Three Stochastic Models for Order Book Dynamics

Qi Gong

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2013

Approved by:

Vidyadhar Kulkarni

Nilay Argon

Amarjit Budhiraja

Shankar Bhamidi

Chuanshu Ji

ABSTRACT

Qi Gong: Three Stochastic Models for Order Book Dynamics
(Under the direction of Vidyadhar Kulkarni)

In this dissertation, we study three stochastic models for order book dynamics. We first consider a double-ended queue with renewal arrivals where buyers and sellers arrive to conduct trades. Because the analytical results of the limiting behavior of the double ended queue is intractable except in very special cases, we apply the diffusion approximation method. We find that the queue length process of the double-ended queue converges to an asymmetric Ornstein-Uhlenbeck process with drift. We use simulation to evaluate the goodness of our approximations.

Next we consider the double-ended queue where the arrival processes of the buyers and sellers are state-dependent. We assume that traders arrive at the queue according to a phase-type renewal process (PH-renewal process), and an arriving trader is a buyer or a seller according to state-dependent probabilities. We derive an explicit algorithm to compute the limiting distribution of this double-ended queue. We study two special cases with Erlang and Hyper-exponential inter-arrival times. The goodness of the algorithm is validated by simulation.

Finally, we study a stochastic model of an order book describing the movement of the market ask and market bid prices. As soon as the market bid price matches the market ask price, a trade occurs. Consequently the market ask and bid prices separate and start a new movement. We use the moment estimation method to estimate the parameters of the model, and apply this model to the real data. One application of this model is forecasting, in which the performance is validated by numerical examples.

ACKNOWLEDGEMENTS

This is a great opportunity to express my gratitude to a number of individuals who have helped me in achieving the task of completing this dissertation.

TABLE OF CONTENTS

CHAPTER 1

Introduction

The main trading problem in the financial market is matching compatible traders. Traders offer bid prices to buy and ask prices to sell. The orders are maintained in an order book and executed by using an execution system. The highest bid price on the order book is called the market bid price, and lowest ask price is called the market ask price. When the market ask price matches the market bid price, a trade occurs, and the market ask and bid prices change to reflect this trade.

Traditional studies of this mechanisms were based on quote-driven markets, where a market maker or dealer centralizes the orders. However, with the development of Electronic Communications Networks (ECN), more and more market participants use the alternative order-driven trading systems. In these electronic platforms, the information of all outstanding limit orders in a limit order book is available to the market participants and the market orders are executed against the market prices. Currently many exchanges such as the NYSE, Nasdaq, the Tokyo Stock Exchange and the London Stock Exchange provide electronic order-driven platforms. Because of the popularity of ECN and the availability of data, stochastic models are increasingly used to model the order-driven markets.

The dynamic models of the order book have been studied extensively in the microstructure literature. Cohen (1981)[6] and Parlour (1998)[26] studied a dynamic programming model that describe the behavior of single trader with multiple period. Foucault (1999) [14] modified the model with non-fixed ask and bid price. Besides, many researchers used queueing models to describe the dynamics of limit order book. Garman (1976)[15] modeled the limit order book as CTMC (Continuous Time Markov Chain) and derived its balance equations. The CTMC is positive recurrent, however the limiting behavior is hard to find. Cont (2010) [10] considered the arrival rate of orders dependent on current market price, which make the analysis more intractable. Cont (2011) [8] proposed a model for the dynamics of a limit order book in a liquid market where buy and sell orders are submitted at high frequency. Subsequently Cont (2013) [9] derived a stochastic model for the dynamics of a limit order book including arrivals of market orders, limit orders, and order cancelations. Luckock (2003)[24] constructed a model similar to Garman's model and studied the limiting distribution of order book. Compared with two-sided model of order book, one-sided model is more analytically tractable, see Kleinrock (1967)[21], Seppi (1997)[32] and Rocsu (2009)[31]. However, one-sided models lack applicability.

At the level of applications, number of models provide a quantitative framework to help market participants optimize their trade execution strategies, such as Domowitz (1994)[12], Bertsimas (1998)[2], Alfonsi (2007)[1]. Besides, some researchers provide statistical view and empirical methods to study the order book, such as Bouchaud (2002)[4], Hollifield (2004)[16] and Farmer (2004)[13].

In the study of the order book dynamics, we aim to consider a limit order book with finite number of trading prices which is similar to Garman's model. Under this situation we find that the order book dynamic forms a multidimensional double-sided queue with dynamic priorities. Although this queue can be described by a CTMC (continuous time Markov chain), its analytical exploration is intractable. Therefore, we reduce the dimension of the queue to be one. Thus this queueing process becomes a one-dimensional double-ended queue with two arrival streams–sellers and buyers.

When the arrival process of the double-ended queue is a Poisson process, there are many existing methods to deal with the analysis of the queueing process. In this thesis, we consider the arrival processes of the double-ended queue to be general renewal processes. In chapter 2, we consider a double-ended queueing system with two independent renewal arrival streams. Whenever there is a pair of customers from both arrival streams, they immediately depart together, and so there cannot exist nonzero customers from both arrival streams simultaneously. We further assume that traders are impatient, that is, if they do not see a matching trader within a trader-specific random time (called the trader's patience time) they leave without completing the trade. By applying the diffusion approximation method, we find that the queue length process of the double-ended queue converges to an asymmetric Ornstein-Uhlenbeck process with drift. The goodness of the approximation is tested by numerical examples.

In chapter 3, we study this double-ended queue with state-dependent arrival mechanism. This extra assumption makes the double-ended queue more realistic. We assume the arrival process of traders is a phase-type renewal process (that is, the inter-arrival time follows a phase-type distribution), and an arriving trader is a buyer or a seller with state-dependent probabilities. We derive an algorithm for the limiting distribution of this double-ended queue, and obtain the explicit formula of limiting distribution for two special inter-arrival distribution–Erlang distribution and hyper-exponential distribution. We also derive several performance measures and analyze the goodness of our method through numerical examples.

In chapter 4, we build a stochastic model focusing on the market ask price and the market bid price. Because of the new arrivals and reneging (that is, traders leaving the market without trading) of traders, the market bid and ask prices can move upwards and downwards. Immediately after the trade the market bid price moves down the the new largest bid price, while the market ask price moves up to the new smallest ask price. We use two independent geometric Brownian motions (GBM) to describe the movement of the market ask and bid prices. From the model we study the inter-trading times and the trading prices, and obtain the explicit estimators of each parameter of our model. Finally we derive a simple forecasting formula by applying this model. The performance of the forecasting is validated through numerical examples.

We present relevant literature review in each chapter separately.

CHAPTER 2

The Diffusion Model for a Double-ended Queue with Renewal Arrival Processes

2.1 Introduction

We consider a double-ended queueing system which consists of two independent renewal arrival streams. Whenever there is a pair of customers from both arrival streams, they immediately depart together, and so there cannot exist nonzero customers from both arrival streams simultaneously (see Figure 2.1). We further assume that traders are impatient, that is, if they do not see a matching trader within a trader-specific random time (called the trader's patience time) they leave without completing the trade.

Such double-ended queues arise in many applications, such as taxi-service system, buyers and sellers in a common market, assembly systems, organ transplant systems, to name a few. The first work on *double-ended queue* was by Kashyap [19] for a taxi service example. Kashyap considers the taxi queueing system as a double-ended queue with limited waiting space. Under the assumptions that arrival processes of taxies and passengers are Poisson processes, he derives the analytical results about the steady state distribution of the system state. Conolly et al. [7] study the effect of impatient behavior primarily in the context of double-ended queues under the assumption of Poisson arrivals and exponential patience times. Researchers also find many other practical applications of the double-ended queues, such as networks with synchronization nodes (Prabhakar et al. [28]), and perishable inventory system (Perry et al. [27]). When renewal arrivals are considered, the explicit form of the stationary distribution becomes intractable. Degirmenci [11] studies the asymptotic behavior of the stationary distribution of the double-ended queue using algebraic approximation methods. Several researchers study the double-ended queue using simulation methods, see Zenios [34] and Kim et al. [20].

The rest of the this chapter is organized as follows. In the next section we present the model of the double-ended queue with renewal arrivals and introduce the relevant notation. In Section 2.3 we collect the results about the special case when the renewal processes are Poisson. Some of these results are known, while some are new. We use these results to approximate the renewal case by replacing the renewal arrivals by Poisson arrivals in Section 2.5. In Section 2.4 we study the fluid and diffusion approximations for the queue length process. (See Kushner [23] for comprehensive references about diffusion approximations). Under suitable conditions (Assumptions 2.4.1 and 2.4.2), the fluid limit satisfies the ODE (2.4.6), and the diffusion limit is given by the SDE (2.4.22). We provide the exact solution to the ODE, and study the moments and stationary distribution of the SDE. We also remark on the connections between the fluid and diffusion approximations and the special case in Section 2.3.

Finally we study a numerical example in Section 2.5, and compare goodness of the two approximations: the Poisson approximation, and the diffusion approximation. We make comments on extensions of the model in Section 2.6.

We use the following notation. Denote by $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{N}$, and $\mathbb{Z}$ the sets of real numbers, nonnegative real numbers, integers, and positive integers, respectively. For a real number a, define $a^+ = \max\{a, 0\}$ and $a^- = \max\{0, -a\}$. Similarly, for a real function defined on $[0, \infty)$, define for $t \in [0, \infty)$, $f^+(t) = \max\{0, f(t)\}$ and $f^-(t) = \max\{0, -f(t)\}$. Let a and b be two nonnegative real numbers. We use $a \gg b$ to denote that a is much larger than b. For $c \in \mathbb{R}_+$, denote by $\lfloor c \rfloor$ the largest integer less than or equal to c. Denote by $\mathcal{C}_0^2(\mathbb{R})$ and $\mathcal{D}([0, \infty), \mathbb{R})$

the space of twice differentiable real-valued functions with compact support and the space of right continuous functions with left limits defined from $[0,\infty)$ to $\mathbb{R}$ with the usual Skorohod topology. For semimartingales $X, Y \in \mathcal{D}([0,\infty) : \mathbb{R})$, we denote by $[X,Y]$ the quadratic covariation of X and Y. For $x \in \mathcal{D}([0,\infty),\mathbb{R})$, for $t \in [0,\infty)$, let

$$\|x\|_t = \sup_{s\in[0,t]} |x(s)|.$$

A mapping $F : \mathcal{D}([0,\infty),\mathbb{R}) \to \mathcal{D}([0,\infty),\mathbb{R})$ is called Lipschitz continuous if for any $t \in [0,\infty)$, there exists $\kappa \in (0,\infty)$ such that for $x_1, x_2 \in \mathcal{D}([0,\infty),\mathbb{R})$,

$$\|F(x_1) - F(x_2)\|_t \leq \kappa \|x_1 - x_2\|_t.$$

Finally, normal distribution with mean μ and variance σ^2 is denoted by $N(\mu,\sigma^2)$, and its density and distribution function are denoted by by $\phi(\cdot;\mu,\sigma^2)$ and $\Phi(\cdot;\mu,\sigma^2)$, respectively.

2.2 Model Formulation

Consider a trading market where the sellers and buyers arrive according to independent renewal processes. When a seller is matched with a buyer, a trade occurs and they both leave the queue. The trading follows according to first-come-first-served principle. If an arriving seller (buyer) cannot be matched with a buyer (seller), he/she will stay in the queue and wait for the upcoming buyers (sellers). Thus there cannot be non-zero number of buyers and sellers simultaneously in the system. We also assume that each seller (buyer) can leave the queue without trading because of impatience. The patience time of each seller (buyer) follows an exponential distribution with rate θ (γ). The patience times of the buyers and sellers are independent of each other. A queueing system forms a double-ended queue is schematically shown in Figure 2.1. Let $X(t)$ be the length of the double-ended queue at time t. We note that $X(t) \in \mathbb{Z}$. If $X(t) > 0$, there are $X(t)$ sellers waiting in the queue, and if $X(t) < 0$, there are $-X(t)$ buyers waiting in the queue.

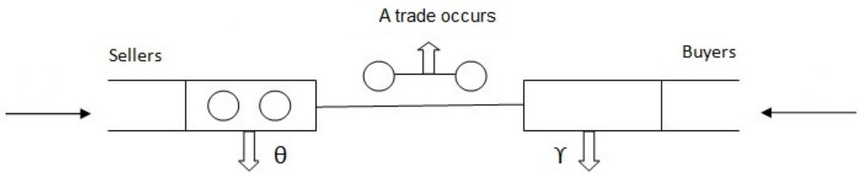


Figure 2.1: Double-ended Queue

Let $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$ be a filtered probability space satisfying the usual conditions. All the random variables and stochastic processes are assumed to be defined on this space. We assume the inter-arrival times for sellers and buyers are independent sequences of i.i.d. random variables $\{U(k) : k \in \mathbb{N}\}$ and $\{V(k) : k \in \mathbb{N}\}$, respectively. The random variable $U(1)$ has mean $1/\alpha$ and standard deviation σ, and $V(1)$ has mean $1/\beta$ and standard deviation ς. Define

$$N_s(t) = \max\left\{k : \sum_{i=1}^{k} U(i) \leq t\right\},$$
$$N_b(t) = \max\left\{k : \sum_{i=1}^{k} V(i) \leq t\right\}.$$

The renewal processes N_s and N_b can be interpreted as the arrival processes for buyers and seller, respectively. As mentioned before, the patience time of each seller (buyer) follows an exponential distribution with rate θ (γ). Let N_{sr} and N_{br} be two independent unit-rate Poisson processes, which are independent of N_s and N_b. Then we have the following evolution equation for $X(t)$: For $t \geq 0$,

$$X(t) = X(0) + N_s(t) - N_b(t) - N_{sr}\left(\theta \int_0^t X^+(s)ds\right) + N_{br}\left(\gamma \int_0^t X^-(s)ds\right). \qquad (2.2.1)$$

where $X(0)$ denotes the initial number of sellers or buyers in the system, which is assumed to be independent of N_s, N_b, N_{sr} and N_{br}.

2.3 Special Case: Poisson Arrivals

When the arrival processes are Poisson, it is easy to see that $\{X(t), t \geq 0\}$ is a birth and death process on $\mathbb{Z}$ with birth parameters $\lambda_i = \alpha + i^-\gamma$ and death parameters $\mu_i = \beta + i^+\theta$ for $i \in \mathbb{Z}$. Using the standard theory (see Kulkarni [22]), we see that this birth and death process is:

- positive recurrent, if $\theta > 0$ and $\gamma > 0$;
- null recurrent, if $\theta = \gamma = 0$ and $\alpha = \beta$;
- transient, if $\theta = \gamma = 0$ and $\alpha \neq \beta$.

In the analysis, we assume $\theta > 0$ and $\gamma > 0$. Therefore, this CTMC has a unique limiting distribution. Let $\{\pi_i,\ i \in \mathbb{Z}\}$ denote the limiting distribution of X. Using the standard theory of birth and death processes we see that the limiting distribution is given by:

$$\pi_i = \frac{\alpha^i}{\prod_{j=1}^{i} (\beta + j\theta)} \pi_0, \quad i = 1, 2, \cdots \tag{2.3.1}$$

$$\pi_{-i} = \frac{\beta^i}{\prod_{j=1}^{i} (\alpha + j\gamma)} \pi_0, \quad i = 1, 2, \cdots \tag{2.3.2}$$

$$\pi_0 = \left(1 + \sum_{i=1}^{\infty} \frac{\alpha^i}{\prod_{j=1}^{i} (\beta + j\theta)} + \sum_{i=1}^{\infty} \frac{\beta^i}{\prod_{j=1}^{i} (\alpha + j\gamma)} \right)^{-1}. \tag{2.3.3}$$

Now we consider the first two moments of $X(t)$. Define

$$\begin{aligned} m(t) &= \mathbb{E}(X(t)), \\ m_+(t) &= \mathbb{E}(X^+(t)), \\ m_-(t) &= \mathbb{E}(X^-(t)), \\ s(t) &= \mathbb{E}(X(t)^2), \\ s_+(t) &= \mathbb{E}(X^+(t)^2), \\ s_-(t) &= \mathbb{E}(X^-(t)^2). \end{aligned}$$

Clearly, $m(t) = m_+(t) - m_-(t)$ and $s(t) = s_+(t) + s_-(t)$. We also note that in general $m_+(t) \neq m^+(t),\ m_-(t) \neq m^-(t)$, and so $|m(t)| \neq m_+(t) + m_-(t)$. The next theorem gives our main result in this section.

Theorem 2.3.1. *Assume that $X(0)$ has finite first two moments. Then the moment functions $m(t)$ and $s(t)$ satisfy the following differential equations. For $t \geq 0$,*

$$\frac{dm(t)}{dt} = (\alpha - \beta) - \theta m_+(t) + \gamma m_-(t), \tag{2.3.4}$$

and

$$\frac{ds(t)}{dt} = -2\theta s_+(t) - 2\gamma s_-(t) + (2\alpha - 2\beta + \theta)m_+(t) + (-2\alpha + 2\beta + \gamma)m_-(t) + \alpha + \beta, \tag{2.3.5}$$

with initial conditions $m(0) = \mathbb{E}(X(0))$ and $s(0) = \mathbb{E}(X(0)^2)$.

Proof. We first consider $m(t) = \mathbb{E}(X(t))$. Taking expectation of equation (2.2.1), we get

$$m(t) = m(0) + \alpha t - \beta t - \theta \int_0^t m_+(s)ds + \gamma \int_0^t m_-(s)ds.$$

Taking derivative on both sides of above equation we get equation (2.3.4).

Next we consider the second moment of $X(t)$. Using the infinitesimal analysis, for a small $h > 0$, we get,

$$X(t+h)^2 = \begin{cases} (X(t)+1)^2, & w.p. \quad (\alpha + \gamma X^-(t))\, h + o(h) \\ X(t)^2, & w.p. \quad (1 - \alpha - \gamma X^-(t) - \beta - \theta X^+(t))\, h + o(h) \\ (X(t)-1)^2, & w.p. \quad (\beta + \theta X^+(t))\, h + o(h). \end{cases}$$

Therefore,

$$\begin{aligned} \mathbb{E}(X(t+h)^2 | X(t)) &= X(t)^2 + 2X(t)\left(\alpha - \beta - \theta X^+(t) + \gamma X^-(t)\right) h \\ &\quad + \left(\alpha + \beta + \theta X^+(t) + \gamma X^-(t)\right) h + o(h). \end{aligned}$$

Since $X(t) = X^+(t) - X^-(t)$, we have

$$\begin{aligned} \mathbb{E}(X(t+h)^2 | X(t)) &= X(t)^2 - 2\theta X^+(t)^2 h - 2\gamma X^-(t)^2 h + (2\alpha - 2\beta + \theta) X^+(t) h \\ &\quad + (-2\alpha + 2\beta + \gamma) X^-(t) h + \alpha h + \beta h + o(h). \end{aligned}$$

Taking expectation on both sides of above equation, we get

$$\begin{aligned} \frac{s(t+h) - s(t)}{h} &= -2\theta s_+(t) - 2\gamma s_-(t) + (2\alpha - 2\beta + \theta) m_+(t) \\ &\quad + (-2\alpha + 2\beta + \gamma) m_-(t) + \alpha + \beta + \frac{o(h)}{h}. \end{aligned}$$

Taking limit $h \to 0$, the equation (2.3.5) follows. ■

Remark 2.3.1. If $\theta = \gamma$, we can further simplify (2.3.4) as follows:

$$\frac{dm(t)}{dt} = (\alpha - \beta) - \theta m(t), \text{ and } m(0) = \mathbb{E}(X(0)) \tag{2.3.6}$$

which immediately yields

$$m(t) = \left(m(0) - \frac{\alpha - \beta}{\theta}\right) e^{-\theta t} + \frac{\alpha - \beta}{\theta}. \tag{2.3.7}$$

Also (2.3.5) can be simplified and we have that

$$\begin{aligned}\frac{ds(t)}{dt} &= -2\theta s(t) + (2\alpha - 2\beta)m(t) + \theta\left(m_+(t) + m_-(t)\right) + \alpha + \beta \\ &\geq -2\theta s(t) + (2\alpha - 2\beta)m(t) + \theta|m(t)| + \alpha + \beta.\end{aligned}$$

Consider the following ODE

$$\frac{d\tilde{s}(t)}{dt} = -2\theta\tilde{s}(t) + (2\alpha - 2\beta)m(t) + \theta|m(t)| + \alpha + \beta, \text{ and } \tilde{s}(0) = s(0). \tag{2.3.8}$$

We have that

$$s(t) \geq \tilde{s}(t), \ t \in [0, \infty). \tag{2.3.9}$$

Using (2.4.16) to solve the ODE in (2.3.8) and, we have that

$$\lim_{t\to\infty} s(t) \geq \left(\frac{\alpha - \beta}{\theta}\right)^2 + \frac{\max\{\alpha, \beta\}}{\theta}. \tag{2.3.10}$$

■

In the following proposition, we study the first moment function $m(t)$ when $\theta \neq \gamma$. We provide a feasible region for the stationary first moment.

Proposition 2.3.1.

(i) When $\alpha \geq \beta$ and $\theta < \gamma$, we have

$$\frac{\alpha-\beta}{\theta} \leq \liminf_{t\to\infty} m(t) \leq \limsup_{t\to\infty} m(t) \leq \min\left\{\frac{\alpha-\beta}{\theta} + \frac{(\gamma-\theta)\beta}{\theta\gamma}, \frac{\alpha-\beta}{\gamma} + \frac{(\gamma-\theta)\alpha}{\theta\gamma}\right\}.$$

(ii) When $\alpha \geq \beta$ and $\theta > \gamma$, we have

$$\max\left\{\frac{\alpha-\beta}{\theta} + \frac{(\gamma-\theta)\beta}{\theta\gamma}, \frac{\alpha-\beta}{\gamma} + \frac{(\gamma-\theta)\alpha}{\theta\gamma}\right\} \leq \liminf_{t\to\infty} m(t) \leq \limsup_{t\to\infty} m(t) \leq \frac{\alpha-\beta}{\theta}.$$

(iii) When $\alpha < \beta$ and $\theta < \gamma$, we have

$$\frac{\alpha-\beta}{\gamma} \leq \liminf_{t\to\infty} m(t) \leq \limsup_{t\to\infty} m(t) \leq \min\left\{\frac{\alpha-\beta}{\theta} + \frac{(\gamma-\theta)\beta}{\theta\gamma}, \frac{\alpha-\beta}{\gamma} + \frac{(\gamma-\theta)\alpha}{\theta\gamma}\right\}.$$

(iv) When $\alpha < \beta$ and $\theta > \gamma$, we have

$$\max\left\{\frac{\alpha-\beta}{\theta}+\frac{(\gamma-\theta)\beta}{\theta\gamma}, \frac{\alpha-\beta}{\gamma}+\frac{(\gamma-\theta)\alpha}{\theta\gamma}\right\} \leq \liminf_{t\to\infty} m(t) \leq \limsup_{t\to\infty} m(t) \leq \frac{\alpha-\beta}{\gamma}.$$

Proof. We first show (i) and (iii). We note that

$$\begin{aligned} m'(t) &= \alpha - \beta - \theta m_+(t) + \gamma m_-(t) \\ &= \alpha - \beta - \theta m(t) + (\gamma-\theta) m_-(t) \\ &\geq \alpha - \beta - \theta m(t). \end{aligned}$$

Consider the following ODE

$$\tilde{m}'(t) = \alpha - \beta - \theta \tilde{m}(t)$$

with initial distribution $\tilde{m}(0) = m(0)$. Then we have

$$\liminf_{t\to\infty} m(t) \geq \lim_{t\to\infty} \tilde{m}(t) = \frac{\alpha-\beta}{\theta}. \tag{2.3.11}$$

We next observe that $m_-(t)$ can be bounded above by the expected queue length in an $M/M/\infty$ queue with arrival rate β and service rate γ. Hence

$$m_-(t) \leq m_-(0)e^{-\gamma t} + \frac{\beta}{\gamma}(1-e^{-\gamma t}),$$

and so

$$m'(t) \leq \alpha - \beta - \theta m(t) + (\gamma-\theta)m_-(0)e^{-\gamma t} + \frac{(\gamma-\theta)\beta}{\gamma}(1-e^{-\gamma t}).$$

Consider now the ODE

$$\check{m}'(t) = \alpha - \beta - \theta \check{m}(t) + (\gamma-\theta)m_-(0)e^{-\gamma t} + \frac{(\gamma-\theta)\beta}{\gamma}(1-e^{-\gamma t})$$

with initial condition $\check{m}(0) = m(0)$. Thus we have

$$\limsup_{t\to\infty} m(t) \leq \lim_{t\to\infty} \check{m}(t) = \frac{\alpha-\beta}{\theta} + \frac{(\gamma-\theta)\beta}{\theta\gamma}. \tag{2.3.12}$$

