High Frequency Trading: Price Dynamics Models and Market Making Strategies

Cheng Lu

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-144
http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-144.html

May 31, 2012

High Frequency Trading:

Price Dynamics Models and Market Making Strategies

Cheng Lu

23269284

Electrical Engineering and Computer Science

In partial fulfillment of the

requirements for the Degree of

Master of Engineering

University of California at Berkeley

Abstract

High Frequency Trading (HFT) has recently drawn public and regulatory attention after the "flash crash" in U.S. stock market on May 6, 2010. Data processing and statistical modeling techniques in finance has been revolutionized by the availability of high frequency data on transactions, quotes and order flow in electronic order-driven markets, which has and brought up new theoretical and computational challenges. Market dynamics at the transaction level cannot be characterized solely in terms the dynamics of a single price and one must also take into account the interaction between buy and sell orders of different types by modeling the order flow at the bid price, ask price and possibly other levels of the limit order book. In this paper, I implemented and improved a queuing model that characterizes the market dynamics as a Discrete Markovian System, which is more suitable for illiquid market. I then propose and examine a few market-making trading strategies & applications of such a model and point to the simulation results.

Keywords: High-Frequency Trading, Markovian Queuing Model, Market Making Strategies.

Table of Content

Abstract i

1 Introduction 1

2 Literature Review 4

3 Methodology 8

3.1 Model Setup 8

3.2 Model Modifications 11

3.2.1 Event Arrival Rate 11

3.2.2 Order Size 15

3.2.3 Event Correlation 16

3.3 Model Application 18

3.3.1 Market Making 18

3.3.2 Market Making with balancing strategy 19

3.3.3 Smoking Strategy 20

4 Discussion 23

4.1 Model Simulation Results 23

4.2 Market Making Simulation Results 24

4.3 Market Making With Balancing Simulation Results 26

4.4 Smoking Strategy Simulation Results 28

5 Conclusion 32

6 Reference 33

1 Introduction

Algorithm trading of stock first became a significant part of Wall Street in the 1980s. Since then, more powerful computers and more sophisticated algorithms have grown vastly. For years, High-Frequency Trading (HFT) firms stepped away from Wall Street, reaping billions of revenue while being criticized as damaging markets and hurting ordinary investors. Now, after the 2008 Crisis, they are stepping into the light.

There are plenty of definitions of High-Frequency Trading. HFT is a strategy that trades for investment horizons of less than a day and seeks to unwind all positions before the end of each trading day. Because they must finish the day flat, high-frequency traders must exhibit balanced bi-dimensional flow, thus HFTs can't accumulate large position and deploy large amount of capital, and they have little need for outside capital, so tend to be proprietary traders. The opportunities of HFT usually come from taking the opposite side of trades of long-term investors, who will impact many securities besides the one they are directly traded, because stocks are correlated. This creates opportunities for HFTs, whose activities keep correlated stocks "fairly priced" with respect to one another. It is worth to notice that the primary driver of growth in HFT market is reducing trading costs, not the technology. As trading cost diminish, including bid/ask spread, commissions, market access fees and SEC fees, smaller and smaller opportunities become profitable to trade, leading to higher HF volume.

There are several distinct characteristics that differentiate HFT from traditional long-term investment. The HFT's average net profit margin, transaction costs, capital requirements and total profit potential are all far smaller than long-term investment, and while HFT have higher consistency of profits than long-term investment too. The opportunities for short-term returns follow a Gaussian distribution, and HF traders target opportunities that are tiny but plentiful. Since HFT opportunities are short-lived, capturing them requires uses of advanced technology. HFT requires speed to capture opportunities before competitors access them.

HFTs are the backbone of market liquidity and serve as an important part of the market's ecosystem for long-term investors. Market makers contribute immediately transactable shares at prevailing prices. Statistical Arbitrageurs make sure that information is efficient transmitted from securities being impacted by long-term investors to other securities that are correlated, resulting in cross-sectional fair prices. HFTs risk their own capital to provide their services, yet earn razor-thin margins for doing so.

High-frequency trading changes the behavior of all market participants, and calls for new models for understanding market dynamics and providing quantitative frameworks for optimal execution of trades and accurate prediction of market variables. In this paper, I implemented and improved the Discrete Markovian Queuing model to characterize the dynamic of HFT market, to HFT data, which recorded the Limit Order Book of a HK-traded stock for one week. I assume that the model could accurately simulate the real market behavior, upon which I apply and test different trading strategies. The final deliverable includes a market simulation model and several feasible trading strategies.

The rest of the paper is organized as follows: In the Literature Review Section, I present the review of state of the art research developments in HFT market. In the Methodology Section, I introduce the hypothesis, implementation and improvement of the Discrete Markovian Queuing Model, and then presented several market-making trading strategies and their associated simulation results. Finally, in the Conclusion Section, I give brief conclusion about my HFT capstone project.

2 Literature Review

The electronic platforms form a limit order book aggregating most trading data in a financial market every day. At the same time, the frequency of order submissions has increased and the time for market order execution on electronic markets has dropped from more than 25 milliseconds to less than a millisecond in the past decade. As a result, the evolution of supply, demand and price behavior in equity markets is being increasingly recorded, and this data is available to all market participants in real time and to researchers in the forms of high frequency database. The analysis of such high frequency data constitutes a challenge. At a fundamental level, statistical modeling of high frequency market provide insightful analysis of the dynamics between order flow, liquidity and price dynamics [4, 5, 6], and might help bridge the gap between market microstructure theories [7, 8, 9]. At the level of applications, models of high frequency data provide a quantitative framework for market making [10] and optimal execution of trades [11, 12, 13]. Another obvious application is the development of statistical models in view of predicting short-term behavior of market variables such as price, trading volume and order flow.

At any given time in a limit order market, outstanding limit orders are represented by the limit order book, which summarizes the price and quantity of supply and demand. Not surprisingly, empirical studies [14] indicates that the state of the order book contains information about short-term price movements so it is of great interest to provide statistical model for the dynamics of the order book.

R. CONT, KUKANOV and STOIKOV [4] suggested a conceptually simple model that relates the price changes to the order flow imbalance (OFI) defined as the imbalance between supply and demand at the best bid and ask prices. Their study reveals a linear relationship between OFI and price changes. However, statistical results reveal that this linear relationship may not be the case, so I will be more interested in other sophisticated models that take into account more factors.

R.CONT and A. DE LARRARD characterized it by a heavy traffic model [1]. In the heavy traffic limit, the possibly complicated discrete dynamics of the queuing system is approximated by a system with a continuous state space, which can be either described by a system of ordinary differential equations or a system of stochastic differential equations. The bid/ask queue sizes follow the diffusive behavior and it is important to consider the diffusion limit of the order book. When the sequences of order sizes at the bid and the ask and inter-event durations are weakly dependent covariance-stationary sequences, the rescaled order book process converges weakly to a two-dimensional Markov process diffusing in the quarter-plane, which is 'renewed' every time it hits one of the axes. And before each time the process renewed, it is a two-dimensional Brownian motion. However, it is noticed that the heavy traffic assumption is problematic for my data because the Hong Kong listed stock is not sufficiently traded, so characterized it by continuous Brownian motion would cast severe deviation from reality. Therefore I need a "fine-grained" model that is able to track individual orders.

