# The Impact of Market Structure on Price Determination: A Simulation Approach Using Multi-Agent Reinforcement Learning in Continuous State and Action Space

by
Buliao (Jerry) Shu
Bachelor of Business Administration (Honours)
in Quantitative Finance and Risk Management
City University of Hong Kong, 2013

Submitted to the MIT Sloan School of Management
in partial fulfillment of the requirements for the degree of
Master of Finance
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2014

Signature redacted
Author ..........................................................................
MIT Sloan School of Management
May 09, 2014

Certified by... Signature redacted ....................
Robert C. Merton
MIT Sloan Distinguished Professor of Finance
Thesis Supervisor

Accepted by.... Signature redacted ...................
Heidi Pickett
Program Director, Master of Finance Program
MIT Sloan School of Management

# The Impact of Market Structure on Price Determination: A Simulation Approach Using Multi-Agent Reinforcement Learning in Continuous State and Action Space

by

Buliao (Jerry) Shu

Submitted to the MIT Sloan School of Management
on May 09, 2014, in partial fulfillment of the
requirements for the degree of
Master of Finance

## Abstract

This thesis proposes a simulation tool to study the question of how market structure and market players' behavior affect price movements. The adaptive market simulation system consists of multiple agents and a centralized exchange. By applying reinforcement learning techniques, agents evolve and become capable of making intelligent trading decisions while adapting to changing market conditions. Trading dynamics in the real world are complex yet compelling. The presence of the human element in trading makes studying it via repeatable scientific models, especially on a large scale, very difficult and almost unfeasible. By making it possible to conduct controlled experiments under various market scenarios, this simulation seeks to help researchers gain a better understanding of how different types of traders affect price formation under distinct market scenarios. The impact of trading frequency on prices is also explored as a test of the simulation tool. Results suggest that the market generates richer information when the frequency of trading is high, and when the market is more frequently accessed, short-term market prices demonstrate higher volatilities and move faster in respond to market sentiments.

Thesis Supervisor: Robert C. Merton
Title: MIT Sloan Distinguished Professor of Finance

# Contents

# List of Figures

# Chapter 1

# Introduction

In order to better understand how market forces affect price formation, a simulation method is developed so that various control experiments can be conducted to analyze the impact of different settings in a market. Prices have historically been and always will be a key parameter of the markets, serving as the crucial signal in all trading activities. However, there remains till this day a limited understanding of how different market structures influence price formation. Researches in this area rely heavily on historical data, the insufficiency and static nature of which greatly restricts our knowledge of the price determination process. The human factor also makes it impossible to conduct interactive research based on large-scale scientific experiments. Moreover, with numerous breakthroughs in trading technology and unprecedented internationalization of trading, the marketplace has never been this complex. The task of understanding price formation has also become more challenging than ever before.

The adaptive market simulation is proposed as a tool to tackle the aforementioned challenges by providing an alternative way to conduct research on price impacts. The tool effectively simulates traders that are adaptive to the market. In the simulation, traders with different goals submit bid-ask orders to a centralized exchange, and will adjust their behavior according to newly generated trading information. This tool allows researchers to construct simulations under different artificial market scenarios; and directly compare and analyze the price impact of factors such as the supply-demand imbalances, assorted numbers of participants, and varying market compositions. As a starting point to test the proposed simulation method, this study investigates one of the key aspects of market structure–trading frequency.

By capturing the characteristics of the real world market, this adaptive market simulation sheds much light on the factors affecting the price formation, and also reveals the extent to which each factor influences price formation. A sound understanding of market participants and the exchange market mechanisms is crucial for construction of a meaningful simulation. The following section details the evolution of behaviors among market participants and the market structure, and also discusses the reinforcement learning method used in designing the simulation.

## 1.1 The Evolution of Market Participants

Information and the behavior of market participants are among the key factors affecting prices. In their study of housing price bubbles, Karl E. Case and Robert J. Shiller (2003) put forth that future expectations of prices and word of mouth are the two main contributors to a housing bubble formation. Andrew Lo (2002) examines the tech bubble and argues that markets are largely driven by the emotions of investors. Information defines the mindsets of market participants, and the participants' behavior subsequently determines prices. These researches indicate that the behavior of market participants generates information, which in return strengthens or deteriorates the behavior itself in a self-fulfilling cycle. Based on these researches, it is evident that a bottom-up approach in modeling the market from the level of individual participants will yield a more comprehensive and accurate picture of the market.

The composition of market participants also changes over time. Under the old investment regime, wealthy individuals were the major market participants. As the investment management industry flourished, institutional investors who held large quantities of capital relative to individual investors then became the leading players. Today, with tremendous advancements in trading technology, high frequency trading firms rein as the new dominant players, contributing over 77% of daily trading volume (Brogaard 2010). Because high frequency traders maintain positions for very short periods of time, price changes are of far greater concern to them than absolute price levels. The different dominant market players naturally cause prices to move in disparate directions.

Given the significant daily trading volume from high frequency traders, it is natural to think about their impact on the market. Even though extensive studies have been carried out, opinions on high frequency trading remain diverse. Yacine Ait-Sahalia and Mehmet Saglam (2013) proposes a model for the scenario where high frequency traders receive imperfect signals and can exploit low frequency traders. Thomas H. McInish and James Upson (2013) measure the revenue that high frequency trading firms earn by taking advantage of market structural issues. Frank Zhang (2010) concludes that high frequency trading is positively correlated with stock price volatility. On the other hand, Eurex (2011) argues that high-frequency traders alone should not be blamed for volatile markets and major price fluctuations. David Easley, Marcos M. Lopez de Prado, and Maureen O'Hara (2012) suggest that high frequency traders do not hold a particular advantage over others because of their greater speed, and it is their decision making skills that give them an edge in the market. These discussions motivate the creation of the proposed simulation system to directly compare a market that has high frequency access with one that only allows low frequency trading activities.