We note that there is an alternative estimate for $m(t)$. In fact, we have that

$$m'(t) = \alpha - \beta - \gamma m(t) + (\gamma - \theta)m_+(t) \geq \alpha - \beta - \gamma m_+(t),$$

and

$$m'(t) = \alpha - \beta - \gamma m(t) + (\gamma - \theta)m_+(t) \leq \alpha - \beta - \gamma m(t) + (\gamma - \theta)m_+(0)e^{-\theta t} + \frac{(\gamma - \theta)\alpha}{\theta}(1 - e^{-\theta t}).$$

Thus we have

$$\frac{\alpha - \beta}{\gamma} \leq \liminf_{t\to\infty} m(t) \leq \limsup_{t\to\infty} m(t) \leq \frac{\alpha - \beta}{\gamma} + \frac{(\gamma - \theta)\alpha}{\theta\gamma}. \tag{2.3.13}$$

Combining (2.3.11) - (2.3.13), we show (i) and (iii). The results in (ii) and (iv) follow similarly. ■

Remark 2.3.2. In Proposition 2.3.1, if

$$|\alpha - \beta| \gg \max\left\{\frac{|\gamma - \theta|\beta}{\gamma}, \frac{|\gamma - \theta|\alpha}{\theta}\right\}, \tag{2.3.14}$$

then when $\alpha \geq \beta$ (see (i) and (ii)),

$$\lim_{t\to\infty} m(t) \approx \frac{\alpha - \beta}{\theta},$$

and when $\alpha < \beta$ (see (iii) and (iv)),

$$\lim_{t\to\infty} m(t) \approx \frac{\alpha - \beta}{\gamma}.$$

■

2.4 Fluid and Diffusion Approximations

In this section we establish fluid and diffusion approximations for the double-ended queue under appropriate conditions (see Assumptions 2.4.1 and 2.4.2). To describe the asymptotic region where such an approximation is valid, we consider a sequence of double-ended queues indexed by $n \in \mathbb{N}$. For the n-th system, all the notation introduced in Section 4.2 is carried forward except that we append a superscript n to all quantities to indicate the dependence of parameters, random variables, and stochastic processes on n. In particular, on the space $(\Omega^n, \mathcal{F}^n, P^n, \{\mathcal{F}_t^n\}_{t\geq 0})$, $\{U^n(k) : k \in \mathbb{N}\}$ and $\{V^n(k) : k \in \mathbb{N}\}$ are the interarrival times, N_s^n and N_b^n are the arrival processes, θ^n and γ^n are the reneging rates, and N_{sr}^n and N_{br}^n are the unit-rate Poisson processes used to formulate the reneging processes. Also $1/\alpha^n, \sigma^n$ and $1/\beta^n, \varsigma^n$ are the means and standard deviations for the inter-arrival times of sellers and buyers, respectively. The expectation operator with respect to P^n will be denoted by $\mathbb{E}^n$, but frequently we will suppress n from the notation. We further assume the following strict positivity and uniform integrability on $\{U^n(1) : n \in \mathbb{N}\}$ and $\{V^n(1) : n \in \mathbb{N}\}$.

$$\mathbb{P}^n(U^n(1) > 0) = \mathbb{P}(V^n(1) > 0) = 1 \text{ for all } n \in \mathbb{N}. \tag{2.4.1}$$

$$\{(U^n(1))^2 : n \in \mathbb{N}\} \text{ and } \{(V^n(1))^2 : n \in \mathbb{N}\} \text{ are uniformly integrable.} \tag{2.4.2}$$

Finally, the queue length process X^n can be described as follows:

For $t \geq 0$,

$$X^n(t) = X^n(0) + N_s^n(t) - N_b^n(t) - N_{sr}^n\left(\theta^n \int_0^t X^{n,+}(s)ds\right) + N_{br}^n\left(\gamma^n \int_0^t X^{n,-}(s)ds\right). \tag{2.4.3}$$

The following assumption describes the asymptotic regime of the parameters.

Assumption 2.4.1.

(i) *There exist* $\alpha, \beta, \sigma, \varsigma \in (0, \infty)$ *such that*

$$\alpha^n \to \alpha,\ \beta^n \to \beta,\ \sigma^n \to \sigma,\ \varsigma^n \to \varsigma.$$

(ii) *For* $\theta, \gamma \in (0, \infty)$, *we have that*

$$n\theta^n \to \theta,\ n\gamma^n \to \gamma.$$

2.4.1 Fluid approximation

We begin by defining the fluid scaled processes. Loosely speaking, we accelerate time by factor n and scale down the queue size by the same factor n. More precisely, for $t \geq 0$, define

$$\bar{X}^n(t) = \frac{X^n(nt)}{n},\ \bar{N}_s^n(t) = \frac{N_s^n(nt)}{n},\ \bar{N}_b^n(t) = \frac{N_b^n(nt)}{n},\ \bar{N}_{sr}^n(t) = \frac{N_{sr}^n(nt)}{n},\ \bar{N}_{br}^n(t) = \frac{N_{br}^n(nt)}{n}. \tag{2.4.4}$$

Recall that for a stochastic process $\{Y(t), t \geq 0\}$,

$$\|Y\|_t = \sup_{0 \leq u \leq t} |Y(u)|\,,\ t \in [0, \infty).$$

We first establish an asymptotic limit of $\bar{X}^n$ as $n \to \infty$ in Theorem 2.4.1. The solution of the fluid limit equation is then given in Proposition 2.4.1.

Theorem 2.4.1. *Assume that for some $x_0 \in \mathbb{R}$, $\mathbb{E}|\bar{X}^n(0) - x_0| \to 0$ as $n \to \infty$, and Assumption 2.4.1 holds. Then we have that for $t \in [0, \infty)$,*

$$\mathbb{E}\left(\|\bar{X}^n - x\|_t\right) \to 0, \quad \text{as } n \to \infty, \tag{2.4.5}$$

where x is the solution of the following integral equation

$$x(t) = x_0 + (\alpha - \beta)t + \int_0^t \left(-\theta x^+(s) + \gamma x^-(s)\right) ds. \tag{2.4.6}$$

Proof. We note from (2.4.3) that for $t \geq 0$,

$$\bar{X}^n(t) = \bar{X}^n(0) + \bar{N}_s^n(t) - \bar{N}_b^n(t) - \bar{N}_{sr}^n\left(n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right) + \bar{N}_{br}^n\left(n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right).$$

For $t \in [0, \infty)$, let $N^n(t) = N_{sr}^n(t) + N_{br}^n(t)$ and $O^n(t) = |\bar{N}_s^n(t) - \alpha^n t - \bar{N}_b^n(t) + \beta^n t|$. Then we have that

$$\begin{aligned}
|\bar{X}^n(t)| &\leq |\bar{X}^n(t) - (\bar{N}_s^n(t) - \alpha^n t - \bar{N}_b^n(t) + \beta^n t)| + O^n(t) \\
&\leq |\bar{X}^n(0)| + |\alpha^n - \beta^n|t + O^n(t) + n^{-1}N^n\left(n^2(\gamma^n + \theta^n)\int_0^t |\bar{X}^n(u)|du\right).
\end{aligned}$$

Define for $t \in [0, \infty)$,

$$Y^n(t) = |\bar{X}^n(0)| + |\alpha^n - \beta^n|t + O^n(t) + n^{-1}N^n\left(n^2(\gamma^n + \theta^n)\int_0^t Y^n(s)ds\right).$$

Then

$$|\bar{X}^n(t)| \leq Y^n(t),\ t \in [0, \infty).$$

Noting that N^n, O^n and $\bar{X}^n(0)$ are mutually independent, we see that

$$M^n(t) := Y^n(t) - \bar{X}^n(0) - |\alpha^n - \beta^n|t - O^n(t) - 2n(\gamma^n + \theta^n)\int_0^t Y^n(s)ds$$

is a $\{\mathcal{F}_t^n\}$ martingale. Using Ito's formula, we have that

$$\begin{aligned}(Y^n(t) - O^n(t))\exp\{-2n(\gamma^n + \theta^n)t\} &= |\bar{x}^n(0)| + \int_0^t \exp\{-2n(\gamma^n + \theta^n)s\}dM^n(s)\\ + \int_0^t 2n(\gamma^n + \theta^n)\exp\{-2n(\gamma^n + \theta^n)s\}O^n(s)ds &+ |\alpha^n - \beta^n|\int_0^t \exp\{-2n(\gamma^n + \theta^n)s\}ds,\end{aligned}$$

and so

$$\begin{aligned}&\left(Y^n(t) + \frac{|\alpha^n - \beta^n|}{2n(\gamma^n + \theta^n)}\right)\exp\{-2n(\gamma^n + \theta^n)t\} - \left(|\bar{X}^n(0)| + \frac{|\alpha^n - \beta^n|}{2n(\gamma^n + \theta^n)}\right)\\ &= \int_0^t \exp\{-2n(\gamma^n + \theta^n)s\}dO^n(s) + \int_0^t \exp\{-2n(\gamma^n + \theta^n)s\}dM^n(s). \qquad (2.4.7)\end{aligned}$$

We observe that from the functional law of large numbers for renewal processes, $O^n \Rightarrow 0$ as $n \to \infty$, and from the continuous mapping theorem, $\|O^n\| \Rightarrow 0$ as $n \to \infty$. For $t \in [0, \infty)$, we have the following claim:

$$\begin{aligned}&\{\bar{N}_s^n(t) : n \in \mathbb{N}\} \text{ and } \{\bar{N}_b^n(t) : n \in \mathbb{N}\} \text{ are uniformly integrable.}\\ &\{\|O^n\|_t : n \in \mathbb{N}\} \text{ is uniformly integrable.}\end{aligned} \qquad (2.4.8)$$

(We show the claim at the end of this proof.) Thus we conclude that for $T \in [0, \infty)$,

$$\mathbb{E}\left(\sup_{0 \leq t \leq T} O^n(t)\right) \to 0, \quad \text{as } n \to \infty. \qquad (2.4.9)$$

We next note that from (2.4.7) and (2.4.9), for any $t \in [0, \infty)$,

$$\mathbb{E}(Y^n(t)) \to \left(x_0 + \frac{|\alpha - \beta|}{2(\gamma + \theta)} \right) \exp\{2(\gamma + \theta)t\} - \frac{|\alpha - \beta|}{2(\gamma + \theta)}, \quad \text{as } n \to \infty. \tag{2.4.10}$$

From Doob's inequality and (2.4.10), for any $T \in [0, \infty)$,

$$\begin{aligned}
&\mathbb{E}\left(\sup_{0 \le t \le T} \left| \int_0^t \exp\{-2n(\gamma^n + \theta^n)s\} dM^n(s) \right| \right)^2 \le 4\mathbb{E}\left(\int_0^T \exp\{-2n(\gamma^n + \theta^n)s\} dM^n(s) \right)^2 \\
&= 4\mathbb{E}\left(\int_0^T \exp\{-4n(\gamma^n + \theta^n)s\} d[M^n, M^n]_s \right) \le 4\mathbb{E}([M^n, M^n]_T) \\
&= 4n^{-2}\mathbb{E}\left(N^n \left(n^2(\gamma^n + \theta^n) \int_0^T Y^n(s) ds \right) \right) = 4n^{-1}(n\gamma^n + n\theta^n) \int_0^T \mathbb{E}(Y^n(s)) ds \\
&\to 0, \quad \text{as } n \to \infty. \qquad (2.4.11)
\end{aligned}$$

Now from (2.4.7), (2.4.9), and (2.4.11), for any $T \in [0, \infty)$,

$$\begin{aligned}
&\mathbb{E}\left(\sup_{0 \le t \le T} \left| Y^n(t) - \left[\left(x_0 + \frac{|\alpha - \beta|}{2(\gamma + \theta)} \right) \exp\{2(\gamma + \theta)t\} - \frac{|\alpha - \beta|}{2(\gamma + \theta)} \right] \right| \right) \\
&\le \mathbb{E}\left(\sup_{0 \le t \le T} \left| Y^n(t) - \left[\left(\bar{X}^n(0) + \frac{|\alpha^n - \beta^n|}{2n(\gamma^n + \theta^n)} \right) \exp\{2n(\gamma^n + \theta^n)t\} - \frac{|\alpha^n - \beta^n|}{2n(\gamma^n + \theta^n)} \right] \right| \right) + o(1) \\
&\le \exp\{2n(\gamma^n + \theta^n)T\} \mathbb{E}\left(\sup_{0 \le t \le T} \int_0^t \exp\{-2n(\gamma^n + \theta^n)s\} dO^n(s) \right) \\
&\quad + \exp\{2n(\gamma^n + \theta^n)T\} \mathbb{E}\left(\sup_{0 \le t \le T} \left| \int_0^t \exp\{-2n(\gamma^n + \theta^n)u\} dM^n(u) \right| \right) + o(1) \\
&\to 0, \quad \text{as } n \to \infty. \qquad (2.4.12)
\end{aligned}$$

We next observe that

$$
\begin{aligned}
&|\bar{X}^n(t) - x(t)| \\
&\leq |\bar{X}^n(0) - x_0| + |\bar{N}^n_s(t) - \alpha^n t - \bar{N}^n_b(s) + \beta^n t| \\
&\quad + \left|\bar{N}^n_{sr}\left(n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right) - n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right| \\
&\quad + \left|\bar{N}^n_{br}\left(n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right) - n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right| \\
&\quad + \left|n\theta^n \int_0^t \bar{X}^{n,+}(s)ds - \theta \int_0^t \bar{X}^{n,+}(s)ds\right| + \left|n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds - \gamma \int_0^t \bar{X}^{n,-}(s)ds\right| \\
&\quad + \left|\theta \int_0^t \bar{X}^{n,+}(s)ds - \theta \int_0^t x^+(s)ds\right| + \left|\gamma \int_0^t \bar{X}^{n,-}(s)ds - \gamma \int_0^t x^-(s)ds\right| \\
&\leq |\bar{X}^n(0) - x_0| + O^n(t) \\
&\quad + \left|\bar{N}^n_{sr}\left(n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right) - n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right| \\
&\quad + \left|\bar{N}^n_{br}\left(n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right) - n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right| \\
&\quad + (|n\theta^n - \theta| + |n\gamma^n - \gamma|) \int_0^t Y^n(s)ds \\
&\quad + (\theta + \gamma) \int_0^t \left|\bar{X}^n(s) - x(s)\right| ds.
\end{aligned}
$$

Gronwall's inequality yields that

$$
\begin{aligned}
\mathbb{E}\left(\sup_{0\leq t\leq T} |\bar{X}^n(t) - x(t)|\right) \leq \mathbb{E}\Bigg(&|\bar{X}^n(0) - x_0| + \sup_{0\leq t\leq T} O^n(t) \\
&+ \sup_{0\leq t\leq T}\left|\bar{N}^n_{sr}\left(n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right) - n\theta^n \int_0^t \bar{X}^{n,+}(s)ds\right| \\
&\quad + \sup_{0\leq t\leq T}\left|\bar{N}^n_{br}\left(n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right) - n\gamma^n \int_0^t \bar{X}^{n,-}(s)ds\right| \\
&\quad + (|n\theta^n - \theta| + |n\gamma^n - \gamma|) \int_0^T Y^n(s)ds\Bigg) e^{(\theta+\gamma)T}.
\end{aligned}
\tag{2.4.13}
$$

Let $\tau^n(t) = \int_0^t Y^n(s)ds$, and then from (2.4.12), for any $T \in [0, \infty)$,

$$
\mathbb{E}\left(\sup_{0\leq t\leq T}\left|\tau^n(t) - \left[\left(x_0 + \frac{|\alpha-\beta|}{2(\gamma+\theta)}\right)\frac{\exp\{2(\gamma+\theta)t\} - 1}{2(\gamma+\theta)} - \frac{|\alpha-\beta|}{2(\gamma+\theta)}\right] t\right|\right) \to 0.
$$

We then note that

$$\mathbb{E}\left(\sup_{0\le t\le T}\left|\bar{N}^n_{sr}\left(n\theta^n\int_0^t \bar{X}^{n,+}(s)ds\right)-n\theta^n\int_0^t \bar{X}^{n,+}(s)ds\right|\right)\le \mathbb{E}\left(\sup_{0\le t\le n\theta^n\tau^n(T)}\left|\bar{N}^n_{sr}(t)-t\right|\right).$$

For $t\in[0,\infty)$, let $\bar{N}^{n,c}_{sr}(t)=\bar{N}^n_{sr}(t)-t$. Using functional law of large numbers for Poisson processes, we have that $\bar{N}^{n,c}_{sr}\Rightarrow 0$ as $n\to\infty$. Now from continuous mapping theorem, we have that $\|\bar{N}^n_{sr}\|\Rightarrow 0$ as $n\to\infty$. Noting that $n\theta^n\tau^n$ converges to a finite limit in probability as $n\to\infty$, and using the random change of time theorem (see Section 3.14 in Billingsley [3]), we have that

$$\|\bar{N}^{n,c}_{sr}\|_{n\theta^n\tau^n}\Rightarrow 0,\quad \text{as } n\to\infty.$$

Finally, we show the uniform integrability of $\|\bar{N}^{n,c}_{sr}\|_{n\theta^n\tau^n(T)}$ for each fixed $T\in[0,\infty)$. In fact, we have that for $T\in[0,\infty)$,

$$\begin{aligned}
&\mathbb{E}(\|\bar{N}^{n,c}_{sr}\|_{n\theta^n\tau^n(T)})^2=\mathbb{E}\left(\sup_{0\le t\le n\theta^n\tau^n(T)}\left|\bar{N}^n_{sr}(t)-t\right|\right)^2\\
&\le 2\mathbb{E}\left(\bar{N}^n_{sr}(n\theta^n\tau^n(T))\right)^2+2\mathbb{E}(n\theta^n\tau^n(T))^2\\
&\le 2n^{-2}\mathbb{E}\left(N^n\left(n^2(\theta^n+\gamma^n)\int_0^T Y^n(s)ds\right)-n^2(\theta^n+\gamma^n)\int_0^T Y^n(s)ds\right)^2\\
&\quad +2[(n\theta^n)^2+(n\theta^n+n\gamma^n)^2]\mathbb{E}(\tau^n(T))^2\\
&=2\mathbb{E}(M^n(T))^2+2[(n\theta^n)^2+(n\theta^n+n\gamma^n)^2]\mathbb{E}(\tau^n(T))^2\\
&=2\mathbb{E}([M^n,M^n]_T)+2[(n\theta^n)^2+(n\theta^n+n\gamma^n)^2]\mathbb{E}(\tau^n(T))^2<\infty,
\end{aligned}$$

uniformly on $n\in\mathbb{N}$. Thus we have that for $T\in[0,\infty)$,

$$\mathbb{E}\left(\sup_{0\le t\le n\theta^n\tau^n(T)}\left|\bar{N}^n_{sr}(t)-t\right|\right)=\mathbb{E}(\|\bar{N}^{n,c}_{sr}\|_{n\theta^n\tau^n(T)})\to 0,\quad \text{as } n\to\infty,$$

and so

$$\mathbb{E}\left(\sup_{0\le t\le T}\left|\bar{N}^n_{sr}\left(n\theta^n\int_0^t \bar{X}^{n,+}(s)ds\right)-n\theta^n\int_0^t \bar{X}^{n,+}(s)ds\right|\right)\to 0,\quad \text{as } n\to\infty. \tag{2.4.14}$$

Similarly, we have that

$$\mathbb{E}\left(\sup_{0\le t\le T}\left|\bar{N}^n_{br}\left(n\theta^n\int_0^t \bar{X}^{n,-}(s)ds\right)-n\gamma^n\int_0^t \bar{X}^{n,-}(s)ds\right|\right)\to 0,\quad \text{as } n\to\infty. \tag{2.4.15}$$

Applying (2.4.9), (2.4.14), (2.4.15), and the convergence $n\gamma^n\to\gamma, n\theta^n\to\theta$ to (2.4.13), (2.4.5) follows immediately. We now show the claim (2.4.8). We first note from Lemma 4.2 in Budhiraja and Ghosh [5], there exist $\epsilon,\delta\in(0,1)$ such that

$$\inf_{n\in\mathbb{N}}\mathbb{P}(U^n(1)>\delta)>\epsilon. \tag{2.4.16}$$

Fix $t\in(0,\infty)$. Following the proof of Theorem 1.1 in Sigman [33], define new interarrival times which are given as follows: For $i\in\mathbb{N}$,

$$\tilde{U}^n(i)=\delta 1_{\{U^n(i)>\delta\}}.$$

We observe that the new arrivals can only occur at the deterministic times $\{\delta k: k\in\mathbb{N}\cup\{0\}\}$. Let C_k denote the number of arrivals at time δk. Then $\{C_k : k\in\mathbb{N}\}$ are i.i.d. random variables with geometric distribution with success probability $p^n=\mathbb{P}(U^n(1)>\delta)$. Defining the new renewal process $\tilde{N}^n_s$ by using the new interarrival times $\{\tilde{U}^n(i): i\in\mathbb{N}\}$, we then note that for $t\in[0,\infty)$,

$$N^n_s(t)\le\tilde{N}^n_s(t)\le\sum_{k=1}^{[t/\delta]}C_k.$$

Finally, we have that for $x\in[0,\infty)$,

$$\begin{aligned}\mathbb{E}(\bar{N}^n_s(t))^2 &\le n^{-2}\mathbb{E}\left(\sum_{k=1}^{[nt/\delta]}C_k\right)^2 = n^{-2}\mathrm{Var}\left(\sum_{k=1}^{[nt/\delta]}C_k\right)+n^{-2}\left(\mathbb{E}\left(\sum_{k=1}^{[nt/\delta]}C_k\right)\right)^2\\ &= n^{-2}[nt/\delta]\frac{1-p^n}{(p^n)^2}+n^{-2}([nt/\delta])^2\left(\frac{1-p^n}{p^n}\right)^2<\infty,\end{aligned}$$

uniformly on $n\in\mathbb{N}$ on noting that $\inf_n p^n\ge\epsilon>0$. This shows $\{\bar{N}^n_s(t): n\in\mathbb{N}\}$ is uniformly integrable. Similarly, we have $\{\bar{N}^n_b(t): n\in\mathbb{N}\}$ is uniformly integrable. The claim follows. ■

Proposition 2.4.1. *Consider the integral equation in* (2.4.6).

(i) If $\alpha \geq \beta$ and $x_0 \geq 0$, then

$$x(t) = \left(x_0 - \frac{\alpha-\beta}{\theta}\right) e^{-\theta t} + \frac{\alpha-\beta}{\theta},\ t \in [0,\infty).$$

(ii) If $\alpha \geq \beta$ and $x_0 < 0$, then

$$x(t) = \begin{cases} \left(x_0 - \frac{\alpha-\beta}{\gamma}\right) e^{-\gamma t} + \frac{\alpha-\beta}{\gamma}, & t \in [0,t_1], \\ \frac{\alpha-\beta}{\theta}\left(1 - e^{-\theta(t-t_1)}\right), & t \in [t_1,\infty), \end{cases}$$

where

$$t_1 = \gamma^{-1} \log\left(\frac{\alpha-\beta-\gamma x_0}{\alpha-\beta}\right).$$

(iii) If $\alpha < \beta$ and $x_0 \leq 0$, then

$$x(t) = \left(x_0 - \frac{\alpha-\beta}{\gamma}\right) e^{-\gamma t} + \frac{\alpha-\beta}{\gamma},\ t \in [0,\infty).$$

(iv) If $\alpha < \beta$ and $x_0 > 0$, then

$$x(t) = \begin{cases} \left(x_0 - \frac{\alpha-\beta}{\theta}\right) e^{-\theta t} + \frac{\alpha-\beta}{\theta}, & t \in [0,t_2], \\ \frac{\alpha-\beta}{\gamma}\left(1 - e^{-\gamma(t-t_2)}\right), & t \in [t_2,\infty), \end{cases}$$

where

$$t_2 = \theta^{-1} \log\left(\frac{\alpha-\beta-\theta x_0}{\alpha-\beta}\right).$$

Proof. We first show (i) and (ii). Assume $\alpha \geq \beta$. We consider the following three situations.

(a) Let $x_0 > 0$. Define $\tau_1 = \inf\{t \geq 0 : x(t) \leq 0\}$. Then for $t \in [0,\tau_1)$, we have $x(t) \geq 0$, and so

$$x(t) = x_0 + (\alpha-\beta)t - \theta \int_0^t x(s)ds. \tag{2.4.17}$$

Solving the above equation, we have for $t \in [0, \tau_1)$,

$$x(t) = \left(x_0 - \frac{\alpha - \beta}{\theta}\right) e^{-\theta t} + \frac{\alpha - \beta}{\theta}. \tag{2.4.18}$$

If $\tau_1 < \infty$, then $x(\tau_1) = \lim_{t \uparrow \tau_1} x(t) > 0$, which contradicts the definition of τ_1. Thus $\tau_1 = \infty$, and so equation (2.4.18) holds for all $t \in [0, \infty)$.