R. CONT and LARRARD [2] recently proposed a discrete stochastic model for the dynamics of a limit order book, in which arrivals of market order, limit orders and order cancellations are characterized in terms of a Markovian queuing system. Through its analytical tractability, the model allows to obtain analytical expressions for various quantities of interest such as the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, conditional on the state of the order book. This model meets my expectations quite well, but I also found that the data I have do not satisfy some assumptions they put, for instance, the duration of events is not exponentially distributed. I then try to modify the original model to accommodate facts from empirical studies of the data.

As for the application level, current states of art researches have been focused on electronic market making and technical analysis. Y. NEVMYVAKA, K. SYCARA, and D. SEPPI [15] established an analytical foundation for electronic market making in which they focused on the normative automation of the market maker's activities. They utilized the "non-predictive" trading strategies to highlight the fundamental issues: depth of quote, quote positioning and timing of updates. They also examined the impact of various parameters on the market maker's performance. Similarly, Y. FENG, R. YU, P. STONE [16] examined an automated stock-trading agent in the context of the Penn-Lehman-Automated-Trading (PLAT) simulator, which devised a market making strategy exploit market volatility without predicting the exact stock price movement direction. Understanding the market maker's activities and exploring the different market making strategies have become the research focus in high-frequency market. Inspired by these ideas, and together with an accurate market dynamics model, I would be able to better analysis the market maker's activities and providing profitable strategies.

3 Methodology

Having argued in the above section that the Hong Kong listed stock I have is not liquid and hence the key assumption of the heavy traffic model is not satisfied, I start from a description of order arrivals for different kinds of orders, the dynamics of a limit order book is naturally described in the language of queuing theory [2]. Motivated by the fact that it is sufficient to focus on the dynamics of the best bid and ask queue if one is primarily interested in the level I order book dynamics, I then decided to follow the Markovian queuing model [2] to test its validity on the data where the limit order book is driven by orders at the bid and ask side, represented as a system of two interacting Markovian queues. I will first introduce the setup of R. CONT's queuing model [2], and then elaborate my modifications according to the empirical analysis.

3.1 Model Setup

To simplify the initial model, we use the following terms to represent the limit order book:

- The bid price s_t^b and the ask price $s_t^a = s_t^b + \delta$, which captures the majority of the market situation.
- The size of the bid queue q_t^b represents the outstanding limit buy orders at the bid.

- The size of the bid queue q_t^a represents the outstanding limit buy orders at the ask.

The state of the limit order book is thus described by the triplet $X_t = (s_t^b, q_t^b, q_t^a)$, which takes values in the discrete state space $\delta.\mathbb{Z}\times\mathbb{N}^2$.

The state X_t of the order book is updated by order book events: limit orders (at the bid or ask), market orders and cancelations. According to the R.CONT's works [2], I got the assumption that these events occur according to independent Poisson processes:

- Market buy (resp. sell) orders arrive at independent, exponential times with rate μ .

- Limit buy (resp. sell) orders at the (best) bid (resp. ask) arrive at independent, exponential times with rate λ .

- Cancellations occur at independent, exponential times with rate θ .

- These events are mutually independent.

- All orders sizes are equal (assumed to be 1 without loss of generality).

- All the previous sequences are independent.

Under these assumptions $q_t = (q_t^b, q_t^a)$ is thus a Markov process, taking values in $\mathbb{N}^2$, whose transitions correspond to the order book events $\{T_i^a, i\geq 1\}\bigcup\{T_i^b, i\geq 1\}$.

When the bid or ask orders is depleted, the price moves up or down to the next level of the order book. Analogous to the heavy traffic model, the new queue sizes are sampled from the empirical PDF $f^{b/a}(x,y)$. I further assume that the order book contains no empty levels so that these price increments are equal to one tick (in our case, 0.05 HKD):

- When the bid queue is depleted, the price decreases by one tick.
- When the ask queue is depleted, the price increases by one tick.

In summary, the process $X_t = (s_t^b, q_t^b, q_t^a)$ is a continuous-time process with right-continuous, piecewise constant sample paths whose transitions correspond to the order book events $\{T_i^a, i \geq 1\} \bigcup \{T_i^b, i \geq 1\}$ [2]. At each event:

- If an order or cancelation arrives on the ask side i.e. $T \in \{T_i^a, i \geq 1\}$:

$$(s_T^b, q_T^b, q_T^a) = (s_{T-}^b, q_{T-}^b, q_{T-}^a + V_i^a) 1_{q_{T-}^a > -V_i^a} + (s_{T-}^b + \delta, R_i^b, R_i^a) 1_{q_{T-}^a \leq V_i^a}$$

- If an order or cancelation arrives on the bid side i.e. $T \in \{T_i^b, i \geq 1\}$:

$$(s_T^b, q_T^b, q_T^a) = (s_{T-}^b, q_{T-}^b + V_i^b, q_{T-}^a) 1_{q_{T-}^b > -V_i^b} + (s_{T-}^b - \delta, R'^b_i, R'^a_i) 1_{q_{T-}^b \leq V_i^b}$$

$(V_i^a)_{i>1}$ and $(V_i^b)_{i>1}$ are sequences of IID variables, $(R_i)_{i>1} = (R_i^b, R_i^a)_{i\geq 1}$ is a sequence of IID variables with (joint) distribution $f^b(x,y)$, and $(R'_i)_{i>1} = (R'^b_i, R'^a_i)_{i\geq 1}$ is a sequence of IID variables with (joint) distribution $f^a(x,y)$.

3.2 Model Modifications

3.2.1 Event Arrival Rate

In R. CONT's original model [2], all kinds of orders arrive at an exponential rate, which I doubt maybe is not the case in our data, so I run some empirical study about the order arrival rate. I first differentiated 6 different kinds of orders from our data, which is: Limit Buy Orders, Limit Sell Orders, Market Buy Orders, Market Sell Orders, Cancellation Buy Orders, and Cancellation Sell Orders. In our terms, Limit Buy means someone posted a buy order at the best bid, and Limit Sell means someone posted a sell order at best ask. Market Buy means someone takes the best bid offers, and Market Sell means someone takes the best ask offers. Cancellation Buy means someone cancelled their bid orders at the best bid, and Cancellation Sell means someone cancelled their ask orders at the best ask. Fig 1 shows the empirical probability distribution of the six different kind orders arrival rate:

