

Article

Detecting the Proportion of Traders in the Stock Market: An Agent-Based Approach

Minh Tran ^{1,2,*}^(D), Thanh Duong ³, Duc Pham-Hi ^{1,4} and Marc Bui ²

- ¹ John von Neumann Institute, Vietnam National University Ho Chi Minh City, Ho Chi Minh City 70000, Vietnam; hi-duc.pham@jvn.edu.vn
- ² CHArt Laboratory EA 4004, EPHE, PSL Research University, 75014 Paris, France; Marc.Bui@ephe.psl.eu
- ³ CEO at QT-Data Inc., Saskatoon, SK S7K 2P7, Canada; henry.duong@qntdata.com
- ⁴ Financial Engineering Department, ECE Paris Graduate School of Engineering, Paris 75015, France; phamhi@ece.fr
- * Correspondence: minh.tran@jvn.edu.vn

Received: 31 December 2019; Accepted: 29 January 2020; Published: 5 February 2020



Abstract: In this research, an agent-based model (ABM) of the stock market is constructed to detect the proportion of different types of traders. We model a simple stock market which has three different types of traders: noise traders, fundamental traders, and technical traders, trading a single asset. Bayesian optimization is used to tune the hyperparameters of the strategies of traders as well as of the stock market. The experimental results on Bayesian calibration with the Kolmogorov–Smirnov (KS) test demonstrated that the proposed separate calibrations reduced simulation error, with plausible estimated parameters. With empirical data of the Dow Jones Industrial Average (DJIA) index, we found that fundamental traders account for 9%–11% of all traders in the stock market. The statistical analysis of simulated data can produce the important stylized facts in real stock markets, such as the leptokurtosis, the heavy tail of the returns, and volatility clustering.

Keywords: Bayesian optimization; artificial stock market; agent-based modeling; traders proportion

1. Introduction

The study of a complex system has always been of great interest because of thier rich properties and behavior, for example, the solar system [1] or biological systems [2]. By analyzing such natural systems, people discovered fundamental laws that help them to build other systems, such as computer networks [3] or power-grid systems [4]. A financial market can also be considered as a complex system. For tens of years, researchers have been faced with the problem of the analysis and modeling of financial markets [5–7]. One of the most important questions is "How can we develop an artificial stock trading system and describe investors' behaviors in a way that is similar to reality?". Previous models are based only on a single fully rational representative agent and fail to reproduce all the properties of real markets [8,9]. To overcome these problems, a new behavioral approach emerged, referred to as agent-based modeling, characterized by markets populated with bounded rational, heterogeneous agents [10].

An agent-based model (ABM) in the financial market is a class of quantitative models to simulate the decisions and interactions of different traders in order to understand their behaviors and their impact on the market [11]. In recent years, ABM in financial markets is still a prominent topic of interest for researchers [6,12]. The behavior of the agents is described in a variety of ways; their interaction and the



impact of exogenous factors in the market are also introduced [13]. Thereby, studies help managers to understand the behavioral characteristics of investors and explain the phenomena occurring in the stock market [13,14].

One of the problems of agent-based modeling is how to tune hyperparameters for the traders' strategies and the market. Many models use large parameter sets and run simulations for long periods of time [15,16]. However, even for small models, exploring the stylized facts through all parameter combinations is not possible or is prohibitively costly [12]. For example, in [15], the authors generated 400,000 initial points to calibrate their model. To solve the above problems, this model introduces an approach with Bayesian optimization and an error measurement function to tune the hyperparameters of the environment.

In this work, we combine the main properties of the pioneer works. An agent-based model is introduced to highlight the stylized facts of the stock market. This research offers a novel perspective on how to detect the proportion of each trader in the stock market. The results can be used in selecting effective trading strategies for traders. The model consists of three types of traders: noise traders, technical traders, and fundamental traders, with various behaviors to increase diversity. A noise trader makes decisions based on price changes. They have different emotions and news reactions which play important roles in making decisions. Technical traders are diversified by trading indicators; they make different decisions based on target return and time. Fundamental traders are introduced to balance the market; they are diversified by their confidence level, which measures the convergence of the stock price to its fundamental value. The interaction is a function that combines the sentiment change when getting information from the market and the influence of other investors' decisions.

The highlight of our model is that Bayesian optimization is introduced to tune the hyperparameters of trading strategies and the environment. The results show that Bayesian optimization can propose an optimal set of parameters and reduce simulation error. This provides stylized facts to compare with the real market, providing a bench-marking method for calibrating parameters of agent-based stock market. We used different methods of comparing the two distributions—Kullback–Leibler (KL) divergence and the Kolmogorov–Smirnov (KS) test—to measure the simulation error. In our model, the KS test showed better results by successfully explaining some important stylized facts in real stock markets, such as the leptokurtosis, the heavy tail of the returns, and the volatility clustering. With extensive simulations, we found that the fundamental traders accounted for 9%–11% of all traders in the stock market. These results are consistent with the real-market research [17].