(b) Let $x_0 = 0$. We first assume $\alpha > \beta$ and note that

$$x'(0) = \alpha - \beta - \theta x_0^+ + \gamma x_0^- = \alpha - \beta > 0.$$

So there exists $\tau_2 > 0$ such that $x(t) > 0$ for $t \in (0, \tau_2]$. Define $\tilde{x}(t) = x(t + \tau_2),\ t \in [0, \infty)$. Then we have for $t \in [0, \infty)$,

$$\tilde{x}(t) = \tilde{x}(0) + (\alpha - \beta)t + \int_0^t -\theta \tilde{x}^+(s) + \gamma \tilde{x}^-(s) ds.$$

Noting that $\tilde{x}(0) = x(\tau_2) > 0$, and using the result in Part (a), we obtain that $\tilde{x}(t) > 0$ for all $t \in (0, \infty)$. Thus $x(t) \geq 0$ for all $t \in [0, \infty)$, and so equations (2.4.17) and (2.4.18) hold for all $t \in [0, \infty)$. If $\alpha = \beta$, then

$$x(t) = 0 \text{ for all } t \in [0, \infty). \tag{2.4.19}$$

Otherwise, if (2.4.19) fails, then there exists $0 < t_1 < t_2 < \infty$ such that $x(t_1) = 0$ and $x(s) > 0$ (or $x(s) < 0$) for all $s \in (t_1, t_2]$. Without loss of generality, we assume $x(s) > 0$ for $s \in (t_1, t_2]$. Then for $x \in (t_1, t_2]$,

$$x(s) = x(t_1) - \theta \int_{t_1}^s x(u) du = -\theta \int_{t_1}^s x(u) du = 0,$$

which is a contradiction.

(c) Let $x_0 < 0$. We note that

$$x'(0) = \alpha - \beta - \theta x_0^+ + \gamma x_0^- = \alpha - \beta - \gamma x_0 > 0.$$

Let $\tau_3 = \inf\{t \geq 0 : x(t) \geq 0\}$. Then for $t \in [0, \tau_3]$,

$$x(t) = x_0 + (\alpha - \beta)t - \gamma \int_0^t x(s)ds,$$

and so

$$x(t) = \left(x_0 - \frac{\alpha - \beta}{\gamma}\right) e^{-\gamma t} + \frac{\alpha - \beta}{\gamma}. \tag{2.4.20}$$

From the fact that $x(\tau_3) = 0$, we have

$$\tau_3 = \gamma^{-1} \log \left(\frac{(\alpha - \beta) - \gamma x_0}{\alpha - \beta} \right) \in (0, \infty).$$

Define $\hat{x}(t) = x(t + \tau_3), t \in [0, \infty)$. We have for $t \in [0, \infty)$,

$$\hat{x}(t) = \hat{x}(0) + (\alpha - \beta)t + \int_0^t \left(-\theta \hat{x}^+(s) + \gamma \hat{x}^-(s) \right) ds.$$

Noting that $\hat{x}(0) = x(\tau_3) = 0$, and using the result in Part (b), we know that $\hat{x}(t) \geq 0$ for all $t \in [0, \infty)$. Hence $x(t) \geq 0$ for all $x \in [\tau_3, \infty)$, and equations (2.4.17) and (2.4.18) hold for $t \in [\tau_3, \infty)$. Combining this with (2.4.20), we obtain that

$$x(t) = \begin{cases} \left(x_0 - \frac{\alpha-\beta}{\gamma}\right) e^{-\gamma t} + \frac{\alpha-\beta}{\gamma}, & t \in [0, \tau_3], \\ \frac{\alpha-\beta}{\theta} \left(1 - e^{-\gamma t}\right), & t \in [\tau_3, \infty). \end{cases}$$

At last, letting $y(t) = -x(t)$ and using the results in (i) and (ii), the results in (iii) and (iv) follow immediately. ■

Remark 2.4.1. We note that when $\theta = \gamma$, the fluid equation is the same as the ODE for $m(t)$ in Section 2.3 (see (2.3.6)). However, when $\theta \neq \gamma$, noting that $m_+(t) \neq m^+(t)$ and

$m_-(t) \neq m^-(t)$ in general, the fluid equation doesn't match the ODE for $m(t)$. Nevertheless, we note that

$$\lim_{t\to\infty} x(t) = \begin{cases} \frac{\alpha-\beta}{\theta}, & \alpha \geq \beta, \\ \frac{\alpha-\beta}{\gamma}, & \alpha < \beta, \end{cases}$$

and hence, under the condition (2.3.14), $\lim_{t\to\infty} x(t) \approx \lim_{t\to\infty} m(t)$. ■

2.4.2 Diffusion approximation

In this subsection, we consider the diffusion approximation and define the diffusion scaled processes. We accelerate time by the same factor n and scale down the queue size by factor $\sqrt{n}$. To be precise, for $t \geq 0$, define

$$\begin{aligned} &\hat{X}^n(t) = \frac{X^n(nt)}{\sqrt{n}}, \ \hat{N}_s^n(t) = \frac{N_s^n(nt) - n\alpha^n t}{\sqrt{n}}, \ \hat{N}_b^n(t) = \frac{N_b^n(nt) - n\beta^n t}{\sqrt{n}}, \\ &\hat{N}_{sr}^n(t) = \frac{N_{sr}^n(nt) - nt}{\sqrt{n}}, \ \hat{N}_{br}^n(t) = \frac{N_{br}^n(nt) - nt}{\sqrt{n}}. \end{aligned} \tag{2.4.21}$$

We next introduce the following "heavy traffic" condition on parameters α^n and β^n. Roughly speaking, in the heavy traffic regime, the arrival processes have the same arrival rate.

Assumption 2.4.2. *For some $c \in \mathbb{R}$,*

$$\sqrt{n}(\alpha^n - \beta^n) \to c, \quad \textit{as } n \to \infty.$$

We note that under Assumptions 2.4.1 and 2.4.2, $\alpha = \beta$.

We now state our main results for the diffusion approximation. Define a diffusion process Z as follows. For a given random variable $Z(0)$ with law ν, and a standard Brownian motion W, let Z be the unique solution to the following stochastic integral equation

$$Z(t) = Z(0) + aW(t) + ct - \theta \int_0^t Z^+(u)du + \gamma \int_0^t Z^-(u)du, \tag{2.4.22}$$

where

$$a = \sqrt{\alpha^3\sigma^2 + \beta^3\varsigma^2}. \tag{2.4.23}$$

The existence and uniqueness of Z is guaranteed by the following lemma.

Lemma 2.4.1 (Reed and Ward [30])**.** *Let $\phi : \mathcal{D}([0,\infty),\mathbb{R}) \to \mathcal{D}([0,\infty),\mathbb{R})$ be Lipschitz continuous. Then for any given $w \in D([0,\infty),\mathbb{R})$, there exists a unique $x \in D([0,\infty),\mathbb{R})$ that satisfies the integral*

$$x(t) = w(t) + \int_0^t \phi(x)(u)du,$$

and $x(0) = w(0)$. Moreover, define the mapping $\mathcal{M}^\phi : D([0,\infty),\mathbb{R}) \to D([0,\infty),\mathbb{R})$ by $\mathcal{M}^\phi(w) = x$, and then $\mathcal{M}^\phi$ is Lipschitz continuous.

Theorem 2.4.2. *Assume that $\hat{X}^n(0)$ converges weakly to a probability measure ν, $\mathbb{E}(|\bar{X}^n(0)|) \to 0$, and Assumptions 2.4.1 and 2.4.2 hold. Then $\hat{X}^n \Rightarrow Z$, where Z is defined by* (2.4.22)*.*

Proof. We first note that, by functional central limit theorem for renewal processes (see Theorem 14.6 in Billingsley [3]),

$$\hat{N}_s^n \Rightarrow W_s,\ \hat{N}_b^n \Rightarrow W_b, \tag{2.4.24}$$

where W_s and W_b are independent Brownian motions with zero drifts and variances $\alpha^3\sigma^2$ and $\beta^3\varsigma^2$, respectively. We also note that $\hat{N}_{sr}^n$ and $\hat{N}_{br}^n$ converge weakly to standard Brownian motions from functional central limit theorem for unit Poisson process. Further noting from Theorem 2.4.1 and Proposition 2.4.1 (when $\alpha = \beta$ and $x_0 = 0$) that $\bar{X}^n \Rightarrow 0$, and using the random change of time theorem (see Section 3.14 in Billingsley [3]), we obtain that

$$\begin{aligned} &\hat{N}_{sr}^n\left(n\theta^n \int_0^\cdot \bar{X}^{n,+}(u)du\right) \Rightarrow 0, \\ &\hat{N}_{br}^n\left(n\gamma^n \int_0^\cdot \bar{X}^{n,-}(u)du\right) \Rightarrow 0. \end{aligned} \tag{2.4.25}$$

Now define for $t \geq 0$,

$$\hat{W}^n(t) = \hat{N}_s^n(t) - \hat{N}_b^n(t) - \hat{N}_{sr}^n \left(n\theta^n \int_0^t \bar{X}^{n,+}(u)du \right) + \hat{N}_{br}^n \left(n\gamma^n \int_0^t \bar{X}^{n,-}(u)du \right).$$

Combining (2.4.24), (2.4.25), and Assumption 2.4.1(ii), we have $\hat{W}^n \Rightarrow aW$, where W is a standard Brownian motion, and a is as in Equation 2.4.23. Furthermore, there exists a random variable $Z(0)$ with law ν such that $(\hat{X}^n(0), \hat{W}^n) \Rightarrow (Z(0), aW)$. By Skorohod representation theorem, without loss of generality, we assume that $(\hat{X}^n(0), \hat{W}^n)$ and $(Z(0), W)$ are defined on the same probability space and $(\hat{X}^n(0), \hat{W}^n) \to (Z(0), W)$ almost surely and uniformly on compact sets of $[0, \infty)$. Define

$$Z(t) = Z(0) + aW(t) + ct - \theta \int_0^t Z^+(s)ds + \gamma \int_0^t Z^-(s)ds, \quad t \geq 0.$$

From Lemma 2.4.1, Z is well-defined. Also note that

$$\hat{X}^n(t) = \hat{X}^n(0) + \hat{W}^n(t) + \sqrt{n}(\alpha^n - \beta^n)t - n\theta^n \int_0^t \hat{X}^{n,+}(s)ds + n\gamma^n \int_0^t \hat{X}^{n,-}(s)ds, \quad t \geq 0.$$

We then have that for $t \geq 0$,

$$\|\hat{X}^n - Z\|_t \leq |\hat{X}^n(0) - Z(0)| + \|\hat{W}^n - W\|_t + (n\theta^n + n\gamma^n + \theta + \gamma) \int_0^t \|\hat{X}^n - Z\|_s ds.$$

By Gronwall's inequality,

$$\|\hat{X}^n - Z\|_t \leq \left(|\hat{X}^n(0) - Z(0)| + \|\hat{W}^n - W\|_t \right) e^{(n\theta^n + n\gamma^n + \theta + \gamma)t}.$$

Using continuous mapping theorem, we have

$$\|\hat{W}^n - W\|_t \to 0, \quad \text{almost surely.}$$

Noting that $n\theta^n + n\gamma^n + \theta + \gamma \to 2\theta + 2\gamma$, we obtain that

$$\|\hat{X}^n - Z\|_t \to 0, \quad \text{almost surely.}$$

The result follows immediately. ■

When $\theta = \gamma$, we derive the first and second moments of $Z(t)$.

Proposition 2.4.2. *Consider the stochastic integral equation in* (2.4.22)*, and assume* $\theta = \gamma$*. Then for* $t \in [0, \infty)$*,*

$$\mathbb{E}(Z(t)) = \left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) e^{-\theta t} + \frac{c}{\theta}, \tag{2.4.26}$$

and

$$\begin{aligned}\mathbb{E}(Z(t)^2) = &\left(\mathbb{E}(Z(0)^2) - \frac{2c}{\theta}\left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) - \left(\frac{c}{\theta}\right)^2 - \frac{a^2}{2\theta}\right) e^{-2\theta t} \\ &+ \frac{2c}{\theta}\left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) e^{-\theta t} + \left(\frac{c}{\theta}\right)^2 + \frac{a^2}{2\theta}.\end{aligned} \tag{2.4.27}$$

Proof. We first note that

$$\mathbb{E}(Z(t)) = \mathbb{E}(Z(0)) + ct - \theta \int_0^t \mathbb{E}(Z(u))du.$$

Thus we have

$$\mathbb{E}(Z(t)) = \left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) e^{-\theta t} + \frac{c}{\theta}, \ t \in [0, \infty). \tag{2.4.28}$$

Using the integration by parts for stochastic integrals (see Corollary 2 in Chaper II.6 in Protter [29]), we have

$$Z(t)^2 = [Z, Z]_t + 2\int_0^t Z(s)dZ(s),$$

where $[Z, Z]$ is the quadratic variation process for Z, and $[Z, Z]_t = Z(0)^2 + a^2 t$. Thus we have

$$\mathbb{E}(Z(t)^2) = \mathbb{E}(Z(0)^2) + a^2 t + 2c\int_0^t \mathbb{E}(Z(s))ds - 2\theta \int_0^t \mathbb{E}(Z(s)^2)ds.$$

Let $v(t) = \mathbb{E}(Z(t)^2)$. From (2.4.28), we have the following ODE with initial condition $v(0) = \mathbb{E}(Z(0)^2)$.

$$\frac{dv(t)}{dt} = a^2 + 2c\left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) e^{-\theta t} + \frac{2c^2}{\theta} - 2\theta v(t).$$

Solving this ODE, we have for $t \in [0, \infty)$,

$$\begin{aligned}v(t) = &\left(v(0) - \frac{2c}{\theta}\left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) - \frac{c^2}{\theta^2} - \frac{a^2}{2\theta}\right) e^{-2\theta t} + \frac{2c}{\theta}\left(\mathbb{E}(Z(0)) - \frac{c}{\theta}\right) e^{-\theta t} \\ &+ \frac{c^2}{\theta^2} + \frac{a^2}{2\theta}.\end{aligned}$$

$\blacksquare$

Remark 2.4.2. We observe that when $\theta = \gamma$,

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}(Z(t)) &= \frac{c}{\theta} \\ \lim_{t\to\infty} \mathbb{E}(Z(t)^2) &= \left(\frac{c}{\theta}\right)^2 + \frac{a^2}{2\theta}.\end{aligned} \tag{2.4.29}$$

Noting that under heavy traffic condition, $\theta = \lim_{n\to\infty} n\theta^n$ and $c = \lim_{n\to\infty} \sqrt{n}(\alpha^n - \beta^n)$, we have that for large enough $n \in \mathbb{N}$,

$$\begin{aligned}&\lim_{t\to\infty} \mathbb{E}(X^n(nt)) \approx \lim_{t\to\infty} \mathbb{E}(\sqrt{n}Z(t)) \approx \frac{\alpha^n - \beta^n}{\theta^n}, \\ &\lim_{t\to\infty} \mathbb{E}(X^n(nt))^2 \approx \lim_{t\to\infty} \mathbb{E}\big(\sqrt{n}Z(t)\big)^2 \approx \left(\frac{\alpha^n - \beta^n}{\theta^n}\right)^2 + \frac{(\alpha^n)^3(\sigma^n)^2 + (\beta^n)^3(\varsigma^n)^2}{2\theta^n}.\end{aligned}$$

In the special case of Poisson arrivals, we have for large enough $n \in \mathbb{N}$,

$$\lim_{t\to\infty} \mathbb{E}(X^n(nt))^2 \approx \left(\frac{\alpha^n - \beta^n}{\theta^n}\right)^2 + \frac{\alpha^n + \beta^n}{2\theta^n} \approx \left(\frac{\alpha^n - \beta^n}{\theta^n}\right)^2 + \frac{\max\{\alpha^n, \beta^n\}}{2\theta^n},$$

which matches the lower bound for $s(t)$ in Section 2.3 (see (2.3.10)). $\blacksquare$

When $\theta \neq \gamma$, the exact moments of $Z(t)$ become intractable. Thus we consider the limiting distribution of the process Z, and compute the stationary moments. We observe that Z is a unique solution of the following SDE

$$dZ(t) = adW(t) + \big(c - \theta Z^+(t) + \gamma Z^-(t)\big)\, dt. \tag{2.4.30}$$

Theorem 2.4.3. *Let a be as in Equation 2.4.23 and denote by $\phi(\cdot;\xi,\eta)$ and $\Phi(\cdot;\xi,\eta)$ the density and distribution function of $N(\xi,\eta)$. Then the density of stationary distribution of the diffusion process satisfying (2.4.30) is given by*

$$\psi(x) = \begin{cases} \frac{C}{\sqrt{\theta}} \exp\left\{\frac{c^2}{\theta a^2}\right\} \phi\left(x; \frac{c}{\theta}, \frac{a^2}{2\theta}\right), & x \geq 0, \\ \frac{C}{\sqrt{\gamma}} \exp\left\{\frac{c^2}{\gamma a^2}\right\} \phi\left(x; \frac{c}{\gamma}, \frac{a^2}{2\gamma}\right), & x < 0, \end{cases} \tag{2.4.31}$$

where C is given by

$$C = \frac{1}{\frac{1}{\sqrt{\theta}}\exp\left\{\frac{c^2}{\theta a^2}\right\}\left(1-\Phi\left(0;\frac{c}{\theta},\frac{a^2}{2\theta}\right)\right)+\frac{1}{\sqrt{\gamma}}\exp\left\{\frac{c^2}{\gamma a^2}\right\}\Phi\left(0;\frac{c}{\gamma},\frac{a^2}{2\gamma}\right)}. \tag{2.4.32}$$

Proof. We follow Section 5 of Chapter 15 in Karlin and Taylor [17] to construct a stationary density. Denote by $\mu(x)$ the infinitesimal drift parameter $c - \theta x^+ + \gamma x^-$. We first note that an indefinite integral of $\frac{2\mu(x)}{a^2}$ is $\frac{2c}{a^2}x - \frac{\theta}{a^2}x^2 1\{x \geq 0\} - \frac{\gamma}{a^2}x^2 1\{x < 0\}$. Define for $x \in \mathbb{R}$,

$$s(x) = \exp\left\{\frac{2c}{a^2}x - \frac{\theta}{a^2}x^2 1\{x \geq 0\} - \frac{\gamma}{a^2}x^2 1\{x < 0\}\right\}.$$

We define a density function as follows:

$$\begin{aligned}
\psi(x) &= \tilde{C}s(x) \\
&= \begin{cases} \tilde{C}\exp\left\{\frac{2c}{a^2}x - \frac{\theta}{a^2}x^2\right\}, & x \geq 0 \\ \tilde{C}\exp\left\{\frac{2c}{a^2}x - \frac{\gamma}{a^2}x^2\right\}, & x < 0 \end{cases} \\
&= \begin{cases} \frac{C}{\sqrt{\theta}}\exp\left\{\frac{c^2}{\theta a^2}\right\}\phi\left(x;\frac{c}{\theta},\frac{a^2}{2\theta}\right), & x \geq 0 \\ \frac{C}{\sqrt{\gamma}}\exp\left\{\frac{c^2}{\gamma a^2}\right\}\phi\left(x;\frac{c}{\gamma},\frac{a^2}{2\gamma}\right), & x < 0, \end{cases}
\end{aligned}$$

where

$$C = a\sqrt{\pi}\tilde{C} = \frac{1}{\frac{1}{\sqrt{\theta}}\exp\left\{\frac{c^2}{\theta a^2}\right\}\left(1-\Phi\left(0;\frac{c}{\theta},\frac{a^2}{2\theta}\right)\right)+\frac{1}{\sqrt{\gamma}}\exp\left\{\frac{c^2}{\gamma a^2}\right\}\Phi\left(0;\frac{c}{\gamma},\frac{a^2}{2\gamma}\right)}.$$

∎

In the following, we calculate the first two moment of the stationary distribution of Z. First, note that if $X \sim N(\xi, \eta)$, the density of truncated normal random variable on (x_1, x_2) is given by $\frac{\phi(x;\xi,\eta)}{\Phi(x_2;\xi,\eta)-\Phi(x_1;\xi,\eta)}$. Let X_1 be a truncated $N(\frac{c}{\theta}, \frac{a^2}{2\theta})$ random variable on $(0, +\infty)$, and X_2 be a truncated $N(\frac{c}{\gamma}, \frac{a^2}{2\gamma})$ random variable on $(-\infty, 0)$. Let V be the random variable with distribution ψ of Equation 2.4.31. Then V can be written as a mixture of two truncated Normal random variables as follows:

$$V = \begin{cases} X_1 & \text{w.p. } d_1, \\ X_2 & \text{w.p. } d_2, \end{cases}$$

where

$$d_1 = \frac{C}{\sqrt{\theta}} \exp\left\{\frac{c^2}{\theta a^2}\right\} \left(1 - \Phi\left(0; \frac{c}{\theta}, \frac{a^2}{2\theta}\right)\right), \tag{2.4.33}$$

$$d_2 = \frac{C}{\sqrt{\gamma}} \exp\left\{\frac{c^2}{\gamma a^2}\right\} \Phi\left(0; \frac{c}{\gamma}, \frac{a^2}{2\gamma}\right). \tag{2.4.34}$$

Now, the first and second moments of the truncated Normals are given by

$$\begin{aligned}
\mathbb{E}(X_1) &= \frac{c}{\theta} + \frac{a}{\sqrt{2\theta}} \frac{\phi\left(-\frac{c}{a}\sqrt{\frac{2}{\theta}}; 0, 1\right)}{1 - \Phi\left(-\frac{c}{a}\sqrt{\frac{2}{\theta}}; 0, 1\right)}, \\
\mathbb{E}(X_2) &= \frac{c}{\gamma} - \frac{a}{\sqrt{2\gamma}} \frac{\phi\left(-\frac{c}{a}\sqrt{\frac{2}{\gamma}}; 0, 1\right)}{\Phi\left(-\frac{c}{a}\sqrt{\frac{2}{\gamma}}; 0, 1\right)}, \\
\mathbb{E}(X_1^2) &= \left(\frac{c}{\theta}\right)^2 + \frac{a^2}{2\theta} + \frac{\sqrt{2}}{2} \frac{ac}{\theta\sqrt{\theta}} \frac{\phi\left(-\frac{c}{a}\sqrt{\frac{2}{\theta}}; 0, 1\right)}{1 - \Phi\left(-\frac{c}{a}\sqrt{\frac{2}{\theta}}; 0, 1\right)}, \\
\mathbb{E}(X_2^2) &= \left(\frac{c}{\gamma}\right)^2 + \frac{a^2}{2\gamma} - \frac{\sqrt{2}}{2} \frac{ac}{\gamma\sqrt{\gamma}} \frac{\phi\left(-\frac{c}{a}\sqrt{\frac{2}{\gamma}}; 0, 1\right)}{\Phi\left(-\frac{c}{a}\sqrt{\frac{2}{\gamma}}; 0, 1\right)}.
\end{aligned}$$

Hence the first and second moments of V are given by

$$\mathbb{E}(V) = d_1 \mathbb{E}(X_1) + d_2 \mathbb{E}(X_2), \tag{2.4.35}$$

and

$$\mathbb{E}(V^2) = d_1 \mathbb{E}(X_1^2) + d_2 \mathbb{E}(X_2^2), \tag{2.4.36}$$

where d_1 and d_2 is given by equation (2.4.33) and (2.4.34).

Remark 2.4.3. In the above theorem, when $\theta = \gamma$,

$$\psi(x) = \phi\left(x; \frac{c}{\theta}, \frac{a^2}{2\theta}\right),$$

and so the first and second moments $\mathbb{E}(V)$ and $\mathbb{E}(V^2)$ are the same as the limits in (2.4.29).