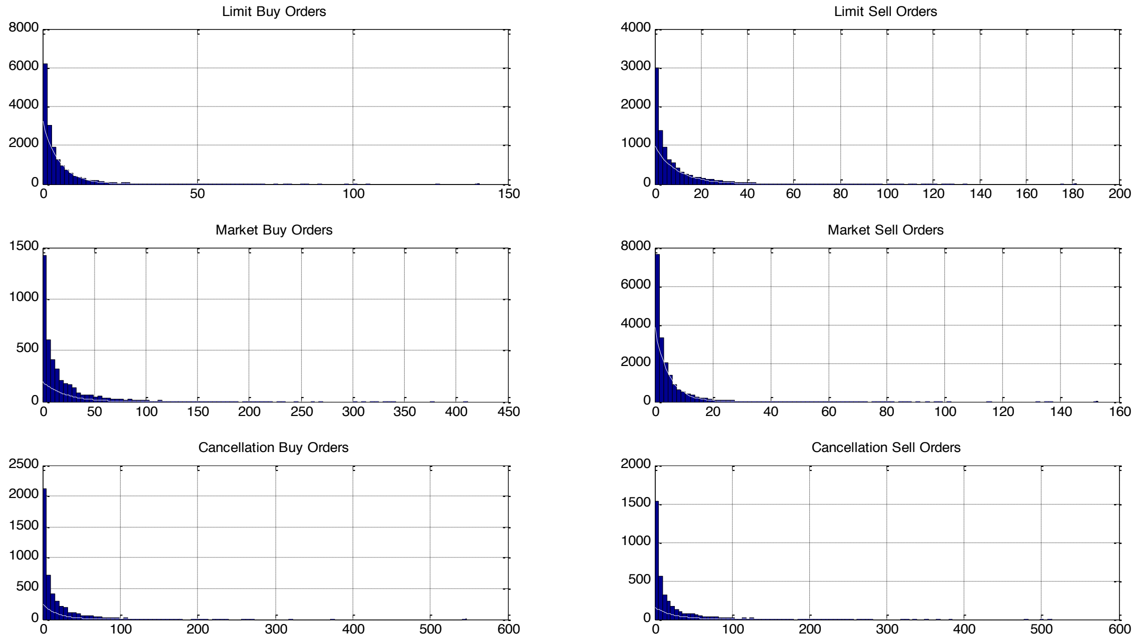

Fig 1. Order arrival distribution and the exponential distribution.

In Fig 1, the white line indicates the best fitted exponential distribution. It is observed that, in some cases, the exponential distribution fits the data well, such as limit buy orders and market sell orders, but in some other cases, such as cancellation buy orders and market buy orders, the exponential distribution obviously fails to fit the data.

This empirical study suggests that one potential threat of R. CONT's Queuing Model [2] might be the exponential arrival rate, so I try to figure out another distribution that may capture the arrival rate better than exponential, and I found one candidate after some trial and error, which is called Weibull Distribution, which is often used to describe the size distribution of particles. The probability density function of a Weibull random variable x is $f(x;\lambda,k) = k/\lambda(x/\lambda)^{k-1}e^{-(x/\lambda)^k}\mathbf{1}(x \geq 0)$

I then try to fit the arrival rate using the Weibull distribution. The results are shown in Fig 2:

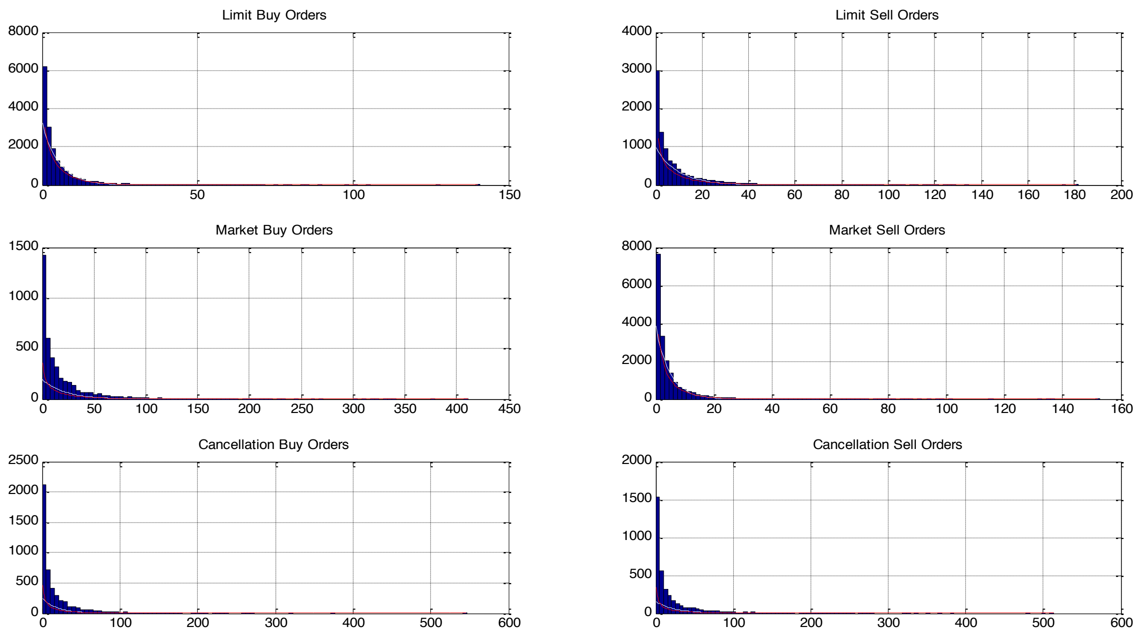

Fig 2. Order arrival distribution and the Weibull distribution.

In Fig 2, the white line indicates the exponential distribution and the red one is the Weibull distribution. To access the goodness of fitting, I also look at the Q-Q plots as shown in Fig 3.

