Journal of Organizational and End User Computing
Volume 36 • Issue 1

Deep Reinforcement Learning for Adaptive Stock Trading:

Tackling Inconsistent Information and Dynamic Decision Environments

Lei Zhao, Southwestern University of Finance and Economics, China
Bowen Deng, Southwestern University of Finance and Economics, China
Liang Wu, Beijing Normal University, China*
https://orcid.org/0000-0002-3606-356X
Chang Liu, Southwestern University of Finance and Economics, China
Min Guo, China Great-Wall Asset Management Co., Ltd., China
Youjia Guo, Sichuan University, China

ABSTRACT

In this study, the authors explore how financial institutions make decisions about stock trading strategies in a rapidly changing and complex environment. These decisions are made with limited, often inconsistent information and depend on the current and future strategies of both the institution itself and its competitors. They develop a dynamic game model that factors in this imperfect information and the evolving nature of decision-making. To model reward transitions, they utilize a combination of t-Copula simulation of a non-stationary Markov chain, probabilistic fuzzy regression, and chaos optimization algorithms. They then apply deep q-network, a method from deep reinforcement learning, to ensure the effectiveness of the chosen strategy during ongoing decision-making. The approach is significant for both researchers across fields and practical professionals in the finance industry.

KEYWORDS

Adaptive Stock Trading, Dynamic Decision, Inconsistent Information, Probabilistic Preference Modeling

1. INTRODUCTION

Decision making is the process of making decision based on imperfect information about environment and opponents. The importance of decision making in many fields makes it receive much attention from scientist. Making decisions with imperfect information under uncertainty in inconsistent and dynamic environments is particularly true for stock institutions in the stock market. What makes things even more complicated is the situation in which the participants in the game (competition)

DOI: 10.4018/JOEUC.335083
*Corresponding Author

need to make their strategic business decisions in a conflicting or cooperative way (A.R. Heidari, 2010; Sun et al,2022a). A cooperative competition strategy evaluates the strategy effects not only for itself, but also for its opponent. On the other hand, a conflicting strategy maximizes the reward only for itself (J.-Y. Kim, J.Y. Kwon, 2017). As the simulation results in this study demonstrate, for financial institutions, in some scenarios, they should adopt the conflicting competition strategy, and in others, they should adopt the cooperative competition strategy. When decision making in a real environment becomes complex, the optimum equilibrium for games would not be easily achieved without the aid of intelligent computational algorithms.

In order to be able to deal with inconsistent information and dynamic decision-making environment in an isolated environment without interference from other conditions, this study chose a small-scale listed company. According to the annual report of this listed company, in a long period of time There are only two institutions trading the company's shares. In this study, a dynamic and imperfect information game model and algorithm are built to address the inconsistency of information and the dynamic nature of the decision-making process toward maximizing rewards under two scenarios, conflicting (*algorithm 1*) and cooperative (*algorithm 2*). Next, in order to model the risk transition probabilities, which would be used in the Markov Decision Process of Reinforced Learning, Probabilistic Fuzzy Regression (PFR), Chaos Optimization Algorithm (COA) and t-Copula simulation of a Non-Stationary Markov Chain model (algorithm 3) would be implemented to study the external risk factors' effect on the transition probabilities. Finally, the Deep Q-Network (DQN) of Reinforced Learning would be used to estimate the optimum actions (strategies) derived from the progressive game decision process, under the two assumptions, conflicting or cooperative. Each institution has three opportunities to adjust its stocks orders placement strategies (*Algorithm 3*) or each has unlimited opportunities (*Algorithm 4*). The former focuses on the analysis of the impact of competition type and reaction speed, while the latter focuses on the comparison of compensation under different competition setups. In addition, in order to study the long-term dynamic decision-making of participants in inconsistent stochastic complex environment, the optimal order placement strategies and the optimal competition strategies would be estimated under the infinite adjustment scenarios. Therefore, the method proposed in this paper is of great significance not only to interdisciplinary research, but also to the practitioners as well.

The approach proposed in this study would solve several obstacles common in dynamical decision-making with inconsistent information under a stochastic decision outcome (T. Zu, M. Wen, R. Kang, 2017; Sun et al., 2021)). First, in real business competition, the information is seldom complete. In most circumstances, the quality of the information is insufficient at best. Participants in the game often release false information (hence inconsistent) about their strategies on purpose in order to strategically mislead their opponents (Sun et al., 2022b; Sun et al.,2022c). As a result, the imperfect dynamical information game model with different competition scenarios (conflicting or cooperative) is proposed to isolate the untrustworthy information and abstract the true behavior pattern from the actual events. Since this is the audited key risk indicator, as well as the well-known performance measurement, even the lagged figure of reward could be quite useful in collecting information. On the other hand, by comparing the simulation results under conflicting scenarios with those under cooperative scenarios, the effect of misleading information could be kept at a minimum. Detailed descriptions and reasoning of the game model and competition scenarios are found in section 2.

The study contributes the existing literatures in three ways: First, the dynamic game designed for conflicting and cooperative competition scenarios is used to overcome problems arising from inconsistent information and a dynamic decision environment. Second, the Deep Reinforcement Learning algorithm is aimed to deal with the complexity and irrationality in decision making by simultaneously evaluating future reward trajectories of each potential strategy for all game participants with intertwined interaction along a variety of factors considered. Finally, in order to keep the uncertain decision results to a minimum, this study implements PFR with COA to confine

the probability distributions of reward states, and *t*-Copula simulation with a NMC model to track the transition probabilities.

The rest of the paper is organized as follows. In Section 2, the existing studies in the literature are summarized. In Section 3, some relevant mathematical models, including the PFR model, and algorithms, including COA and t-Copula simulations, are summarized. In Section 4, numerical examples are used to determine empirical results. The paper concludes in Section 5 with some conclusions.

2. LITERATURE REVIEW

The use of fintech tools such as artificial intelligence is becoming increasingly widespread in the financial sector. Zhao et al., (2022) focus on deep learning digital economy scale measure based on a big data cloud platform and its application. They find that big data cloud platforms would improve the ability of digital media and digital transactions. Dai (2022) improves the existing CNN and applies it to financial credit from a different perspective. The study by Srivastava et al. (2021) deploys machine learning algorithms such as support vector machines, random forests, gradient boosting, and deep neural networks to predict stock movements, demonstrating the scope of future applications of deep learning in multi-parameter time series forecasting.

In the available studies, various approaches for developing empirical models have been attempted. These approaches include Probabilistic Fuzzy Regression (PFR), Chaos Optimization Algorithm (COA), and Intelligent Decision-Making Technology (IDT). For PFR, de Hierro et al., (2016) provides fuzzy regression models using fuzzy distances. Su et al., (2013) introduced kernel-based nonlinear fuzzy regression. Otadi (2014) introduced fuzzy polynomial regression based on fuzzy neural networks. Li et al., (2016) introduced fuzzy regression models based on the least absolute deviation. Lau et al., (2006) introduced the fuzzy logic approach. Han et al., (2000) introduced multiple statistical regressions. Sekkeli et al., (2010) introduced fuzzy linear regression. Feng & Cheng., (2022) introduced an Artificial Neural Network Analysis on Retail E-Commerce Service Quality Evaluation. Zhao(2022) introduced Deep Learning Algorithm to deal risk prediction for internet financial enterprises. Sun et al., (2023) introduced integrating a multifactor model to build a medium-term investment strategy. Chang et al.,(2022) based on data mining combination to predict stock price.

For IDT, Morsalin et al., (2016) studied IDT based on an artificial neural network to determine electric vehicle charge scheduling. Pombeiro et al., (2017) applied the genetic algorithm to IDT to control an HVAC (Heating, Ventilation, and Air Conditioning) system. Weng et al., (2017) introduced an expert system to IDT to build an intelligent decision-making tool assisting investors in making trading decisions. Salih et al., (2015) introduced random Forest to IDT to provide support for a real-time health care monitoring system. A few applied cluster analyses to intelligent decision support systems (IDSS) for medical radioisotope diagnostics Dovbysh et al., (2015) and Sen (2015) introduced probabilistic reasoning to IDT to design an intelligent framework for multiple robots and human coalition formation. Zhao and Wei (2017) studied IDT based on the Bayesian decision to develop a novel algorithm of human-like motion planning for robotic arms. Weidong and Shiping (2009) studied IDT based on D/S evidence theory and its application in science research project selection. IDT based on uncertainty-decision was initially proposed at the IFIP TC 8 WG 8.9 International Conference Xu et al., (2008). Among others, Tan et al., (2017) applied fuzzy set theory to intelligent decision-making technology. Vagin et al., (2015) studied intelligent decision-making based on rough sets theory. Hou et al., (2022) studied Deep Learning to Rural Financial Development and Rural Governance. Du & Shu (2022) focused on using Deep Learning and Bionic Algorithm to explore of financial market credit scoring and risk management. The results of in-depth research in this area include quality of big data marketing analytics (Haverila et al., 2022); financial risk early warning model (Li et al., 2022); intelligent employee retention system(Srivastava et al.,2021); analyze the market risks of A+H shares by BPNN algorithm(Wu et al.,2022).

In general, the existing literature uses artificial neural networks for intelligent decision making in many places but rarely in games between investment institutions in uncertain environments. We improve on the studies of Weng et al., (2017), Srivastava et al., (2021) by first specifying two scenarios, i.e., competition and cooperation, in which all games are played and institutional investors either choose to compete to maximize their own interests or to cooperate to achieve mutual The institutional investors either choose to compete to maximize their own interests or choose to cooperate to maximize the sum of their interests. Secondly, this study develops the study of Tan et al. (2017) to deal with the complexity and irrationality of decision making by simultaneously evaluating the future reward trajectories of each potential strategy for all game participants. Finally, in order to keep the outcome of the decision within an acceptable range, this study integrates models such as COA and NMC to set a reasonable probability distribution of reward states. Wu, Z et al.,(2022) based on convolutional neural network to predict investment portfolio. Yang & Wu (2022) utilizing artificial intelligence in financial risk management.

3. MATHEMATICAL MODELS AND ALGORITHMS

Deep Reinforcement Learning algorithms are particularly suitable for solving decision-making problems in dynamic games with incomplete information. First, the dynamic decision-making process is assumed to be captured by the reinforcement learning Q network algorithm and the profit at each moment is related to the decision-making action at that moment. In other words, the DQN provides a set of action $A_t \;\; t = 1,2...T$, the optimal returns are obtained at each decision point in the set $R_{A_t} = \sup_{i=1,2...m} \left(R^i_{A_t} \right), t = 1,2...T$. This process is a close mimic of a decision deduction process of an actual human being. The uncertain and stochastic aspect of the decision context is modeled by Probabilistic Fuzzy Regression (PFR) and Non-stationary Markov Chains (NMC) with *t*-Copula simulation, in which the transition probabilities at time *t* are explicit functions of the following factors: The lagged prices of stock trading orders of the decision-maker, the lagged prices of stock trading orders of the opponent, the lagged reward of the decision-maker, and the lagged reward of the opponent. On the other hand, the transition probabilities over pre-defined states for each of the factors can be modelled based on the Probabilistic Fuzzy Regression (PFR) and Chaos Optimization Algorithm (COA), which is helpful for the next simulation. Finally, the inconsistent environment in which the game participants have to make decision is presented in an imperfect information dynamic game model in order to capture the full complexity of competition in the real world.

