha-1} E_{\alpha,\alpha}(-\lambda t^\alpha), \qquad t > 0.$$

Also, let

$$F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(s) ds, \qquad t \geq 0$$

Lemma F.1. (A.4 in [16], P278)

For $\alpha \in (\frac{1}{2}, 1)$, $f^{\alpha,\lambda}$ is square-integrable and its Laplace transform is given for $z \geq 0$ by

$$\hat{f}^{\alpha,\lambda}(z) = \int_0^\infty f^{\alpha,\lambda}(s) e^{-as} ds = \frac{\lambda}{\lambda + z^\alpha}$$

Lemma F.2. (A.1 in [17], P37)

Below are some properties of $f^{\alpha,\lambda}$:

1. $I^{1-\alpha} f^{\alpha,\lambda}(t) = \lambda(1 - F^{\alpha,\lambda}(t))$
2. $f^{\alpha,\lambda}(t) \underset{t \to 0^+}{\sim} \dfrac{\lambda}{\Gamma(\alpha)} t^{\alpha-1}, \qquad f^{\alpha,\lambda}(t) \underset{t \to \infty}{\sim} \dfrac{\alpha}{\lambda \Gamma(1-\alpha)} t^{-(\alpha+1)}$
3. $F^{\alpha,\lambda}(t) = 1 - E_{\alpha,1}(-\lambda t^\alpha), \qquad F^{\alpha,\lambda}(t) \underset{t \to 0^+}{\sim} \dfrac{\lambda}{\Gamma(\alpha+1)} t^\alpha, \qquad 1 - F^{\alpha,\lambda}(t) \underset{t \to \infty}{\sim} \dfrac{1}{\lambda \Gamma(1-\alpha)} t^{-\alpha}$

Lemma F.3. (Proposition 3.1 in [30], P2868)

$f^{\alpha,\lambda}$ is C^∞ on $(0,1]$ and

$$f^{\alpha,\lambda}(x) \underset{x\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha)} x^{\alpha-1},$$

$$(f^{\alpha,\lambda})'(x) \underset{x\to 0^+}{\sim} \frac{\lambda(\alpha-1)}{\Gamma(\alpha)} x^{\alpha-2}.$$

Furthermore, $f^{\alpha,\lambda}x^{1-\alpha}$ has Hölder regularity α on $(0,1]$.

For $\nu < \alpha$, $f^{\alpha,\lambda}$ is ν fractionally differentiable and

$$D^\nu f^\alpha(x) = \lambda x^{\alpha-1-\nu} E_{\alpha,\alpha-\nu}(-\lambda x^\alpha)$$

Therefore,

$$D^\nu f^\alpha(x) \underset{x\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha-\nu)} \frac{1}{x^{1-\alpha+\nu}}$$

and

$$(D^\nu f^\alpha)'(x) \underset{x\to 0^+}{\sim} \frac{\lambda(\alpha-1-\nu)}{\Gamma(\alpha-\nu)} \frac{1}{x^{2-\alpha+\nu}}.$$

For $\nu' > 0$, f^α is ν' fractionally integrable and

$$I^{\nu'} f^\alpha(x) = \lambda \frac{1}{x^{1-\alpha-\nu'}} E_{\alpha,\alpha+\nu'}(-\lambda x^\alpha).$$

Therefore,

$$I^{\nu'} f^\alpha(x) \underset{x\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha+\nu')} \frac{1}{x^{1-\alpha-\nu'}}$$

and for $\alpha + \nu' \neq 1$,

$$(I^{\nu'} f^\alpha)'(x) \underset{x\to 0^+}{\sim} \frac{\lambda(\alpha-1+\nu')}{\Gamma(\alpha+\nu')} \frac{1}{x^{2-\alpha-\nu'}}$$

APPENDIX G: AFFINE VOLTERRA PROCESS

Consider the following d-dimenstional stochastic Volterra equation:

$$\mathscr{Y}(t) = \mathscr{Y}_0 + \int_0^t \mathscr{K}(t-s)b(\mathscr{Y}(s))ds + \int_0^t \mathscr{K}(t-s)\sigma(\mathscr{Y}(s))d\mathscr{B}(s), \tag{G.1}$$

where $\mathscr{K} \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$, initial condition $\mathscr{Y}_0 \in \mathbb{R}^d$, the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, and $\mathscr{B}$ an m-dimensional Brownian motion. The following results are from [27] and [20] for the situation when $a(y) := \sigma(y)\sigma(y)^\top$ and $b(y)$ are affine of the form

$$a(y) = A^0 + y_1 A^1 + \cdots + y_d A^d$$

$$b(y) = b^0 + y_1 b^1 + \cdots + y_d b^d$$

for some d-dimensional symmetric matrix A^i and vectors b^i. Let $B = (b^1 \cdots b^d)$ be a $d \times d$ matrix and

$$A(u) = (uA^1u^\top, \cdots, uA^du^\top)$$

a row vector for any row vector $u \in (\mathbb{C}^d)^*$.