## 1.2 A Review of the Market Structure

A sensible simulation of the market should be an abstraction of the real world, capturing the major characteristics of the two major components in the market: the traders and the exchange.

Traders in the market can be categorized according to their purpose of trading, market functions and/or behavior characteristics. Based on objective for trading, traders can be classified as hedgers, speculators or liquidity suppliers. Speculators can be further divided into informational traders and noise traders depending on the amount of information they possess. Liquidity providers are those who submit limit orders to the exchange and are willing to take urgent market orders in relatively large quantities. Liquidity providers normally aim to maintain a low level of inventory and profit solely from market making activities. Many high frequency traders play the role of liquidity providers.

From another perspective, traders can be put into three brackets characterized by their daily trading volume and net positions. Andrei Kirilenko (2011) characterizes the three brackets as follows: (1) high frequency traders with high volume and low net positions, (2) intermediaries and opportunistic traders with low volume and low net positions, and (3) fundamental buyers and sellers with low volume and high net positions.

Although the behavior of traders under different categories may vary, their activities can nevertheless be simply abstracted as taking and exiting positions. The duration of a trade cycle depends on the specific objective of each type of trader. Market makers aim for a net zero position normally at the end of the day; high frequency firms submitting market-making orders close out their positions within an even shorter time interval; while long-term institutional fundamental traders take a much longer time to build up their positions and hold the investment for months or even years before sale. The review of the similarities and differences between traders suggests that the simulated traders should be distinguished according to their goals.

The exchange is a venue that facilitates trading activities. The exchange can be either a continuous market or a call market depending on the way its trading sessions operate. In a continuous market, traders can trade at any time. In a call market, trades for a specific security are only allowed when the security is called. Depending on the trade execution system, exchanges can be a quote-driven dealer market where the counterparty of all trades is the respective designated dealer, or an order-driven market where traders are treated equally and all transactions are based on submitted orders. Different auction models may be adopted in an order-driven market to match and execute orders.

## 1.3 Techniques to Model the Marketplace

The aim of the simulation is to construct traders that can formulate intelligent trading strategies and adapt to changes in information. Reinforcement learning may be applied to construct smart traders who can build proper strategies and eventually achieve their goals. The learning technique is a procedure whereby a policy plan that outputs actions to maximize cumulative expected reward is built. An agent builds up an optimal plan by learning through its interactions with the environment, and then updates its reward score corresponding to an action under a particular state. The method works in environments that can be formulated as Markov decision process (Howard, 1960). Many interesting applications of reinforcement learning have been developed to solve engineering problems. Classic examples include the mountain car (Sutton, R. S. 1996), robot learning (Connell and Mahadevan, 1993) and elevator dispatching (Crites and Barto, 1996). In finance, M. A. H. Dempster, Tom W. Payne, Yazann Romahi, and G. W. P. Thompson (2001) apply the reinforcement learning methodology for intraday FX trading. Common reinforcement learning techniques such as Q-learning are designed for environments where states and actions are finite and discrete.

In many real world applications, states and actions in an environment cannot be partitioned into finite pieces and therefore both have to be treated as continuous variables. Function approximation provides a possible solution for modeling states that have continuous values (Barto, A. G. 1998, Kenji Doya, 2000). To further model environments in which actions are also continuous, Baird and Klopf (1993) and Prokhorov and Wunsch (1997) and Hado van Hasselt and Marco A. Wiering (2007) proposes algorithms that may be applied. While most of the research focus on a single agent in an environment, Michael L. Littman (1994) studied interactions between agents who share one environment, and proves the agents' ability to learn under reinforcement learning.

A market can be successfully simulated once the exchange and agents are properly constructed. The simulation of a trading system can be categorized as a complex adaptive system. Complex adaptive systems refer to systems whose components learn and adapt as they interact (John H. Holland 2005). Any system that has many individual components interacting with each other fits the definition of a complex adaptive system. Michael J. Mauboussin (2002) uses the concept of a complex adaptive system to study market efficiency. W. Brian Arthur, John H. Holland, Blake LeBaron, Richard Palmer, and Paul Tayler (1996) simulate the stock market with heterogeneous agents. In their research, agents are modeled to predict future stock returns. The concept of complex adaptive system is applicable and valuable to many fields such as economics, biology, and sociology.

# Chapter 2

# Methodology

This chapter discusses the construction of the adaptive market simulation. It first provides an abstraction of the real-world exchange and addresses the process that the proposed simulation method focuses on. It then gives a detailed description of the structure as well as the algorithms used in the proposed simulation.