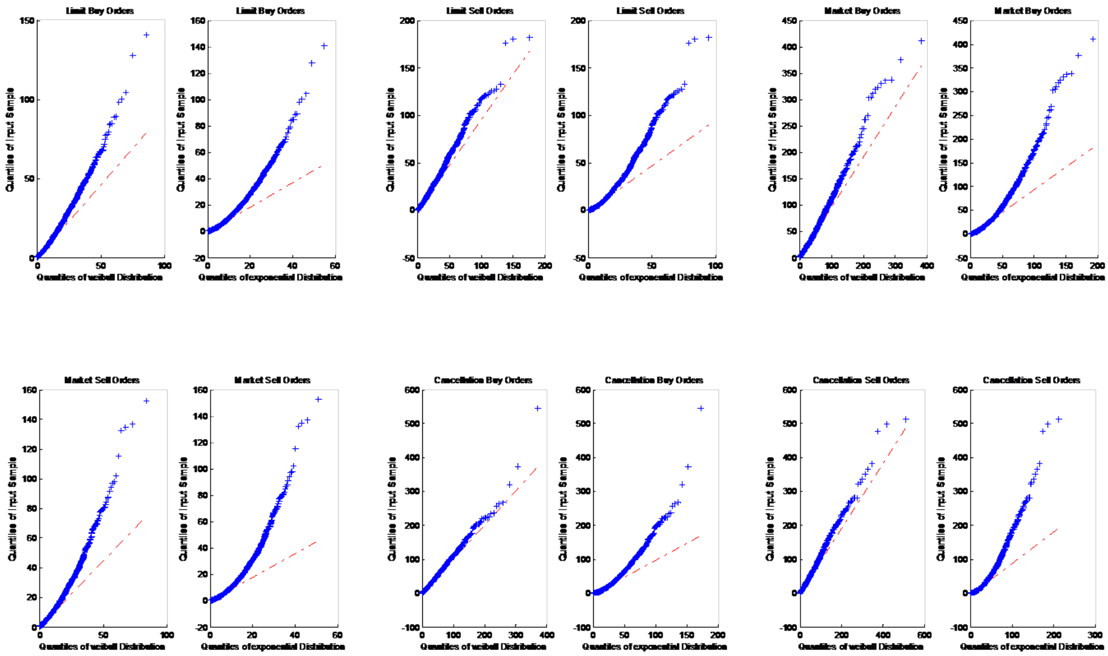

Fig 3. Q-Q plots of the real data against the Weibull distribution or the exponential distribution.

In Fig 3, it shows the comparison between the Weibull distribution and the exponential distribution in each order figure. The blue lines indicate the authentic data, while the red lines are the calculated distribution. The closeness indicates the accuracy of the distribution to the real data. One can then clearly see that nearly for each figure, the Weibull distribution outperforms Exponential, especially for Limit Sell Orders, Market Buy Orders, Cancellation Buy Orders and Cancellation Sell Orders, where we say that the Weibull distribution is really a good fit to the data. After these empirical studies, I decided to use the Weibull distribution instead of the exponential distribution to characterize the arrival rate of different orders.

3.2.2 Order Size

Another concern I have about the original model is the assumption that all coming orders have unit lot size (400 shares), which I believe is quite biased from the reality. I first run a study to analyze the order size distribution, as is shown in Fig 4.

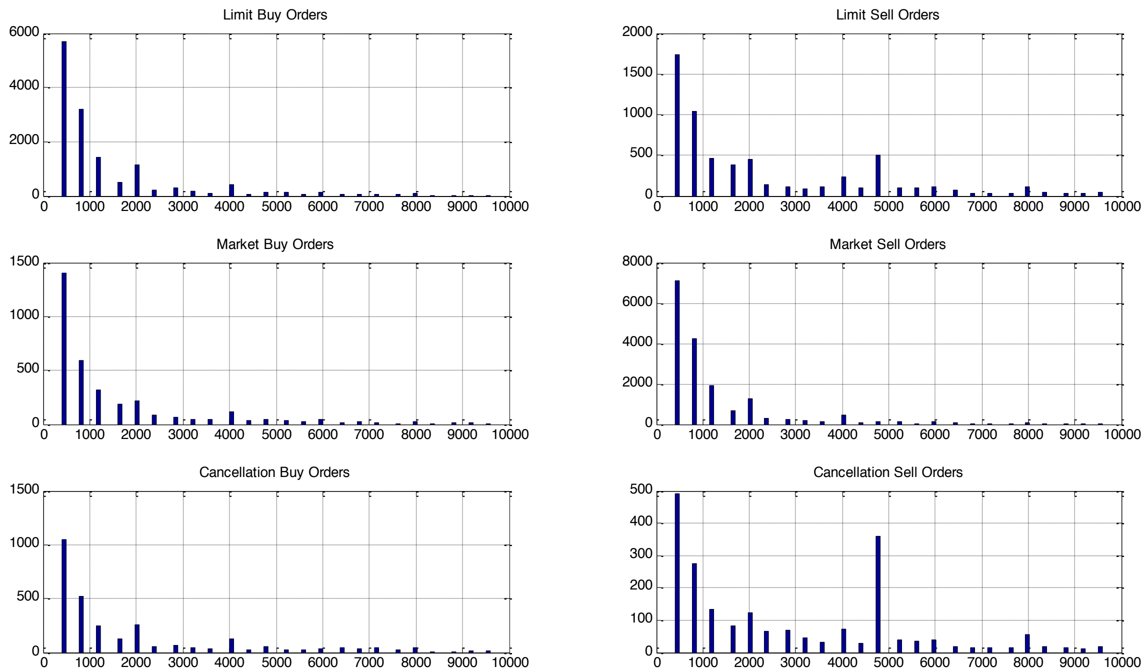

Fig 4. Order size distribution.

It's obvious to see that the order size follows an exponential distribution as well, which I believe is close to the reality, so I decided to add order size as another random variable to form a compound renewal process, in which we use the Weibull distribution again to characterize its dynamics.

3.2.3 Event Correlation

The last doubt I have is that I believe all the events should have some inner connection between each other, whereas the original model claims that orders are independent. I first calculated the correlation between all kind of orders arrival duration and its related size in Chart 1.

1	-0. 09015	0. 455609	-0. 18231	0. 294676	-0. 09951	0. 542464	-0. 08124	0. 414945	-0. 02833	0. 332325	-0. 10009
-0. 09015	1	-0. 02315	0. 120038	-0. 06451	0. 029154	0. 035758	0. 28201	-0. 13633	0. 280543	-0. 04331	0. 022291
0. 455609	-0. 02315	1	-0. 23662	0. 33788	-0. 26002	0. 402279	-0. 02452	0. 366383	0. 030002	0. 350888	-0. 09453
-0. 18231	0. 120038	-0. 23662	1	-0. 13246	0. 188782	-0. 16282	0. 169525	-0. 19938	0. 029456	-0. 0874	0. 302906
0. 294676	-0. 06451	0. 33788	-0. 13246	1	-0. 09493	0. 25532	-0. 12832	0. 165132	-0. 05571	0. 190941	0. 015835
-0. 09951	0. 029154	-0. 26002	0. 188782	-0. 09493	1	-0. 06947	0. 003814	-0. 13788	-0. 07008	-0. 19063	0. 024543
0. 542464	0. 035758	0. 402279	-0. 16282	0. 25532	-0. 06947	1	-0. 10245	0. 444973	0. 090028	0. 26974	-0. 04697
-0. 08124	0. 28201	-0. 02452	0. 169525	-0. 12832	0. 003814	-0. 10245	1	-0. 06431	0. 047303	-0. 06717	0. 05845
0. 414945	-0. 13633	0. 366383	-0. 19938	0. 165132	-0. 13788	0. 444973	-0. 06431	1	0. 175695	0. 204552	-0. 12019
-0. 02833	0. 280543	0. 030002	0. 029456	-0. 05571	-0. 07008	0. 090028	0. 047303	0. 175695	1	0. 009468	-0. 05362
0. 332325	-0. 04331	0. 350888	-0. 0874	0. 190941	-0. 19063	0. 26974	-0. 06717	0. 204552	0. 009468	1	-0. 03486
-0. 10009	0. 022291	-0. 09453	0. 302906	0. 015835	0. 024543	-0. 04697	0. 05845	-0. 12019	-0. 05362	-0. 03486	1

Chart 1. Order duration and size correlation matrix.