Most of the following results require the same conditions on $\mathscr{K}$. We list the conditions below to avoid repetition.

$$\begin{gathered}\mathscr{K} \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}) \text{ and there is } \gamma \in (0,2] \text{ such that } \int_0^h \mathscr{K}(t)^2 dt = \mathcal{O}(h^\gamma) \\ \text{and } \int_0^T (\mathscr{K}(t+h) - \mathscr{K}(t))^2 dt = \mathcal{O}(h^\gamma) \text{ for every } T < \infty.\end{gathered} \tag{G.2}$$

$$\begin{aligned}&\text{The shifted kernel } \Delta_h\mathscr{K}(t) := \mathscr{K}(t+h) \text{ is nonnegative, not identically zero,}\\ &\text{non-increasing and continuous on } (0,\infty), \text{ and its resolvent of the}\\ &\text{first kind } \mathscr{L} \text{ is nonnegative and non-increasing in that}\\ &s \mapsto \mathscr{L}([s, s+t]) \text{ is non-increasing in all } t \geq 0.\end{aligned} \tag{G.3}$$

Lemma G.1. (Theorem 5.5.4 in [20], P159)

Let $\mathscr{K} \in L^1_{\text{loc}}(\mathbb{R}_+, \mathbb{C}^{d\times d})$ be completely monotone on $(0, \infty)$, and suppose that $\langle v, \mathscr{K}(t)v\rangle > 0$ for some $t > 0$ and all nonzero vectors $v \in \mathbb{C}^d$. Then $\mathscr{K}$ has a resolvent of the first kind. This resolvent is the sum of a point mass at zero and a completely monotone function. The point mass at zero is invertible iff $\limsup_{t\downarrow 0} |\mathscr{K}|(t) < \inf$, and it is absent iff $\limsup_{t\downarrow 0}\langle v, \mathscr{K}(t)v\rangle = \inf$ for all nonzero vectors $v \in \mathbb{C}^d$.

Lemma G.2. (Theorem 6.1 in [27], P3181)

Consider the d-dimensional stochastic Volterra equation (G.1). If the following conditions hold

(i) $\sigma(y) = C_\sigma\sqrt{y}$ with $C_\sigma > 0$.

(ii) $\mathscr{K}$ satisfies (G.2) and (G.3).

(iii) $b^0 \in \mathbb{R}^d_+$, and $B_{ij} \geq 0$ and $i \neq j$.

Then the stochastic Volterra equation (G.1) has a unique in law $\mathbb{R}^d_+$-valued continuous weak solution $\mathscr{Y}$ for any initial condition $\mathscr{Y}_0 \in \mathbb{R}^d$. For each i, the paths of $\mathscr{Y}_i$ are Hölder continuous of any order less than $\gamma_i/2$, where γ_i is the constant associated with $\mathscr{K}_i$ in (G.2).

Lemma G.3. (Lemma 3.1 in [27], P3165)

Consider the d-dimensional stochastic Volterra equation (G.1). Assume b and σ are continuous and satisfy the linear growth condition

$$|b(y)| \vee |\sigma(y)| \leq C_{LG}(1+|y|), \qquad y \in \mathbb{R}^d,$$

for some constant C_{LG}. Let $\mathscr{Y}$ be a continuous solution of (1.1) with initial condition $\mathscr{Y}_0 \in \mathbb{R}^d$. Then for any $p \geq 2$ and $T < \infty$, one has

$$\sup_{t\leq T} \mathbb{E}[|\mathscr{Y}(t)|^p] \leq c$$

for some constant c that only depends on $|\mathscr{Y}(0)|$, $\mathscr{K}|_{[0,T]}$, C_{LG}, p and T.