As a result, section 2.1 presents the detailed idea for the Imperfect Information Dynamic Model, followed by the presentation of Probabilistic Fuzzy Regression (PFR) and Chaos Optimization Algorithm (COA) in section 2.2. In section 2.3 the idea of t-Copula Simulation of Non-stationary Markov Chain (NMC) is presented. Finally, section 2.4 introduces the Markov Decision Process and Reinforced Learning DQN algorithm.

3.1 The Imperfect Information Dynamic Game Model

In order to establish decision-making processes under different decision-making scenarios, institution actions and decisions are divided into two categories: Cooperation and Conflicting. Specifically, decision-making under conflicting scenarios will choose actions to maximize only their own benefits, while decision-making under cooperative scenarios will choose actions to maximize the benefits for all institutional decision-making participants. Defining $R^i_t \left(P^i_t, n^i_t \right) i = 1,2...m \;\; t = 1,2...T$ as the reward of institution *i* at time (round) *t*, reward of institutions, given as $R^i_t = P^i_t \times n^i_t$, where P^i_t is the profit for order placed at time *t* and n^i_t is the numbers of the stocks placed at time *t* for institution *i*. Meanwhile, institutional decision-makers are expected to smoothly reduce the possibility of

withdrawal of orders. If they decide to buy more than the supply of the market or sell more than the market demands, the transaction will probably not go smoothly. Therefore, the orders are placed at the market, which means the buy orders placed at time *t* are equal to the sell orders listed at time *t-1* and the sell orders placed at time *t* are equal to the buy orders listed at time *t-1*. This assumption could not only simplify the gaming setup, but also is parallel to the stealth trading strategies mentioned in Liu, et al (2019), in which the institutions limit their orders only the listed orders amounts in order not to send extra information about their attentions. It can be seen as related to the price information obtained by all institutional decision-makers, so the quantity of buy or sell of the orders itself is a function of the current price of the order, the lagged price of the counterparty and the current asymmetric price information of the order. The price orders of the institution's optimal order places can be written in the following forms:

$$P_t^{i*} = \arg\max_{P_t^i} \left\{ R_t^i \left(P_t^i, n_t^i \left(P_t^i, P_{t-1}^{-i}, P_t^{-i'} \right) \right) \right\} \tag{1}$$

Since the supreme for R_t^i is only reached for its own maximized reward, the R_t^i given in Eq.(1) is under a conflicting scenario. On the other hand, under a cooperative scenario, the prices of the orders for institution *I* is given as:

$$P_t^{i*} = \arg\max_{P_t^i} \left\{ R_t^1, R_t^2 \ldots R_t^i \left(P_t^i, n_t^i \left(P_t^i, P_{t-1}^{-i}, P_t^{-i\prime} \right) \right) \ldots R_t^m \right\} \tag{2}$$

in which the decision for adjusting its price needs to consider the optimum rewards both for itself and for the opponent as well.

For simplicity, we assume only two institutions, institution 1 and institution 2, and the strategy for maximizing their rewards are prices they placed. Defining $P_t^i, P_{t-1}^i \;\; i = 1,2$ as the prices for institution 1 and institution 2 at time(or round) *t* and *t-1* respectively, and $P_t^{1'}, P_t^{2'}$ as the imperfect information about the prices for institution 1 and institution 2 at time(or round) *t* and *t-1* respectively, then the optimum prices at time t for institution *i*, *i=1,2*, is given by:

$$P_t^{i*} = \begin{cases} \arg\max_{P_t^i} \left\{ R_t^i \left(P_t^i, n_t^i \right) \right\}, n_t^i = \begin{cases} n_t^1 \left(P_t^1, P_{t-1}^2, P_t^{2'} \right), i = 1 \\ n_t^2 \left(P_t^2, P_{t-1}^1, P_t^{1'} \right), i = 2 \end{cases}, Conflicting \\ \arg\max_{P_t^i} \left\{ R_t^1 \left(P_t^1, n_t^1 \right), R_t^2 \left(P_t^2, n_t^2 \right) \right\}, n_t^i = \begin{cases} n_t^1 \left(P_t^1, P_{t-1}^2, P_t^{2'} \right), i = 1 \\ n_t^2 \left(P_t^2, P_{t-1}^1, P_t^{1'} \right), i = 2 \end{cases}, Cooperative \end{cases} \tag{3}$$

which shows that the reward for institution is a function of its provided price, the opponent's lagged price, and the imperfect speculation of the opponents provided price at time *t*, which is confidential information. Let c_t^{i*}, c_t^i be fixed and variable cost, then the institution's total cost is given as $C_t^i = c_t^{i*} + c_t^i \times n_t^i, \;\; i = 1,2$. The following algorithms give the equilibrium rewards for all decision points *t=1,2…T* for conflicting and cooperative scenarios, respectively.

The main difference between the two algorithm is that the algorithm 2 estimates and records the theoretical optimum actions and rewards for institution 2 at Step1, but the opponent might not adopt the actions since this is a speculative action from the viewpoint of the first institution. The actual optimum actions that the opponent institution will take and therefore the true best rewards that the institution could get are the ones estimated by Step 3 of algorithm 2 from its own angle such as P_t^{2*} and R_t^2 for institution 2.

Algorithm 1. Equilibrium prices under the conflicting scenario

Input: $P_t^1, P_t^2, P_{t-1}^1, P_{t-1}^2, P_t^{1\prime}, P_t^{2\prime}, c_t^*, c_t$

Output: Equilibrium prices (P_t^{1*}, P_t^{2*})

For t=1,2…T

Step 1: institution 2 takes the first action, then

$$\begin{aligned} Max: \quad & R_t^2 = R_t^2(P_t^2, n_t^2) \\ S.t. \quad & n_t^2 = n_t^2(P_t^2, P_{t-1}^1, P_t^{1\prime}) \\ & C_t^2 = c_t^* - c_t \times n_t^2 \\ & \frac{dR_t^2}{dP_t^2} = 0 \end{aligned}$$

Output: Optimal strategy for institution 2 $P_t^{2*} = P_t^2(P_{t-1}^1, P_t^{1\prime})$

Step 2: institution 2 takes the counter-measure action

$$\begin{aligned} Max: \quad & R_t^1 = R_t^1(P_t^1, n_t^1) \\ S.t. \quad & n_t^1 = n_t^1(P_t^1, P_{t-1}^2, P_t^{2*}) \\ & C_t^1 = c_t^* - c_t \times n_t^1 \\ & \frac{dR_t^1}{dP_t^1} = 0 \end{aligned}$$

Output: Optimal strategy for institution 1 $P_t^{1*} = P_t^1(P_{t-1}^2, P_t^{2*})$

Step 3: Record the optimal strategies and rewards for both institutions at time t

$$P_t^* = (P_t^{1*}, P_t^{2*})$$
$$R_t^* = (R_t^1, R_t^2)$$

Stop

3.2 Probabilistic Fuzzy Regression (PFR) and Chaos Optimization Algorithm (COA)

In the current study, probabilistic fuzzy regression (PFR) for preference modeling, is proposed. PFR can address the fuzziness caused by human subjective judgment and the randomness caused by random variables. Probability density functions (PDFs) are adopted in the proposed approach to model the randomness of independent (random) variables. A chaos optimization algorithm (COA) is employed to determine the parameter settings of the PDFs, and PDFs are then generated.

The general form of a fuzzy liner regression model can be expressed as follows:

Algorithm 2. Equilibrium prices under the cooperative scenario

Input: $P_t^1, P_t^2, P_{t-1}^1, P_{t-1}^2, P_t^{1\prime}, p_t^{2\prime}, c_t^*, c_t$

Output: Equilibrium prices $(P_t^1{*}, P_t^2{*})$

For t=1,2…T

Step 1: Institution 2 takes the first action

$$Max:\ P_t^2 = arg\max_{P_t^2}\{R_t^1(P_t^1, n_t^1), R_t^2(P_t^2, n_t^2)\}$$

$$S.t.\quad n_t^1 = n_t^1(P_t^1, P_{t-1}^2, P_t^{2\prime})$$
$$n_t^2 = n_t^2(P_t^2, P_{t-1}^1, P_t^{1\prime})$$
$$C_t^i = c_t^* - c_t \times n_t^i,\ i = 1, 2$$
$$\frac{dR_t^2}{dP_t^2} = 0$$

Output: Optimal strategy and rewards $(P_t^2{*},\ P_{t-1}^1{*}), (R_t^2{*}, R_t^1{*})$

Step 2: Institution 1 takes the counter-measure action

$$Max:\ P_t^1 = arg\max_{P_t^1}\{R_t^1(P_t^1, n_t^1), R_t^2(P_t^2, n_t^2)\}$$

$$S.t.\quad n_t^1 = n_t^1(P_t^1, P_{t-1}^2, P_t^{2\prime})$$
$$n_t^2 = n_t^2(P_t^2, P_{t-1}^1, P_t^{1\prime})$$
$$C_t^i = c_t^* - c_t \times n_t^i,\ i = 1, 2$$
$$\frac{dR_t^1}{dP_t^1} = 0$$

Output: Optimal strategy and rewards $(P_t^1{*},\ P_{t-1}^2{*}), (R_t^1{*}, R_t^2{*})$