The above 12×12 correlation matrix is distributed as following: Limit Buy Orders Duration at the first row, Limit Buy Orders Size at the second row, Limit Sell Orders Duration at the third row, Limit Sell Orders Size at the forth row, and so on. The highlighted items are those with high correlation, and I pick them up in Chart 2.

- **Corr(Limit Buy Duration, Market Sell Duration)=0.54**
- Corr(Limit Buy Duration, Limit Sell Duration)=0.46
- Corr(Market Sell Duration, Cancellation Buy Duration)=0.45
- Corr(Limit Buy Duration, Cancellation Buy Duration)=0.42
- Corr(Limit Sell Duration, Market Sell Duration)=0.40

Chart 2. High correlation entries.

One thing I found really interesting is the correlation between limit buy duration and market sell duration. The high correlation could be explained by the basic economic principles, where limit buy orders represented demand. The high demand will boost the price, according to the economic principles, so people will tend to buy more shares to gain profit, and that's why market sell orders appears so soon. Another interesting fact is that the size correlation is relatively small, so we could say that the size is an independent factor in the queuing system. So I integrated the correlation of limit buy duration and market sell duration using the same empirical distribution function $f(x, y)$, as is describes in the previous session.

3.3 Model Application

After having the Markovian queuing model discussed above, I hypothesized that the model could accurately simulation the dynamics of the real market. Having this hypothesis in mind, I start to test multiple trading strategies based on this model, especially for market makers.

3.3.1 Market Making

In practice, it is of great interest to understand the mechanism of basic market making activities and the impact of possible strategies. Intuitively, market makers reap profits by posting same size ask and bid order simultaneously, selling highly and buying low. Here, I consider the following model problem. Suppose the market maker posts N shares at time t, and he is obliged to close his position with completion time T, that is, close N share position in $[t, t+T]$. The market maker could post limit order on any level of order book he want, and he is expecting other traders in the market could hit both his bid order and ask order. The market maker must close his position at time $t+T$ by purchasing (selling) the remaining position at market price. The most important decision the market maker has to make is which level he should put the order at. It's conceptual obvious that posting orders at a deeper position (closer to market price) will increase the execution probability while acquiring small profit, and posting orders further will get more profit by taking more risk. In reality, most market maker will have a targeting profit margin and risk tolerance threshold. Then it is not obvious quantify the results of putting orders on different levels. I tested this market making strategies with different completion time T, number of shares N and level of order placement by running independent Monte Carlo simulation. The results will be analyzed in the Discussion session.

3.3.2 Market Making with balancing strategy

The previous study on the market making activities is the simplest version. Here, we considered a more complex situation. Normally, rather than leaving the portfolio alone after posting the initial shares at time t (which is the assumption in the previous strategy), most market makers will follow the trend of the market and modify the portfolio according to the market performance in real time. One normal strategy is to rebalance the position around the current price, since it assures the equal probability for both sides to be executed. Here, besides the model problems in the previous session, I also integrated the balancing strategy, which shows as follows:

At each end of time interval Δt

- If currently no position is on the best level and both bid and ask size hasn't been depleted yet, rearrange the remaining shares to be balanced around the current market price.

- If both sides haven't been depleted and the current price is already balanced, decrease the margin of both sides by one tick.

- If one size already depleted, try to increase the other side by one tick.

In the model setup, I ignore the cancellation fee for cancellation the original orders, since in most exchanges, there will be a "liquidity award" for posting limit order, which is approximately the same amount of the cancellation fee. I tested this balancing strategies again with different completion time T, number of shares N and level of order placement by running independent Monte Carlo simulation. The results will be analyzed in the next session.

3.3.3 Smoking Strategy

It is highly interesting to understand the high frequency market maker's trick beyond the basic market making strategy. One of the widely used strategies is called Smoking Strategy. High-Frequency traders place small alluring quotes to attract other side's market order. Suppose another party detects this alluring quote (usually through automated trading system), they will then send out a market order. Since market orders are only executed at the best market price, HF traders can cancel the quotes before the market orders arrive, thus letting the other party hits the large shares previous posted by HF traders. Fig 5 illustrated one example of the smoking strategy.

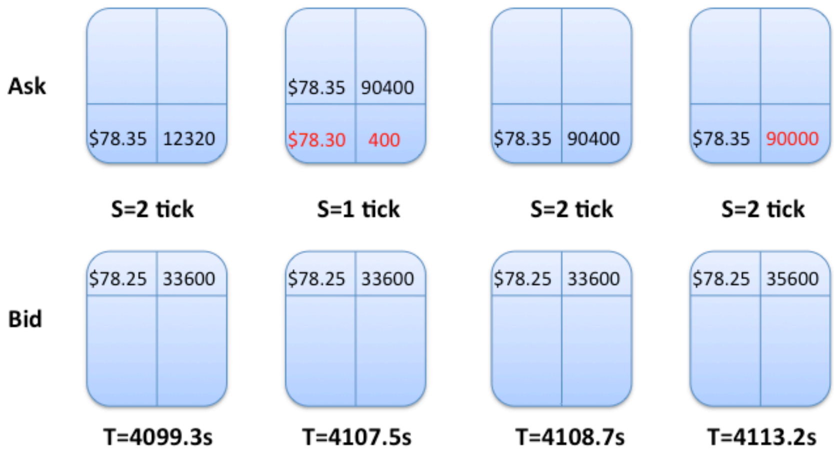

Fig 5. An example of smoking strategy.

The smoking strategy could be happened in both ask and bid side. The key for the smoking strategy is the speed, which is exactly the advantage of high-frequency traders. I examined the 5 days Hong Kong traded stock data I have, and found out that there are 20 occurrences happened in 5 days (7 bid side, 13 ask side). The average alluring quote size is 7,280 shares, the average cancellation duration is 1.98s, and the average trading volume is 2,160 shares. The histogram of cancellation duration is shown in Fig 6.

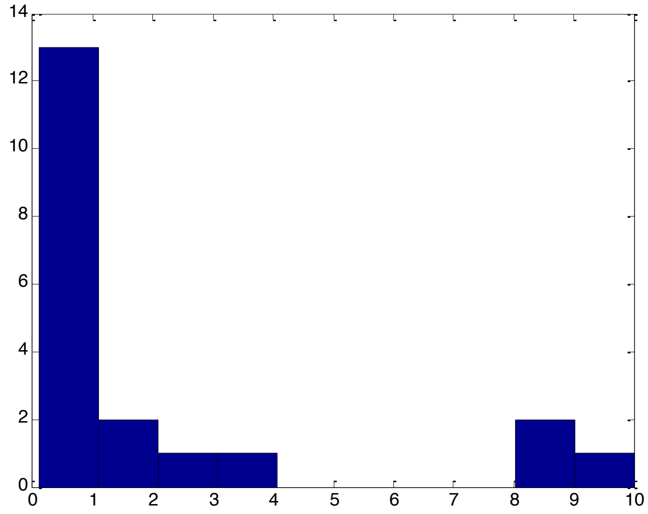

Fig 6. Cancellation duration histogram.