■

2.5 Numerical Example

In this section, we use simulations to evaluate the performance of the Poisson and the diffusion approximation under different arrival processes. We first note that the diffusion limit Z satisfies the stochastic integral equation given by (see (2.4.22))

$$Z(t) = Z(0) + aW(t) + ct - \theta \int_0^t Z^+(u)du + \gamma \int_0^t Z^-(u)du,$$

where $W(t)$ is a standard Brownian Motion. Hence, using (2.4.21), we get for large $N \in \mathbb{N}$, approximately,

$$\frac{X^N(Nt)}{\sqrt{N}} = \frac{X^N(0)}{\sqrt{N}} + aW(t) + ct - \theta \int_0^t \frac{X^{N,+}(Nu)}{\sqrt{N}} du + \gamma \int_0^t \frac{X^{N,-}(Nu)}{\sqrt{N}} du.$$

Letting $s = Nt$, we have approximately,

$$X^N(s) = X^N(0) + aW(s) + \frac{c}{\sqrt{N}} s - \frac{\theta}{N} \int_0^s X^{N,+}(v)dv + \frac{\gamma}{N} \int_0^s X^{N,-}(v)dv.$$

Using Assumptions 2.4.1 and 2.4.2, we get approximately,

$$\begin{aligned} X^N(s) = X^N(0) &+ \sqrt{(\alpha^N)^3(\sigma^N)^2 + (\beta^N)^3(\varsigma^N)^2} W(s) + \left(\alpha^N - \beta^N\right) s \\ &- \theta^N \int_0^s X^{N,+}(v)dv + \gamma^n \int_0^s X^{N,-}(v)dv. \end{aligned}$$

Therefore, if we have a double-ended queueing system with parameters α, β, σ, ς, θ and γ, the dynamics of the queue length process $\{X(t), t \geq 0\}$ can be approximated by an asymmetric Ornstein- Uhlenbeck process

$$X(t) = X(0) + aW(t) + (\alpha - \beta)\, t - \theta \int_0^t X^+(v)dv + \gamma \int_0^t X^-(v)dv, \; t \geq 0, \qquad (2.5.1)$$

or by the stochastic differential equation (SDE)

$$dX(t) = adW(t) + \big((\alpha-\beta) - \theta X^+(t) + \gamma X^-(t)\big)\, dt,\ t \geq 0. \tag{2.5.2}$$

We now consider a double-ended queue length process $\{X(t), t \geq 0\}$ with the buyer inter-arrival time distribution $F_s(\cdot)$ and seller inter-arrival time distribution $F_b(\cdot)$. Let m_s, m_b, sd_s and sd_b be the mean and standard deviation of the inter-arrival time for sellers and buyers. We consider the following inter-arrival time distributions:

- Exponential: $F_s(x) = 1 - e^{-\alpha x}$, $F_b(x) = 1 - e^{-\beta x}$, $m_s = \frac{1}{\alpha}$, $m_b = \frac{1}{\beta}$, $sd_s = \frac{1}{\alpha}$, $sd_b = \frac{1}{\beta}$
- Uniform: $F_s(x) = \frac{\alpha x}{2}$ $(x \in [0, \frac{2}{\alpha}])$, $F_b(x) = \frac{\beta x}{2}$ $(x \in [0, \frac{2}{\beta}])$, $m_s = \frac{1}{\alpha}$, $m_b = \frac{1}{\beta}$, $sd_s = \frac{1}{\sqrt{3}\alpha}$, $sd_b = \frac{1}{\sqrt{3}\beta}$
- Erlang(2): $F_s(x) = 1 - e^{-2\alpha x} - 2\alpha x e^{-2\alpha x}$, $F_b(x) = 1 - e^{-2\beta x} - 2\beta x e^{-2\beta x}$, $m_s = \frac{1}{\alpha}$, $m_b = \frac{1}{\beta}$, $sd_s = \frac{1}{\sqrt{2}\alpha}$, $sd_b = \frac{1}{\sqrt{2}\beta}$
- Hyper-exponential: $F_s(x) = \frac{1}{3}(1 - e^{-\frac{1}{2}\alpha x}) + \frac{2}{3}(1 - e^{-2\alpha x})$, $F_b(x) = \frac{1}{3}(1 - e^{-\frac{1}{2}\beta x}) + \frac{2}{3}(1 - e^{-2\beta x})$, $m_s = \frac{1}{\alpha}$, $m_b = \frac{1}{\beta}$, $sd_s = \frac{\sqrt{2}}{\alpha}$, $sd_b = \frac{\sqrt{2}}{\beta}$

Note that the means of inter-arrival time of the above distributions are the same, while their standard deviations are different, with the following ordering: Hyper-exponential > Exponential > Erlang > Uniform.

We consider the following values for $(\alpha, \beta) = (1, 1)$, $(1, 1.5)$ and $(1, 2)$. We choose the following reneging rates $(\theta, \gamma) = (\alpha, \beta)$, $0.1(\alpha, \beta)$ and $0.01(\alpha, \beta)$. For example, when $(\theta, \gamma) = 0.1(\alpha, \beta)$, the seller's (buyer's) expected reneging time is 10 times the buyer's (seller's) expected inter-arrival time. Thus we consider a total of $4 \times 3 \times 3 = 36$ different parameter settings.

In each parameter setting, we use simulation, Poisson approximation, and diffusion approximation (which will be made precise below) to estimate the following performance measures:

$$\pi_i = \lim_{t\to\infty} \Pr\{X(t) = i\}, \quad -1000 \le i \le 1000 \tag{2.5.3}$$

$$L_1 = \lim_{t\to\infty} E(X(t)) \tag{2.5.4}$$

$$L_2 = \lim_{t\to\infty} E(X(t)^2). \tag{2.5.5}$$

Simulation. We compute the performance measure by using N replications of the simulation using Matlab. Each replication consists of simulating the system for $0 \le t \le T$ and the estimates are computed by using the sample paths over $t \in [\tau, T]$, where $\tau < T$ is a given warmup period. Let $X^k(t)$ be the state of the system at time t in the k-th replication, $k = 1, 2, \cdots, N$, $0 \le t \le T$. Using these sample paths, we compute

$$\pi_i^s = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{T-\tau}\int_{\tau}^{T} 1_{\{X^k(t)=i\}} dt, \quad -1000 \le i \le 1000.$$

Using these we compute the following simulation estimates of the first and second moments of the queue-length:

$$L_1^s = \sum_{i=-1000}^{1000} i\pi_i^s, \tag{2.5.6}$$

$$L_2^s = \sum_{i=-1000}^{1000} i^2\pi_i^s. \tag{2.5.7}$$

Poisson approximation. In this approximation we replace the renewal arrival processes by Poisson arrival processes with the same arrival rates. Clearly, this approximation is exact in the exponential case. From equations (3.3.2) to (3.3.3), we have:

$$\pi_i^p = \frac{\alpha^i}{\prod_{j=1}^{i} (\beta + j\theta)} \pi_0, \quad i = 1, 2, \cdots, 1000$$

$$
\begin{aligned}
\pi_{-i}^p &= \frac{\beta^i}{\prod_{j=1}^{i}(\alpha + j\gamma)}\pi_0, \quad i = 1, 2, \cdots, 1000 \\
\pi_0^p &= \left(1 + \sum_{i=1}^{1000} \frac{\alpha^i}{\prod_{j=1}^{i}(\beta + j\theta)} + \sum_{i=1}^{1000} \frac{\beta^i}{\prod_{j=1}^{i}(\alpha + j\gamma)}\right)^{-1}.
\end{aligned}
$$

Let L_1^p and L_2^p be the Poisson approximation of L_1 and L_2 respectively. Using π_i^p we compute

$$L_1^p = \sum_{i=-1000}^{1000} i\pi_i^p, \tag{2.5.8}$$

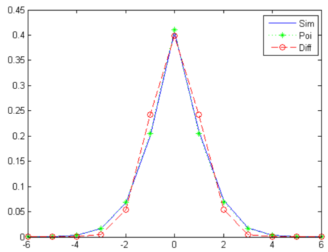
$$L_2^p = \sum_{i=-1000}^{1000} i^2\pi_i^p. \tag{2.5.9}$$

We compute the relative error of the above moments to the ones from simulation method. For example, the relative error of L_1^p to L_1^s is given by $(L_1^p - L_1^s)/L_1^s \times 100\%$.

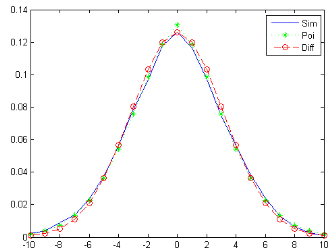
Diffusion approximation. As the second method, we use diffusion approximation to compute these performance measures. We know that the stationary density function of the diffusion approximation of the queue-length process is given by equation (2.4.31). In addition, we directly compute the first moment L_1^d and second moment L_2^d by using equations (2.4.35) and (2.4.36). We also compute the relative errors of the moments to the ones from simulation method.

The comparisons of limiting density are shown in Figure 2.2-3.4. In the figures, we compare the density graphs derived from the simulation method and the two approximation methods. When using simulation method, we evaluate the performance measures using the parameter $(N, \tau, T) = (400, 1000, 4000)$ and obtain the 90% confidence interval. The comparisons of L_1 and L_2 are shown in Tables 2.1-2.8. In the columns of L_1^p and L_2^p, we evaluate the performance measures by Poisson approximation method, and obtain the relative error of each performance measure to the one from simulation method. In the columns of L_1^d and

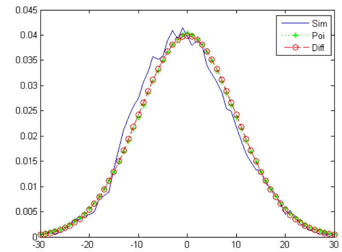
L_2^d, we obtain the performance measures by diffusion approximation method, and also obtain the relative error of each performance measure to the one from simulation method.



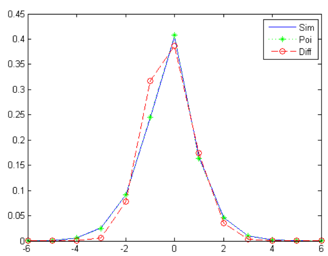
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1, 1, 1)$



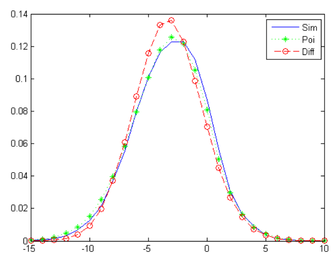
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.1, 0.1)$



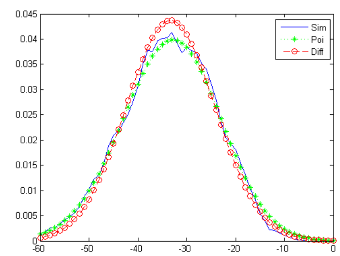
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.01, 0.01)$



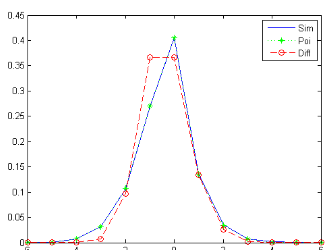
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 1, 1.5)$



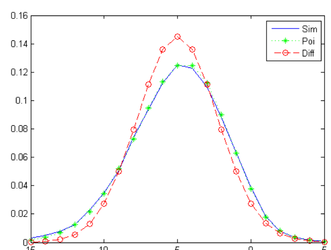
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.1, 0.15)$



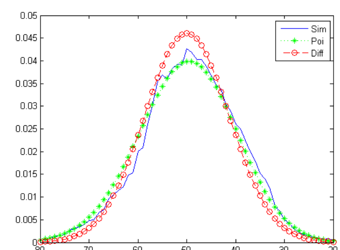
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.01, 0.015)$



(a) $(\alpha, \beta, \theta, \gamma) = (1, 2, 1, 2)$

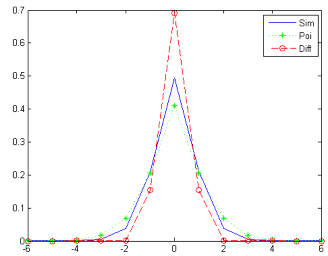


(b) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.1, 0.2)$

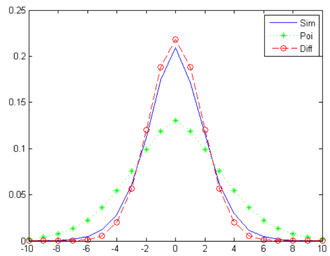


(c) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.01, 0.02)$

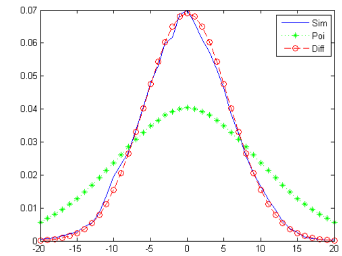
Figure 2.2: Density functions by simulation method, Poisson approximation and diffusion approximation, when inter-arrival times follow exponential distribution



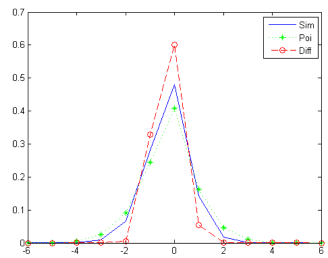
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1, 1, 1)$



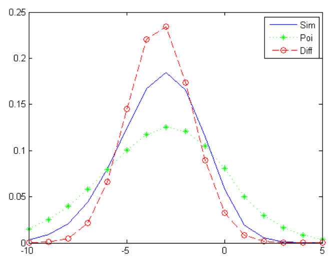
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.1, 0.1)$



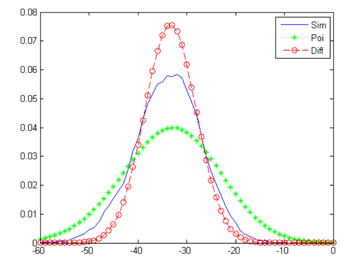
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.01, 0.01)$



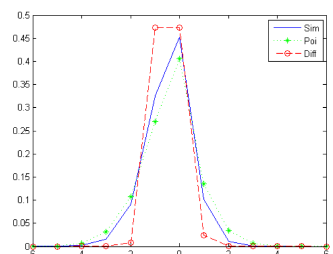
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 1, 1.5)$



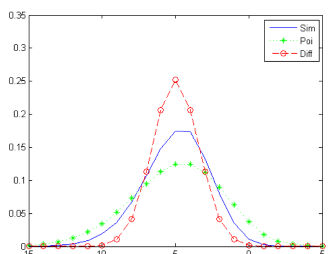
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.1, 0.15)$



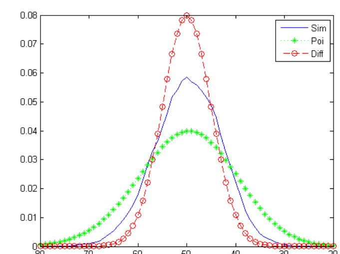
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.01, 0.015)$



(a) $(\alpha, \beta, \theta, \gamma) = (1, 2, 1, 2)$

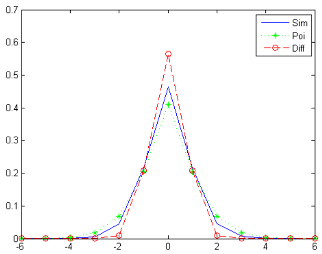


(b) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.1, 0.2)$

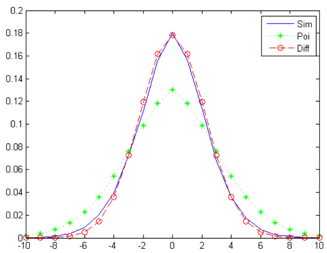


(c) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.01, 0.02)$

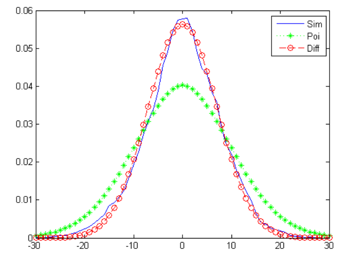
Figure 2.3: Density functions by simulation method, Poisson approximation and diffusion approximation, when inter-arrival times follow uniform distribution



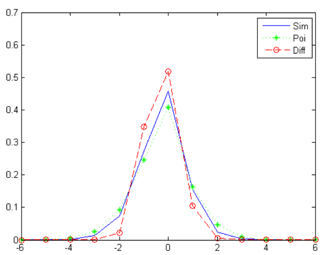
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1, 1, 1)$



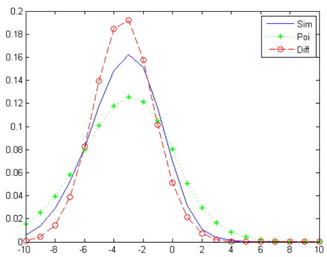
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.1, 0.1)$



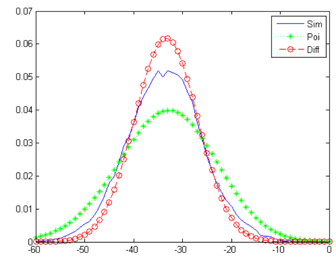
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.01, 0.01)$



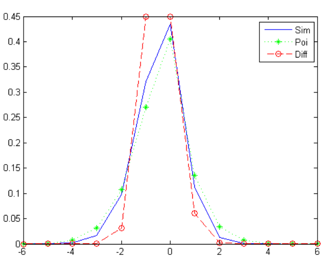
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 1, 1.5)$



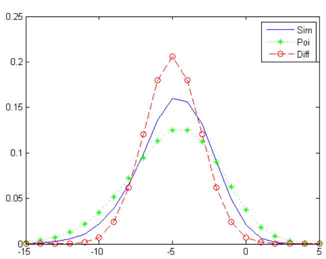
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.1, 0.15)$



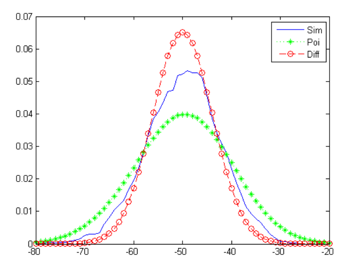
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.01, 0.015)$



(a) $(\alpha, \beta, \theta, \gamma) = (1, 2, 1, 2)$

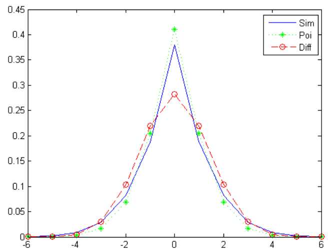


(b) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.1, 0.2)$

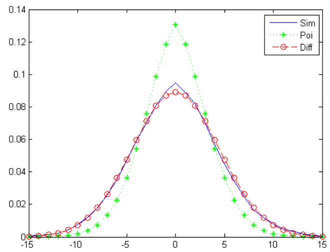


(c) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.01, 0.02)$

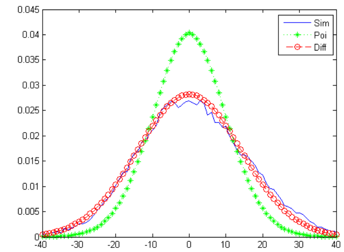
Figure 2.4: Density functions by simulation method, Poisson approximation and diffusion approximation, when inter-arrival times follow Erlang distribution



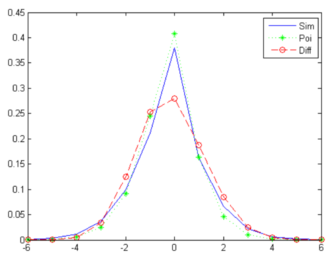
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1, 1, 1)$



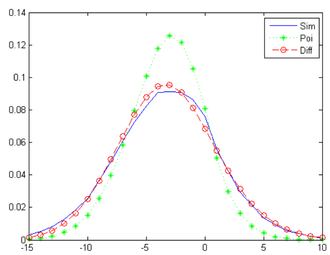
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.1, 0.1)$



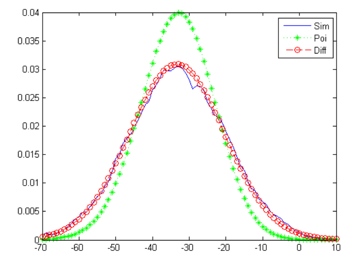
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1, 0.01, 0.01)$



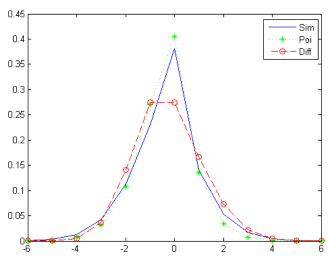
(a) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 1, 1.5)$



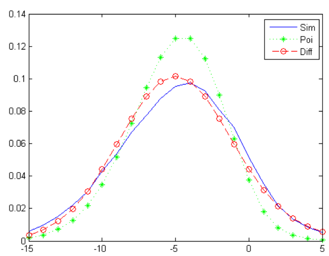
(b) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.1, 0.15)$



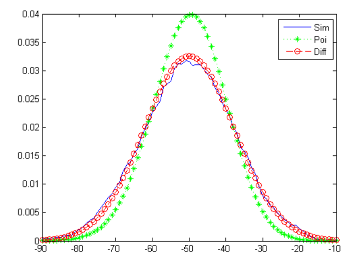
(c) $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.01, 0.015)$



(a) $(\alpha, \beta, \theta, \gamma) = (1, 2, 1, 2)$



(b) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.1, 0.2)$



(c) $(\alpha, \beta, \theta, \gamma) = (1, 2, 0.01, 0.02)$

Figure 2.5: Density functions by simulation method, Poisson approximation and diffusion approximation, when inter-arrival times follow hyper-exponential distribution

(α, β)	(θ, γ)	L_1^s	L_1^p	L_1^d
(1, 1)	(1, 1)	0.0001	0	0
		±0.0024	NA	NA
	(0.1, 0.1)	-0.0178	0	0
		±0.0243	NA	NA
	(0.01, 0.01)	0.1234	0	0
		±0.2084	NA	NA
(1, 1.5)	(1, 1.5)	-0.2352	-0.2343	-0.2375
		±0.0022	-0.41%	0.98%
	(0.1, 0.15)	-3.248	-3.2532	-3.2625
		±0.0192	0.16%	0.44%
	(0.01, 0.015)	-33.1485	-33.3332	-33.3332
		±0.1754	0.56%	0.56%
(1, 2)	(1, 2)	-0.3876	-0.3858	-0.3878
		±0.002	-0.47%	0.04%
	(0.1, 0.2)	-4.9779	-4.9719	-4.9776
		±0.0157	-0.12%	-0.01%
	(0.01, 0.02)	-49.9609	-50	-50
		±0.142	0.08%	0.08%

Table 2.1: The first moment of the stationary distribution when the arrival process is a Poisson process

(α, β)	(θ, γ)	L_1^s	L_1^p	L_1^d
(1, 1)	(1, 1)	0.0004	0	0
		±0.0017	NA	NA
	(0.1, 0.1)	-0.0009	0	0
		±0.0141	NA	NA
	(0.01, 0.01)	-0.1309	0	0
		±0.1231	NA	NA
(1, 1.5)	(1, 1.5)	-0.2736	-0.2343	-0.2979
		±0.0015	-14.39%	8.87%
	(0.1, 0.15)	-3.3315	-3.2532	-3.328
		±0.0114	-2.35%	-0.10%
	(0.01, 0.015)	-33.4634	-33.3332	-33.3333
		±0.1132	-0.39%	-0.39%
(1, 2)	(1, 2)	-0.4375	-0.3858	-0.4714
		±0.0013	-11.82%	7.76%
	(0.1, 0.2)	-4.9946	-4.9719	-4.9998
		±0.0109	-0.45%	0.10%
	(0.01, 0.02)	-50.0716	-50	-50
		±0.1036	-0.14%	-0.14%

Table 2.2: The first moment of the stationary distribution when the inter-arrival times follow Uniform distribution

(α, β)	(θ, γ)	L_1^s	L_1^p	L_1^d
(1, 1)	(1, 1)	0.0006	0	0
		±0.0024	NA	NA
	(0.1, 0.1)	-0.0113	0	0
		±0.0186	NA	NA
	(0.01, 0.01)	0.0409	0	0
		±0.1848	NA	NA
(1, 1.5)	(1, 1.5)	-0.2624	-0.2343	-0.2804
		±0.002	-10.74%	6.84%
	(0.1, 0.15)	-3.2975	-3.2532	-3.3165
		±0.0155	-1.34%	0.57%
	(0.01, 0.015)	-33.1629	-33.3332	-33.3333
		±0.1613	0.51%	0.51%
(1, 2)	(1, 2)	-0.4285	-0.3858	-0.4493
		±0.0018	-9.96%	4.87%
	(0.1, 0.2)	-4.9832	-4.9719	-4.9983
		±0.015	-0.23%	0.30%
	(0.01, 0.02)	-50.089	-50	-50
		±0.1507	-0.18%	-0.18%

Table 2.3: The first moment of the stationary distribution when the inter-arrival times follow Erlang distribution

(α, β)	(θ, γ)	L_1^s	L_1^p	L_1^d
(1, 1)	(1, 1)	0.0022	0	0
		±0.0032	NA	NA
	(0.1, 0.1)	-0.0169	0	0
		±0.0321	NA	NA
	(0.01, 0.01)	0.016	0	0
		±0.3177	NA	NA
(1, 1.5)	(1, 1.5)	-0.2039	-0.2343	-0.1735
		±0.0028	14.89%	-14.92%
	(0.1, 0.15)	-3.1406	-3.2532	-3.1368
		±0.0271	3.59%	-0.12%
	(0.01, 0.015)	-33.2392	-33.3332	-33.3261
		±0.237	0.28%	0.26%
(1, 2)	(1, 2)	-0.3383	-0.3858	-0.2866
		±0.0026	14.04%	-15.26%
	(0.1, 0.2)	-4.8819	-4.9719	-4.8822
		±0.0214	1.84%	0.01%
	(0.01, 0.02)	-50.1134	-50	-50
		±0.1959	-0.23%	-0.23%