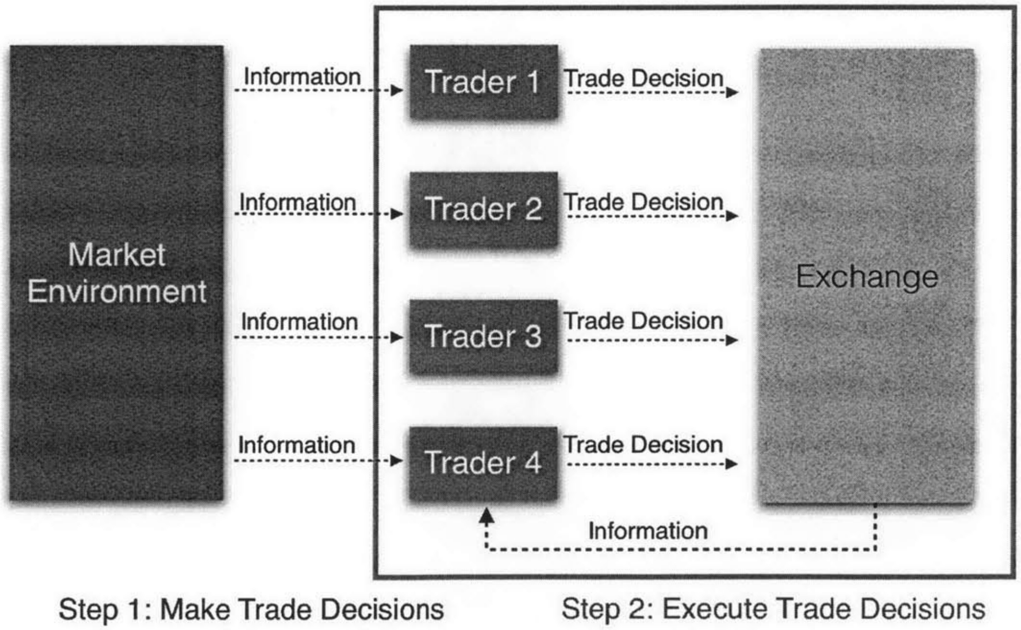
## 2.1 Abstraction of the Market

The behavior of traders in the exchange market can be viewed as a two-step cycle. The first step is converting information into decisions. The second step is executing the decision. In the first step, traders collect market information including but not limited to: news, company announcements, related reports, prices of other assets, and observations on the market behavior. After careful analysis, traders decide on a level of market exposure and the corresponding number of shares that they would like to trade.

The proposed simulation is designed to imitate the second step of the market: execution of the decision. In reality, this second step is completed in a relatively small time period. Within this short time window, it is reasonable to assume that the fundamentals of an asset are unchanged, meaning that there is no new information. The traders' decision (from the first step) is thus assumed to remain unchanged, and they are able to concentrate on executing their original decision. During the second phase, although no new fundamental information hits the market, there is still technical information forthcoming: transaction data. The market aggregates information from the interactions between traders. For example, the price and order book reflect the balance between supply and demand as well as the execution strategies of traders.

Using equity of Company A as an example to illustrate the two-step process: the first step occurs when Company A releases news about a merger deal. Traders evaluate all the available information and come to decisions on their desired exposures to Stock A. The market then moves to the second step. In this step, traders make trades with each other. As no other new information is revealed within the short time window of the second step (execution of the trade decision), the fundamental value of Company A is assumed to remain constant then. However, the price fluctuations of Stock A reflect the judgment of traders, and traders will adjust their strategies accordingly.

Figure 2-1: Two Steps in the Market



The market serves two purposes in the second step. The first function is to serve as a centralized place for facilitating transactions. In other words, it connects buyers with sellers and reduces search costs. Furthermore, the centralization of all transactions stimulates competition between the counterparties and makes it easier for traders to find the best price.

The second function of the market is to generate information. When traders execute their trade decisions, price, trading volume, and order flows all allude strongly to the current market sentiment. Traders may originally have limited knowledge about how other participants view the market, but through trading they are able to estimate the overall market sentiment and make judgment calls on important issues (e.g. total demand and supply, the level of market fragmentation and the presence of dominant players, etc.). In short, the process of trading causes the entire market to become better informed.

## 2.2 Major Components of the Simulation

The simulated marketplace adopts a continuous order-driven market model, allowing participants to submit orders at any time. A double auction model is used for order matching. Traders with different objectives will participate in the market. The simulation consists of three major components:

- The Exchange

  The Exchange is the environment in which traders submit their orders. It executes trades and stores all the related data including account information, historical prices, and historical bid/ask prices.

- Type A Trader

  Type A traders are traders who want to build up a net exposure to the security. At the beginning of the simulation, they will be assigned a target position, and their goal is to achieve that target position with minimal costs. The assignment of their position can be regarded as the decision outcome from information analysis in the first step of the market. The performance score for Type A traders is based on both the level of the target achievement and the cost level for building up the position.

- Type B Trader

  Type B traders are market makers. The goal of Type B traders is to make profits while maintaining a zero final net exposure to the security.

## 2.3 Key Parameters

**Trading Session N:**

The number of sessions controls the frequency of trading. One round of trading simulation corresponds to a very short period of time, $T$, in real world. Therefore, $\frac{N}{T}$ represents the frequency of trading, meaning that one can trade $\frac{N}{T}$ times per second. Increasing the number of sessions $N$ is equivalent to having a market with higher frequency in trading.

**Goal and Number of Type A Traders:**

Each Type A trader has a target position to reflect the disparate views of traders in the real marketplace when hit with an ambiguous event. This distinction also sets in motion the forces of supply and demand for the asset.

By setting goals and numbers, it is possible to have different levels of supply and demand. Given the same aggregate supply and demand, markets can also be differently fragmented. For example, a market can have ten traders wanting to buy 500 shares each and another ten traders hoping to sell 450 shares each. The same imbalance can happen when one trader has a dominant position compared to others, and is assigned to buy 5000 shares while another ten traders are assigned 450 shares each to sell.