Understood that the prime parameter for the success of smoking strategy is the cancellation duration, I tested the smoking strategy based on the model by running independent Monte Carlo simulation. The result will be discussed in the next session.

4 Discussion

4.1 Model Simulation Results

To sum up the methodology, I employed a Markovian Queuing Model [2] whose transition corresponds to the order book events $\{T_i^a, i \geq 1\} \cup \{T_i^b, i \geq 1\}$. There are six kinds of orders in the system. For limit buy orders and market sell orders, I use the empirical probability distribution to generate the simulation queue. Other than that, the corresponded Weibull distribution is used. The order size and duration are seen as independent to each other. The final simulation queue contains around 12,000 events in a day. The details are documented in Chart 3.

	Num of Events	Mean Duration	Mean Size
Limit Buy Orders	3319	5.42s	5600
Limit Sell Orders	1962	9.20s	12800
Market Buy Orders	1114	16.16s	8800
Market Sell Orders	3829	4.69s	3200
Cancellation Buy Orders	994	18.12s	13200
Cancellation Sell Orders	734	24.63s	13600

Chart 3. Discrete event simulation statistics.

Based on this queue, I ran a simple simulation to test the validity of this model. The simulation starts with the first price of the first day, and one sample path is shown in Fig 7 where the red line is the real first day data, and the blue line is the simulated result. One can tell that the simulation results catch the sense of our market pretty good.

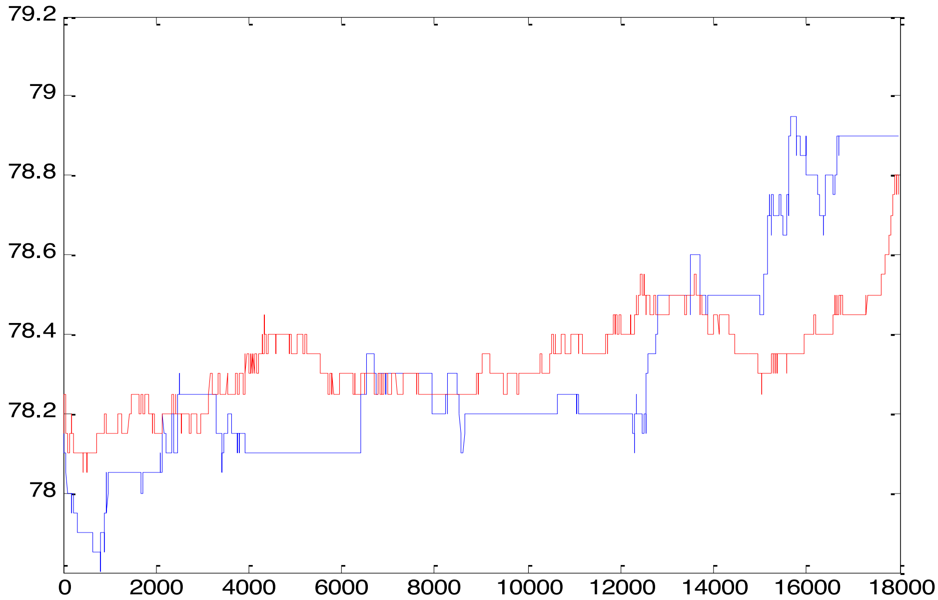

Fig 7. A sample path of simulated stock prices.

4.2 Market Making Simulation Results

According to the previous discussion, I tested the market making strategy based on the Markovian Queuing System. The parameter is *nLevel*, which range from 0 to 5, representing the level market maker put the orders; Completion Time T, which range from 3600s to 18000s (1 to 5 hours), representing the total amount of time for closing the position; Number of shares N, which range from 400 to 2,000, representing the total shares used by the market making strategy. The results are shown in the following figures:

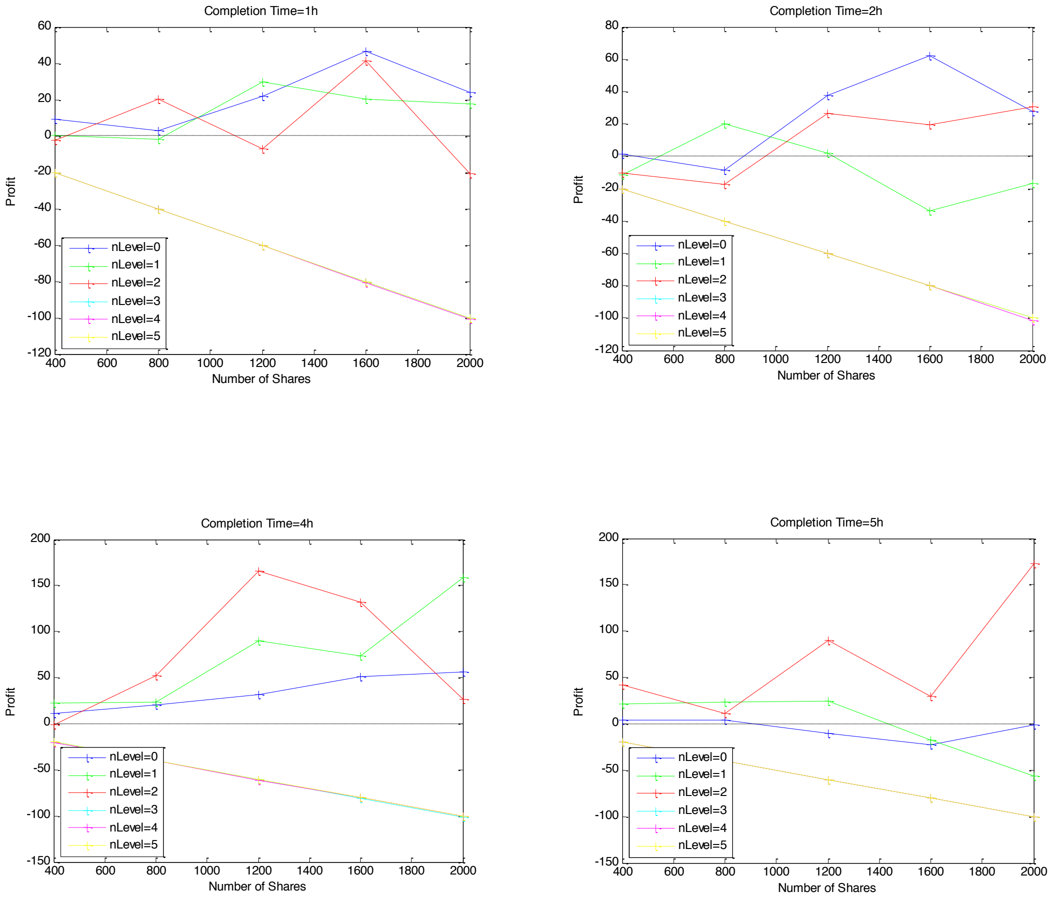

Fig 8. Simulation results for market making strategies.