Table 2.4: The first moment of the stationary distribution when the inter-arrival times follow hyper-exponential distribution

(α, β)	(θ, γ)	L_2^s	L_2^p	L_2^d
(1, 1)	(1, 1)	1.409	1.4104	1
		±0.0042	0.10%	-29.03%
	(0.1, 0.1)	11.3894	11.3045	10
		±0.0838	-0.74%	-12.20%
	(0.01, 0.01)	103.2893	104.0397	100
		±2.2995	0.73%	-3.18%
(1, 1.5)	(1, 1.5)	1.4354	1.4372	1.028
		±0.0038	0.12%	-28.38%
	(0.1, 0.15)	21.2369	21.2498	19.4084
		±0.1458	0.06%	-8.61%
	(0.01, 0.015)	1218.2624	1211.1069	1194.4415
		±12.3607	-0.59%	-1.96%
(1, 2)	(1, 2)	1.4828	1.4841	1.0755
		±0.0036	0.09%	-27.47%
	(0.1, 0.2)	34.8606	34.956	32.4389
		±0.1677	0.27%	-6.95%
	(0.01, 0.02)	2601.2009	2600	2575
		±15.2948	-0.05%	-1.01%

Table 2.5: The second moment of the stationary distribution when arrival process is a Poisson process

(α, β)	(θ, γ)	L_2^s	L_2^p	L_2^d
(1, 1)	(1, 1)	0.8254	1.4104	0.3333
		±0.002	70.87%	-59.62%
	(0.1, 0.1)	4.3492	11.3045	3.3333
		±0.0336	159.92%	-23.36%
	(0.01, 0.01)	34.6831	104.0397	33.3333
		±0.7472	199.97%	-3.89%
(1, 1.5)	(1, 1.5)	0.8961	1.4372	0.3993
		±0.002	60.38%	-55.45%
	(0.1, 0.15)	15.775	21.2498	13.8778
		±0.0952	34.71%	-12.03%
	(0.01, 0.015)	1148.5144	1211.1069	1138.8889
		±7.7166	5.45%	-0.84%
(1, 2)	(1, 2)	1.0102	1.4841	0.5019
		±0.0021	46.91%	-50.32%
	(0.1, 0.2)	30.1933	34.956	27.4992
		±0.1226	15.77%	-8.92%
	(0.01, 0.02)	2551.7944	2600	2525
		±10.7995	1.89%	-1.05%

Table 2.6: The second moment of the stationary distribution when the inter-arrival times follow Uniform distribution

(α, β)	(θ, γ)	L_2^s	L_2^p	L_2^d
(1, 1)	(1, 1)	1.9852	1.4104	2
		±0.0064	-28.95%	0.74%
	(0.1, 0.1)	20.899	11.3045	20
		±0.1485	-45.91%	-4.30%
	(0.01, 0.01)	201.652	104.0397	200
		±4.4079	-48.41%	-0.82%
(1, 1.5)	(1, 1.5)	1.9944	1.4372	2.0092
		±0.0056	-27.94%	0.74%
	(0.1, 0.15)	29.1292	21.2498	28.0112
		±0.1894	-27.05%	-3.84%
	(0.01, 0.015)	1293.001	1211.1069	1277.5943
		±16.5071	-6.33%	-1.19%
(1, 2)	(1, 2)	2.0089	1.4841	2.0252
		±0.0052	-26.12%	0.81%
	(0.1, 0.2)	41.7095	34.956	39.8704
		±0.2351	-16.19%	-4.41%
	(0.01, 0.02)	2684.7348	2600	2649.9986
		±20.576	-3.16%	-1.29%

Table 2.7: The second moment of the stationary distribution when the inter-arrival times follow Erlang distribution

(α, β)	(θ, γ)	L_2^s	L_2^p	L_2^d
(1, 1)	(1, 1)	1.9943	1.4104	2
		±0.0063	-29.28%	0.29%
	(0.1, 0.1)	20.8656	11.3045	20
		±0.1625	-45.82%	-4.15%
	(0.01, 0.01)	205.774	104.0397	200
		±4.7111	-49.44%	-2.81%
(1, 1.5)	(1, 1.5)	1.9962	1.4372	2.0092
		±0.0057	-28.01%	0.65%
	(0.1, 0.15)	29.4329	21.2498	28.0112
		±0.1921	-27.80%	-4.83%
	(0.01, 0.015)	1307.8225	1211.1069	1277.5943
		±16.5066	-7.40%	-2.31%
(1, 2)	(1, 2)	2.0048	1.4841	2.0252
		±0.0048	-25.97%	1.02%
	(0.1, 0.2)	41.5526	34.956	39.8704
		±0.2282	-15.88%	-4.05%
	(0.01, 0.02)	2660.7665	2600	2649.9986
		±20.4434	-2.28%	-0.40%

Table 2.8: The second moment of the stationary distribution when the inter-arrival times follow hyper-exponential distribution

From the numerical example, we find that the performance of the limiting behavior of diffusion approximation method improves when the reneging rate becomes smaller. Besides, the variance of the inter-arrival time also affects the performance. From the four distributions we have tried, we conclude that the larger variance results in better performance. The Poisson approximation works well only in the exponential case, when it is in fact exact. In most other cases, the diffusion approximation outperforms the Poisson approximation. This implies that ignoring the variance in the inter-arrival time leads to significant errors.

When $(\alpha, \beta, \theta, \gamma) = (1, 1.5, 0.01, 0.05)$ and $(1, 2, 0.01, 0.02)$, we find that $(\alpha - \beta)/\gamma$ is -10 and -50, respectively. In this case Remark 3.2 is applicable. Using Equations (2.4.35) and (2.4.36), we see that

$$L_1^p = L_1^d = \frac{\alpha - \beta}{\gamma},$$

and

$$L_2^p = \left(\frac{\alpha - \beta}{\gamma}\right)^2 + \frac{\beta}{\gamma},$$

and

$$L_2^d = \left(\frac{\alpha - \beta}{\gamma}\right)^2 + \frac{a^2}{2\gamma}.$$

The results about the first moment are shown in the tables 2.1 to 2.4 and those about the second moment are shown in tables 2.5 to 2.8. Figures 2.2 to 3.4 show the limiting distributions of the queue-length for the simulation, the Poisson approximation, and the diffusion approximation. It is clear that the diffusion approximation outperforms the Poisson approximation (except in the Exponential case, where the Poisson approximation is exact) since it is better able to capture the effects of the variance. Furthermore, the diffusion approximation improves as the reneging rates become small, since that brings the parameters closer to the asymptotic region.

2.6 Extensions

We end this chapter with suggestions for three extensions.

1. In this chapter we have assumed that the patience times of the buyers and sellers are exponentially distributed. It would be interesting to study the situation when the distribution is general. In this case $\{X(t), t \geq 0\}$ is no longer Markov. It would be interesting to see if we can derive the diffusion approximation under this setting.

2. In this chapter we have assumed that the arrival processes of the buyers and sellers are independent of the state of the system. It would be interesting to consider an extension where the arrival processes are Poisson whose parameters depend on the state of the double-ended queue. For example, the parameters could simply depend upon the sign of the $X(t)$. Thus we could capture the situation where the buyer arrival rate exceeds the seller arrival rate when there are sellers waiting (its a buyers' market), and the seller arrival rate exceeds the buyer arrival rate when there are buyers waiting (its a sellers' market). We think this extension is doable, and the methods developed in this chapter will be useful in its study.

3. In stock market order book dynamics, the buyers and sellers have their own bid and ask prices. Thus we can model the order book as a multi-dimensional double-ended queue. It would be interesting to extend our analysis to such a model. However, this extension promises to be hard, since the state-space of the multi-dimensional double-ended queue is typically not convex.

CHAPTER 3

Double-ended Queue with State-dependent Phase-type Arrival Process

3.1 Introduction

In this chapter we consider a double-ended queue with state-dependent phase-type arrival process. In detail We model the combined arrival process of the buyers and sellers as a single renewal process, with iid Phase-type(PH) inter-arrival times. An arrival from this common stream may be a buyer or seller with given probabilities that depend on the state of the queue. The buyers and seller may also depart without conducting a trade due to impatience. We assume that the impatience times of the buyers and sellers are independent exponential random variables. The form of this dependence will be made precise in the next section. We also assume that there is a limited waiting space for the buyers and the sellers, and the arrivals to the system are lost when it is full.

In section 3.2 we present the details of the model of the double-ended queue with arrivals generated by a PH-renewal process and introduce the relevant notation. In Section 3.3 we collect the results about the special case when the PH-renewal process is a Poisson process. In Section 3.4 we derive the algorithm for the limiting distribution of the double-ended queue. In section 3.5 we derive the explicit formula of limiting distribution for the two special cases. In section 3.6 we study the numerical examples considering the inter-arrival time follows Erlang distribution and Hyper-exponential distribution, derive several performance measures and validate the algorithm using simulation. We also and make comments on the possible extensions of the model in Section 3.7.

3.2 Model Formulation

Consider a trading market where the traders arrive at the queue according to a PH-renewal processes. The inter-arrival times form a sequences of i.i.d. Phase-type random variables $\{A(k) : k \in \mathbb{N}\}$. Assume that $A(1)$ follows a Phase-type distribution with parameters (p, T) (see Chapter 2 section 2.2 from Neuts [25]). When a seller is matched with a buyer (we say a trade occurs), they both leave the queue. The trading follows according to first-come-first-served principle. If an arriving seller (buyer) cannot be matched with a buyer (seller), he/she will stay in the queue and wait for the upcoming buyers (sellers). Thus there cannot be non-zero number of buyers and sellers simultaneously in the queue. Also, each seller (buyer) may leave the queue without trading because of impatience. The patience time of each seller (buyer) follows an exponential distribution with rate θ (γ). The patience times of the buyers and sellers are independent of each other. Thus queueing system forms a double-ended queue as shown in Figure 3.1.

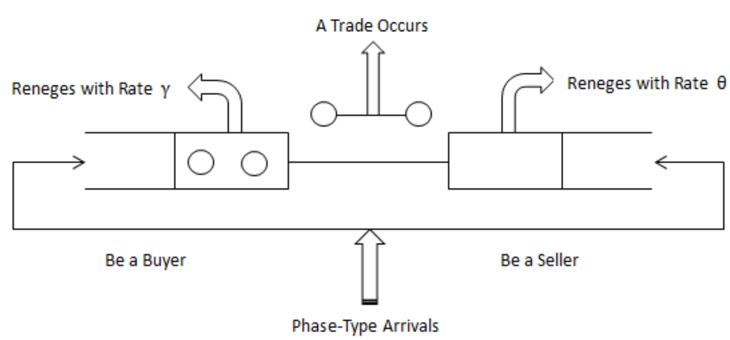


Figure 3.1: Double-ended Queue

Let $X(t)$ be the length of the double-ended queue at time t ($X(t) \in \mathbb{Z}) \doteq \{\cdots, -2, -1, 0, 1, 2, \cdots\}$). If $X(t) > 0$, there are $X(t)$ sellers waiting in the queue, and if $X(t) < 0$, there are $-X(t)$ buyers waiting in the queue. An arriving trader is a buyer or a seller with state-dependent probabilities as given below:

$$\Pr\{\text{Customer arriving at time } t \text{ is a seller}|X(t-)=n\} = \begin{cases} \alpha, & \text{if } n>0 \\ \beta, & \text{if } n<0 \\ \eta, & \text{if } n=0 \end{cases}$$

$$\Pr\{\text{Customer arriving at time } t \text{ is a buyer}|X(t-)=n\} = \begin{cases} 1-\alpha, & \text{if } n>0 \\ 1-\beta, & \text{if } n<0 \\ 1-\eta, & \text{if } n=0 \end{cases}$$

In practice, if $X(t) > 0$, we say the market state is a buyers' market. In this case, the buyers are more likely to come to queue, since they need not wait for sellers. On the contrary, if $X(t) < 0$, we say the market state is a sellers' market. In this case, the sellers are more willing to come to the queue, since they need not wait for buyers. Therefore, we are likely to have $\alpha < 0.5$ and $\beta > 0.5$.

3.3 Special Case: Poisson Arrivals

When the arrival process is a Poisson process, the arrivals of sellers and buyers can be considered as two independent Poisson processes (with state-dependent arrival rates) by splitting the original arriving Poisson process. Thus it is obvious that $\{X(t), t \geq 0\}$ is a birth and death process on $\mathbb{Z}$. In this case, the queueing system forms a double-ended queue as shown in Figure (3.2).

Assume the arrival process of the traders is Poisson with λ. Thus the birth rates λ_n $(n \in \mathbb{Z})$and death rates μ_n $(n \in \mathbb{Z})$are given by:

$$\lambda_n = \begin{cases} \alpha\lambda, & n>0 \\ \eta\lambda, & n=0 \\ \beta\lambda - n\gamma, & n<0 \end{cases}$$

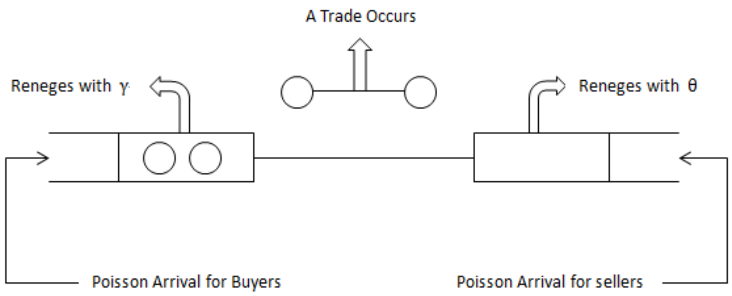


Figure 3.2: Double-ended Queue

$$\mu_n = \begin{cases} (1-\alpha)\lambda + n\theta, & n > 0 \\ (1-\eta)\lambda, & n = 0 \\ (1-\beta)\lambda, & n < 0 \end{cases}$$

Using the standard theory (see Kulkarni [22]), we see that this Birth and Death process is:

- positive recurrent, if

$$\begin{aligned} & \theta > 0, \gamma > 0, \\ or \quad & \theta = 0, \gamma > 0, \alpha < 0.5, \\ or \quad & \theta > 0, \gamma = 0, \beta > 0.5, \\ or \quad & \theta = 0, \gamma = 0, \alpha < 0.5, \beta > 0.5. \end{aligned}$$

- null recurrent, if

$$\begin{aligned} & \theta = 0, \gamma > 0, \alpha = 0.5, \\ or \quad & \theta > 0, \gamma = 0, \beta = 0.5, \\ or \quad & \theta = 0, \gamma = 0, \alpha = \beta = 0.5. \end{aligned}$$

- transient, if

$$\begin{aligned} & \theta = 0, \gamma > 0, \alpha > 0.5, \\ or \quad & \theta > 0, \gamma = 0, \beta < 0.5, \\ or \quad & \theta = 0, \gamma = 0, \alpha > 0.5, \beta < 0.5. \end{aligned}$$

In the analysis, we assume $\theta > 0$ and $\gamma > 0$ for simplicity. Therefore, this birth and death process is irreducible and positive recurrent and has a unique limiting distribution. Let

$$\pi_n = \Pr\left\{\lim_{t\to\infty} X(t) = n\right\}, \quad i \in \mathbb{Z}.$$

Using the standard theory of birth and death processes we see that the limiting distribution is given by:

$$\pi_n = \frac{\eta\alpha^{n-1}\lambda^n}{\prod_{j=1}^{n}((1-\alpha)\lambda + j\theta)}\pi_0, \quad n = 1, 2, \cdots \tag{3.3.1}$$

$$\pi_{-n} = \frac{(1-\eta)(1-\beta)^{n-1}\lambda^n}{\prod_{j=1}^{n}(\beta + j\gamma)}\pi_0, \quad n = 1, 2, \cdots \tag{3.3.2}$$

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty}\frac{\eta\alpha^{n-1}\lambda^n}{\prod_{j=1}^{n}((1-\alpha)\lambda + j\theta)} + \sum_{n=1}^{\infty}\frac{(1-\eta)(1-\beta)^{n-1}\lambda^i}{\prod_{j=1}^{n}(\beta + j\gamma)}\right)^{-1}. \tag{3.3.3}$$

3.4 PH-Renewal Arrival

In this section we consider the double-ended queue with a PH-Renewal arrival process and finite capacity C, that is, there can be at most C buyers or sellers in the system at any time. Any buyer (seller) arrivals when the system is full of buyers (sellers) are lost. We can approximate the double-ended queue with unlimited capacity by a double-ended queue with a sufficiently large finite capacity C.

Suppose the inter-arrival times are iid with phase type distribution with parameters (p, T) of order K. Here T is an invertible K by K matrix such that (using e to denote a K-dimensional column vector with all the entries equal to 1)

$$\begin{bmatrix} T & -Te \\ 0 & 0 \end{bmatrix}$$

is an infinitesimal generator matrix of a CTMC with state space $\{1, 2, \cdots, K+1\}$, with absorbing state $K+1$. The vector $p = [p_1, p_2, \cdots, p_K]$ is a probability mass function. We know that the mean inter-arrival time is $-pT^{-1}e$.

Let $X(t)$ be the number of traders in the system at time t, with the convention that $X(t) > 0$ means that there are $X(t)$ sellers in the system, while $X(t) < 0$ means that there are $-X(t)$ buyers in the system. Thus $-C \le X(t) \le C$ for all $t \ge 0$. Let $Y(t)$ be the phase of the arrival process at time t. That is $\{Y(t), t \ge 0\}$ is an irreducible continuous-time Markov chain (CTMC) on state space $\{1, 2, \cdots, K\}$ with generator matrix $T - Tep$. Although $\{X(t), t \ge 0\}$ is not a CTMC, the bivariate process $\{(X(t), Y(t)), t \ge 0\}$ is a CTMC on the state space $\Omega = \{(i, j) : -C \le i \le C, 1 \le j \le K\}$.

Borrowing notation from chapter 3 of Neuts(1981) [25], we define:

$$T^0 = -Te, \quad T^{00} = T^0 e', \quad A^0 = diag(p), \quad A^{00} = ep.$$

Thus we have $T^{00} A^0 = T^0 p$.

Since $\{Y(t), t \ge 0\}$ is irreducible and jumps in $\{X(t), t \ge 0\}$ are ± 1, $\{(X(t), Y(t)), t \ge 0\}$ is irreducible. Furthermore, it has finite state space Ω, hence it has a unique stationary distribution

$$\pi_{ij} = \lim_{t \to \infty} \Pr\{X(t) = i, Y(t) = j\}, \quad (i, j) \in \Omega.$$

Let $\pi_i = [\pi_{i,1}, \pi_{i,2}, \cdots, \pi_{i,K}]$ $(-C \le i \le C)$. Following the analysis in Neuts [25], we first introduce the following notation $(-C \le i \le C)$:

$$\begin{aligned} S_i &= \left(i\theta I + (1-\alpha)T^{00}A^0\right)\left((i-1)\theta I - T - (i-1)\theta A^{00} - (1-\alpha)T^{00}A^0\right)^{-1}, && (3.4.1) \\ B_i &= \left(i\gamma I + \beta T^{00}A^0\right)\left((i-1)\gamma I - T - (i-1)\gamma A^{00} - \beta T^{00}A^0\right)^{-1}, && (3.4.2) \end{aligned}$$

and

$$\begin{aligned} \tilde{S}_i &= S_C S_{C-1} \cdots S_i, && (3.4.3) \\ \tilde{B}_i &= B_C B_{C-1} \cdots B_i. && (3.4.4) \end{aligned}$$

In addition, we denote

$$\begin{aligned} D_i &= i\gamma I + \beta T^{00}A^0, \quad -C \le i \le -1, \\ D_0 &= \eta T^{00}A^0, \\ D_i &= \alpha T^{00}A^0, \quad 1 \le i \le C-1, \\ E_{-C} &= T - C\gamma I + (1-\beta)T^{00}A^0, \\ E_i &= T + i\gamma I, \quad -C+1 \le i \le 0, \\ E_i &= T - i\theta I, \quad 1 \le i \le C-1, \\ E_C &= T - C\theta I + \alpha T^{00}A^0, \\ F_i &= (1-\beta)T^{00}A^0, \quad -C+1 \le i \le -1, \\ F_0 &= (1-\eta)T^{00}A^0, \\ F_i &= i\theta I + (1-\alpha)T^{00}A^0, \quad 1 \le i \le C. \end{aligned}$$

Using this notation we see that the generator matrix Q of this Markov process $\{(X(t), Y(t)), t \ge$

0} is given by the following tridiagonal block matrix:

$$\begin{pmatrix} E_{-C} & D_{-C} & & & \\ F_{-C+1} & E_{-C+1} & D_{-C+1} & & \\ & \ddots & \ddots & \ddots & \\ & & F_{C-1} & E_{C-1} & D_{C-1} \\ & & & F_C & E_C \end{pmatrix}$$

.

Theorem 3.4.1. *When the arrival process is a PH-renewal process, the limiting probability vector π_i $(-C \leq i \leq C)$ are given by*

$$\pi_C = ds\tilde{\pi}_C, \tag{3.4.5}$$

$$\pi_{-C} = db\tilde{\pi}_{-C}, \tag{3.4.6}$$

$$\pi_i = \pi_C \tilde{S}_{i+1}, \quad 1 \leq i \leq C-1, \tag{3.4.7}$$

$$\pi_{-i} = \pi_{-C} \tilde{B}_{i+1}, \quad 1 \leq i \leq C-1, \tag{3.4.8}$$

$$\pi_0 = -\left(\pi_{-1}(\gamma I + \beta T^{00} A^0) + \pi_1(\theta I + (1-\alpha) T^{00} A^0)\right) T^{-1}, \tag{3.4.9}$$

where $\tilde{\pi}_{-C}$ and $\tilde{\pi}_C$ satisfy

$$\tilde{\pi}_{-C}(T - C\gamma I + T^{00} A^0 + C\gamma A^{00}) = 0, \tag{3.4.10}$$

$$\tilde{\pi}_C(T - C\theta I + T^{00} A^0 + C\theta A^{00}) = 0; \tag{3.4.11}$$

and the scalar constants s, b and d are given by

$$s = \eta \tilde{\pi}_{-C} \tilde{B}_2(\beta T^{00} A^0 + \gamma A^{00})(1), \tag{3.4.12}$$

$$b = (1-\eta)\tilde{\pi}_C \tilde{S}_2((1-\alpha) T^{00} A^0 + \theta A^{00})(1); \tag{3.4.13}$$

$$d = \left[b\tilde{\pi}_{-C}(I + \sum_{i=1}^{C} \tilde{B}_i)e + s\tilde{\pi}_C(I + \sum_{i=1}^{C} \tilde{S}_i)e \right]^{-1}. \tag{3.4.14}$$

Proof. By using the generator matrix, the steady-state equations are given by

$$\pi_{-C}(T - C\gamma I + (1-\beta)T^{00}A^0) + \pi_{-C+1}((1-\beta)T^{00}A^0) = 0, \tag{3.4.15}$$

$$\pi_{-i-1}((i+1)\gamma I + \beta T^{00}A^0) + \pi_{-i}(T - k\gamma I) + \pi_{-i+1}((1-\beta)T^{00}A^0) = 0, \ \ (2 \le i \le C-1), \tag{3.4.16}$$

$$\pi_{-2}(2\gamma I + \beta T^{00}A^0) + \pi_{-1}(T - \gamma I) + \pi_0((1-\eta)T^{00}A^0) = 0, \tag{3.4.17}$$

$$\pi_{-1}(\gamma I + \beta T^{00}A^0) + \pi_0(T) + \pi_1(\theta I + (1-\alpha)T^{00}A^0) = 0, \tag{3.4.18}$$

$$\pi_0(\eta T^{00}A^0) + \pi_1(T - \theta I) + \pi_2(2\theta I + (1-\alpha)T^{00}A^0) = 0, \tag{3.4.19}$$

$$\pi_{i-1}(\alpha T^{00}A^0) + \pi_i(T - j\theta I) + \pi_{i+1}((i+1)\theta I + (1-\alpha)T^{00}A^0) = 0, \ \ (2 \le i \le C-1), \tag{3.4.20}$$

$$\pi_{C-1}(\alpha T^{00}A^0) + \pi_C(T - C\theta I + \alpha T^{00}A^0) = 0. \tag{3.4.21}$$

Multiply both sides of equation (3.4.15) by e, we get

$$\pi_{-C}(\beta T^0 + C\gamma e) = \pi_{-C+1}((1-\beta)T^0).$$

Applying this result in equation (3.4.17) when $i = C-1$, we obtain

$$\pi_{-C+1}(\beta T^0 + (C-1)\gamma e) = \pi_{-C+2}((1-\beta)T^0).$$

Recursively, we have

$$\pi_{-i}(\beta T^0 + i\gamma e) = \pi_{-i+1}((1-\beta)T^0), \ \ 2 \le i \le C. \tag{3.4.22}$$

Similarly, we obtain

$$\pi_i((1-\alpha)T^0 + i\theta e) = \pi_{i-1}(\alpha T^0), \ \ 2 \le i \le C. \tag{3.4.23}$$

Besides, using these recursive equations we have

$$\pi_{-1}(\beta T^0 + \gamma e) = \pi_0((1-\eta)T^0), \tag{3.4.24}$$

$$\pi_1((1-\alpha)T^0 + \theta e) = \pi_0(\eta T^0). \tag{3.4.25}$$

Multiply p on the right of equations (3.4.22) to (3.4.25), we have:

$$\begin{aligned}
\pi_{-i}(\beta T^{00}A^0 + i\gamma A^{00}) &= \pi_{-i+1}((1-\beta)T^{00}A^0), \;\; 2 \le i \le C, && (3.4.26)\\
\pi_i((1-\alpha)T^{00}A^0 + i\theta A^{00}) &= \pi_{i-1}(\alpha T^{00}A^0), \;\; 2 \le i \le C, && (3.4.27)\\
\pi_{-1}(\beta T^{00}A^0 + \gamma A^{00}) &= \pi_0((1-\eta)T^{00}A^0), && (3.4.28)\\
\pi_1((1-\alpha)T^{00}A^0 + \theta A^{00}) &= \pi_0(\eta T^{00}A^0). && (3.4.29)
\end{aligned}$$

Substituting the above equations into the original steady state equations and simplifying, we have

$$\begin{aligned}
\pi_{-i-1}\left((i+1)\gamma I + \beta T^{00}A^0\right) + \pi_{-i}\left(T - i\gamma I + \beta T^{00}A^0 + i\gamma A^{00}\right) &= 0, \;\; 1 \le i \le C-1\\
\pi_{i+1}\left((i+1)\theta I + (1-\alpha)T^{00}A^0\right) + \pi_i\left(T - i\theta I + (1-\alpha)T^{00}A^0 + i\theta A^{00}\right) &= 0, \;\; 1 \le i \le C-1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\pi_{-i} &= \pi_{-i-1}\left((i+1)\gamma I + \beta T^{00}A^0\right)\left(i\gamma I - T - \beta T^{00}A^0 - i\gamma A^{00}\right)^{-1}, \;\; 1 \le i \le C-1\\
\pi_i &= \pi_{i+1}\left((i+1)\theta I + (1-\alpha)T^{00}A^0\right)\left(i\theta I - T - (1-\alpha)T^{00}A^0 - i\theta A^{00}\right)^{-1}, \;\; 1 \le i \le C-1
\end{aligned}$$

With the notations defined in equations (3.4.2) to (3.4.3), we have the results for π_i and π_{-i} in equations (3.4.7) and (3.4.8).