Step 3: Record the optimal strategies and rewards for both institutions at time t

$$P_t^* = (P_t^1{*}, P_t^2{*})$$
$$R_t^* = (R_t^1, R_t^2)$$

Stop

$$\widetilde{Y}_i = \widetilde{A}_0 + \widetilde{A}_1 x_{i1} + \ldots + \widetilde{A}_k x_{ik} = \widetilde{A} X_i,\ i = 1, 2, \ldots, n \tag{4}$$

where $\widetilde{Y}_i$ is the predicted output, which is a fuzzy number; n is the number of data sets; $x_{ij}, j = 1, 2, \ldots, k$ is the *j*-th independent variable of the *i*-th data set; *k* is the number of independent variable; and $\widetilde{A}_j, j = 1, 2, \ldots, k$ is the coefficient of the *j*-th independent variable. $\widetilde{A}_j = \left(s_j^L, a_j^c, s_j^R\right),\ j = 1, 2, \ldots, k$ where s_j^L, s_j^c, s_j^R are the central value, left-side and right-side spreads of the fuzzy coefficients,

respectively. The predicted output of Eq. (4) can be presented as $\widetilde{Y}_i = \left(\widetilde{Y}_i^{sL}, \widetilde{Y}_i^{c}, \widetilde{Y}_i^{sR}\right)$, where $\widetilde{Y}_i^{sL}, \widetilde{Y}_i^{c}, \widetilde{Y}_i^{sR}$ are the central, left- and right-side spreads values of the output, respectively. The fuzzy regression model, Eq. (4) can be rewritten as follows:

$$\begin{aligned}\widetilde{Y}_i &= \left(\widetilde{Y}_i^{sL}, \widetilde{Y}_i^{c}, \widetilde{Y}_i^{sR}\right) \\ &= \left(s_0^L, s_0^c, s_0^R\right) + \left(s_1^L, s_1^c, s_1^R\right) x_{i1} + \ldots + \left(s_k^L, s_k^c, s_k^R\right) x_{ik},\ \ i = 1,2,\ldots,n\end{aligned} \tag{5}$$

With the type of PDFs determined, the expected value function of a random variable *X*, *E[X]*, can be generated as show in Eq. (6) to replace the corresponding random variables in the model shown in Eq. (5) and become a probabilistic term.

$$E\left[X\right] = \int_{x\min}^{x\max} xf(x)dx \tag{6}$$

Considering the random variables, the model in Eq. (5) can be rewritten as follows:

$$\begin{aligned}\widetilde{Y}_i &= \left(\widetilde{Y}_i^{sL}, \widetilde{Y}_i^{c}, \widetilde{Y}_i^{sR}\right) \\ &= \left(s_0^L, s_0^c, s_0^R\right) + \left(s_1^L, s_1^c, s_1^R\right) x_{i1}{}' + \ldots + \left(s_k^L, s_k^c, s_k^R\right) x_{ik}{}',\ \ i = 1,2,\ldots,n\end{aligned} \tag{7}$$

where $x_{ij}{}' = E\left[X\right]$ if x_{ij} is a random variable and is defined as a probabilistic term; otherwise, $x_{ij}{}' = x_{ij}, i = 1,2,\ldots,n,\ j = 1,2,\ldots,k$. For example, if five variables are involved in preference modeling and $x_1\ and\ x_4$ are random variables, the PFR model to be generated can be expressed as follows:

$$\begin{aligned}\widetilde{Y}_i &= \left(s_0^L, s_0^c, s_0^R\right) + \left(s_1^L, s_1^c, s_1^R\right) E\left(x_{i1}\right) + \left(s_2^L, s_1^c, s_2^R\right) x_{i2} \\ &+ \left(s_3^L, s_3^c, s_3^R\right) x_{i3} + \left(s_4^L, s_4^c, s_4^R\right) E\left(x_{i4}\right) + \left(s_5^L, s_5^c, s_5^R\right) x_{i5}\end{aligned} \tag{8}$$

Furthermore, fuzzy regression is employed to determine the fuzzy coefficients for each term of the PFR model. The predicted output of Eq. (7) are calculated as follows:

$$\begin{aligned}\widetilde{Y}_i^{c} &= \sum_{j=0}^{k} a_j^c x_{ij}{}' \\ \widetilde{Y}_i^{sL} &= s_0^L + \sum_{j=1, x_{ij}{}' \geq 0}^{k} s_j^L x_{ij}{}' + \sum_{j=1, x_{ij}{}' < 0}^{k} s_j^R \left(-x_{ij}{}'\right) \\ \widetilde{Y}_i^{sR} &= s_0^R + \sum_{j=1, x_{ij}{}' \geq 0}^{k} s_j^R x_{ij}{}' + \sum_{j=1, x_{ij}{}' < 0}^{k} s_j^L \left(-x_{ij}{}'\right)\end{aligned} \tag{9}$$

The asymmetric fuzzy coefficients with central point and spread values can be determined by solving the following linear programming (LP) problem (Ishibuchi and Nii, 2001; Fung et al,. 2005):

$$\begin{aligned} Min \quad & J = \sum_{j=0}^{k}\left(\left(s_j^L + s_j^R\right)\sum_{i=1}^{n}\left|x_{ij}{}'\right|\right) \\ S.t. \quad & -\sum_{j=0}^{k} a_j^c x_{ij}{}' + \left(1-h\right)\widetilde{Y}_i^{sL} \geq -\left[y_i\right]_{hL}, i = 1,2,\ldots,n \\ & \sum_{j=0}^{k} a_j^c x_{ij}{}' + \left(1-h\right)\widetilde{Y}_i^{sR} \geq \left[y_i\right]_{hR}, i = 1,2,\ldots,n \\ & s_j^L, s_j^R \geq 0, a_j^c \in \mathbb{R}, j = 0,1,2,\ldots,k \\ & x_{i0}{}' = 1 \; for \; all \; i \; and \; 0 \leq h \leq 1 \end{aligned} \tag{10}$$

where J is the objective function that represents the total width of the fuzzy outputs of the model shown in Eq. (7); $1+k$ is the number of terms of the fuzzy regression model; $x_{ij}{}'$ is the j-th term of the i-th data set in the model; $|\cdot|$ refers to the absolute value; $\left[y_i\right]_{hL}$ and $\left[y_i\right]_{hR}$ are the values of the h-level of the i-th output of the data set; and h refers to the degree to which the fuzzy model fits the given data and is located between 0 and 1.

With PFR model obtained from above, the parameter settings of PDFs can be determined by using chaos optimization algorithm (COA). The form of PDF depends on the probability distribution of a continuous random variation. The parameter settings of PDFs are determined using COA. COA is a stochastic search algorithm in which chaos is introduced into the optimization strategy to accelerate the optimum seeking operation and determine the global optimal solution (Ren and Zhong, 2011). Compare with conventional optimization methods, it has faster convergence with more accurate estimation (Nanba et al., 2002). COA employs chaotic dynamics to solve optimization problems and it has been applied successfully in various areas such as robot optimization control, function optimization and supply chain optimization (Mishra et al., 2008).

COA uses the carrier wave method to linearly map the selected chaos variables onto the space of optimization variables and then searches for the optimal solution based on the ergodicity of the chaos variables. The processes of applying COA in this study are described as follows.

First, the number of iterations of COAs is defined. The parameter settings of PDFs are represented by each chaos variable, and the number of parameters to be determined is equal to the number of elements of chaos variable. The chaos variable is initialized in which the values are selected randomly in the range [0, 1]. The range of parameters [a, b] is initialized, in which a and b are the lower and upper limits of the optimization variable, respectively.

Second, the iteration number is set as m. Based on the initialized chaos variable, the logistic model used in COA is shown in Eq. (11), and logistic mapping can generate chaos variables through iteration:

$$c_m = f\left(c_{m-1}\right) = uc_{m-1}\left(1-c_{m-1}\right) \tag{11}$$

where u is a control parameter; $c_m \in \left[0,1\right]$ is the m^{th} iteration value of the chaos variable c; and c_0 is the initialized chaos variable. While the linear mapping for converting chaos variables into optimization variables is formulated as follows:

$$q_m = a + \left(b-a\right)\cdot c_m \tag{12}$$

where q_m is the optimization variable and the value of q_m is the parameter settings of PDFs. Based on the iteration, the chaos variables traverse between [0, 1], and the corresponding optimization variables traverse in the corresponding rage *[a, b]*. In this case, the optimal solution can be identified in the area of feasible solution. The model can be developed based on *f(x)* and fuzzy coefficients by which the predicted output $\widetilde{Y}_i = \left(\widetilde{Y}_i^{sL}, \widetilde{Y}_i^{c}, \widetilde{Y}_i^{sR}\right)$ could be obtained. The predicted crisp output of $\widetilde{Y}_i$ is denoted as $\hat{y}_i$, which is equal to the center value $\widetilde{Y}_i^{c}$ if symmetric triangular member function are used in PFR. The mean absolute percentage error (MAPE) is defined as the average of percentage errors, which is scale-independent and is a popular measure for evaluating prediction accuracy (Gilliland et al., 2015). Thus, MAPE was adopted in this study as the fitness function in COA, which is defined as follow:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{y}_i - y_i\right|}{y_i}\cdot 100 \tag{13}$$

where *n* is the number of data sets; $\hat{y}_i$ is the *i*-th predicted crisp output of $\widetilde{Y}_i$ and y_i is the *i*-th actual crisp output. The values of MAPE and q_m in the first iteration are recorded as the best fitness value $fv^* = MAPE_1$ and the best solution $q^* = q_1$, respectively.

Third, the iteration continues. The chaos variable and optimization variable are updated by Eq. (11) and Eq. (12), respectively. The MAPE in the *m+1*-th iteration, $MAPE_{m+1}$, is obtained using Eq. (13). If $MAPE_{m+1} < fv^*$, then $fv^* = MAPE_{m+1}$ and $q^* = q_{m+1}$. Otherwise, fv^* and q^* remain the same.

Finally, after the number of iterations reaches the predefined number, the iteration of COA stops. fv^* is the best fitness value and the value of q^* are the determined parameter settings of PDFs.

3.3 *t*-Copula Simulation of the Non-Stationary Markov Chain

Under the condition of incomplete information, the risk management is an important factor that affects decision-making in complex environment. More specifically, we define A as the set of stocks Buy, B as the set of Overweight, C as the set of stocks Hold, D as the set of stocks Underweight, E as the set of stocks Sell, the transition matrix is given as follows:

$$P_{ih} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} & P_{AE} \\ P_{BA} & P_{BB} & P_{BC} & P_{BD} & P_{BE} \\ P_{CA} & P_{CB} & P_{CC} & P_{CD} & P_{CE} \\ P_{DA} & P_{DB} & P_{DC} & P_{DD} & P_{DE} \\ P_{EA} & P_{EB} & P_{EC} & P_{ED} & P_{EE} \end{bmatrix} \end{array} \tag{14}$$

Transition only happened $from\ i \in \{A,B,C,D,E\}\ to\ h \in \{A,B,C,D,E\}$. Under most circumstances, institutions usually use high frequency data to provide the real time risk report. Since every state of the stock is possible, there is $P_{ih}^{t} > 0\ \ i,h = A,B,C,D,E$. Because different institutions have different information about the company's stocks, different institutions have heterogeneous criteria for stock rating. Detailed one definition of states for the institution *i* is given in Table 1.