As is shown in fig 8, the completion time determined the optimal order placement of market making strategies. If the completion time is one hour, the optimal strategy is to put the order on best level, while placing orders on level 2 will results in best profit performance if the completion time is more than 4 hours. This result generally reflects the reality. If the time frame is short, placing orders deeper will results in quicker order execution, while there is not much time for market actually moving to the further orders. In the longer time frame, market fluctuates more severely, thus enabling placing order further to reap more profit. This simulation results reveal the relationship between general profit, level of placement, completion time and number of shares used, and it could provide a benchmark for market making activities.

4.3 Market Making With Balancing Simulation Results

As discussed above, I explored a more complex market making strategy, which re-balancing the position every other Δt . In the simulation stage, I fixed the $\Delta t = 1800s$, which is half hour. Ran similar simulation as above, I got the following results:

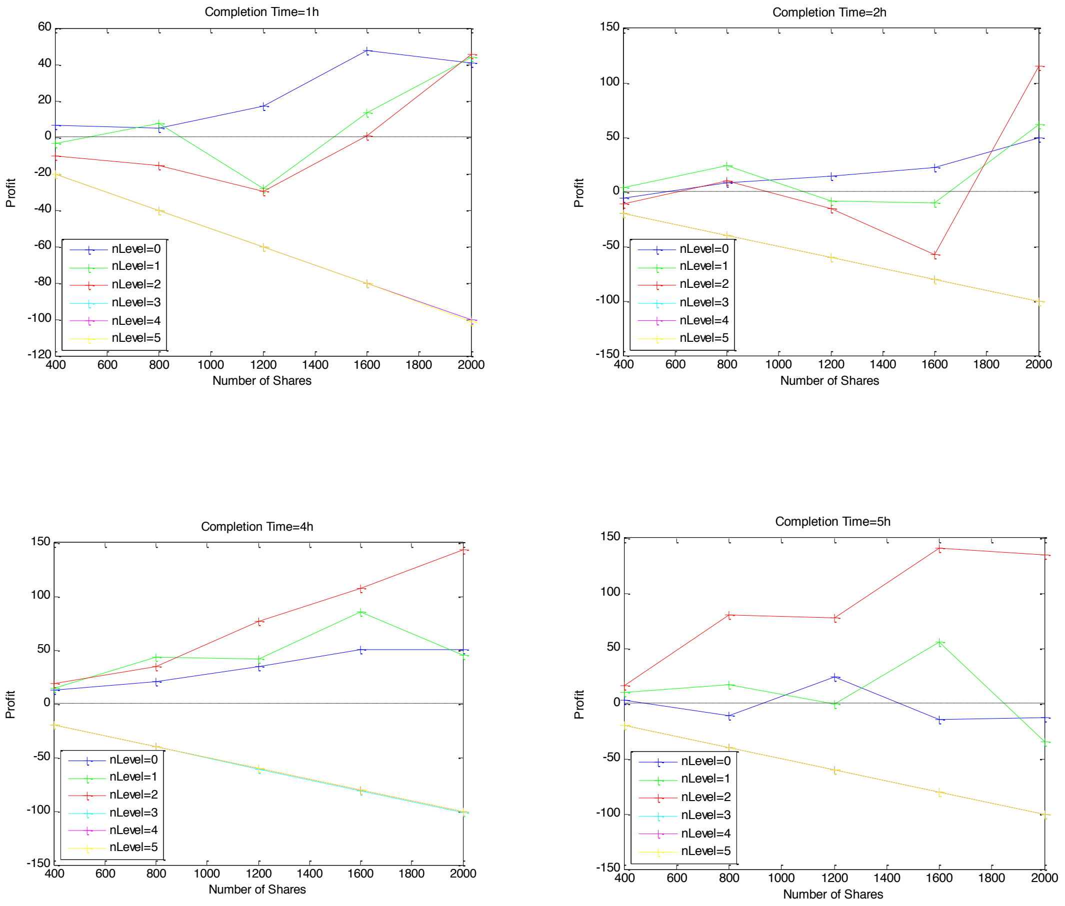

Fig 9. Simulation results for market making strategies (with balancing).

Having integrated the balancing option into the market making strategy significantly enhances the performance of market making strategies, as is shown in fig 9. Comparing to non-balancing strategies, the optimal trade points are moved upper-right side, which means that market makers could utilize more shares, posting further and reaping more profits by posting shares on both bid/ask sides and keeping the position balanced. This simulation result validates the balancing option in market making strategies.

4.4 Smoking Strategy Simulation Results

One of the drawbacks of the Markovian Queuing System is that it only focuses on the best level, that is, the best ask orders and best bid orders. The system addresses the level II information by integrating the empirical probability spaces only when the best level orders have been depleted. Because of this limit, the smoking strategy cannot be completed simulated, cause we cannot differentiate the market orders after successful alluring orders and the limit orders which fill in the two tick gaps. Fortunately however, we can simulate one of the most important factors for smoking strategy, that is, the cancellation duration. It is obvious that the shorter the cancellation duration is, the lower probability that the other party will accidentally hit the alluring orders. If the cancellation duration goes to zero, then there is 100% probability that the alluring orders will not be hit, though it won't be able to be detected as well. Therefore, I defined the probability of the alluring orders successfully cancelled as the *Success Probability*, and ran 10000 independent Monte Carlo simulations against the model. The simulation results are shown as follows:

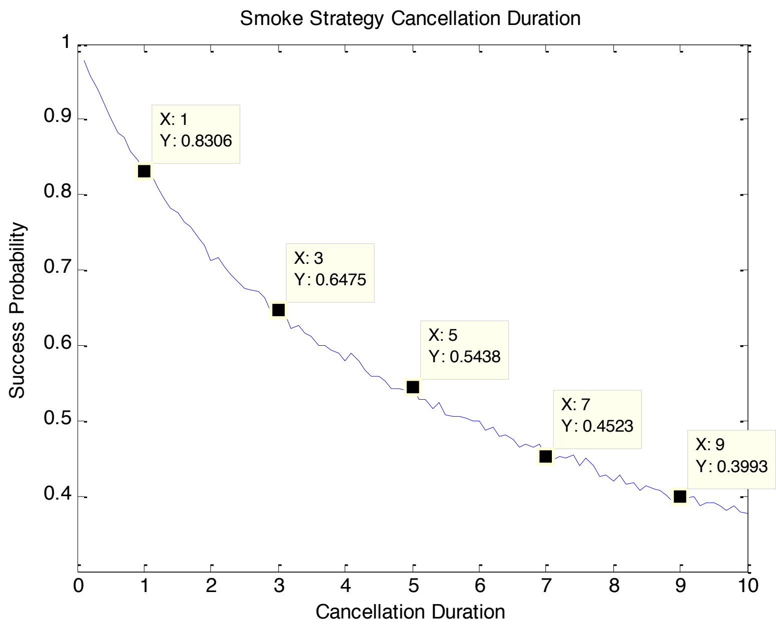

Fig 10. Simulation results for smoking strategy.