Substituting equation (3.4.26) and (3.4.27) when $i = C$ into the equations (3.4.15) and (3.4.21) respectively, we have

$$\pi_{-C}(T - C\gamma I + T^{00}A^0 + C\gamma A^{00}) = 0,$$

$$\pi_C(T - C\theta I + T^{00}A^0 + C\theta A^{00}) = 0.$$

Also the matrix $T - C\gamma I + T^{00}A^0 + C\gamma A^{00}$ and $T - C\theta I + T^{00}A^0 + C\theta A^{00}$ are stochastic, in which the diagonal entries are negative and other entries are nonnegative, and the sum of each row is equal to zero. Hence $\pi_{-C} = \tilde{b}\tilde{\pi}_{-C}$ and $\pi_C = \tilde{s}\tilde{\pi}_C$, where $\tilde{b}$ and $\tilde{s}$ are constants.

Since we have equations (3.4.24) and (3.4.25), we get

$$\frac{1}{1-\eta}\pi_{-1}(\beta T^{00}A^0+\gamma A^{00})=\frac{1}{\eta}\pi_1((1-\alpha)T^{00}A^0+\theta A^{00}).$$

Thus,

$$\frac{1}{1-\eta}\tilde{b}\pi_{-C}\tilde{B}_2(\beta T^{00}A^0+\gamma A^{00})=\frac{1}{\eta}\tilde{s}\pi_C\tilde{S}_2((1-\alpha)T^{00}A^0+\theta A^{00}).$$

Using the notations of s and b in equations (3.4.12) and (3.4.13), $\tilde{s}$ and $\tilde{b}$ are given by $\tilde{s}=ds$ and $\tilde{b}=db$, where d is a constant.

Finally, we determine d by normalization. From equation (3.4.18) and the normalization equation

$$\left(\sum_{i=-C}^{-1}\pi_i+\pi_0+\sum_{i=1}^{C}\pi_i\right)e=1,$$

we have

$$d\left[b\tilde{\pi}_{-C}\left(I+\sum_{i=-C}^{-1}\tilde{B}_i\right)+s\tilde{\pi}_C\left(I+\sum_{i=1}^{C}\tilde{S}_i\right)\right]e=1.$$

Thus the equation (3.4.14) follows. ■

Corollary 3.4.1. *From the results of Theorem 3.4.1, the limiting distribution of* $\{X(t),t\geq 0\}$, $\hat{\pi}_i(-C\leq i\leq C)$, *is given by*

$$\hat{\pi}_i=\lim_{t\to+\infty}\mathbb{P}\{X(t)=i\}=\pi_i e,\ \ -C\leq i\leq C. \tag{3.4.30}$$

Proof. This is obvious from the definition of $\{(X(t),Y(t)),t\geq 0\}$ and π_i. ■

3.5 Special Case: Erlang and Hyper-exponential Arrivals

In this section, we consider the special case when the inter-arrival time of the PH-renewal process follows an Erlang distribution or Hyper-exponential distribution, and obtain the explicit expressions of $\tilde{\pi}_C$ and $\tilde{\pi}_{-C}$ as given in the following theorems:

Theorem 3.5.1. *If the inter-arrival times follow an Erlang($k, k\lambda$) distribution, $\tilde{\pi}_{-C}$ and $\tilde{\pi}_C$ are given by the following equations:*

$$\tilde{\pi}_{-C} = \left(1, \frac{k\lambda}{k\lambda + C\gamma}, \left(\frac{k\lambda}{k\lambda + C\gamma}\right)^2, \cdots, \left(\frac{k\lambda}{k\lambda + C\gamma}\right)^k\right), \tag{3.5.1}$$

$$\tilde{\pi}_{C} = \left(1, \frac{k\lambda}{k\lambda + C\theta}, \left(\frac{k\lambda}{k\lambda + C\theta}\right)^2, \cdots, \left(\frac{k\lambda}{k\lambda + C\theta}\right)^k\right). \tag{3.5.2}$$

Proof: When inter-arrival time follows an Erlang($k, k\lambda$) distribution, we have

$$T = \begin{pmatrix} -k\lambda & k\lambda & & \\ & \ddots & \ddots & \\ & & -k\lambda & k\lambda \\ & & & -k\lambda \end{pmatrix}, \; T^0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ k\lambda \end{pmatrix}, \; T^{00} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ k\lambda & k\lambda & \cdots & k\lambda \end{pmatrix}, \; A^0 =$$

$$\begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \vdots & \\ & & & 0 \end{pmatrix}, p = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}, e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, A^{00} = ep = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Therefore from the equation (3.4.10), we have

$$k\lambda\tilde{\pi}_{-C}(1) = (k\lambda + C\gamma)\,\tilde{\pi}_{-C}(2),$$

$$k\lambda\tilde{\pi}_{-C}(2) = (k\lambda + C\gamma)\,\tilde{\pi}_{-C}(3),$$

$$\cdots\cdots$$

$$k\lambda\tilde{\pi}_{-C}(k-1) = (k\lambda + C\gamma)\,\tilde{\pi}_{-C}(k),$$

$$-k\lambda\tilde{\pi}_C(1) + C\gamma\left(\sum_{i=2}^{k-1}\tilde{\pi}_C(i)\right) + (C\gamma + k\lambda)\,\pi_C(k) = 0.$$

Thus we obtain an explicit solution for π_{-C} from above equations, which is shown in equation 3.5.1. Similarly we obtain an explicit solution for π_C, which is shown in equation 3.5.2. ■

Theorem 3.5.2. *If the inter-arrival times follow an Hyper-exponential distribution with parameters* $(\lambda_1, \cdots, \lambda_k)$ *and* $(p_1, \cdots, p_k)$, $\tilde{\pi}_{-C}$ *and* $\tilde{\pi}_C$ *are given by the following equations:*

$$\tilde{\pi}_{-C} = \left(\frac{p_1}{C\gamma + \lambda_1}, \frac{p_2}{C\gamma + \lambda_2}, \cdots, \frac{p_k}{C\gamma + \lambda_k}\right), \tag{3.5.3}$$

$$\tilde{\pi}_{C} = \left(\frac{p_1}{C\theta + \lambda_1}, \frac{p_2}{C\theta + \lambda_2}, \cdots, \frac{p_k}{C\theta + \lambda_k}\right). \tag{3.5.4}$$

Proof: When inter-arrival time follows an Hyper-exponential distribution with parameters $(\lambda_1, \cdots, \lambda_k)$ and $(p_1, \cdots, p_k)$ $(\sum_{i=1}^k p_k = 1)$, we have

$$T = \begin{pmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_k \end{pmatrix}, \ T^0 = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix}, \ T^{00} = \begin{pmatrix} \lambda_1 & \lambda_1 & \cdots & \lambda_1 \\ \lambda_2 & \lambda_2 & \cdots & \lambda_2 \\ \vdots & \vdots & & \vdots \\ \lambda_k & k\lambda_k & \cdots & \lambda_k \end{pmatrix}, \ A^0 =$$

$$\begin{pmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \vdots & \\ & & & \alpha_k \end{pmatrix}, p = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_k \end{pmatrix}, e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, A^{00} = ep = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_k \\ \alpha_1 & \alpha_2 & \cdots & \alpha_k \\ \vdots & \vdots & & \vdots \\ \alpha_k & \alpha_k & \cdots & \alpha_k \end{pmatrix}.$$

Therefore from the equation (3.4.10), we have

$$\begin{aligned}
&\sum_{i=1}^{k} \tilde{\pi}_{-C}(i) \left(C\gamma + \lambda_i\right) p_1 = \tilde{\pi}_{-C}(1) \left(C\gamma + \lambda_1\right), \\
&\sum_{i=1}^{k} \tilde{\pi}_{-C}(i) \left(C\gamma + \lambda_i\right) p_2 = \tilde{\pi}_{-C}(2) \left(C\gamma + \lambda_2\right), \\
&\cdots\cdots \\
&\sum_{i=1}^{k} \tilde{\pi}_{-C}(i) \left(C\gamma + \lambda_i\right) p_k = \tilde{\pi}_{-C}(k) \left(C\gamma + \lambda_k\right).
\end{aligned}$$

Thus we obtain an explicit solution for π_{-C}, which is given by equation 3.5.3. Similarly we obtain an explicit solution for π_C, which is shown in equation 3.5.4. ■

3.6 Numerical Example

In this section, we use simulations to test the performance of our algorithm. We consider the following inter-arrival time distributions:

- Erlang$(2, 2\lambda)$: $F(x) = 1 - e^{-2\lambda x} - 2\lambda x e^{-2\lambda x}$,
- Hyper-exponential with parameter $(\lambda/2, 2\lambda; 1/3, 2/3)$: $F(x) = \frac{1}{3}(1 - e^{-\frac{1}{2}\lambda x}) + \frac{2}{3}(1 - e^{-2\lambda x})$.

We want to capture the situation where the buyer arrival rate exceeds the seller arrival rate when there are sellers waiting (its a buyers' market), and the seller arrival rate exceeds the buyer arrival rate when there are buyers waiting (its a sellers' market). Therefore, we choose $\alpha < \frac{1}{2}$ and $\beta > \frac{1}{2}$. Without loss of generality, we choose $\lambda = 1$.

We consider the following values for $(\alpha, \beta, \eta) = (0.1, 0.7, 0.5)$, $(0.3, 0.6, 0.5)$ and $(0.45, 0.55, 0.5)$. We choose the following reneging rates $(\theta, \gamma) = \lambda(\alpha, \beta)$, $0.1\lambda(\alpha, \beta)$ and $0.01\lambda(\alpha, \beta)$. Thus when $(\theta, \gamma) = 0.1(\alpha, \beta)$, the seller's (buyer's) expected reneging time is 10 times the buyer's (seller's) expected inter-arrival time.

3.6.1 Limiting Probability Mass Function

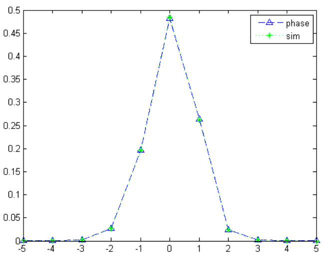
We compare the limiting probability mass function (pmf) of $X(t)$ derived from the simulation and our method. When applying simulation method, we compute the limiting pmf of $X(t)$ by using N replications of the simulation using Matlab. Each replication consists of simulating the system for $0 \leq t \leq T$ and the limiting probabilities $\{\pi_i\}$ are computed by using the sample paths over $t \in [\tau, T]$, where $\tau < T$ is a given warmup period. Here we choose $i \in [-1000, 1000]$. Since the range of this interval is large enough so that the capacity of the simulated queue can be considered to be infinity. Let $X^k(t)$ be the state of the system at time t in the k-th replication, $k = 1, 2, \cdots, N$, $0 \le t \le T$. Using these sample paths, we compute

$$\pi_i^s = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{T-\tau} \int_{\tau}^{T} 1_{\{X^k(t)=i\}} dt, \quad -1000 \le i \le 1000.$$

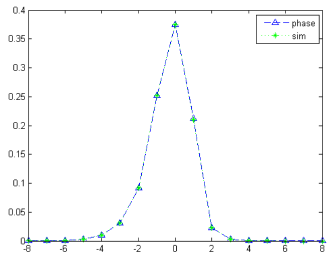
We evaluate the limiting pmf using the parameter $(N, \tau, T) = (400, 5000, 20000)$.

Note that our algorithm assume a finite capacity C for buyers and sellers. One can approximate an infinite capacity ($C = \infty$) queue by using a sufficient large C. Here we use $C = 20$.

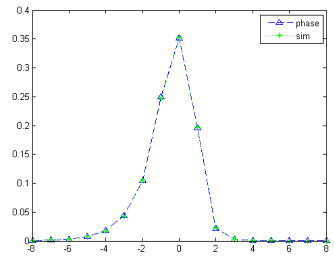
Next we use simulation to verify this approximation. The comparisons of limiting density are shown in Figure 3.3 and 3.4. From these graphs, we observe that the pmf from our algorithm is almost exact the same as the one from simulation.



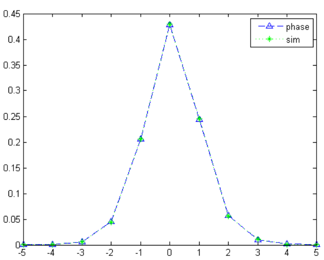
(a) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.1, 0.7)$



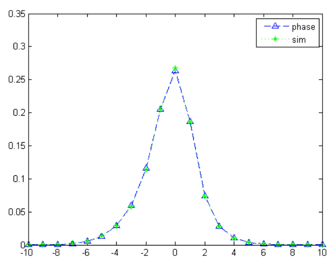
(b) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.01, 0.07)$



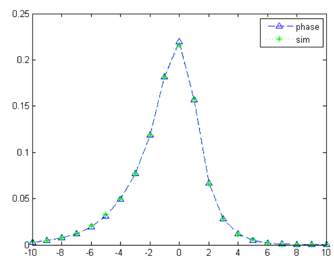
(c) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.001, 0.007)$



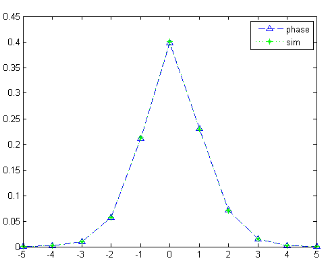
(a) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.1, 0.7)$



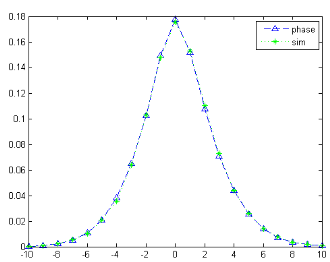
(b) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.01, 0.07)$



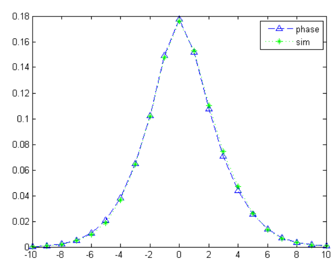
(c) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.001, 0.007)$



(a) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.1, 0.7)$

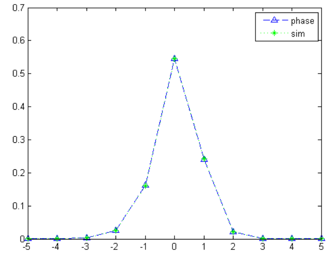


(b) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.01, 0.07)$

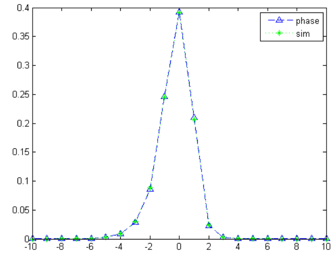


(c) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.001, 0.007)$

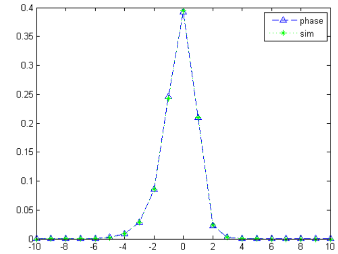
Figure 3.3: Limiting pmf functions by simulation method and our method, when inter-arrival times follow Erlang distribution and $\eta = 0.5$



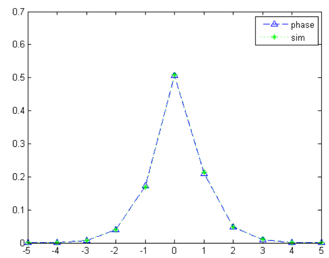
(a) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.1, 0.7)$



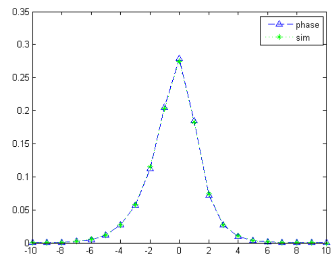
(b) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.01, 0.07)$



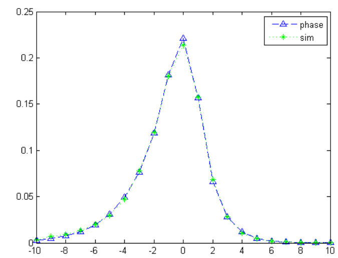
(c) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.001, 0.007)$



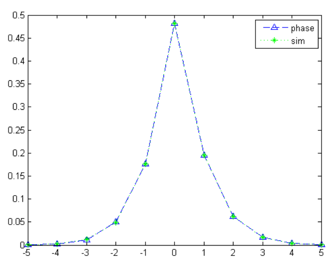
(a) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.1, 0.7)$



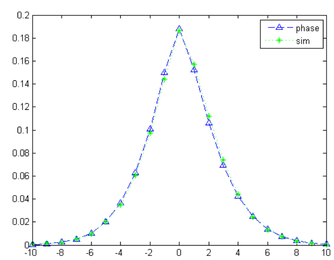
(b) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.01, 0.07)$



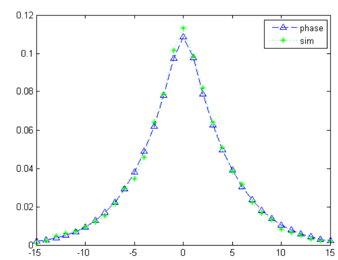
(c) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.001, 0.007)$



(a) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.1, 0.7)$



(b) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.01, 0.07)$



(c) $(\alpha, \beta, \theta, \gamma) = (0.1, 0.7, 0.001, 0.007)$

Figure 3.4: Limiting pmf functions by simulation method and our method, when inter-arrival times follow Hyper-exponential distribution and $\eta = 0.5$

3.6.2 Expected queue length

In this subsection, we test the performance of the expected queue length. More precisely, we aim to find the queue length of sellers (L_s), queue length of buyers (L_b) and the total queue length L, that is

$$L_s = \sum_{i=1}^{C} i\hat{\pi}_i, \quad L_b = \sum_{i=-C}^{-1} -i\hat{\pi}_i, \quad L_s = \sum_{i=-C}^{C} |i|\hat{\pi}_i.$$

These performance measures are compared with the ones from simulation, shown in the table 3.1-3.2. In the columns of L_s^s, L_b^s and L^s, we evaluate the the above performance measures by simulation method, and obtain the 95% confidence interval. In the columns L_s^a, L_b^a and L^a, we obtain the performance measures by our method, and obtain the relative error of each performance measure to the one from simulation method. From the table we find that if the reneging rate is very small, the expected queue length largely depends on the value of α and β. Thus if α is closer to 0.5, the expected queue length for sellers is larger. On the other hand, if β is closer to 0.5, the expected queue length for buyers is larger. This is true since the $\alpha(\beta)$ closer to 0.5 will result in a less frequent trades and a larger queue length. Besides, we find that the expected queue length for Erlang inter-arrival is larger than the one for the hyper-exponential arrival.

(α, β, η)	(θ, γ)	L_b^s	L_b^a	L_s^s	L_s^a	L^s	L^a
(0.1, 0.7, 0.5)	(0.1, 0.7)	0.26 ±0.00	0.26 0.09%	0.32 ±0.00	0.32 0.00%	0.58 ±0.00	0.58 0.01%
	(0.01, 0.07)	0.58 ±0.00	0.58 0.05%	0.27 ±0.00	0.27 -0.04%	0.84 ±0.00	0.84 -0.05%
	(0.001, 0.007)	0.73 ±0.00	0.73 0.02%	0.25 ±0.00	0.25 0.15%	0.98 ±0.00	0.98 -0.16%
(0.3, 0.6, 0.5)	(0.3, 0.6)	0.32 ±0.00	0.32 -0.05%	0.40 ±0.00	0.40 0.07%	0.71 ±0.00	0.71 -0.02%
	(0.03, 0.06)	0.84 ±0.00	0.84 0.09%	0.49 ±0.00	0.49 0.17%	1.34 ±0.00	1.33 -0.2%
	(0.003, 0.006)	1.39 ±0.01	1.39 -0.03%	0.47 ±0.00	0.47 -0.58%	1.87 ±0.01	1.86 -0.61%
(0.45, 0.55, 0.5)	(0.45, 0.55)	0.36 ±0.00	0.36 0.25%	0.43 ±0.00	0.43 -0.15%	0.80 ±0.00	0.80 0.02%
	(0.045, 0.055)	0.93 ±0.01	0.94 0.41%	1.08 ±0.01	1.07 -0.67%	2.01 ±0.01	2.01 0.08%
	(0.0045, 0.0055)	1.83 ±0.02	1.82 -0.35%	1.95 ±0.03	1.93 -0.71%	3.78 ±0.02	3.76 -0.64%

Table 3.1: Measures of queue length when the inter-arrival time follows the Erlang distribution

(α, β, η)	(θ, γ)	L_b^s	L_b^a	L_s^s	L_s^a	L^s	L^a
(0.1, 0.7, 0.5)	(0.1, 0.7)	0.22	0.22	0.29	0.29	0.51	0.51
		±0.00	0.05%	±0.00	-0.17%	±0.00	0.01%
	(0.01, 0.07)	0.56	0.56	0.26	0.26	0.82	0.82
		±0.00	0.19%	±0.00	0.07%	±0.00	0.04%
	(0.001, 0.007)	0.73	0.73	0.25	0.25	0.97	0.98
		±0.00	0.00%	±0.00	0.27%	±0.00	0.11%
(0.3, 0.6, 0.5)	(0.3, 0.6)	0.28	0.28	0.34	0.34	0.63	0.63
		±0.00	0.04%	±0.00	-0.08%	±0.00	0.00%
	(0.03, 0.06)	0.83	0.82	0.48	0.48	1.30	1.30
		±0.00	-0.10%	±0.00	-0.53%	±0.00	0.12%
	(0.003, 0.006)	1.39	1.38	0.47	0.47	1.86	1.85
		±0.01	-0.49%	±0.00	0.39%	±0.01	-0.19%
(0.45, 0.55, 0.5)	(0.45, 0.55)	0.32	0.32	0.39	0.39	0.71	0.71
		±0.00	-0.17%	±0.00	-0.07%	±0.00	-0.07%
	(0.045, 0.055)	0.92	0.92	1.06	1.06	1.98	1.98
		±0.01	-0.46%	±0.01	-0.31%	±0.01	-0.27%
	(0.0045, 0.0055)	1.82	1.82	1.94	1.93	3.79	3.75
		±0.02	-0.46%	±0.03	-0.58%	±0.03	-1.11%

Table 3.2: Measures of queue length when the inter-arrival time follows the Hyper-exponential distribution

3.6.3 Fraction of reneging

In this subsection, we test the performance of the fraction of sellers/buyers reneging from the queue. Let the fraction of sellers reneging be f_s and fraction of buyers reneging be f_b respectively. From our method, we have

$$\begin{aligned} f_s &= \frac{\sum_{i=1}^{c} \hat{\pi}_i i\theta}{\sum_{i=-c}^{-1} \hat{\pi}_i \beta\lambda + \hat{\pi}_0 \eta\lambda + \sum_{i=1}^{c} \hat{\pi}_i \alpha\lambda} \\ f_b &= \frac{\sum_{i=-c}^{-1} \hat{\pi}_i (-i)\gamma}{\sum_{i=-c}^{-1} \hat{\pi}_i (1-\beta)\lambda + \hat{\pi}_0 (1-\eta)\lambda + \sum_{i=1}^{c} \hat{\pi}_i (1-\alpha)\lambda}. \end{aligned}$$

These performance measures are compared with the ones from simulation, shown in the table 3.3-3.4. In the columns of f_s^s and f_b^s, we evaluate the the above performance measures by simulation method, and obtain the 95% confidence interval. In the columns f_s^a and f_b^a, we obtain the performance measures by our method, and also obtain the relative error of each performance measure to the one from simulation method. From the table, we find that when the reneging rate becomes smaller, the fraction of reneging will becomes smaller. The fraction of reneging for Erlang inter-arrival is larger than the one for the hyper-exponential arrival.