Table 1. Stock states and distribution

$\Pi \in \{A,B,C,D,E\}$ **Definition of the Stock States**			$P^t(\Pi), \Pi \in \{A,B,C,D,E\}$ **The Last Observed State Probabilities**	
A	Buy	The return of the stock trading is higher than the return of market index by more than 20% since last order placed.	$P^t(A)$	Probability of The Buy State
B	Overweight	The return of the stock trading is higher than the return of market index by 10%-20% since last order placed.	$P^t(B)$	Probability of The Overweight State
C	Hold	The return of the stock trading is higher than the return of market index by no more than 10%-20% since last order placed.	$P^t(C)$	Probability of The Hold State
D	Underweight	The return of the stock trading is lower than the return of market index by no more than 10%-20% since last order placed.	$P^t(D)$	Probability of The Underweight State
E	Sell	The return of the stock trading is lower than the return of market index by more than 20% since last order placed.	$P^t(E)$	Probability of The Sell State

$P^{t+1}(\Pi), \Pi \in \{A,B,C,D,E\}$ can be expressed as follows (Smith et al., 1996):

$$P^t(h) = \sum_{i=A}^{E} P^{t-1}(i) \cdot P_{ih}^t, \ \ h = A,B,C,D,E \tag{15}$$

where P_{ih}^{t-1} is the estimated transition probability from state *i* to state *h*, which can be calculated by simulation. As a result, the relation between the transition probabilities and external factors ($P_{t-1}^j, n_{t-1}^j, R_{t-1}^j, \ j=1,2$ in the conflicting scenario and $P_{t-1}^j, n_{t-1}^j, R_{t-1}^j, \ j=1,2$ in the cooperative scenario) can be presented as:

$$\begin{cases} P_{ih}^t = f\left(X_t\right) = f\left(\left(P_{t-1}^j, n_{t-1}^j, R_{t-1}^j, R_{t-1}^{-j}\right)_t^T\right), \ \ Cooperative \\ P_{ih}^t = f\left(X_t\right) = f\left(\left(P_{t-1}^J, n_{t-1}^j, R_{t-1}^j\right)_t^T\right), \ \ Conflicting \\ i,h = A,B,C,D,E \\ j = 1,2 \end{cases} \tag{16}$$

No particular model type would be specified in this study. Instead, the simulation based on t-Copula would be used to estimate the effects of external factors on the transition probabilities in order to overcome issues such as "Model Error". Multivariate t-Copula is given as follows (Demarta and McNeil, 2004):

$$\begin{cases} C^t\left(P_{ih}, X_t, \tau\right) = C^t_{\tau,\nu}\left(f_1\left(P^j_{t-1}\right), f_2\left(n^j_{t-1}\right), f_3\left(R^j_{t-1}\right), f_4\left(R^{-j}_{t-1}\right)\right), \ \textit{Cooperative} \\ C^t_{\tau,\nu}\left(X_t\right) = \int_{-\infty}^{f_1\left(P^j_{t-1}\right)} \int_{-\infty}^{f_2\left(n^j_{t-1}\right)} \int_{-\infty}^{f_3\left(R^j_{t-1}\right)} \int_{-\infty}^{f_4\left(R^{-j}_{t-1}\right)} \dfrac{\Gamma\left(\dfrac{\nu+N}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{\left(\pi\nu\right)^n\left|\tau\right|}}\left(1+\dfrac{X'\tau^{-1}X}{\nu}\right)^{-\frac{\nu+n}{2}} dX \\ X_t = \left(P^j_{t-1}, n^j_{t-1}, R^j_{t-1}, R^{-j}_{t-1}\right)^T_t, \ i,h = A,B,C,D,E, \ j = 1,2 \end{cases} \tag{17}$$

$$\begin{cases} C^t\left(P_{ih}, X_t, \tau\right) = C^t_{\tau,\nu}\left(f_1\left(P^j_{t-1}\right), f_2\left(n^j_{t-1}\right), f_3\left(R^j_{t-1}\right)\right), \ \textit{Conflicting} \\ C^t_{\tau,\nu}\left(X_t\right) = \int_{-\infty}^{f_1\left(P^j_{t-1}\right)} \int_{-\infty}^{f_2\left(n^j_{t-1}\right)} \int_{-\infty}^{f_3\left(R^j_{t-1}\right)} \dfrac{\Gamma\left(\dfrac{\nu+N}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\sqrt{\left(\pi\nu\right)^n\left|\tau\right|}}\left(1+\dfrac{X'\tau^{-1}X}{\nu}\right)^{-\frac{\nu+n}{2}} dX \\ X_t = \left(P^j_{t-1}, n^j_{t-1}, R^j_{t-1}\right)^T_t, \ i,h = A,B,C,D,E, \ j = 1,2 \end{cases} \tag{18}$$

where denotes the degrees of freedom for the *i*-th univariate *t*-distribution. $f\left(X_t\right)$ are the PDFs of the variables, ν is a vector of the degree of freedom for the t-distribution $\nu = \left(\nu_0, \nu_1 \ldots \nu_k\right)$, $k\left(conflicting\right) = 3, k\left(cooperative\right) = 4$, and τ is a nonparametric correlation coefficient matrix. Eq. (17) and Eq. (18) estimate the transition probabilities P_{ih} by considering the correlation between them and their corresponding external variables $P^j_{t-1}, n^j_{t-1}, R^j_{t-1}, \ j = 1,2$ in the conflicting scenario and $P^j_{t-1}, n^j_{t-1}, R^j_{t-1}, \ j = 1,2$ in the cooperative scenario.

Based on the probabilistic fuzzy regression (PFR), chaos optimization algorithm (COA), *t*-Copula simulation and models obtained above, the following algorithm gives the estimation of states probabilities at time t.

3.4 The Markov Decision Process and Reinforced Learning DQN Algorithms

Markov decision process (MDP) provides a mathematical architecture model for the problem of making decisions in a state that is partly random and partly controllable by trading institutions. Markov decision process is a quintuple $\left\{S_t, A_t, P_{S_tS_{t+1}|A_t}, \gamma, R_{t|A_t}\right\}$. The five elements are introduced as follows: S_t is the state of financial institution at time *t*. A_t is a set of available actions at time *t*. $P_{S_tS_{t+1}|A_t}$ is the transition probability from state S_t to state S_{t+1} given that the action at time *t* is A_t. $\gamma \in \left[0,1\right]$ is the discount factor, which represents the difference between future and present rewards. $R_{t|A_t}$ is the reward of financial institutions given that the action at time t is A_t.

Markov decision process is that financial institutions periodically or continuously observe the stochastic dynamic process with Markov property and make decisions sequentially. The detailed process is as follows: The original state is S_0. An action A_0 is chosen from the set of available actions A. S_0 and A_0 are substituted into the process of *t*-Copula simulation and then the transition probability $P_{S_tS_{t+1}}$ is estimated. Then, the next state S_1 is determined by the following equation:

Algorithm 3. Estimation of $P^t(i), i = A,B,C,D,E$ with t-Copula simulation

Input: $P^{t-1}(i), R_t^j, P_t^j, n_t^j,\ j=1,2\ t=1,2,...,T\ i=A,B,C,D,E$

Output: $P^t(i)\ \ i=A,B,C,D,E$

Initialize: number of iterations, types of PDFs, number of PDF parameters, *[a, b]*.

Step 1: $PFR = \widetilde{Y}_i$

Step 2:

$$\widetilde{Y}_i = \left(\widetilde{Y}_i^{sL}, \widetilde{Y}_i^{c}, \widetilde{Y}_i^{sR}\right)$$
$$= (s_0^L, s_0^c, s_0^R) + (s_1^L, s_1^c, s_1^R)x_{i1}' + ... + (s_k^L, s_k^c, s_k^R)x_{ik}',\ i=1,2,...,n$$

Step 3: $x_{ij}' = E[X]$

$$E[X] = \int_{x\min}^{x\max} x f(x)\,dx$$

Step 4:

$$Min \quad J = \sum_{j=0}^{k}\left((s_j^L + s_j^R)\sum_{i=1}^{n}|x_{ij}'|\right)$$

$$S.t. \quad -\sum_{j=0}^{k} a_j^c x_{ij}' + (1-h)\widetilde{Y}_i^{sL} \geq -[y_i]_{hL},\ i=1,2,...,n$$

$$\sum_{j=0}^{k} a_j^c x_{ij}' + (1-h)\widetilde{Y}_i^{sR} \geq [y_i]_{hR},\ i=1,2,...,n$$

$$s_j^L,\ s_j^R \geq 0,\ a_j^c \in \mathbb{R},\ j=0,1,2,...,k$$

$$x_{i0}' = 1\ for\ all\ i\ and\ 0 \leq h \leq 1$$

Step 5: $MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i}\cdot 100$

Step 6: Calculate $f_1(R_t^j), f_2(P_t^j), f_3(n_t^j)$

$$q_m = a + (b-a)\cdot c_m,\ \ fv^* = MAPE_1,\ \ q^* = q_1$$

If $MAPE_{m+1} < fv^*$

$$fv^* = MAPE_{m+1} \text{ and } q^* = q_{m+1}.$$

Else

continue

Step 7: Calculate the multivariate cumulative t-distribution function:

$$U = CDF(t, t_1, ...t_n).$$

Step 8: $P^t(h) = \sum_{i=A}^{E} P^{t-1}(i)\cdot P_{ih}^t,\ h=A,B,C,D,E$

Stop

$$\begin{cases} P_{S_1} = P_{S_0} \cdot \sum_{S_1} P_{S_0 S_1} \\ S_1 \subset \arg\max_{S_1} \left(P_{S_1} \right) \end{cases}, \; S_1 \in \left\{ s_1, \ldots, s_m \right\} \tag{19}$$

After updating the status of the financial institution at time *1* through Eq.(19), the process is transferred to the next state S_1. Then the action A_1 is taken and the state of financial institution is updated to S_2 through the same process. The sum of the rewards after the whole process is:

$$Q\left(S_t, A_t\right) = R_0\left(S_0, A_0\right) + \gamma^1 R_1\left(S_1, A_1\right) + \gamma^2 R_2\left(S_2, A_2\right) + \ldots + \gamma^t R_t\left(S_t, A_t\right) \tag{20}$$

where γ^t is the discount factor at time *t*, $t = 0,1,2,\ldots,T$. The discount factor means that early rewards are given lower weightings, while later rewards are given higher weightings.

Action is chosen to maximize rewards, which means choosing the optimal game strategy (order placing action) through greedy selection of *q*. Therefore, the method to choose the optimal action is as follows:

$$\begin{aligned} \alpha^* &= greedy\left(Q_\alpha\right), \; or \\ \alpha^* &= \arg\max_{A_t \in A} Q_\alpha\left(S_t, A_t\right), \; t = 0,1,2,\ldots,T \end{aligned} \tag{21}$$

where α^* is the optimal action at each stage and Q_α is the corresponding rewards. A is the set of available actions. Algorithms 1 and 2 are used in conflict or cooperative competition situations, respectively. The reinforcement learning algorithm based on Markov decision process could be used in general cases.