Lemma G.4. (Lemma 2.4 in [27], P3161)

Assume $\mathscr{K}$ satisfies (G.2) and consider a process $\mathscr{Y} = \mathscr{K} * (bdt + dM)$, where b is an adapted process and M is a continuous local Martingale with $\langle M\rangle_t = \int_0^t a(s)ds$ for some adapted process a. Let $T \geq 0$, and $p > \max\{2, 2/\gamma\}$ be such that $\sup_{t\leq T} \mathbb{E}[|a(t)|^{p/2} + |b(t)|^p]$ is finite. Then $\mathscr{Y}$ admits a version which is Hölder continuous on $[0,T]$ of any order $\alpha < \gamma/2 - 1/p$. Denoting this version again by $\mathscr{Y}$, one has

$$\mathbb{E}\Big[\Big(\sup_{0\leq s<t\leq T} \frac{|\mathscr{Y}(t)-\mathscr{Y}(s)|}{|t-s|^\alpha}\Big)^p\Big] \leq c \sup_{t\leq T} \mathbb{E}[|a(t)|^{p/2} + |b(t)|^p]$$

for all $\alpha \in [0, \gamma/2 - 1/p)$, where c is a constant that only depends on $p, \mathscr{K}, T$. As a consequence, if a and b are locally bounded, then $\mathscr{Y}$ admits a version which is Hölder continuous for any order $\alpha < \gamma/2$.

LIST OF REFERENCES

[1] R. Almgren and N. Chriss, *Optimal execution of portfolio transactions,* The journal of risk, 3 (2)(2001), 5-39.

[2] A. Andersen, R. Cont, and E. Vinkovskaya, *A point process model for the high-frequency dynamics of a limit order book,* Ph.D. thesis, Columbia University, 2014.

[3] E. Bacry, k. Dayri, and J. F. Muzy, *Non-parametric kernel estimation for symmetric hawkes processes. Application to high frequency financial data.* European Physics Journal B, 85(5)(2012), 157–172.

[4] E. Bacry, I. Mastromatteo, J. F. Muzy, *Hawkes Process in Finance,* Market microstructure and liquidity, 1(1)(2015), Article No. 1550005.

[5] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy, *Some limit theorems for Hawkes processes and application to financial statistics,* Stochastic Processes and their Applications, 123(7)(2013), 2475–2499.

[6] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy, *Modelling microstructure noise with mutually exciting point processes,* Quantitative Finance, 13(2013), 65–77.

[7] E. Bacry, T. Jaisson, and J. Muzy *Estimation of slowly decreasing Hawkes kernels: Application to high frequency order book modelling,* Quantitative Finance, 16(8)(2016), 1179-1201.

[8] J. P. Bouchaud, M. Mezard, and M. Potters, *Statistical properties of stock order books: empirical results and models,* Quantitative finance, 2(4)(2004), p.251-256.

[9] P. Bremaud and L. Massoulie, *Stability of Nonlinear Hawkes Process,* The Annals of Probability, 24(3)(1996), 1563–1588.

[10] L. Caestensen, *Hawkes processes and combinatorial transcriptional regulation. Ph.D. thesis,* University of Copenhagen, 2010.

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

[12] R. Cont, A. Kukanov, and S. Stoikov, *The price impact of order book events, J. Financial Econometrics,* 12(2014), 47–88, doi:10.1093/jjfinec/nbt003.

[13] R. Cont and M. S. Müeller, *A stochastic PDE model for limit order book dynamics, SIAM Journal on Financial Mathematics,* 12(2)(2021), 744–787.

[14] G. Da Prato, A Jentzen, and M Röckner *A mild Itô formula for SPDEs, Transactions of the American Mathematical Society,* 372 (2019), 3755-3807.

[15] L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces,* 3rd edn. (Elsevier, 2005)

[16] O. E. Euch, M. Fukasawa, and M.Rosenbaum, *The microstructural foundations of leverage effect and rough volatility, Finance and Stochastics,* 22(2)(2018), 243–280.

[17] O. E. Euch, and M.Rosenbaum, *The characteristic function of rough Heston models, Mathematical Finance,* 29(1)(2018), 3–38.

[18] T. Foucault, O. Kadan, and E. Kandel, *Limit Order Book as a Market for Liquidity, The Review of financial studies,* 18 (4)(2005), 1171-1217.

[19] J. Gatheral, T. Jaisson, and M. Rosenbaum. *Volatility is rough, Quantitative Finance,* 18(6)(2018), 933–949.

[20] G. Gripenberg, S. Londen, O. Staffans *Volterra integral and functional equations, volume 34 of Encyclopedia of Mathematics and its Applications,* Cambridge University Press, Cambridge, 1990. ISBN 0-521-37289-5.

[21] S. J. Hardiman, N. Bercot, and J. P. Bouchaud, *Critical reflexivity in financial markets: a Hawkes process analysis.* The European Physical Journal B, 86(10)(2014), 1—9.

[22] A. G. Hawkes, *Spectra of Some Self-Exciting and Mutually Exciting Point Processes,* Biometrika, 58(1)(1971), 83-90.