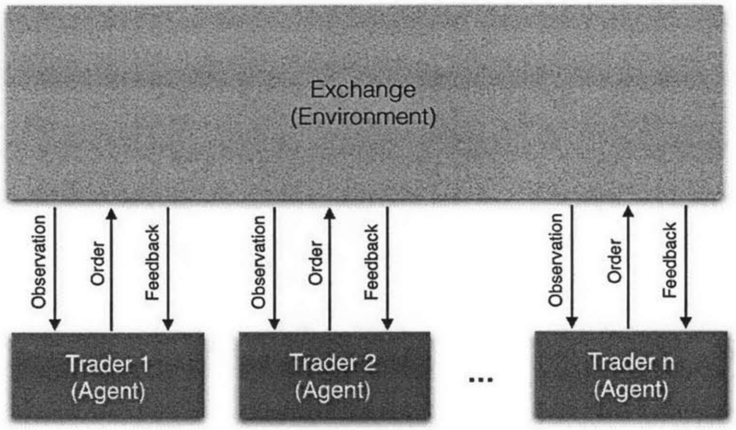
**Goal and Number of Type B Traders:**

The number of Type B traders is indicative of the competitiveness of market making: the more Type B traders a market has, the more competitive the market making business is. Collectively, Type B traders provide liquidity to the market and maintain a net zero exposure. As individuals focus only on personal profit, it is interesting to see how well they provide liquidity to the market as a group, as well as how profitability varies when the number of market makers changes.

## 2.4 Construction of Reinforcement Learning Traders

In reinforcement learning, traders interact with the exchange based on a policy plan. In each trading session, traders retrieve information about the current state of the market from the exchange and then submit bid or ask orders. The exchange executes orders, stores the information and sends feedback to traders. Traders then update their policy plan based on the feedback, and move on to the next trading session. The process can be illustrated in the following graph:

Figure 2-2: Reinforcement Learning with Multiple Agents



As both the state and action space in the environment are continuous variables, the classic Q-learning method utilizing the Bellman equation for policy update is not applicable. To deal with continuous variables, parameterized function approximation is used to store information of observed states and estimate actions for the unknown scenarios.

The policy plan of an agent can be divided into two parts: one value function to calculate the value of a state at time t, $V_t(s_t)$, and two action estimation functions to estimate the optimal order price and order size, $Acp_t(s_t)$ and $Acs_t(s_t)$. All approximation functions take the state observation $s_t$ as input. In order to reduce the complexity of the simulated system, agents adopt linear function approximation, and each component of the state observation is used as one input in the linear function. The coefficient $\theta^V$ controls the prediction of state values. The weights, $\theta^{Acp}$ and $\theta^{Acs}$, are used to estimate the optimal order price and size. Specifically, the value of states are calculated as:

$$V_t(s_t) = \theta_1^V * f_1(s_t) + \theta_2^V * f_2(s_t) + \ldots + \theta_n^V * f_n(s_t)$$

$$f_i(s_t) = s_{i,t}$$

$V_t$ stands for the estimated value of state $s$ at time $t$. $s_t$ consists of a vector of information from the market, such as historical price, order book, current account position, remaining session. $s_{i,t}$ represents the $i$th component in the state observation. The $f_i(s_{i,t})$ outputs the $i$'s component of the state observation $s$. The estimated order price and size are:

$$Acp_t(s_t) = \theta_1^{Acp} * f_1(s_t) + \theta_2^{Acp} * f_2(s_t) + \ldots + \theta_n^{Acp} * f_n(s_t)$$

$$Acs_t(s_t) = \theta_1^{Acs} * f_1(s_t) + \theta_2^{Acs} * f_2(s_t) + \ldots + \theta_n^{Acs} * f_n(s_t)$$

, , where $Acp_t(s_t)$ denotes the estimated optimal order price, and $Acs_t(s_t)$ denotes the estimated optimal order size.

To explore unknown states, a Gaussian distribution is used. The submitted order, $[as, ap]$, is obtained from a normal distribution with the estimated optimal price and order size as input parameters. The output order equals to:

$$order_t = [ap_t, as_t] \sim \mathcal{N}(\mu_t, \Sigma)$$

$$\mu_t = \begin{bmatrix} Acp_t(s_t) \\ Aps_t(s_t) \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_s^2 \end{bmatrix}$$

, where $ap_t$ and $as_t$ denotes the submitted price and order size respectively, and $\sigma_p$ and $\sigma_s$ denote the exploration rate for price and order size.

The idea is that there is a probability associated with each possible order. The probability of an order being closer to the estimated optimal order is larger than that of it being far away. The probability of submitting a selected bid and ask order is:

$$\pi_t(s_t, order_t) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp(-\frac{1}{2}(order_t - \mu_t)^T \Sigma^{-1}(order_t - \mu_t))$$

, where $\pi_t(s_t, order_t)$ denotes the policy, and $\pi$ denotes the mathematical constant.