In order to show the dynamic process of decision making, the game between two financial institutions is set as the situation that two financial institutions have unlimited opportunities to change their pending order strategies. Two financial institutions in these two different circumstances determine their optimal decision to get the maximized rewards. The two financial institutions are respectively called financial institution *a* and financial institution *b*. Algorithm 4 implements Eq. (21) for its general form and present the unlimited strategy adjusting arrangements. Algorithm 5 produces the optimum strategies and rewards at each time *t* in the sense that the strategies produce not only the maximum reward for the immediate state, but also for all the states prior to time *t+1*. The entire workflow of the estimation is given in Fig. 1.

4. EMPIRICAL RESULTS

This section shows the process and outcomes of intelligent decision making under two competition scenarios (conflicting and cooperative) and antecedence setups (which institution makes the first move). The game is assumed to be played between Ningbo Ningju asset management center (NJ, institution *a*) and Shenzhen Dahe investment management co. LTD (DH, institution *b*). The two private equity firms differ in many ways, such as size, volume of assets, and business strategy. The Return on Investment (ROI) is selected as the reward variables, which refers to the economic daily return of an institution from an investment for an institution' order placement strategies. The ROI data of NJ and DH from the firms' annual reports, from June 21th 2018 to June 30th 2018, have been obtained. The data is shown in Table 2.

Algorithm 4. Financial institutions have unlimited opportunities to adjust their strategies

Input: the strategy α to be evaluated derived from Q
Output: max $Q(S,A)$, α^* under conflict or cooperative competition situations
Initialize R, A, ξ
Repeat (for each step):
If $(|areward_t - acreward_{t-1}| > \xi) \&\& (|breward_t - breward_{t-1}| > \xi))$
Initialize $Q(S) = 0$
Repeat (for each episode):
Initialize S
Choose action A from S using strategy derived from Q
Repeat (for each step of episode):
Take action A given by α for S
observe R, S^*
Choose A^* from S^* using strategy derived from Q
Call Algorithm 3 to get the transition probability $P_{S_t S_{t+1}}$

$$\{[S_0, A_0, R_0], [S_1, A_1, R_1], ..., [S_n, A_n, R_n]\} \rightarrow Q(S_t, A_t) = E\left[\sum_{t=0} \gamma^t R(S_t, A_t) | S = S_0, A = A_0\right]$$

If competition scenarios==Conflicting
Call Algorithm 1

$$Q(S_t, A_t) \leftarrow (1-\omega) \times Q(S_t, A_t) + \omega \times (R_t + \gamma^* \max Q(S_{t+1}, A_t))$$

$$\alpha^* = \max Q(S, A^*) = E\left[\sum_{1}^{\infty} \gamma^t R(S_t, A_t^*) | S = S_0, A = A_0\right]$$

$$S \leftarrow S^*;\ A \leftarrow A^*$$

Else
Call Algorithm 2

$$Q(S_t, A_t) \leftarrow (1-\omega) \times Q(S_t, A_t) + \omega \times (R_t + \gamma^* \max Q(S_{t+1}, A_t))$$

$$\alpha^* = \max Q(S, A^*) = E\left[\sum_{1}^{\infty} \gamma^t R(S_t, A_t^*) | S = S_0, A = A_0\right]$$

$$S \leftarrow S^*;\ A \leftarrow A^*$$

Until S is null
Output R, A
Stop

ROIs of NJ and DH are compared in four situations to show the optimal decisions of these two institutions. These four situations are as follows: NJ makes the decision first in the conflicting competition scenario; DH makes the decision first in the conflicting competition scenario; NJ makes the decision first in the cooperative competition scenario; DH makes the decision first in the cooperative competition scenario.

Algorithm 5. Optimal strategies for both financial institutions

Input: the strategy α to be evaluated derived from Q
Output: max $Q(S,A)$, α^* for all steps
Initialize R, A, ξ, K_areward$_t$, K_breward$_t$, KA_a, KA_b
Repeat for each step
Call Algorithm 4
If R_a$_t$>K_areward$_t$
K_areward$_t$=R_a$_t$
Appendix A$_t$_a *to* KA_a$_t$
If R_b$_t$>K_breward$_t$
K_breward$_t$=R_b$_t$
Appendix A$_t$_b *to* KA_b$_t$
Output K_areward$_t$, K_breward$_t$, KA_a, KA_b
Stop

4.1 Results for Unlimited Adjustment Opportunities

Under the assumption of an unlimited number of adjustments, when the difference between the values of ROI in two consecutive simulations is less than the preset value, the game reaches equilibrium. The data of every periods and adjustment process of the two financial institutions is shown in Table 3 and Figures 2-3.

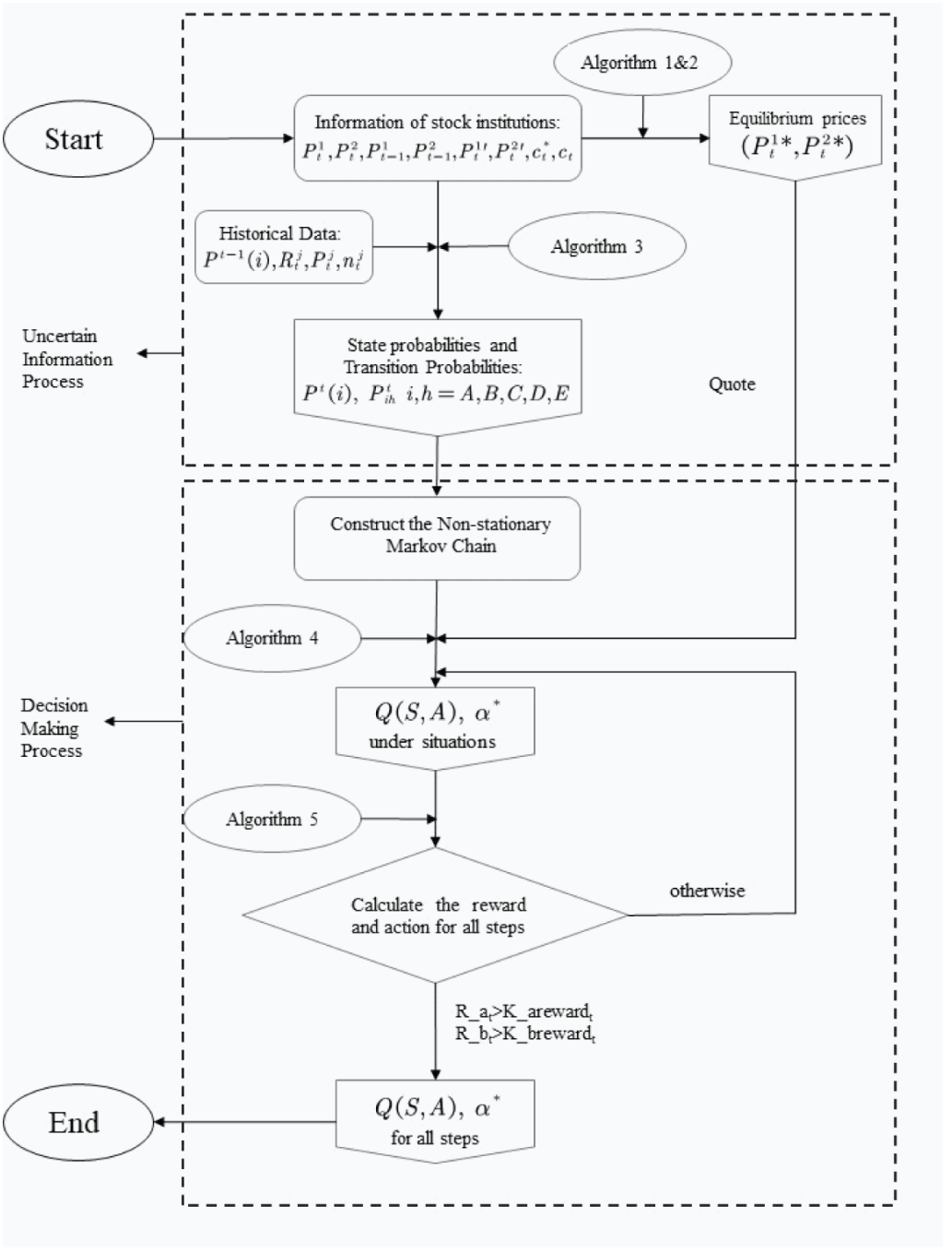
Figures 2 and 3 show several phenomena: 1. The game stops at 14 rounds until balance is reached. The maximum time span for the two financial institutions to change the pending order strategy is 7 rounds. This simulation covers the entire possible time span for financial institutions to make smart decisions on the optimal pending order strategy. 2. Cooperative competition strategy completes the game earlier than conflicting strategy. Because conflict can cause financial institutions to continue changing strategies several times. 3. For large fund companies, no matter who makes the decision first, the conflicting strategy will bring higher returns. 4. In the case of conflicting, large fund companies should let small fund companies take the initiative and then decide on the appropriate unit price to guide retail investors to follow. 5. For smaller fund companies, conflicting strategy generates higher returns than cooperative strategy. Blindly following orders of large fund companies as free riders may lead to high losses. In other words, adopting the conflicting strategy can achieve a higher reward domain.

4.2 Analysis on the Optimal Competition Decision of the Two Institutions

In this section, both institutions have two states, opening positions (OP) and raising order placement prices for shipment (RP). One party does not know the status of the other party when making decisions. According to the results of the simulation, the benefit matrix for NJ and DH first making decision is shown in the following table respectively:

The p and q refer to the probabilities that NJ and DH are in the stage of opening positions respectively. The benefits in the matrix are the value of ROIs. From table 4, the strategy Nash equilibriums are (OP, Cooperation) and (RP, Conflict). However, the ROI of NJ in the strategy (OP, Cooperation) is 0.11148, which is higher than that in the strategy (RP, Conflict), so DH will not choose the conflicting competition strategy with the risk that ROI decreases from 0.06502 to 0.04807. In this case, the optimal strategy for both NJ and DH is (OP, Cooperation). From table 5,

Figure 1. The workflow of the entire estimation



there is no pure strategy Nash equilibrium. Then the expected benefits will be used to analyze the Nash equilibrium of the game. The expected benefits when NJ choose cooperative and conflicting competition strategies are:

Table 2. ROIs of NJ and DH

Date	DH (%)	NJ (%)
2018-06-21	6.6144	12.9471
2018-06-22	3.5815	9.2016
2018-06-23	4.2051	5.8098
2018-06-24	4.7811	6.3604
2018-06-25	6.1340	3.0131
2018-06-26	4.9156	2.0461
2018-06-27	3.8171	4.0668
2018-06-28	10.0077	4.2466
2018-06-29	6.4034	3.6098
2018-06-30	3.3937	11.8633