As is shown in fig 10, if the market maker cancel the alluring orders at 1 seconds, he will get 83.06% probability that the orders will be successfully cancelled; and if the market maker cancel the alluring orders at 3 seconds, the probability drops to 64.75%; if the market maker cancel the alluring orders at 9 seconds, there will be only 39.93% probability to successfully cancellation. Because the cancellation duration is the most important factor, this conceptually simple simulation results could be served as guidance for those market makers who want to employ the smoking strategy.

5 Group Summary

Formed as a group of five members, we also investigated the High frequency trading from other aspects and gain meaningful results as well.

To better understand the nature of random stock price, we investigated the price fluctuation phenomenon, which, according to J. P. Bouchaud [17], strictly follows a diffusive behavior. It is encouraging to found out that the mean-square fluctuation of our data shows a linear behavior at a longer time domain, which indicating a strong diffusive behavior. Furthermore, we were also surprised by the long-term correlation of trading signs, which could be described using Power Law. This assumption was justified by our empirical analysis of the data. Having these notions about price fluctuation in mind, we proposed a superposition model of price fluctuation, where the price at a certain time point is described as a sum over all past trades. This price fluctuation model takes account into the long-term correlation of the trading sign and captures the diffusive behavior of the stock prick effectively, especially in Hong Kong exchange. Although this model is still an over-simplified model that doesn't take into account of the limit order and cancellation order as price shifting momentum, it is a good starting point for understanding the extreme sophisticated price fluctuation behavior.

We also investigated the inter temporal dependence models to explain the empirical observation of volatility clustering, which is a natural application for the GARCH model introduced by Bollerslev in 1986. The GARCH specification allows the current condition variance to be a function of past conditional variances. Good parameter estimation is the key for developing GARCH model. We estimate the parameter by adopting a $L_{1/2}$ regularizer algorithm that has unbiasedness, sparsity and oracle characteristics. The experiments show that the $L_{1/2}$ regularization method is very useful and efficient to solve the GARCH model. We further developed the GARCH model by incorporating transaction volume and turnover as additional factors. Experiments show that the developed GARCH model outperformed the original one by 30%-40%. This algorithm provides an interpretable model while reducing the computation complexity.

Other than these, we also spent some time looking into the regulation issue of High frequency trading market. There are already some audit system designed to regulate this market, but the major debates of these audit system is the diploma of reducing volatility and increasing risks. We believe that the government regulation is one of the most important factors of High frequency market and every practitioner should not overlook this issue in any case. We explored the High frequency market from several different prospective, each provides us at least a little insight into this promising yet challenging field.

6 Conclusion

In summary, this paper has explored the price dynamics and its application in high frequency trading market. I have adopted discrete Markovian queuing model to describe the limit order book dynamics of the stock in Hong Kong Exchange. In the discrete Markovian queuing model, order events (limit order, market order and cancelation) are characterized as by Weibull distribution for different arrival rate, respectively. Weibull distribution is also used to characterize the arrival order size, which is an additional parameter of the original model setting. Empirical correlations between different types of orders are also employed. Simulation results reveal that the Markovian Queuing Model is more suitable for characterizing the price dynamics of illiquid market than the continuous models. Several market-making strategies are explored based on the queuing model, and the simulation results could provide guidance for market making activities.

There remain lots of spaces for improvement. First, the Markovian queuing system only takes the level I order book into account, where the level II information is characterized as empirical distribution. Having modeled the dynamics of level II order book will enable more sophisticated researches into the market dynamics and more accurate simulations such as smoking strategy simulation. Second, large order impact are also significant in reality, which has been ignored in the previous model setting. Finally, more accurate parameter estimation method could be applied to fit the mode.

7 Reference

[1] R. CONT AND A. DE LARRARD, *Linking volatility with order flow: heavy traffic approximations and diffusion limits of order book dynamics*, working paper, 2010

[2] R. CONT AND A. DE LARRARD, *Price dynamics in a markovian limit order market*, tech. report, Available at SSRN: http://ssrn.com/abstract=1735338, 2010.

[3] R.CONT, A. KUKANOV AND S. STOIKONOV, *The price impact of order book events*, working paper, SSRN, 2010

[4] J.P. BOUCHAUD, M. ME´ZARD, AND M. POTTERS, *Statistical properties of stock order books: empirical results and models*, Quantitative Finance, 2 (2002), pp. 251–256.

[5] J. D. FARMER, L. GILLEMOT, F. LILLO, S. MIKE, AND A. SEN, *What really causes large price changes?*, Quantitative Finance, 4 (2004), pp. 383–397.

[6] E. SMITH, J. D. FARMER, L. GILLEMOT, AND S. KRISHNAMURTHY, *Statistical theory of the continuous double auction*, Quantitative Finance, 3 (2003), pp. 481–514.

[7] B. BIAIS, L. GLOSTEN, AND C. SPATT, *Market microstructure: A survey of microfoundations, empirical results and policy implications*, Journal of Financial Markets, 8 (2005), pp. 217–264.

[8] T. FOUCAULT, O. KADAN, AND E. KANDEL, *Limit order book as a market for liquidity*, Review of Financial Studies, 18 (2005), pp. 1171–1217.

[9] L. GLOSTEN, *Is the limit order book inevitable?*, Journal of Finance, 49 (1994), pp. 1127–1161.

[10] M. AVELLANEDA AND S. STOIKOV, *High frequency trading in a limit order book*, Quantitative Finance, 8 (2008), pp. 217–224.

[11] A. ALFONSI, A. SCHIED, AND A. SCHULZ, *Optimal execution strategies in limit order books with general shape function*, Quantitative Finance, 10 (2010), pp. 143–157.

[12] R. ALMGREN AND N. CHRISS, *Optimal execution of portfolio transactions*, Journal of Risk, 3 (2000), pp. 5– 39.

[13] D. BERTSIMAS AND A. LO, *Optimal control of execution costs*, Journal of Financial Markets, 1 (1998), pp. 1– 50.

[14] R. CONT, S. STOIKOV, AND R. TALREJA, *A stochastic model for order book dynamics*, Operations Research, 58 (2010), pp. 549–563.

[15] Y. NEVMYVAKA, K. SYCARA, AND D. SEPPI, *Electronic Market Making: Initial Investigation*, In the Proceedings of Third International Workshop on Computational Intelligence in Economics and Finance, 2003

[16] Y. FENG, R. YU, AND P. STONE, *Two Stock Trading Agents: Market Making and Technical Analysis*, Lecture Notes in Artificial Intelligence, Springer Verlag, 2004.

[17] J. P. BOUCHAUD, Y. GEFEN, M. POTTERS, AND M. WYART, *Fluctuations and response in financial markets: the subtle nature of "random" price changes,* Quantitative Finance, 4(2), p176-190, 2004