The coefficient is updated by the Gradient Descent method, and the state value is updated after every trading session. The update function for the parameters of the state value is:

$$\theta^V_{i,t+1} = \theta^V_{i,t} + \alpha(V_{t+1}(s_t) - V_t(s_t))\frac{\partial V_t(s_t)}{\partial \theta_{i,t}{}^V}$$

$$V_{t+1}(s) - V_t(s) = (reward + \gamma * V_t(s'))$$

, where $\alpha$ is the learning rate, $\gamma$ is the discount rate for the value, and $s'$ is the next state observation. *reward* measures the change of performance score, which is discussed in the next section. As the linear function approximation takes in the components of $s$ directly, the updates can be written as:

$$\theta^V_{i,t+1} = \theta^V_{i,t} + \alpha(V_{t+1} - V_t)s_{i,t}$$

The idea for updating the coefficients in the action estimation is to increase the probability of the action that brings positive change to the state value:

If $V_{t+1}(s) - V_t(s) > 0$: increase $\pi_t(s_t, order_t)$

The coefficients for action estimation are updated as:

$$\theta^{Acp}_{i,t+1} = \theta^{Acp}_{i,t} + \alpha(ap_t - Acp_t(s_t))\frac{\partial Acp_t(s_t)}{\partial \theta^{Acp}_{i,t}}$$

$$\theta^{Acs}_{i,t+1} = \theta^{Acs}_{i,t} + \alpha(as_t - Acs_t(s_t))\frac{\partial Acs_t(s_t)}{\partial \theta^{Acs}_{i,t}}$$

By adopting linear function approximation and feeding in the components of $s$

directly, the update function can be written as:

$$\theta^{Acp}_{i,t+1} = \theta^{Acp}_{i,t} + \alpha(ap_t - Acp_t(s_t))s_{i,t}$$

$$\theta^{Acs}_{i,t+1} = \theta^{Acs}_{i,t} + \alpha(as_t - Acs_t(s_t))s_{i,t}$$

## 2.5 Reward and Performance Evaluation

After every trading session, the traders' behavior in the last session can be evaluated. In general, if the last order bring the trader closer to its target position or reduces the overall cost for its inventory, the order is a good play. It is also important to note that traders may sacrifice short-term profit for benefits in the long run, in which case they will submit 'bad' orders intentionally.

To properly measure the merit of each trade decision, a scoring system is developed. The scoring system provides an overview of the distance a player is to its goal. The incremental change to the score provides information on the merit of the latest order, and is used as the *reward* for updating approximation functions.

For type A traders, the level of target achievement is measured by an exponential function to set a boundary to the performance score:

$$S1_t = \exp(-|\frac{Ps_t - PT}{PT}|) - \exp(-1)$$

$$\delta_{S1} = S1_t - S1_{t-1}$$

, where $S1_t$ denotes the score for the level of target position achievement at time t, $Ps_t$ denotes the current stock position, and $PT$ denotes the target share of the agent.

The cost score is calculated by:

$$S2_t = \begin{cases} \exp(-pa_t) & : pa_t > 0 \\ (1 - \exp(pa_t)) & : pa_t < 0 \end{cases}$$

$$pa_t = \frac{Pc_t}{Ps_t}$$

, where $S2_t$ denotes the score for the average cost at time t, $pa_t$ denotes the average price estimated by the account information, $Pc_t$ denotes the amount of cash in the current account, and $Ps_t$ denotes the current number of shares in the account.

The total performance score for type A traders is:

$$performance_A = S1_t + c_1 * S2_t$$

, where $c_1$ is the constant that weighs the scores from different sources.

For type B traders, the performance function measures how much cash remains after the asset position is liquidated at a large discount:

$$performance_B = Pc_t - c_2 * Ps_t * p_t$$

, where $c_2$ is the discount rate to liquidate the position of type B traders, and $p_t$ is the current market price for the asset.

# Chapter 3

# Results

## 3.1 Behavior of Traders

The traders trained via reinforcement learning demonstrate strong capabilities in recognizing their assigned trading targets and in interacting with the market to achieve their goals. All the controlling coefficients of agents are first initialized as 0. Agents are supposed to accumulate market knowledge and build their own action policies in the training rounds. After completing the training rounds, formal simulation then begins and market data are recorded.

The following section uses a sample simulation to see if the agents can learn about their goals and build correct action plans after training. The sample simulation has the following settings:

- Total trading session = 10
- Number of Buyers (Type A Trader) = 20
- Target of Buyers (shares) = 200
- Number of Sellers (Type A Trader) = 20
- Target of Seller (shares) = −200
- Number of Market Maker (Type B Trader) = 100

- Number of Training rounds = 50
- Number of Testing rounds = 10

For type A traders, the ending share positions after the first training round and the 50th training round are illustrated in the following figure.

Figure 3-1: Ending Shares of Type A Traders after 1st Training

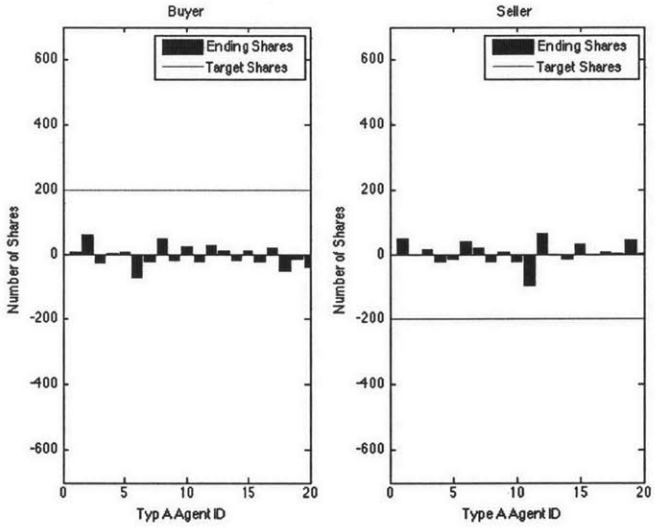
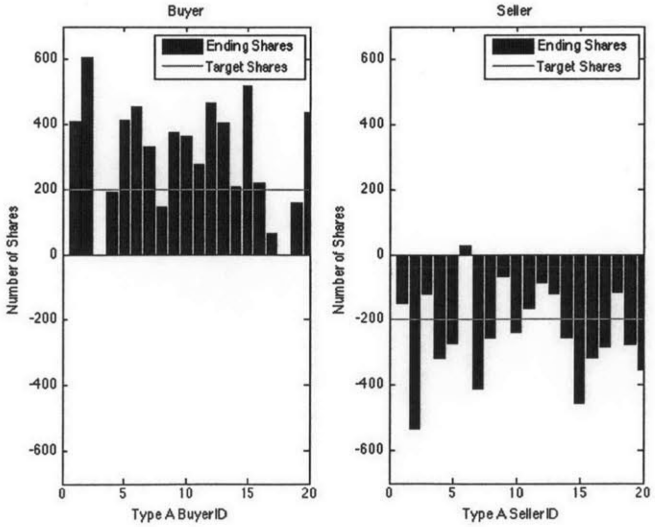