(α, β, η)	(θ, γ)	f_b^s	f_b^a	f_s^s	f_s^a
(0.1, 0.7, 0.5)	(0.1, 0.7)	0.32	0.32	0.08	0.07
		±0.00	0.63%	±0.00	-0.9%
	(0.01, 0.07)	0.08	0.08	0.01	0.01
		±0.00	0.06%	±0.00	-0.07%
	(0.001, 0.007)	0.01	0.01	0.00	0.00
		±0.00	0.76%	±0.00	-1.02%
(0.3, 0.6, 0.5)	(0.3, 0.6)	0.36	0.36	0.26	0.26
		±0.00	0.06%	±0.00	0.45%
	(0.03, 0.06)	0.10	0.10	0.03	0.03
		±0.00	-0.13%	±0.00	0.09%
	(0.003, 0.006)	0.02	0.02	0.00	0.00
		±0.00	0.62%	±0	-0.19%
(0.45, 0.55, 0.5)	(0.45, 0.55)	0.4	0.4	0.39	0.39
		±0.00	0.11%	±0.00	-0.02%
	(0.045, 0.055)	0.10	0.10	0.10	0.10
		±0.00	0.18%	±0.00	0.19%
	(0.0045, 0.0055)	0.02	0.02	0.02	0.02
		±0.00	0.29%	±0.00	-0.78%

Table 3.3: Fraction of reneging when the inter-arrival time follows the Erlang distribution

(α, β, η)	(θ, γ)	f_b^s	f_b^a	f_s^s	f_s^a
(0.1, 0.7, 0.5)	(0.1, 0.7)	0.28	0.28	0.07	0.07
		±0.00	-0.62%	±0.00	0.81%
	(0.01, 0.07)	0.08	0.07	0.01	0.01
		±0.00	-0.42%	±0.00	0.39%
	(0.001, 0.007)	0.01	0.01	0.00	0.00
		±0.00	-0.04%	±0.00	3.32%
(0.3, 0.6, 0.5)	(0.3, 0.6)	0.32	0.32	0.22	0.22
		±0.00	0.36%	±0.00	-0.36%
	(0.03, 0.06)	0.10	0.10	0.03	0.03
		±0.00	-0.27%	±0.00	-0.37%
	(0.003, 0.006)	0.02	0.02	0.00	0.00
		±0.00	1.06%	±0.00	-0.36%
(0.45, 0.55, 0.5)	(0.45, 0.55)	0.35	0.35	0.35	0.35
		±0.00	0.02%	±0.00	-0.01%
	(0.045, 0.055)	0.10	0.10	0.10	0.10
		±0.00	-0.10%	±0.00	-0.18%
	(0.0045, 0.0055)	0.02	0.02	0.02	0.02
		±0.00	0.49%	±0.00	-0.55%

Table 3.4: Fraction of reneging when the inter-arrival time follows the Hyper-exponential distribution

3.6.4 Expected waiting time

In this subsection, we test the performance of expected waiting time of sellers/buyers. Let the expected waiting time of buyers and selelrs be w_b and w_s respectively. Applying Little's law, we have

$$\begin{aligned} w_b &= \frac{\sum_{i=-C}^{-1} \hat{\pi}_i(-i)}{\sum_{i=-C}^{-1} \hat{\pi}_i(1-\beta)\lambda + \hat{\pi}_0(1-\eta)\lambda + \sum_{i=1}^{C} \hat{\pi}_i(1-\alpha)\lambda}, \\ w_s &= \frac{\sum_{i=1}^{C} \hat{\pi}_i i}{\sum_{i=-C}^{-1} \hat{\pi}_i\beta\lambda + \hat{\pi}_0\eta\lambda + \sum_{i=1}^{C} \hat{\pi}_i\alpha\lambda}. \end{aligned}$$

These performance measures are compared with the ones from simulation, shown in the table 3.5-3.6. In the columns of w_s^s and w_b^s, we evaluate the the above performance measures by simulation method, and obtain the 95% confidence interval. In the columns w_s^a and w_b^a, we obtain the performance measures by our method, and also obtain the relative error of each performance measure to the one from simulation method. From the tables, we find that with the same arrival process, the smaller reneging rate results in a larger expected waiting time. And the expected waiting time for the Erlang inter-arrival is larger than the one for the hyper-exponential arrival.

(α, β, η)	(θ, γ)	w_b^s	w_b^a	w_s^s	w_s^a
(0.1, 0.7, 0.5)	(0.1, 0.7)	0.45 ±0.00	0.45 0.64%	0.75 ±0.00	0.75 -0.88%
	(0.01, 0.07)	1.11 ±0.00	1.11 0.19%	0.55 ±0.00	0.55 -0.34%
	(0.001, 0.007)	1.45 ±0.01	1.46 0.23%	0.50 ±0.00	0.50 0.09%
(0.3, 0.6, 0.5)	(0.3, 0.6)	0.59 ±0.00	0.59 0.01%	0.85 ±0.00	0.86 0.24%
	(0.03, 0.06)	1.63 ±0.01	1.63 0.10%	1.02 ±0.00	1.02 0.03%
	(0.003, 0.006)	2.74 ±0.02	2.76 0.55%	0.95 ±0.01	0.95 -0.26%
(0.45, 0.55, 0.5)	(0.45, 0.55)	0.72 ±0.00	0.72 0.16%	0.87 ±0.00	0.87 -0.13%
	(0.045, 0.055)	1.86 ±0.01	1.87 0.35%	2.16 ±0.01	2.15 -0.26%
	(0.0045, 0.0055)	3.64 ±0.05	3.64 -0.14%	3.89 ±0.05	3.87 -0.35%

Table 3.5: Expected waiting time when the inter-arrival time follows the Erlang distribution

$(\alpha,\, \beta,\, \eta)$	$(\theta,\, \gamma)$	w_b^s	w_b^a	w_s^s	w_s^a
(0.1, 0.7, 0.5)	(0.1, 0.7)	0.40	0.40	0.66	0.67
		±0.00	-0.66%	±0.00	0.72%
	(0.01, 0.07)	1.08	1.07	0.55	0.55
		±0.00	-0.40%	±0.00	0.49%
	(0.001, 0.007)	1.46	1.45	0.50	0.50
		±0.01	-0.44%	±0.00	-0.03%
(0.3, 0.6, 0.5)	(0.3, 0.6)	0.53	0.53	0.74	0.73
		±0.00	0.17%	±0.00	-0.44%
	(0.03, 0.06)	1.59	1.59	1.00	0.99
		±0.01	-0.03%	±0.00	-0.22%
	(0.003, 0.006)	2.75	2.75	0.94	0.95
		±0.02	-0.02%	±0.01	0.10%
(0.45, 0.55, 0.5)	(0.45, 0.55)	0.64	0.64	0.77	0.78
		±0.00	-0.22%	±0.00	0.22%
	(0.045, 0.055)	1.84	1.84	2.13	2.12
		±0.01	-0.07%	±0.01	-0.80%
	(0.0045, 0.0055)	3.66	3.63	3.86	3.86
		±0.05	-0.78%	±0.05	-0.05%

Table 3.6: Expected waiting time when the inter-arrival time follows the Hyper-exponential distribution

3.7 Extensions

We end this chapter with suggestions for four extensions.

1. In this chapter we have assumed that the patience time of the traders follows an exponential distribution. It may be interesting if the distribution is general. In this case $\{X(t), t \geq 0\}$ is no longer Markov.

2. We have assumed that the arrival processes of the buyers and sellers are PH-renewal process. It would be interesting to consider an extension where the inter-arrival time follows an general distribution.

3. In this chapter we only consider the limiting behavior of the double-ended queueing process. In order to consider the transient behavior, it would be interesting to derive the

diffusion approximation to the system.

4. We only consider the double-ended queue with one dimension. It would be interesting to extend our analysis to multidimensional queues.

CHAPTER 4

A Stochastic Model of Order Books with Bouncing Geometric Brownian Motions

4.1 Introduction

A trading system consists of two types of arrivals, namely, the "sellers" and "buyers". Each buyer and seller is interested in trading one unit of a security (such as a stock). Every trader has his/her own trading price. The largest trading price of buyers is the market bid price, while the smallest trading price of sellers is the market ask price. Because of the new arrivals and reneging (that is, leaving the market without trading) of traders, the market bid and ask prices can move upwards and downwards. When a market bid price reaches market ask price, a trade occurs. Immediately after the trade the market bid price moves down the the new largest bid price, while the market ask price moves up to the new smallest ask price. In this chapter, we consider to use two independent geometric Brownian motions (GBM) to describe the movement of market prices. From the model we study the inter-trading time and the trading price, and obtain the explicit estimators of each parameter of our model. Finally we derive a simple forecasting formula by applying this model.

In this chapter, we first formulate the trading system as a bouncing GBMs in section 4.2. We also obtain a joint Laplace transform of inter-trading time and logarithmic increase of trading price in this section. In section 4.3, we derive the estimators of the parameters of the model. In section 4.4, we apply this model to real data and derive a simple forecasting formula. The performance of the forecasting is validated through numerical examples. We also and make comments on the possible extensions in section 4.5.

4.2 Model Formulation

In this section, the prices offered by sellers or buyers are assumed to be continuous, taking values in $[0, \infty)$. Because trading occurs when the market ask price is matched with the market bid price, we are only concerned about the dynamics of the market ask price and market bid price. Let $A(t)$ and $B(t)$ be the market ask price and market bid price at time t respectively ($A(0) > B(0)$). Until the market ask price is matched with the market bid price, the market ask price can decrease because of the arrivals of new sellers with smaller ask prices; it can also increase because of the reneging of the seller with current market ask price (that is, the seller can leave without a trade). Similarly the market bid price can also increase and decrease.

Let T_1 be the first time the two processes meet:

$$T_1 = \min\{t > 0 : A(t) = B(t)\}. \tag{4.2.1}$$

Thus we call $(0, T_1]$ as the first trading period. Let P_1 be the first trading price, hence

$$P_1 = A(T_1) = B(T_1).$$

After the trading occurs, the market ask price and the market bid price will separate so that $A(T_1+) > B(T_1+)$. Consequently, we define for $k > 1$:

$$\begin{aligned} T_k &= \min\{t > T_{k-1} : A(t) = B(t)\}, \\ P_k &= A(T_k) = B(T_k). \end{aligned}$$

Therefore, T_k is the kth trading time and P_k is the kth trading price. The dynamics of the market ask price and the market bid price is shown in Fig (4.1).

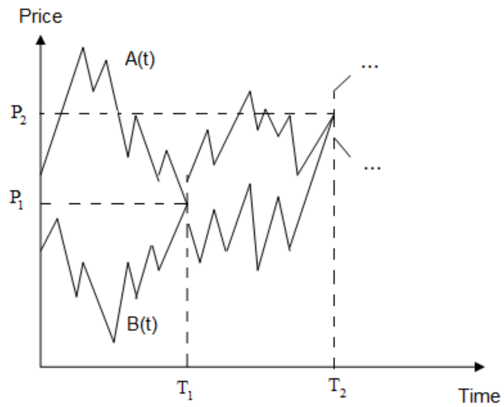


Figure 4.1: Dynamics of the ask price and the bid price

4.2.1 First trading period

Assume that $\{A(t), t \geq 0\}$ and $\{B(t), t \geq 0\}$ behave like two independent GBMs with $A(0) > B(0)$, until they meet for the first time at time T_1 in equation (4.2.1). That is we assume that, for $t \in [0, T_1]$,

$$A(t) = \exp\{a + (\tilde{\mu}_a - \sigma_a^2/2)t + \sigma_a W_a(t)\},$$

$$B(t) = \exp\{b + (\tilde{\mu}_b - \sigma_b^2/2)t + \sigma_b W_b(t)\},$$

where $\{W_a(t), t \geq 0\}$ and $\{W_b(t), t \geq 0\}$ are two independent standard Brownian motions, and $A(0) = e^a$ and $B(0) = e^b$ are the initial market ask price and market bid price. We assume that $a > b$, so that $A(0) > B(0)$. We assume that $\tilde{\mu}_a - \sigma_a^2/2 < \tilde{\mu}_b - \sigma_b^2/2$ so that $T_1 < \infty$ with probability 1. We aim to compute the joint Laplace transform of the bivariate random variable $(\ln P_1, T_1)$. For convenience, we first denote

$$\mu_a = \tilde{\mu}_a - \sigma_a^2/2, \quad \mu_b = \tilde{\mu}_b - \sigma_b^2/2$$

.

Theorem 4.2.1. *Given $A(0) > B(0)$, the joint Laplace transform of $(\ln P_1, T_1)$ are given by*

$$E[\exp(-s \ln P_1 - tT_1)] = \exp\{\theta_1(s,t)a + \theta_2(s,t)b\}, \tag{4.2.2}$$

where $\theta_1(s,t)$ *and* $\theta_2(s,t)$ *are given by*

$$\theta_1(s,t) = \frac{(\mu_b - \mu_a - s\sigma_b^2) - \sqrt{(\mu_b - \mu_a - s\sigma_b^2)^2 - (\sigma_a^2 + \sigma_b^2)(s^2\sigma_b^2 - 2t - 2s\mu_b)}}{\sigma_a^2 + \sigma_b^2}, \quad (4.2.3)$$

$$\theta_2(s,t) = -s - \frac{(\mu_b - \mu_a - s\sigma_b^2) - \sqrt{(\mu_b - \mu_a - s\sigma_b^2)^2 - (\sigma_a^2 + \sigma_b^2)(s^2\sigma_b^2 - 2t - 2s\mu_b)}}{\sigma_a^2 + \sigma_b^2} \quad (4.2.4)$$

Proof. Let

$$\begin{aligned} Y_a(t) &= \exp\{\theta_1(\mu_a t + \sigma_a W_a(t)) - (\theta_1\mu_a + \frac{1}{2}\theta_1^2\sigma_a^2)t\}, \\ Y_b(t) &= \exp\{\theta_2(\mu_b t + \sigma_b W_b(t)) - (\theta_2\mu_b + \frac{1}{2}\theta_2^2\sigma_b^2)t\}, \end{aligned}$$

where θ_1 and θ_2 are any real numbers. Therefore, $\{Y_a(t), t \geq 0\}$ and $\{Y_b(t), t \geq 0\}$ are two independent martingales (see Karlin (1975)[18]), and T_1 is a stopping time for the bivariate martingale $\{(Y_a(t), Y_b(t)), t \geq 0\}$. Because $\mu_a < \mu_b$ and $A(0) > B(0)$, $\Pr\{T_1 < +\infty\} = 1$. Hence we can apply the Optional Stopping Theorem, which yields

$$(E(Y_a(T_1)), E(Y_b(T_1))) = (E(Y_a(0)), E(Y_b(0))).$$

Now denote $Y(t) = Y_a(t)Y_b(t)$, we obtain $E(Y(T_1)) = E(Y(0))$. This yields

$$E\left\{\exp\{(\theta_1 + \theta_2)\ln P_1 - (\theta_1\mu_a + \frac{1}{2}\theta_1^2\sigma_a^2 + \theta_2\mu_b + \frac{1}{2}\theta_2^2\sigma_b^2)T_1\}\right\} = \exp\{\theta_1 a + \theta_2 b\}.$$

Let

$$\begin{aligned} &\theta_1 + \theta_2 = -s, \\ &\theta_1\mu_a + \frac{1}{2}\theta_1^2\sigma_a^2 + \theta_2\mu_b + \frac{1}{2}\theta_2^2\sigma_b^2 = t. \end{aligned}$$

Solving θ_1 and θ_2 in terms of s and t, we obtain equation (4.2.3) and (4.2.4). Therefore, the Laplace transform of (P_1, T_1) is given by equation (4.2.2) ■

Remark: When solving θ_1 and θ_2 in terms of s and t, we should have two groups of solutions. But we only choose above solution, since the other solution will produce the negative value when we calculate the expectations in next section.

4.2.2 Multiple trading periods

In this section, we consider multiple trading periods. Let T_n be the nth trading time and the P_n be the nth trading price ($n = 1, 2, \cdots$). In this section we assume that

$$(A(T_n+), B(T_n+)) = (P_n e^{\delta}, P_n e^{-\delta}),$$

where $\delta > 0$ is a given constant. Thus if the current trading price is P_n, the immediate new ask price will increase to $P_n e^{\delta}$ and the immediate bid price will decrease to $P_n e^{-\delta}$. For convenience, we denote

$$U_n = \ln P_{n+1} - \ln P_n, \quad V_n = T_{n+1} - T_n.$$

Proposition 4.2.1. *$\{(U_n, V_n), n \geq 0\}$ is a sequence of i.i.d. bivariate random variables with common joint Laplace transform given by*

$$E\left[\exp\{-s(\ln P_{n+1} - \ln P_n) - t(T_{n+1} - T_n)\}\right] = \exp\{\theta_1(s,t)\delta - \theta_2(s,t)\delta\},$$

where θ_1 and θ_2 are given by equation (4.2.3) and (4.2.4).

Proof. From the assumption and the results of Theorem 4.2.1, we obtain

$$E\left[\exp\left\{-s\ln P_{n+1} - t(T_{n+1} - T_n)\right\} | P_n\right] = \exp\left\{\theta_1(s,t)(\ln P_n + \delta) + \theta_2(s,t)(\ln P_n - \delta)\right\}.$$

Since $\theta_1(s,t) + \theta_2(s,t) = -s$, we obtain

$$E\left[\exp\{-s(\ln P_{n+1} - \ln P_n) - t(T_{n+1} - T_n)\}|P_n\right] = \exp\{\theta_1(s,t)\delta - \theta_2(s,t)\delta\}.$$

Hence,

$$E\left[\exp\{-s(\ln P_{n+1} - \ln P_n) - t(T_{n+1} - T_n)\}\right] = \exp\{\theta_1(s,t)\delta - \theta_2(s,t)\delta\}.$$

The result follows. ■

Define $N(t) = \min\{n \geq 0 : T_n \leq t\}$ which gives the number of trades up to time t. We also define $P(t) = P_{N(t)}$.

Proposition 4.2.2. *The process* $\{P(t), t \geq 0\}$ *is a Semi-Markov Process (SMP) with state space* $[0, +\infty)$*.*

Proof. Since $\{(P_n, T_n), n \geq 0\}$ is a Markov renewal sequence, from the definition of $N(t)$ and $P(t)$, the stochastic process $\{P(t), t \geq 0\}$ is a semi-Markov process. ■

The process $\{P(t), t \geq 0\}$ is observable, while the $\{(A(t), B(t)), t \geq 0\}$ process may not be publicly observable. The ask and bid processes may be accessible to the brokers and dealers, but not to common traders. The question becomes how to find the parameters of $\{(A(t), B(t)), t \geq 0\}$ by observing $\{P(t), t \geq 0\}$.

4.3 Parameter Estimation

4.3.1 Moments

We shall estimate the parameters μ_a, μ_b, σ_a, σ_b and δ by the method of moment. Now assume that $\{(U_n, V_n), n \geq 0\}$ has common distribution of the random variable (U, V), whose bivariate Laplace transform is given by

$$E\left[\exp\{-sU - tV\}\right] = \exp\{\theta_1(s,t)\delta - \theta_2(s,t)\delta\}.$$

Next we find the moments of U and V, namely, $E(V)$, $E(V^2)$, $E(U)$, $E(U^2)$ and $E(UV)$.

Theorem 4.3.1. *$E(U)$, $E(U^2)$, $E(V)$, $E(V^2)$ and $E(UV)$ are given as follows:*

$$
\begin{aligned}
E(V) &= \frac{2\delta}{\mu_b - \mu_a}, \quad E(V^2) = \frac{4\delta^2}{(\mu_b - \mu_a)^2} + \frac{2(\sigma_a^2 + \sigma_b^2)\delta}{(\mu_b - \mu_a)^3} \\
E(U) &= \frac{\delta(\mu_b + \mu_a)}{(\mu_b - \mu_a)}, \quad E(U^2) = \frac{\delta^2(\mu_a + \mu_b)^2}{(\mu_b - \mu_a)^2} + \frac{2(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)\delta}{(\mu_b - \mu_a)^3}, \\
E\left(UV\right) &= \frac{2\delta^2(\mu_b + \mu_a)}{(\mu_b - \mu_a)^2} + \frac{2(\mu_b\sigma_a^2 + \mu_a\sigma_b^2)\delta}{(\mu_b - \mu_a)^3}.
\end{aligned}
$$

Proof. For convenience, we first need some simple results about $\theta_1(s,t)$ and $\theta_2(s,t)$. For $\theta_1(s,t)$,

$$
\begin{aligned}
&\theta_1(0,0) = 0, \\
&\left.\frac{\partial\theta_1(s,t)}{\partial t}\right|_{s=0,t=0} = -\frac{1}{\mu_b - \mu_a}, \quad \left.\frac{\partial^2\theta_1(s,t)}{\partial t^2}\right|_{s=0,t=0} = \frac{\sigma_a^2 + \sigma_b^2}{(\mu_b - \mu_a)^3}, \\
&\left.\frac{\partial\theta_1(s,t)}{\partial s}\right|_{s=0,t=0} = -\frac{\mu_b + \mu_a}{\mu_b - \mu_a}, \quad \left.\frac{\partial^2\theta_1(s,t)}{\partial s^2}\right|_{s=0,t=0} = \frac{\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2}{(\mu_b - \mu_a)^3}, \\
&\left.\frac{\partial^2\theta_1(s,t)}{\partial s\partial t}\right|_{s=0,t=0} = \frac{\mu_b\sigma_a^2 + \mu_a\sigma_b^2}{(\mu_b - \mu_a)^3}.
\end{aligned}
$$

For $\theta_2(s,t)$,

$$
\begin{aligned}
&\theta_2(0,0) = 0, \\
&\left.\frac{\partial\theta_2(s,t)}{\partial t}\right|_{s=0,t=0} = \frac{1}{\mu_b - \mu_a}, \quad \left.\frac{\partial^2\theta_2(s,t)}{\partial t^2}\right|_{s=0,t=0} = -\frac{\sigma_a^2 + \sigma_b^2}{(\mu_b - \mu_a)^3}, \\
&\left.\frac{\partial\theta_2(s,t)}{\partial s}\right|_{s=0,t=0} = \frac{\mu_a + \mu_b}{\mu_b - \mu_a}, \quad \left.\frac{\partial^2\theta_2(s,t)}{\partial s^2}\right|_{s=0,t=0} = -\frac{\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2}{(\mu_b - \mu_a)^3}, \\
&\left.\frac{\partial^2\theta_2(s,t)}{\partial s\partial t}\right|_{s=0,t=0} = -\frac{\mu_b\sigma_a^2 + \mu_a\sigma_b^2}{(\mu_b - \mu_a)^3}.
\end{aligned}
$$

Therefore,

$$\begin{aligned}
E(V) &= -\frac{\partial}{\partial t} E\left[\exp\{-sP - tT\}\right]\big|_{s=0,t=0} \\
&= -\frac{\partial}{\partial t} \exp\{\theta_1(s,t)\delta - \theta_2(s,t)\delta\}\big|_{s=0,t=0} \\
&= -\exp\{\theta_1(s,t)\delta - \theta_2(s,t)\delta\}\left\{\frac{\partial\theta_1(s,t)}{\partial t}\delta - \frac{\partial\theta_2(s,t)}{\partial t}\delta\right\}\Bigg|_{s=0,t=0} \\
&= \frac{2\delta}{\mu_b - \mu_a}.
\end{aligned}$$

Similarly, we can obtain

$$\begin{aligned}
E(V^2) &= \frac{\partial^2}{\partial t^2} E\left[\exp\{-sP - tT\}\right]\big|_{s=0,t=0} \\
&= \frac{4\delta^2}{(\mu_b - \mu_a)^2} + \frac{2(\sigma_a^2 + \sigma_b^2)\delta}{(\mu_b - \mu_a)^3} \\
E(U) &= -\frac{\partial}{\partial s} E\left[\exp\{-sP - tT\}\right]\big|_{s=0,t=0} \\
&= \frac{\delta(\mu_b + \mu_a)}{(\mu_b - \mu_a)} \\
EU^2 &= \frac{\partial^2}{\partial s^2} E\left[\exp\{-sP - tT\}\right]\big|_{s=0,t=0} \\
&= \frac{\delta^2(\mu_a + \mu_b)^2}{(\mu_b - \mu_a)^2} + \frac{2(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)\delta}{(\mu_b - \mu_a)^3} \\
E(UV) &= \frac{\partial^2}{\partial s \partial t} E\left[\exp\{-sP - tT\}\right]\big|_{s=0,t=0} \\
&= \frac{2\delta^2(\mu_b + \mu_a)}{(\mu_b - \mu_a)^2} + \frac{2(\mu_b\sigma_a^2 + \mu_a\sigma_b^2)\delta}{(\mu_b - \mu_a)^3}.
\end{aligned}$$

The results follows. ■

4.3.2 Estimation of parameters

In this subsection, we assume that the sample data for the ith trading time t_i and the ith trading price p_i are given for $i = 1, 2, \cdots, n$. Let

$$u_i = \ln p_{i+1} - \ln p_i, \quad v_i = t_{i+1} - t_i,$$

hence the sample data for (U,V) is given by $(\{u_i\}_{i=1}^n, \{v_i\}_{i=1}^n)$. Denote

$$\begin{aligned} x_1 &= \sum_{i=1}^n v_i/n, \quad x_2 = \sum_{i=1}^n u_i/n, \\ x_3 &= \sum_{i=1}^n v_i^2/n, \quad x_4 = \sum_{i=1}^n u_i^2/n, \quad x_5 = \sum_{i=1}^n v_i u_i/n. \end{aligned}$$

We aim to derive explicit estimators of the five parameters μ_1, μ_2, σ_1, σ_2, δ using moment estimation.