Table 3. ROIs of NJ and DH under different scenarios

	NJ's ROIs under different scenarios				DH's ROIs under different scenarios			
	NJ First_handed Cooperative	DH First_handed Cooperative	NJ First_handed Conflicting	DH First_handed Conflicting	NJ First_handed Cooperative	DH First_handed Cooperative	NJ First_handed Conflicting	DH First_handed Conflicting
1	0.09544	0.09544	0.09544	0.09544	0.06637	0.06637	0.06637	0.06637
2	0.09714	0.09667	0.09022	0.09105	0.07628	0.07252	0.05701	0.05505
3	0.10095	0.09819	0.10274	0.08828	0.08122	0.07731	0.04854	0.04253
4	0.10143	0.10204	0.09775	0.09973	0.04851	0.08067	0.04156	0.06354
5	0.10049	0.10072	0.09071	0.09174	0.05205	0.09752	0.06857	0.07057
6	0.10614	0.10216	0.08921	0.10422	0.05853	0.09954	0.06602	0.06158
7	0.10739	0.10539	0.08846	0.09673	0.06302		0.04303	0.04702
8	0.11094		0.08214	0.09425	0.06904		0.04105	0.08151
9	0.11148		0.07823	0.09374	0.08012		0.06156	0.07202
10			0.07607	0.10526			0.03551	0.09253
11			0.07994	0.10024			0.06552	0.07754
12			0.09159	0.11327			0.04255	0.09852
13			0.08548	0.10424			0.09104	0.09555
14			0.08429	0.10178			0.08502	0.09054

$$E_{cooperation} = 0.10539 \times q + 0.08023 \times \left(1 - q\right)$$

$$= 0.08023 + 0.02516 \times q \tag{22}$$

$$E_{conflict} = 0.10557 \times q + 0.10178 \times \left(1 - q\right)$$

$$= 0.10178 + 0.00379 \times q \tag{23}$$

Figure 2. Unlimited adjusting opportunities: NJ's ROIs under different scenarios

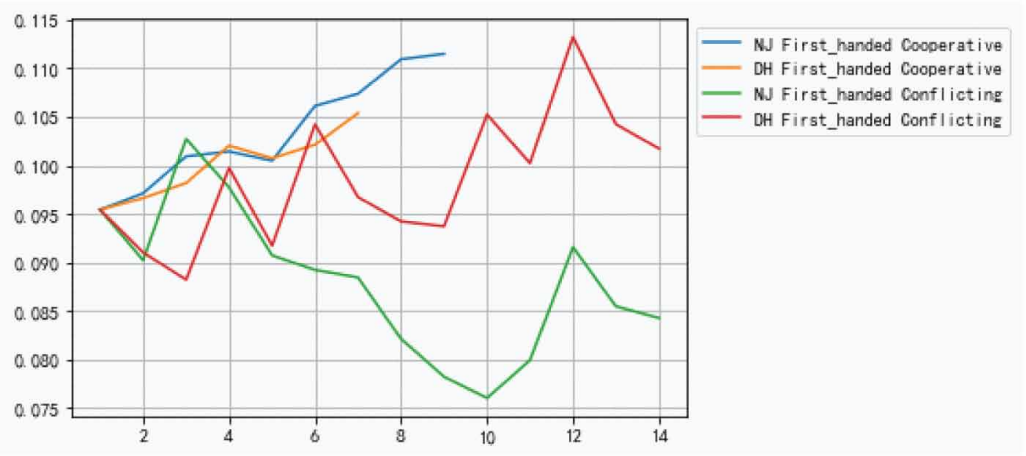


Figure 3. Unlimited adjusting opportunities: DH's ROIs under different scenarios

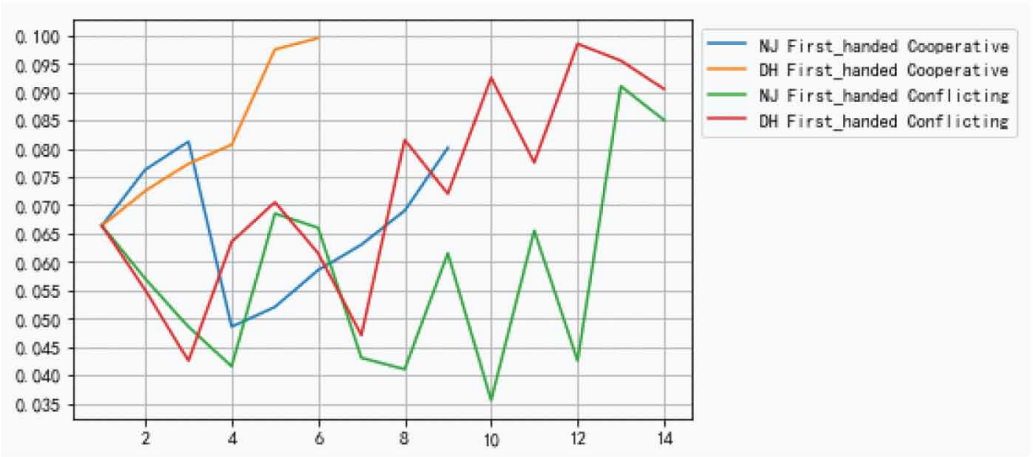


Table 4. The benefit matrix for NJ first making decision

		DH	
		Cooperation	**Conflict**
NJ	OP (p)	0.11148,0.08012	0.08315,0.04807
	RP (1-p)	0.09513,0.04267	0.08429,0.06502

Table 5. The benefit matrix for DH first making decision

		DH	
		OP (q)	**RP (1-q)**
NJ	Cooperation	0.10539,0.09954	0.08023,0.09372
	Conflict	0.10557,0.09562	0.10178,0.09054

Since $E_{cooperation} - E_{conflict} = 0.02137 \times q - 0.02155$, which is only positive when $q > 1.00842$. The *q* refers to probability which is always less than 1, so $E_{cooperation} < E_{conflict}$ and the optimal choice for NJ is the conflicting competition strategy.

Next, the study will introduce the cooperative and conflicting competition strategy in pending orders in detail. Choosing the best strategy is complicated by the fact that companies can only get the information on the trading book without knowing the real intention of other parties. First of all, when NJ is in the stage of opening a position, the cooperative competition strategy is optimal for DH, because NJ will prevent DH from placing large buy or sell orders. When NJ is pulling up the stock price, DH is best to test the NJ's attention by continuing to raise the stock price. If NJ knocks out the DH sell order, it indicates that NJ is bound to raise the stock price for the next round, and DH is better to buy it again. If DH adopts a conflicting competition strategy and buys at time *t*, then DH will bear the risk that NJ will take the opportunity to encounter after it has pended the buy orders. When NJ wants to place the price up, DH should adopt a cooperative competition strategy. At this time, NJ will often pend buy orders to attract the buy orders. DH can choose a lower selling position than NJ to place orders, and then NJ will withdraw its buy order. DH can buy the stocks at a lower price. Fig. 4 and Fig. 5 present the dynamic game process for the HJ initiative and the DH initiative, respectively.

4.3 Robust Tests for Two Factors: Corporate Tax and Market State

This study would use robust tests to assess the following 2 factor effects on the intelligent decision making process: tax and market state. Tax has important effects on fund firms' decisions of pending orders and operational strategies. The change of the tax rate would cause the management to adjust the business decision. The market state is divided into bull market and bear market. The relative position of the 24-tick moving average and the market index can be used to distinguish between bull and bear markets (Chen, 2009). Since both NJ and DH both trade the stock listed on the Shenzhen Stock Exchange, the market index used in this paper is the Shenzhen 100 Index. Market state is a crucial factor when the management makes strategy adjustment decisions. Thus, the above two factors are selected to perform robust tests.

Figure 4. Decision tree when NJ first decides

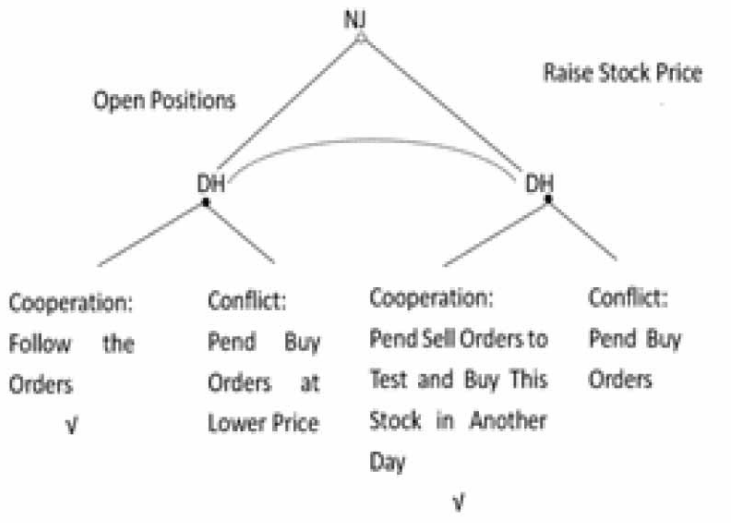
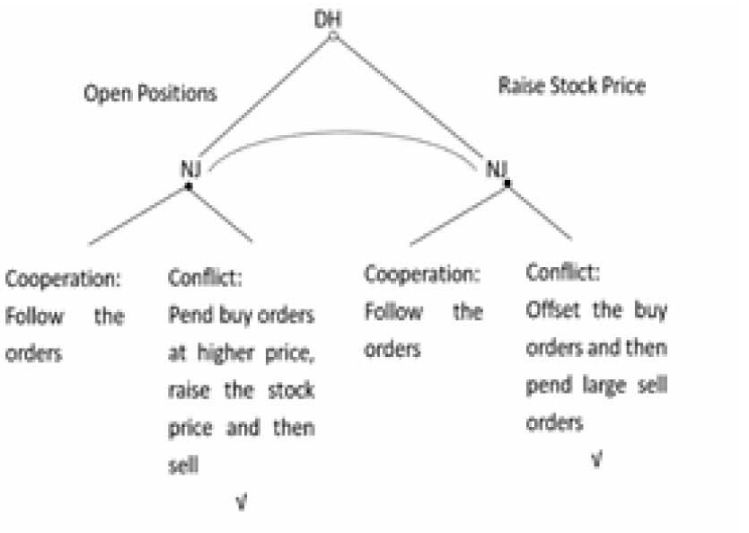


Figure 5. Decision tree when DH first decides



4.3.1 Robust Test Results for Tax

As shown in figure 6 and figure 7, the effects of taxation on NJ and DH are completely different. Higher tax rates give smaller fund firm DH less return, but for larger fund firm NJ more return. The impact of taxation on the decision-making of fund firms' pending orders is not only reflected in the corporate income tax imposed on profits, but also on the handling fees of pending orders. The increase in the tax rate directly leads to an increase in the handling rate, which makes the fund firms pay a higher cost per transaction pending order, so the fund firms' matched order operation will be reduced. For the larger fund firm NJ, due to the large amount of pending order, each transacted order will have more profit, which can appropriately offset the impact of the handling fee. However, when the tax rate rises more, DH pays more for the execution of the matched order operation, which leads to a bigger loss. In addition, big companies usually have more ways to avoid paying taxes. As a result, the impact of the tax on small fund firms is usually greater than that on large fund firms.