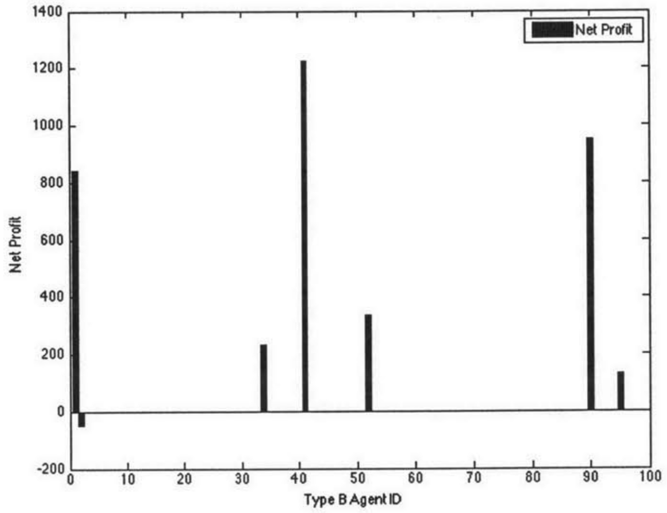
Figure 3-2: Ending Shares of Type A Traders after 50th Training



The graph shows that in the early stages of the training, the actions of all type A traders are random. Almost all the type A traders are able to purchase/sell shares to achieve their goals after the 50 rounds of training. Moreover, the number of ending shares for each agent is very close to their assigned values. This success holds true on other market settings.

The type B traders, who are known as market makers, also behave in a desirable way. The following figure shows the net cash position of market makers after liquidating all shares at a large price discount. It is noteworthy that the market making skills are difficult to pick up. Among the 100 type B traders, only 6 of them can effectively create a large cash position while keeping the inventory level low.

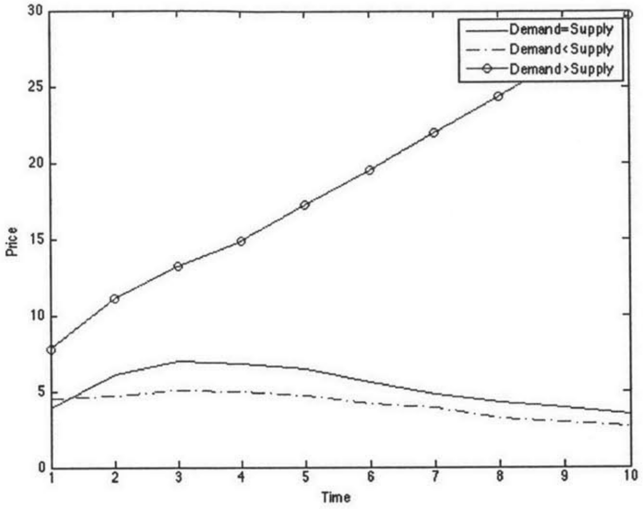
Figure 3-3: Net Profit of Type B Traders



At an aggregate level, the market reflects the demand and supply balances very well. The following figure shows the price path in three scenarios in which supply and demand pairs are (1) target of buyer = 200 & target of seller = -200 (2) target of buyer = 500 & target of seller = -200, and (3) target of buyer = 200 & target of seller = -500.

It should be noted that the reward function is critical in determining the behavior of the agents. One episode during the development of the simulation was that a cost score, $S2$, which measures the cost by the average price, was used for market makers. This generates markets in which prices always go down. It was later found that the

Figure 3-4: Price Paths under Different Scenarios



market makers were smart enough to recognize that by pushing the prices low they can always achieve a higher performance score as their average inventory cost per share falls. After redesigning the cost function, the cost score for market makers now measures the total profit after liquidating all shares. This change corrects the behavior of market makers.

## 3.2 Analysis of High Frequency and Low Frequency Market

Having constructed the traders, the market simulation can now be used as a tool to analyze how price paths are different under varied market structures. The question concerning how the trading frequency in a market impacts the price is explored as an application of the proposed adaptive market simulation method.

The simulation assumes that the intrinsic value of the traded asset remains constant as no new information arrives during the simulation. Given this assumption, every simulation corresponds to a short period of time and the number of trading sessions in each simulation measures how frequently traders can make trades.