[23] A. G. Hawkes and D. Oakes, *A cluster process representation of a self-exciting process,* Journal of Applied Mathematics, 11(1971), 493–503.

[24] A. G. Hawkes and D. Oakes, *A Cluster Process Representation of a Self-Exciting Process,* Journal of Applied Probability, 11(3)(1974), 493–503.

[25] H. J. Haubold, A. M. Mathai, and R. K. Saxena, *Mittag-Lefller Functions and Their Applications,* Journal of Applied Mathematics, 2011.

[26] S. L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options,* Review of Financial Studies, 6(2)(1993), 327-–343.

[27] E.A. Jaber, M. Larsson, S. Pulido *Affine Volterra processes,* Annals of Applied Probability, 29 (5)(2019), 3155–3200.

[28] J. Jacod, A. N. Shiryaev, *Limit Theorems for Stochastic Processes, vol. 288*, Springer-Verlag, Berlin, 1987.

[29] T. Jaisson and M. Rosenbaum, *Limit theorems for nearly unstable Hawkes processes.* The Annals of Applied Probability, 25(2)(2015), 600–631.

[30] T. Jaisson and M. Rosenbaum, *Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes,* The Annals of Applied Probability, 26(5)(2015), 2860–2882.

[31] D. G. Kendall, *Stochastic Processes and Population Growth.* J.R.Statist.Soc., 11(1949), 230–264.

[32] A. Kirilenko, A. Kyle, M. Samadi, T. Tuzun, *The Flash Crash: The Impact of High Frequency Trading on an Electronic Market,* Journal of Finance, 72(3)(2014), 967-998.

[33] P. Laub, T. Taimre, and P. Pollet, *Hawkes Processes,* arxiv.org/abs/1507.02822, 2015.

[34] P. J. Laub, Y. Lee, and T. Taimre, *The Elements of Hawkes Processes,* 1st edn. (Springer, 2021)

[35] C. A. Lehalle and S. Lareuelle, *Market microstructure in practice,* World Scientific, 2013.

[36] X. Lu and F. Abergel, *High dimensional Hawkes processes for limit order books Modelling, empirical analysis and numerical calibration* Quantitative Finance, 18(2)(2018), 249-264.

[37] V. Manahov, *High-frequency trading order cancellations and market quality: Is stricter regulation the answer?* International journal of finance and economics, 26 (4)(2021), p.5385-5407.

[38] B. B. Mandelbrot and J. W. Van Ness, *Fractional brownian motions, fractional noises and applications.* SIAM Review, 10(4)(1968), p.422–437.

[39] A. J. Menkveld, *High frequency trading and the new market makers,* Journal of financial markets, 16 (4)(2013), p.712-740.

[40] C. Parlour, *Price Dynamics in Limit Order Markets,* The Review of financial studies, 11 (4)(1998), 789-816.

[41] P. Protter, Q. Wu, and S. Yang, *Order Book Queue Hawkes-Markovian Modeling,* arxiv.org/pdf/2107.09629.pdf, 2021.

[42] J. G. Rasmussen, *Applied Probability of Queues,* 2nd edn. Applications of Mathematics: Stochastic Modelling and Applied Probability (Springer, 2003).

[43] D. Revus and M. Yor *Continuous martingales and Brownian motion, volume 293,* Springer Science & Business Media, 2013.

[44] M. A. Rizoiu, Y. Lee, S. Mishra, and L. Xie *A Tutorial on Hawkes Processes for Events in Social Media,* *arxiv.org/abs/1708.06401*, 2017.

[45] S. M. Ross, *Introduction to Probability Models,* 11th edn. (Elsevier, 2014).

[46] I. Rosu, *A Dynamic Model of the Limit Order Book,* Review of Financial Studies, 22 (11)(2009), p.4601-4641.

[47] E. Said, A. B. H. Ayed, A. Husson, F. Abergel. *Market Impact: A Systematic Study of Limit Orders,* Market microstructure and liquidity, 3 (3-4)(2018), pp.1850008.

[48] G. Samorodnitsky, *Long Range Dependence,* Foundations and Trends in Stochastic Systems, 1(3)(2006), 163–257.

[49] A. Subrahmanyam and H. Zheng, *Limit order placement by high-frequency traders,* Borsa Istanbul Review, 16 (4)(2016), p.185-209.

[50] G. Warfield, *The Investors' Guide to Stock Quotations and Other Financial Listings, Harper & Row,* 1983.

[51] M. Westcott, *A Note on the Non-Homogeneous Poisson Cluster Process,* Journal of Applied Probability, 14(2)(1977), 396–398.