4.3.2 Robust Test Results for Market State

The vertical axis of Figures 8 and 9 represents the difference between the Shenzhen 100 Index and the 24-tick moving average. When the difference is greater than zero, the market is a bull market. When it is less than zero, the market is a bear market. Figure 8 and figure 9 show that changes in market conditions can have the same effect on NJ and DH, but the extent of the impact is different. First of all, both NJ and DH can benefit from the high-priced shipments in the bull market, but NJ can raise the stock price to its desired level by the way of pending large buy orders. NJ has strong control over stock prices so it can get more benefits. Second, when it is a bear market, both NJ and DH have losses due to stock price declines. When NJ wants to raise the stock price but the stock price drops a lot, even the difference between the Shenzhen 100 Index and the 24-day moving average is equal to -400, DH can distinguish NJ's fake buy order and then buy the stock at a lower price. But for NJ, the situation is reversed. As the market index falls, retail investors will sell a lot of stocks, and the next sell orders will be closed. Therefore, NJ faces the risk that its own buy orders will be dumped, which will lead to NJ a greater loss.

Figure 6. NJ's robust test results for tax

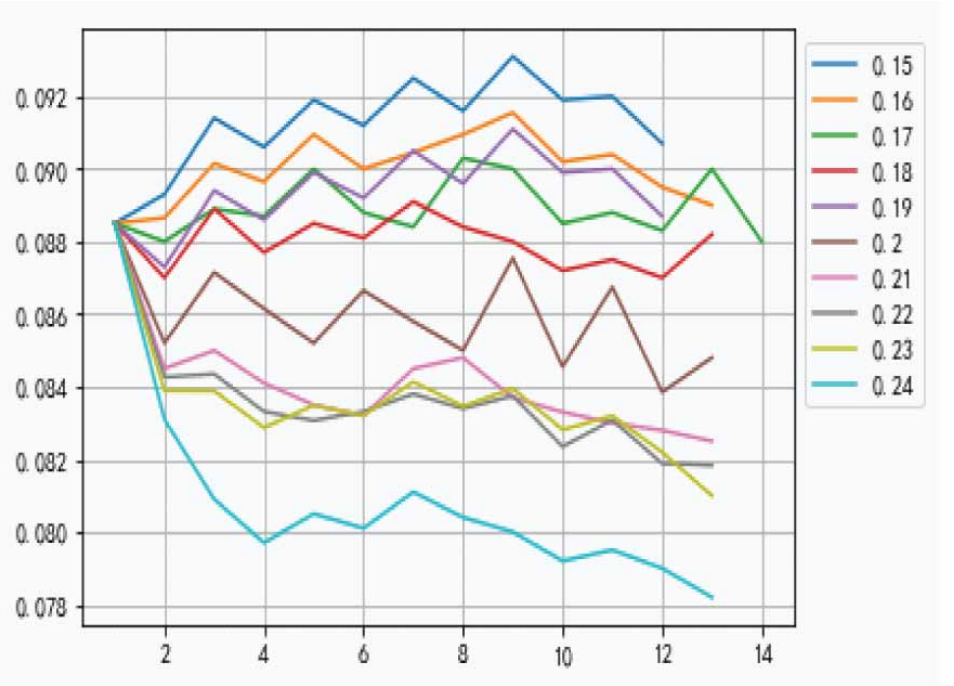


Figure 7. DH's robust test results for tax

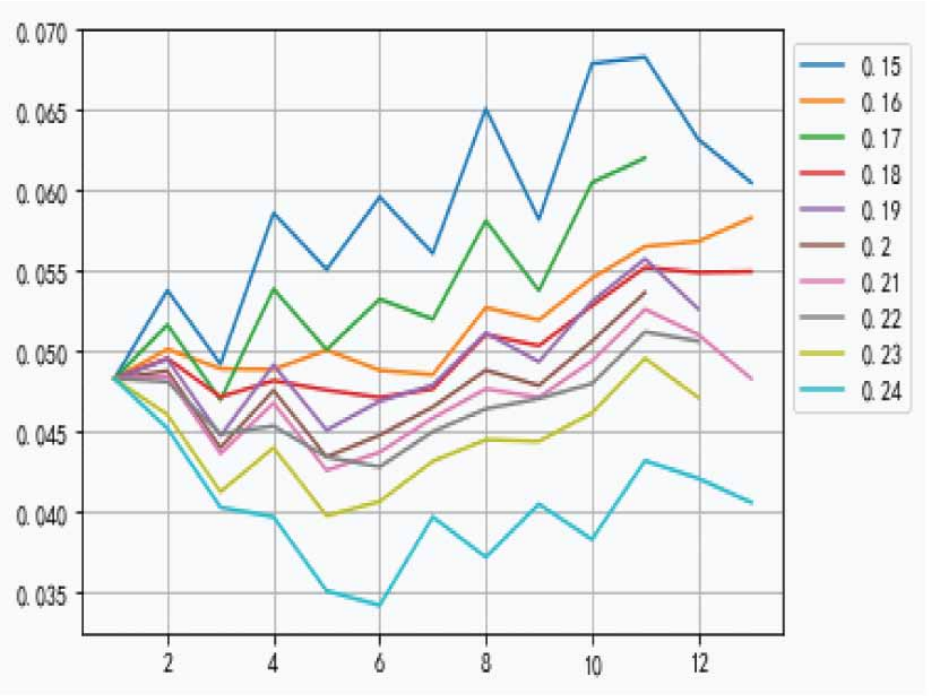


Figure 8. NJ's robust test results for market state

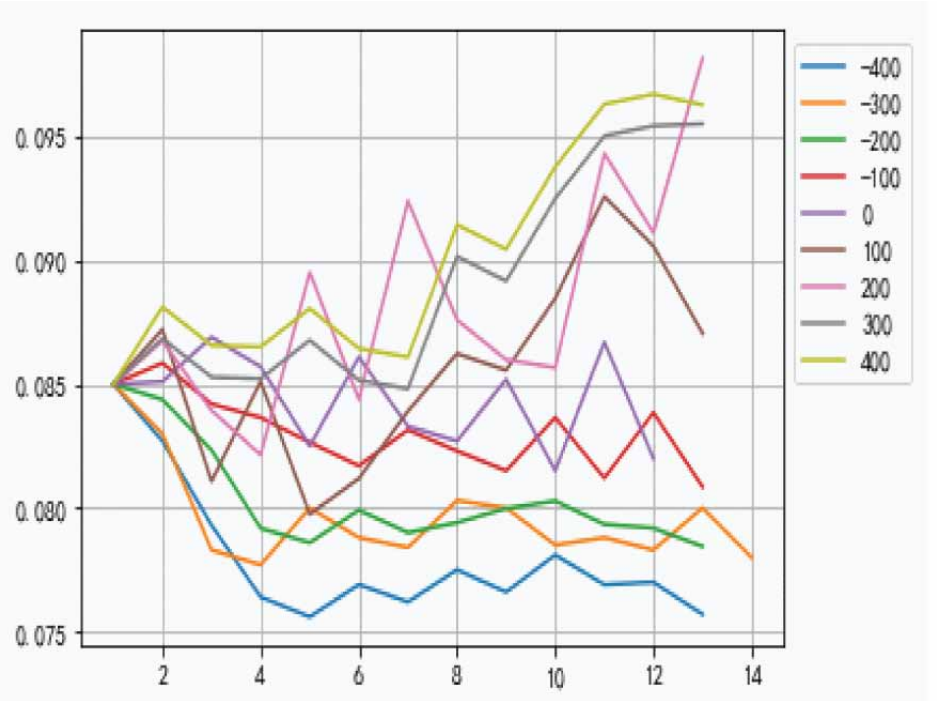
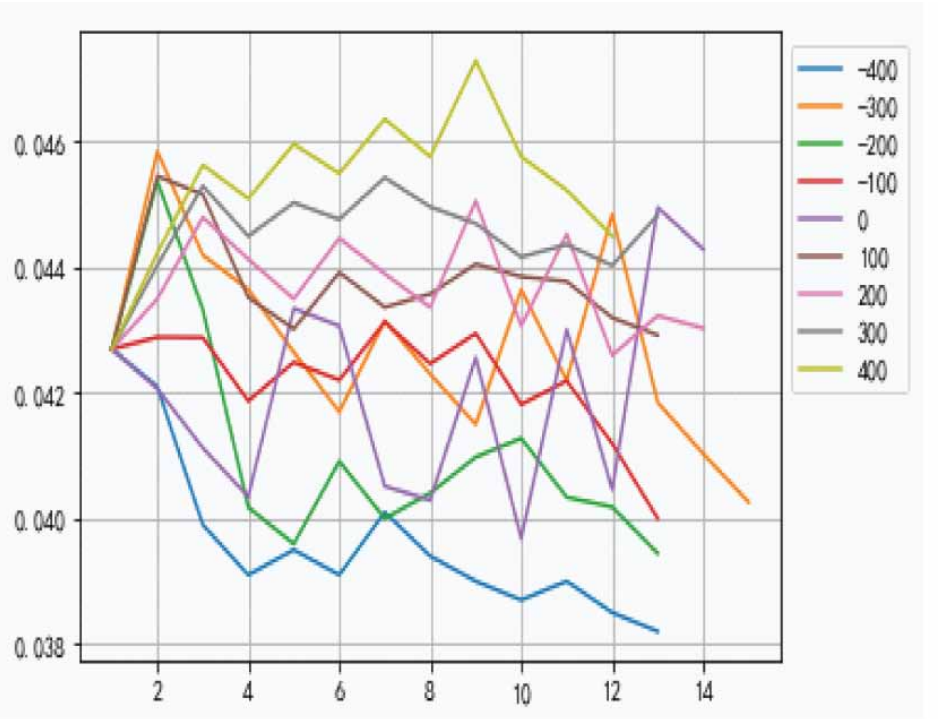


Figure 9. DH's robust test results for market state



5. CONCLUSION

In this study we build adaptive deep reinforcement learning algorithms for financial institutions trading stocks in an inconsistent information and dynamic decision environment. More specifically, the incomplete information dynamic game model and the non-stationary Markov chain are combined to solve the problem of the optimal decision for pending orders of two financial institutions in the case of information inconsistency. The Non-stationary Markov chains (NMC) model is built by using Probabilistic Fuzzy Regression (PFR), Chaos Optimization Algorithm (COA), *t*-Copula simulation, and Reinforced Learning DQN algorithm. First, a dynamic and imperfect information game model and algorithm are built to address the inconsistency of information and the dynamic nature of the decision-making process toward maximizing rewards under two scenarios, conflicting and cooperative. Second, in order to simulate the transition probabilities of non-stationary Markov chain, which would be used in the Markov Decision Process of Reinforced Learning, Probabilistic Fuzzy Regression (PFR), Chaos Optimization Algorithm (COA) and *t*-Copula simulation of a Non-Stationary Markov Chain model would be implemented to study the external risk factors' effect on the transition probabilities. Finally, the Reinforced Learning DQN Algorithm would be used to verify the validity of the optimum actions derived from the progressive game decision process, under the two assumptions, conflicting or cooperative. The process of the complete adjustment strategy of two financial institutions are obtained with both financial institutions having unlimited opportunities to change their strategies before the game starts. The method presented in this paper is of importance not only for interdisciplinary research, but also for practitioners such as fund managers and regulators as well.