In the simulation, markets with two different numbers of the total sessions are

Figure 3-5: Price Paths in Low Frequency Setting

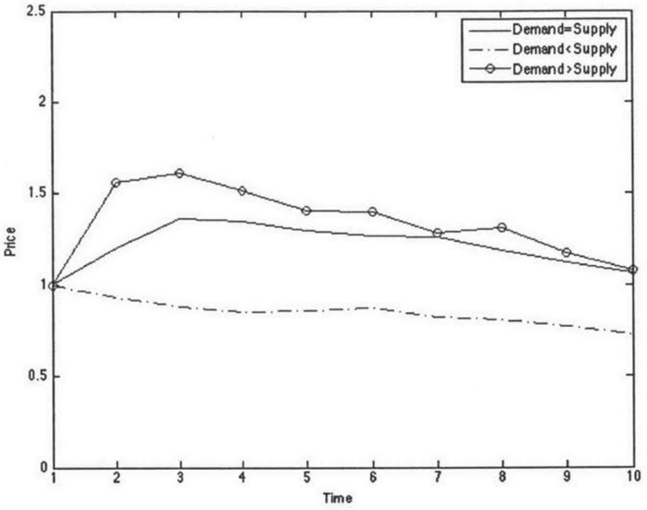
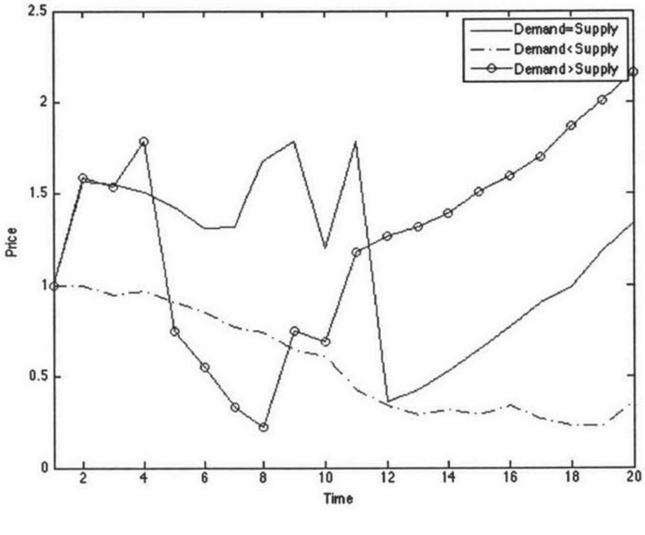


Figure 3-6: Price Paths in High Frequency Setting



constructed to transact the same amount net demand and supply between type A traders. The market sessions are $N_1 = 10$ and $N_2 = 20$ respectively. The market has 20 buyers, 20 sellers and 100 market makers. Three scenarios are simulated to mimic the situations where demand=supply (target of buyer = 200, target of seller = -200), demand>supply (target of buyer = 500, target of seller = -200), and demand<supply (target of buyer = 200, target of seller = -500) respectively. Samples of the simulated price path are shown in the following figures. Each price path is divided by the

corresponding first-day trading price for comparing purpose.

The simulation suggests that the standard deviation of the returns in the high frequency market with 20 trading sessions is higher than that in the low frequency market. It also reveals that the price paths in high frequency settings are more volatile compared to low frequency setting. One possible reason is that as the number of trading sessions increases, order sizes from Type A traders in each session shrink, and the market is more influenced by the orders from market maker. It can also be found that the high frequency market has a better ability to reflect the information between supply and demand. Prices are then able to adjust quickly when mispricings occur.

When judging whether the results from the simulation suggest that one market is superior to another, one should not draw conclusions based solely on volatility in the price paths. Prices in low frequency markets may be more stable, but high frequency markets provide a far richer set of price and order data, which is valuable information. Further analysis is required of the existing tradeoff between the value of the excess information and market stability in order to conclusively decide which feature is more desirable for a market.

# Chapter 4

# Further Research and Application

## 4.1 Improvements to the Current Approach

Although agents trained via reinforcement learning demonstrate a certain level of intelligence, their performance is not yet optimal. One way of improving them is to add complexity via function approximation. The current linear function may fail to capture critical non-linear information, thereby yielding output actions that are not entirely right. Kernels and neural network may be used in function approximation to deal with higher order relationships among observations in states, value of state and corresponding actions. These methods may generate agents with better performance, and should be implemented in further research.

The neural network can also be used independently as an alternative method for constructing agents. The challenge for constructing traders in neural network is to find the optimal coefficients in the network through learning algorithm. Neural network is usually used as a supervised learning. However, in the agent case, there is no optimal output action that pairs with input observations.

Genetic algorithm may be used to solve the optimization problem. This algorithm makes it possible to only evaluate the agents one time at the end of each round of the simulation (not every session). A fitness function will examine the overall performance of each trader. Whoever has better performance will have a higher chance to produce the next generation. The production procedure can include mutation and cross over, which imitate the biological processes that happen in breeding offspring. However, one possible outcome is that the genetic algorithm cannot improve the overall performance of the traders after many generations, which indicates a failure of this method.

Another way to optimize the coefficients in the neural network is to convert the simulation to a supervised learning problem by designing pseudo objective values. Since the objective of an agent is known, it is possible to reverse engineer a policy plan based on the price/volume path. However, it is difficult to tell what kind of order is an optimal decision, as its decision highly depends on the behavior of other agents. Different methods of constructing pseudo values may be tested. However, pseudo objective method can lead to over fitting problems and no trade happens in the following trading session. It also assumes that the revised action of the agent will not affect behaviors of the population, which is incorrect. To mitigate this issue, a lower learning rate for updating the coefficients of an agent can be set so that the expected output would not directly converge to pseudo values. Overall, neural network is a promising approach if the above questions can be solved.