Theorem 4.3.2. *The estimators of μ_a, μ_b, σ_a, σ_b, δ are given by*

$$\begin{aligned} \hat{\mu}_a &= \frac{y_1 - \left(y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2}, \quad \hat{\sigma}_a = \sqrt{(y_4 - z_1 y_2)(z_2 - z_1)}, \\ \hat{\mu}_b &= \frac{y_1 + \left(y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2}, \quad \hat{\sigma}_b = \sqrt{(z_2 y_2 - y_4)(z_2 - z_1)}, \\ \hat{\delta} &= (z_2 - z_1)x_1, \end{aligned}$$

where

$$\begin{aligned} y_1 &= \frac{2x_2}{x_1}, \quad y_2 = \frac{x_3 - x_1^2}{x_1}, \quad y_3 = \frac{x_4 - x_2^2}{x_1}, \quad y_4 = \frac{x_5 - x_1 x_2}{x_1}, \\ z_1 &= \frac{y_1 - \left(y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2}, \quad z_2 = \frac{y_1 + \left(y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2}. \end{aligned}$$

Proof. Using the method of moment estimation and the results of Theorem 4.3.1, we obtain

$$\begin{aligned} x_1 &= \frac{2\hat{\delta}}{\hat{\mu}_b - \hat{\mu}_a} \\ x_2 &= \frac{\hat{\delta}(\hat{\mu}_b + \hat{\mu}_a)}{(\hat{\mu}_b - \hat{\mu}_a)} \\ x_3 &= \frac{4\hat{\delta}^2}{(\hat{\mu}_b - \hat{\mu}_a)^2} + \frac{2(\hat{\sigma}_a^2 + \hat{\sigma}_b^2)\hat{\delta}}{(\hat{\mu}_b - \hat{\mu}_a)^3} \\ x_4 &= \frac{\hat{\delta}^2(\hat{\mu}_a + \hat{\mu}_b)^2}{(\hat{\mu}_b - \hat{\mu}_a)^2} + \frac{2(\hat{\mu}_b^2\hat{\sigma}_a^2 + \hat{\mu}_a^2\hat{\sigma}_b^2)\hat{\delta}}{(\hat{\mu}_b - \hat{\mu}_a)^3} \\ x_5 &= \frac{2\hat{\delta}^2(\hat{\mu}_b + \hat{\mu}_a)}{2(\hat{\mu}_b - \hat{\mu}_a)^2} + \frac{2(\hat{\mu}_b\hat{\sigma}_a^2 + \hat{\mu}_a\hat{\sigma}_b^2)\hat{\delta}}{(\hat{\mu}_b - \hat{\mu}_a)^3}. \end{aligned}$$

Next we solve $\hat{\mu}_a$, $\hat{\mu}_b$, $\hat{\sigma}_a$, $\hat{\sigma}_b$, $\hat{\delta}$ in terms of x_1 to x_5.

Let,

$$\begin{aligned}
y_1 &= \frac{2x_2}{x_1} = \hat{\mu}_b + \hat{\mu}_a \\
y_2 &= \frac{x_3 - x_1^2}{x_1} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_b^2}{(\hat{\mu}_b - \hat{\mu}_a)^2} \\
y_3 &= \frac{x_4 - x_2^2}{x_1} = \frac{\hat{\mu}_b^2\hat{\sigma}_a^2 + \hat{\mu}_a^2\hat{\sigma}_b^2}{(\hat{\mu}_b - \hat{\mu}_a)^2} \\
y_4 &= \frac{x_5 - x_1x_2}{x_1} = \frac{\hat{\mu}_b\hat{\sigma}_a^2 + \hat{\mu}_a\hat{\sigma}_b^2}{(\hat{\mu}_b - \hat{\mu}_a)^2}.
\end{aligned}$$

Note that,

$$y_1^2 - 4\frac{y_1y_4 - y_3}{y_2} = (\hat{\mu}_b - \hat{\mu}_a)^2.$$

Since $\hat{\mu}_b > \hat{\mu}_a$, we obtain

$$\begin{aligned}
\hat{\mu}_a &= \frac{y_1 - \left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2} \\
\hat{\mu}_b &= \frac{y_1 + \left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2}
\end{aligned}$$

Next denote

$$\begin{aligned}
z_1 &= \frac{y_1 - \left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2}, \\
z_2 &= \frac{y_1 + \left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2},
\end{aligned}$$

we obtain

$$\begin{aligned}
\hat{\sigma}_a &= \sqrt{(y_4 - z_1y_2)(z_2 - z_1)}, \\
\hat{\sigma}_b &= \sqrt{(z_2y_2 - y_4)(z_2 - z_1)}.
\end{aligned}$$

Finally we obtain

$$\hat{\delta} = (z_2 - z_1)x_1.$$

The result follows. ■

Next we verify that these estimators are well-defined.

Proposition 4.3.1. *The estimators produce real valued estimates, i.e.*

$$y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2} \geq 0,$$
$$(y_4 - z_1 y_2)(z_2 - z_1) \geq 0,$$
$$(z_2 y_2 - y_4)(z_2 - z_1) \geq 0.$$

Proof. First we prove the term $y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2}$ is nonnegative. Note that

$$y_1^2 - 4\frac{y_1 y_4 - y_3}{y_2} = \frac{4}{x_1^2(x_3 - x_1^2)}\left[x_2^2(x_3 - x_1^2) - 2x_1 x_2(x_5 - x_1 x_2) + x_1^2(x_4 - x_2^2)\right].$$

Since

$$\begin{aligned} x_1^2 &= (\sum_{i=1}^n v_i/n)^2 > 0, \\ x_3 - x_1^2 &= \frac{n\sum_{i=1}^n v_i^2 - \sum_{i=1}^n v_i}{n^2} > 0, \end{aligned}$$

it suffices to show

$$x_2^2(x_3 - x_1^2) - 2x_1 x_2(x_5 - x_1 x_2) + x_1^2(x_4 - x_2^2) \geq 0.$$

Note that

$$\begin{aligned}
& x_2^2(x_3 - x_1^2) - 2x_1x_2(x_5 - x_1x_2) + x_1^2(x_4 - x_2^2) \\
\geq\ & 2x_1x_2\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)} - 2x_1x_2(x_5 - x_1x_2) \\
=\ & 2x_1x_2(\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)} - (x_5 - x_1x_2)) \\
=\ & 2\frac{\sum v_i}{n}\frac{\sum u_i}{n}\left(\sqrt{\left(\frac{\sum v_i^2}{n} - \left(\frac{\sum v_i}{n}\right)^2\right)\left(\frac{\sum u_i^2}{n} - \left(\frac{\sum u_i}{n}\right)^2\right)} - \left(\frac{\sum v_i \Delta p_i}{n} - \frac{\sum v_i}{n}\frac{\sum u_i}{n}\right)\right) \\
=\ & 2\frac{\sum v_i}{n}\frac{\sum u_i}{n}\left(\sqrt{\frac{\sum(v_i - \sum v_i/n)^2}{n}\frac{\sum(u_i - \sum u_i/n)^2}{n}} - \left(\frac{\sum v_i u_i}{n} - \frac{\sum v_i}{n}\frac{\sum u_i}{n}\right)\right) \\
\geq\ & 2\frac{\sum v_i}{n}\frac{\sum u_i}{n}\left(\frac{\sum(v_i - \sum v_i/n)(u_i - \sum u_i/n)}{n} - \left(\frac{\sum v_i u_i}{n} - \frac{\sum v_i}{n}\frac{\sum u_i}{n}\right)\right) \\
=\ & 0
\end{aligned}$$

Next we show $y_4 - z_1y_2$ and $z_2y_2 - y_4$ are nonnegative. Note that

$$\begin{aligned}
y_4 - z_1y_2 &= \frac{y_2\left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2} + \left(y_4 - \frac{y_1y_2}{2}\right), \\
z_2y_2 - y_4 &= \frac{y_2\left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)^{\frac{1}{2}}}{2} - \left(y_4 - \frac{y_1y_2}{2}\right).
\end{aligned}$$

Hence it is enough to show

$$\frac{y_2^2\left(y_1^2 - 4\frac{y_1y_4 - y_3}{y_2}\right)}{4} \geq \left(y_4 - \frac{y_1y_2}{2}\right)^2.$$

After simplifying above inequality, it suffices to show that $y_2y_3 \geq y_4^2$.

Note that

$$y_2y_3 \geq y_4^2 \Leftrightarrow (x_3 - x_1^2)(x_4 - x_2^2) \geq (x_5 - x_1x_2)^2$$

This is proved in previous inequality. Hence the result follows. ∎

4.4 Numerical Example

In this section we apply our model to the real data, and aim to forecast the trading price movement over a short period. Here we select the stock SUSQ (Susquehanna Bancshares Inc). The data is chosen from 01/04/2010 9:30AM to 01/04/2010 4:00PM, including the trading price and trading time. The unit of trading price is dollars and the unit of the difference of consecutive trading time is seconds. From the formula

$$E(\ln p_{n+1} - \ln p_n) = \frac{\delta(\mu_a + \mu_b)}{\mu_b - \mu_a},$$

we obtain

$$E(\ln p_{n+1} - \ln p_0) = \frac{\delta(\mu_a + \mu_b)}{\mu_b - \mu_a}(n+1).$$

Therefore given the initial trading price $P_0 = p_0$, we have

$$E(\ln P(t)) = \frac{\delta(\mu_a + \mu_b)}{\mu_b - \mu_a}E(N(t)) + \ln p_0.$$

Since this stock is frequently traded, we use t/τ to approximate $E(N(t))$, where $\tau = E(V) = \frac{2\delta}{\mu_b - \mu_a}$. Hence we have the prediction formula for $\ln P(t)$ as following

$$\tilde{f}(t) = \frac{\mu_a + \mu_b}{2}t + \ln p_0. \tag{4.4.1}$$

Similarly, from the formula

$$E(\ln P_{n+1} - \ln P_n) = \frac{\delta(\mu_b + \mu_a)}{(\mu_b - \mu_a)}, \quad E(\ln P_{n+1} - \ln P_n)^2 = \frac{\delta^2(\mu_a + \mu_b)^2}{(\mu_b - \mu_a)^2} + \frac{2(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)\delta}{(\mu_b - \mu_a)^3},$$

we obtain

$$Var(\ln P_{n+1} - \ln P_n) = \frac{2(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)\delta}{(\mu_b - \mu_a)^3}.$$

Since $\{(\ln P_{n+1} - \ln P_n), n \geq 0\}$ is an i.i.d. sequence, we have

$$Var(\ln P_{n+1} - \ln P_0) = \frac{2(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)\delta}{(\mu_b - \mu_a)^3}(n+1).$$

Therefore, given the initial trading price $P_0 = p_0$, we have

$$Var(\ln P(t)) = \frac{2(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)\delta}{(\mu_b - \mu_a)^3} E(N(t)).$$

Hence by using the approximate formula of $E(N(t))$, we have the prediction formula for $Var(\ln P(t))$ as following

$$\tilde{g}(t) = \frac{(\mu_b^2\sigma_a^2 + \mu_a^2\sigma_b^2)}{(\mu_b - \mu_a)^2} t. \tag{4.4.2}$$

From above results, we use $\tilde{f}(t)+3\sqrt{\tilde{g}(t)}$ and $\tilde{f}(t)-3\sqrt{\tilde{g}(t)}$ to estimate the upper bound and lower bound of the predicted $\ln P(t)$.

Now we perform the back test for above formula to evaluate the performance of the prediction. In detail we predict the logarithmic trading price at each trading time using the 10-minute data 1 minute before the trading time. For example, we observe there is a trade at 10:34:56, then we use the data from 10:23:56 to 10:33:56 to estimate the parameters and predict the logarithmic trading price at 10:34:56. At the same time we calculate the upper bound and lower bound of the prediction at that trading time. We compare this predicted logarithmic trading price with the real one in Figure 4.2. We do the similar prediction for each trading time but using the 10-minute data 5/10/15 minute before the trading time respectively. The comparisons are shown in Figure 4.3-4.5.

4.5 Extensions

We end this chapter with suggestions for two extensions.

1. In this chapter we only consider the data of trading price and trading time. It would be interesting to include the information of trading size (that is, the number of shares of the traded stock for each trade).

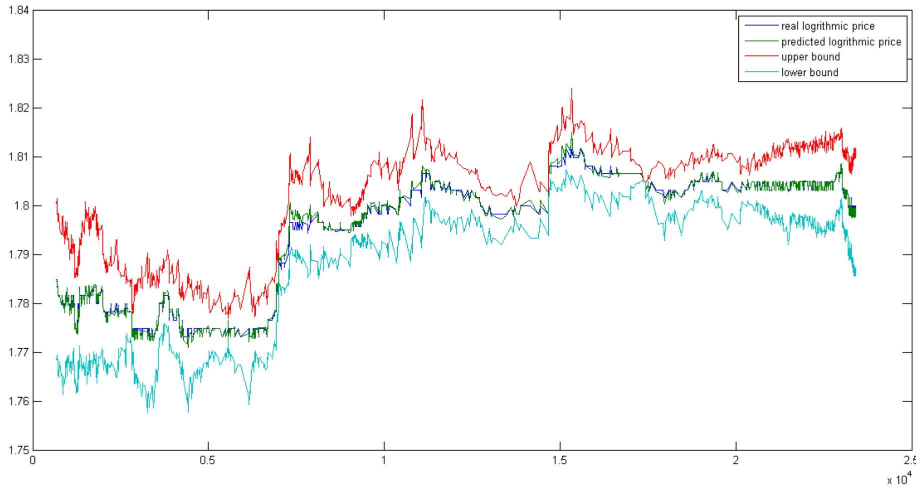


Figure 4.2: Prediction of trading price using the data 1 minute before each trading time

2. We have assumed that the market ask price and the market bid price separate in a proportionally manner immediately after they meet each other. It would be interesting to consider a price-dependent separation mechanism.

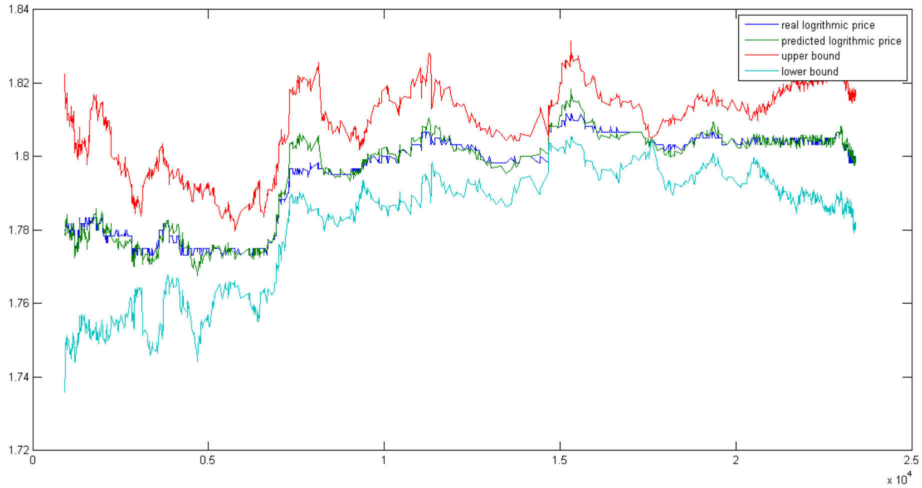


Figure 4.3: Prediction of trading price using the data 5 minute before each trading time

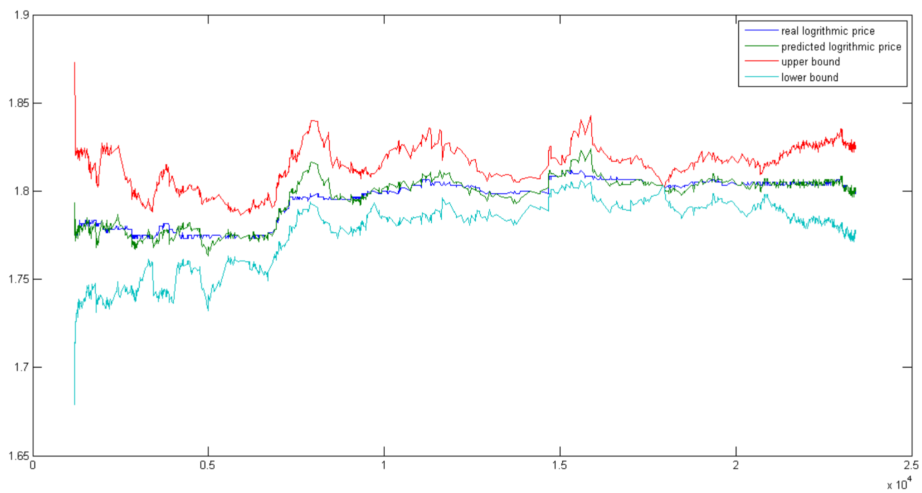


Figure 4.4: Prediction of trading price using the data 10 minute before each trading time

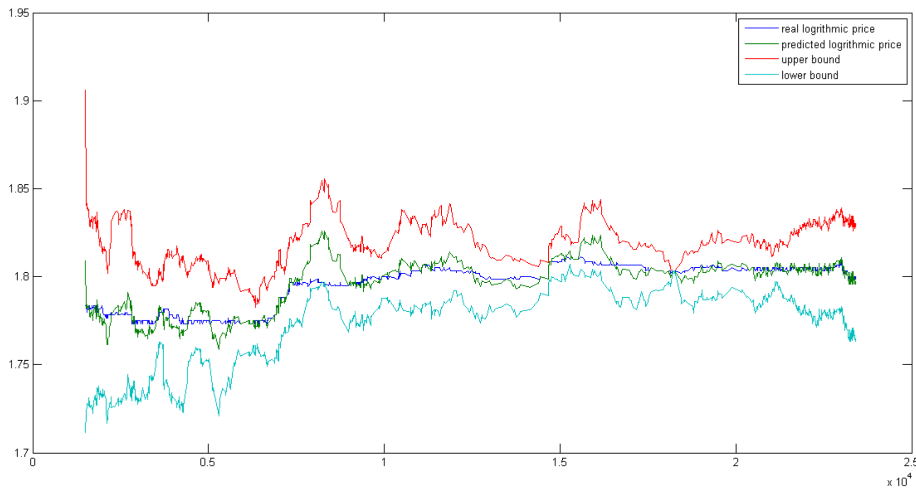


Figure 4.5: Prediction of trading price using the data 15 minute before each trading time

BIBLIOGRAPHY

[1] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

[2] D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[3] P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 1999.

[4] J.-P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.

[5] A. Budhiraja and A. Ghosh. Diffusion approximations for controlled stochastic network: An asymptotic bound for the value function. *The Annals of Applied Probability*, 16(4):1962–2006, 2006.

[6] K. J. Cohen, S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. Transaction costs, order placement strategy, and existence of the bid-ask spread. *The Journal of Political Economy*, pages 287–305, 1981.

[7] B. Conolly, P. Parthasarathy, and N. Selvaraju. Double-ended queues with impatience. *Computers & Operations Research*, 29(14):2053–2072, 2002.

[8] R. Cont and A. De Larrard. Order book dynamics in liquid markets: limit theorems and diffusion approximations. *Available at SSRN 1757861*, 2011.

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

[10] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.

[11] I. Degirmenci. Asymptotic analysis and performance-based design of large scale service and inventory systems. *Ph.D. dissertation, Department of Business Administration, Duke University*, 2010.

[12] I. Domowitz and J. Wang. Auctions as algorithms: Computerized trade execution and price discovery. *Journal of economic dynamics and control*, 18(1):29–60, 1994.

[13] J. Doyne Farmer 5, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? *Quantitative finance*, 4(4):383–397, 2004.

[14] T. Foucault. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial markets*, 2(2):99–134, 1999.

[15] M. B. Garman. Market microstructure. *Journal of financial Economics*, 3(3):257–275, 1976.

[16] B. Hollifield, R. A. Miller, and P. Sandås. Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4):1027–1063, 2004.

[17] S. Karlin and H. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.

[18] S. Karlin and H. E. Taylor. *A first course in stochastic processes*. Access Online via Elsevier, 1975.

[19] B. Kashyap. The double-ended queue with bulk service and limited waiting space. *Operations Research*, 14(5):822–834, 1966.

[20] W. Kim, K. Yoon, G. Mendoza, and M. Sedaghat. Simulation model for extended double-ended queueing. *Computers & Industrial Engineering*, 59(2):209–219, 2010.

[21] L. Kleinrock. Optimum bribing for queue position. *Operations Research*, 15(2):304–318, 1967.

[22] V. Kulkarni. *Modeling and analysis of stochastic systems*. Chapman & Hall/CRC, 1996.

[23] H. Kushner. *Heavy Traffic Analysis of Controlled Queueing and Communication Networks*. Springer-Verlag, New York, 2003.

[24] H. Luckock. A steady-state model of the continuous double auction. *Quantitative Finance*, 3(5):385–404, 2003.

[25] M. F. Neuts. *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Courier Dover Publications, 1981.

[26] C. A. Parlour. Price dynamics in limit order markets. *Review of Financial Studies*, 11(4):789–816, 1998.

[27] D. Perry and W. Stadje. Perishable inventory systems with impatient demands. *Mathematical methods of operations research*, 50(1):77–90, 1999.

[28] B. Prabhakar, N. Bambos, and T. Mountford. The synchronization of poisson processes and queueing networks with service and synchronization nodes. *Advances in Applied Probability*, 32(3):824–843, 2000.

[29] P. Protter. *Stochastic Integration and Differential Equations*. Springer, 2008.

[30] J. Reed and A. Ward. A diffusion approximation for a generalized jackson network with reneging. In *Proceedings of the 42nd annual Allerton conference on communication, control, and computing*, 2004.

[31] I. Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, 2009.

[32] D. J. Seppi. Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies*, 10(1):103–150, 1997.

[33] K. Sigman. Lecture notes on stochastic modeling 1. *Online lecture note at* ***http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-RRT.pdf***, 2009.

[34] S. Zenios. Modeling the transplant waiting list: a queueing model with reneging. *Queueing systems*, 31(3):239–251, 1999.