REFERENCES

Chang, T., Wang, N., & Chuang, W. (2022). Stock price prediction based on data mining combination model. *Journal of Global Information Management*, *30*(7), 1–19. doi:10.4018/JGIM.296707

Chen, S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, *33*(2), 211–223. doi:10.1016/j.jbankfin.2008.07.013

Dai, W. (2022). Application of improved convolution neural network in financial forecasting. *Journal of Organizational and End User Computing*, *34*(3), 1–16. doi:10.4018/JOEUC.289222

de Hierro, A. F. R. L., Martinez-Moreno, J., Aguilar-Pena, C., & de Hierro, C. R. L. (2016). Estimation of a fuzzy regression model using fuzzy distances. *IEEE Transactions on Fuzzy Systems*, *4*(2), 344–359. doi:10.1109/TFUZZ.2015.2455533

Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, *73*(1), 111–129. doi:10.1111/j.1751-5823.2005.tb00254.x

Dovbysh, A. S., Moskalenko, V. V., Rizhova, A. S., & Dyomin, O. V. (2015). Intelligent Decision Support System for Medical Radioisotope Diagnostics with Gamma-camera. *Journal of Nano-and Electronic Physics*, *7*(4), 1–7.

Du, P., & Shu, H. (2022). Exploration of financial market credit scoring and risk management and prediction using deep learning and bionic algorithm. *Journal of Global Information Management*, *30*(9), 1–29. doi:10.4018/JGIM.293286

Feng, Z., & Chen, M. (2022). Platformance-based cross-border import retail e-commerce service quality evaluation using an artificial neural network analysis. *Journal of Global Information Management*, *30*(11), 1–17. doi:10.4018/JGIM.306271

Gilliland, M., Tashman, L., & Sglavo, U. (2015). *Business forecasting: Practical problems and solutions*. John Wiley & Sons. doi:10.1002/9781119244592

Han, S. H., Yun, M. H., Kim, K. J., & Kwahk, J. (2000). Evaluation of product usability: Development and validation of usability dimensions and design elements based on empirical models. *International Journal of Industrial Ergonomics*, *26*(4), 477–488. doi:10.1016/S0169-8141(00)00019-6

Haverila, M., Haverila, K. C., Mohiuddin, M., & Su, Z. (2022). The impact of quality of big data marketing analytics (BDMA) on the market and financial performance. *Journal of Global Information Management*, *30*(1), 1–21. doi:10.4018/JGIM.315646

Heidari, A. R. (2010). Generation expansion planning in pool market: A hybrid DP/GT model. *International Conference on Systems–Proceedings*, 87–92.

Hou, H., Tang, K., Liu, X., & Zhou, Y. (2022). Application of artificial intelligence technology optimized by deep learning to rural financial development and rural governance. *Journal of Global Information Management*, *30*(7), 1–23. doi:10.4018/JGIM.289220

Ishibuchi, H., & Nii, M. (2001). Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks. *Fuzzy Sets and Systems*, *119*(2), 273–290. doi:10.1016/S0165-0114(98)00370-4

Kim, J. Y., & Kwon, J. Y. (2017). Strategic delegation and second mover advantage in duopoly. Economic research-. *Ekonomska Istrazivanja*, *30*(1), 732–744. doi:10.1080/1331677X.2017.1311227

Lau, T. W., Hui, P. C., Ng, F. S., & Chan, K. C. (2006). A new fuzzy approach to improve fashion product development. *Computers in Industry*, *57*(1), 82–92. doi:10.1016/j.compind.2005.04.003

Li, C., Jin, K., Zhong, Z., Zhou, P., & Tang, K. (2022). Financial risk early warning model of listed companies under rough set theory using BPNN. *Journal of Global Information Management*, *30*(7), 1–18. doi:10.4018/JGIM.300742

Li, J., Zeng, W., Xie, J., & Yin, Q. (2016). A new fuzzy regression model based on least absolute deviation. *Engineering Applications of Artificial Intelligence*, *52*, 54–64. doi:10.1016/j.engappai.2016.02.009

Mishra, N., Choudhary, A. K., & Tiwari, M. K. (2008). Modeling the planning and scheduling across the outsourcing supply chain: a Chaos-based fast Tabu–SA approach. *International Journal of Production Research, 46*(13), 3683-3715.

Morsalin, S., Mahmud, K., & Town, G. (2016). Electric vehicle charge scheduling using an artificial neural network. *IEEE*, 276–280.

Nanba, R., Hasegawa, M., Nishita, T., & Aihara, K. (2002). Optimization using chaotic neural network and its application to lighting design. *Control and Cybernetics*, *31*(2), 249–269.

Otadi, M. (2014). Fully fuzzy polynomial regression with fuzzy neural networks. *Neurocomputing*, *142*, 486–493. doi:10.1016/j.neucom.2014.03.048

Pombeiro, H., Machado, M. J., & Silva, C. (2017). Dynamic programming and genetic algorithms to control an HVAC system: Maximizing thermal comfort and minimizing cost with PV production and storage. Sustainable Cities and Society, 34, 228-238.

Ren, B., & Zhong, W. (2011). Multi-objective optimization using chaos based PSO. Information Technology Journal, 10(10), 1908-1916.

Salih, A. S. M., & Abraham, A. (2015). Intelligent decision support for real time health care monitoring system. In A. Abraham, P. Krömer, & V. Snasel (Eds.), Afro-European Conference for Industrial Advancement (pp. 183–192). Springer.

Şekkeli, G., Köksal, G., Batmaz, I., & Türker Bayrak, Ö. (2010). Classification models based on Tanaka's fuzzy linear regression approach: The case of customer satisfaction modeling. *Journal of Intelligent & Fuzzy Systems*, *21*(5), 341–351. doi:10.3233/IFS-2010-0466

Sen, S. D. (2015). *An intelligent and unified framework for multiple robot and human coalition formation*. Vanderbilt University.

Smith, L. D., Sanchez, S. M., & Lawrence, E. C. (1996). A comprehensive model for managing credit risk on home mortgage portfolios. *Decision Sciences*, *27*(2), 291–317. doi:10.1111/j.1540-5915.1996.tb01719.x

Srivastava, P. R., & Eachempati, P. (2021). Intelligent employee retention system for attrition rate analysis and churn prediction: An ensemble machine learning and multi-criteria decision-making approach. *Journal of Global Information Management*, *29*(6), 1–29. doi:10.4018/JGIM.20211101.oa23

Su, Z. G., Wang, P. H., & Song, Z. L. (2013). Kernel based nonlinear fuzzy regression model. *Engineering Applications of Artificial Intelligence*, *26*(2), 724–738. doi:10.1016/j.engappai.2012.05.009

Sun, Y., Liu, L., Fang, J., Zeng, X., & Wan, Z. (2023). MuSu: A medium-term investment strategy by integrating Multifactor model with industrial Supply chain. *International Journal of Financial Engineering*, *10*(02), 2250034. doi:10.1142/S2424786322500347

SunY.LiuL.XuY.ZengX.ShiY. (2022). *A Survey on Alternative Data in Finance and Business: Emerging Applications and Theory Analysis*. https://ssrn.com/abstract=4148628

SunY.ZengX. (2022). Efficient Markets: Information or Sentiment? https://ssrn.com/abstract=4293484

Sun, Y., Zeng, X., Cui, X., Zhang, G., & Bie, R. (2021). An active and dynamic credit reporting system for SMEs in China. *Personal and Ubiquitous Computing*, *25*(6), 989–1000. doi:10.1007/s00779-019-01275-4

Sun, Y., Zeng, X., Zhao, H., Simkins, B., & Cui, X. (2022). The impact of COVID-19 on SMEs in China: Textual analysis and empirical evidence. *Finance Research Letters, 45*, 102211.

Tan, X., Wu, J.-J., Mao, T.-T., & Tan, Y.-J. (2017). Multi-attribute intelligent decision-making method based on triangular fuzzy number hesitant intuitionistic fuzzy sets. *Journal of Systems Engineering and Electronics*, *39*(4), 829–836.

Vagin, V., Fomina, M., & Morosin, O. (2015). Argumentation in inductive concept formation. IEEE, 133–137.

Weidong, Z., & Shiping, G. (2009). Intelligent decision support system and its application in science research project selection. IEEE, 858–862.

Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications*, *79*, 153–163. doi:10.1016/j.eswa.2017.02.041

Wu, Y., Zhu, D., Liu, Z., & Li, X. (2022). An Improved BPNN algorithm based on deep learning technology to analyze the market risks of A+H shares. *Journal of Global Information Management*, *30*(7), 1–23. doi:10.4018/JGIM.313188

Xu, L. D., Tjoa, A. M., & Chaudhry, S. S. (2008). Research and practical issues of enterprise information systems. In A. Min Tjoa, M. Raffai, P. Doucek, & N. Maarit Novak (Eds.), *Lecture Notes in Business Information Processing* (Vol. 327). Springer.

Zhao, J., & Wei, Y. (2017). A novel algorithm of human-like motion planning for robotic arms. *International Journal of HR; Humanoid Robotics*, *14*(1), 1–27. doi:10.1142/S0219843616500237

Zhao, Y. (2022). Risk prediction for internet financial enterprises by deep learning algorithm and sustainable development of business transformation. *Journal of Global Information Management*, *30*(7), 1–16. doi:10.4018/JGIM.300741

Zhao, Y., & Zhou, Y. (2022). Measurement method and application of a deep learning digital economy scale based on a big data cloud platform. *Journal of Organizational and End User Computing*, *34*(3), 1–17. doi:10.4018/JOEUC.295092

Zu, T., Wen, M., & Kang, R. (2017). An optimal evaluating method for uncertainty metrics in reliability based on uncertain data envelopment analysis. *Microelectronics and Reliability*, *75*, 283–287. doi:10.1016/j.microrel.2017.03.033