## 4.2 Potential Applications

The trading simulation provides researchers with an opportunity to conduct interactive experiments. Researchers will be able to control desired variables and obtain information from the simulated market. Currently, quantitative research is mostly conducted with historical aggregated datasets from which it is difficult to test complex ideas requiring consideration of other traders' behaviors. The simulation makes it possible to analyze markets through different perspectives, test varied hypothesis, and understand market behavior in a more thorough manner.

The simulation may be applied to estimate trading costs and price impacts. Under the current convention, traders believe that a target position higher than a certain percentage of daily trading volume will inevitably have a price impact. However, there is no agreement on the maximum size that one can trade without affecting the prices. The simulation may provide a way of estimating the price impact associated with different trading volumes. Going forward, products that transfer risk in executing trades may be priced using this approach.

The interactive simulation system can also be used as a viable platform for training traders. Individuals can utilize the platform to learn trading, interact with agents, and build up their own market acumen. The adaptive nature of the system makes the simulation more realistic, meaning that a currently great strategy may not persist going forward once the market catches on to it. Moreover, trainers can simulate varied markets under different scenarios, be it markets with different demand and supply, or markets in which they have dominant or trivial market power. This will encourage users to focus on the behavior of opponents and keep generating new strategies.

# Chapter 5

# Conclusion

Modeling a complex system via simulation provides us with a useful tool to understand the behavior of its components, as well as many properties of the system itself. Building reliable and smart agents is challenging but also fruitful as it makes the simulated system more adaptive and realistic. The process of constructing intelligent traders sheds much light on the challenge of mimicking human behavior computationally. However, nascent developments in areas such as reinforcement learning, neural network, and genetic algorithm put us in good stead to tackle this challenge in the future.

Among the many issues that can be analyzed through the proposed tool, the result from the simulation suggests that when high frequency trading is available, market can respond to supply and demand disequilibrium and correct mispricings quickly. The exchange is capable of generating more information to reflect current market conditions. On this note, high frequency trading can be viewed positively as improving the process of price discovery. Meanwhile, faster markets exhibit higher volatility, as the trading activity at each time point is less stable. Therefore, trade-off may need to be made in order to design a well-functioning market.

# Chapter 6

# Reference

Garber, P. M. (1989). Who put the mania in tulipmania?. The Journal of Portfolio Management, 16(1), 53-60.

Case, K. E., Shiller, R. J. (2003). Is there a bubble in the housing market?.Brookings Papers on Economic Activity, 2003(2), 299-362.

Lo, A. W. (2002). Bubble, rubble, finance in trouble?. The Journal of Psychology and Financial Markets, 3(2), 76-86.

Brogaard, J. (2010). High frequency trading and its impact on market quality.Northwestern University Kellogg School of Management Working Paper.

AÃŕt-Sahalia, Y., Saglam, M. (2013). High frequency traders: Taking advantage of speed (No. w19531). National Bureau of Economic Research.

McInish, T. H., Upson, J. (2013). The quote exception rule: Giving high frequency traders an unintended advantage. Financial Management, 42(3), 481-501.

Zhang, F. (2010). High-frequency trading, stock volatility, and price discovery.Available at SSRN 1691679.

Kirilenko, A., Kyle, A. S., Samadi, M., Tuzun, T. (2011). The flash crash: The impact of high frequency trading on an electronic market. Manuscript, U of Maryland.

Eurex, 2011, High-frequency trading in volatile markets –an examination

Easley, D., Lopez de Prado, M., O'Hara, M. (2012). The volume clock: Insights into the high frequency paradigm. The Journal of Portfolio Management,(Fall, 2012) Forthcoming.

Howard, R. A. (1970). Dynamic programming and Markov processes.

J. Connell, S. Mahadevan, 1993, Robot Learning, Kluwer Academic, Dordrecht, pp. 141–170

Barto, A., Crites, R. H. (1996). Improving elevator performance using reinforcement learning. Advances in neural information processing systems, 8, 1017-1023.

Dempster, M. A. H., Payne, T. W., Romahi, Y., Thompson, G. W. (2001). Computational learning techniques for intraday FX trading using popular technical indicators. Neural Networks, IEEE Transactions on, 12(4), 744-754.

Barto, A. G. (1998). Reinforcement learning: An introduction. MIT press.

Doya, K. (2000). Reinforcement learning in continuous time and space. Neural computation, 12(1), 219-245.

Van Hasselt, H., Wiering, M. A. (2007, April). Reinforcement learning in continuous action spaces. In Approximate Dynamic Programming and Reinforcement Learning, 2007. ADPRL 2007. IEEE International Symposium on(pp. 272-279). IEEE.

Baird, L. C., Klopf, A. H. (1993). Reinforcement learning with high-dimensional, continuous actions. US Air Force Technical Report WL-TR-93-1147, Wright Laboratory, Wright-Patterson Air Force Base, OH.

Prokhorov, D. V., Wunsch, D. C. (1997). Adaptive critic designs. Neural Networks, IEEE Transactions on, 8(5), 997-1007.

Littman, M. L. (1994, July). Markov games as a framework for multi-agent reinforcement learning. In ICML (Vol. 94, pp. 157-163).

Holland, J. H. (2006). Studying complex adaptive systems. Journal of Systems Science and Complexity, 19(1), 1-8.

Mauboussin, M. J. (2002). Revisiting market efficiency: the stock market as a complex adaptive system. Journal of Applied Corporate Finance, 14(4), 47-55.

Arthur, W. B. (1996). Asset pricing under endogenous expectations in an artificial stock market (Doctoral dissertation, Brunel University, London).