The remainder of this paper is organized as follows. The literature on ABM of stock markets is introduced in Section 2. Section 3 outlines the construction of the model. Section 4 presents the description of the data and the calibration of model. Statistical analysis of the model's results and the stylized facts are given in Section 5. Finally, Section 6 concludes and proposes some extensions of our work.

2. Background

In the financial market, there are many types of agents (investors, banks, central banks, etc.). These agents can interact directly through trading, exchanging derivatives. They can also interact indirectly, because any decision or behavior of an agent affects others [18]. Moreover, the agents have different sensitivities to exogenous factors in the market and thus different news can influence and impact them [19]. Therefore, financial market modeling needs to understand the characteristics of agents such as behavior and interaction between investors and the impact of exogenous factors such as public news on the market [13,20]. The studies in this paper will focus on clarifying these properties.

The first approach addresses the types of agents and how they impact the dynamics of the market. Pioneering works have proposed a number of interesting agent-based financial market models. The concept provides a stabilizing impact on market dynamics [24].

The second approach that should be considered is the diversity of agents. To make every fundamental agent and technical agent unique, they are assigned with different memory lengths or reaction intensities to price and fundamental changes [13]. Exogenous factors such as news, social media, and insider information [25] also have different impacts on each agent. In addition, the sentiments of agents such as risk aversion play an important role in the decision-making process [14,26].

The third approach is the interaction between agents. Besides the interactions that directly affect market prices, investors' decisions also have an indirect effect on each other. In [27,28], the authors introduced the propagation in the decision of investors with a fixed probability. In another study [13], the interactions and diversity among fundamental traders and technical traders were introduced. The average utility associated with each type was calculated and agents could rely on profit to come up with an effective investment strategy.

Each of these approaches has been extended in various interesting directions. From that, the stylized facts of financial markets are highlighted, such as heavy tails in stock return distribution and volatility clustering [10,28]. Another proposed stylized fact is the positive autocorrelation in volatility and trading volume [13,29].

One important issue is how to tune hyperparameters for traders' strategies and the market. Grid search is a method to list all combinations of hyperparameters, then perform a model test with this list [30]. Random search only randomly selects a finite number of hyperparameters from the list to conduct a model test [31]. Theoretically, grid search can find the global value, but as the number of parameters increases, this becomes increasingly impossible due to the time and cost of implementation. Random search does not run as many cases as grid search, so it is significantly faster [31]. However, depending on its randomness, each run will receive a different optimal value because it does not guarantee the optimum value to be global or local only. Bayesian optimization is an adaptive approach to parameter optimization, trading off between exploring new areas of the parameter space and exploiting historical information to find the parameters that maximize the function quickly [30]. Thus, Bayesian optimization is introduced to tune hyperparameters for the trading strategies of technical traders and the environment of our model.

3. The Model

This section outlines the construction of a new agent-based model. This model adds the understanding of the reason for the stylized facts and the estimated proportion of traders in the stock market.

3.1. Environment

The agent-based model in this paper focuses on analyzing the behavior of traders and the log return properties which are reflected in the simulation process. The model considers a fixed amount of agents, denoted as N, trading a single asset. The current time is denoted with t and the corresponding price is S(t). At each time step t, trader i can choose between three actions: buy/sell one unit of the stock or hold, $s_i(t) \in \{+1, -1, 0\}$, respectively. For liquidity purposes, the model assumes that every trader always has enough wealth and shares—in other words, they are always able to buy or sell. The market price and market return are then calculated. The market price S(t) is updated according to

$$S(t) = S(t-1)[1+r(t)],$$

and return r(t) is determined according to

$$r(t) = g\left(\frac{D(t)}{N}\right),$$

where the excess demand $D(t) = \sum_{i=1}^{N} s_i(t)$, the price impact function $g(z) = \arctan(z/\lambda)$, and λ measures the market depth or liquidity.

Due to the tendency to form discrete clusters of traders [32], we define different types of traders in this model based on their predefined trading rules. There are three different types of traders operating in the market. Noise traders are people who make decisions at random according to their news reactions and their propensity to sentiment contagion. Technical traders analyze charts and make decisions based on current patterns and trends analysis, and fundamental traders trade based on the fundamental profit-generating potential of the stock.

3.2. Noise Agent

Noise agents are traders who trade on the basis of misunderstanding information and news regarding future prices. They make decisions and trade based on inaccurate analysis of the market [33]. In this model, noise agents rely only on the current market situation to make trading decisions: buy, sell, or hold [13]. As the key variable, the sentiment of the noise agents implies their reaction to the news they get [20]. Good sentiment means that agents see the news as good news and hope future prices will increase (bullish) and bad sentiment means that the agents consider the news to be bad and hope prices in the future will decrease (bearish).

The decision of noise agents is determined by the following process. At each time step *t*:

- Agents receive public information as a signal I(t).
- Each agent *i* compares the signal to his threshold $\theta_i(t)$ to make a decision.
- After calculating the market return, he updates his threshold with a probability *p*_i.

The inflow of news arrives to the market as a signal I(t) with $I(t) \sim N(0, \sigma_I^2)$. Each agent will decide whether the news is significant or not, in which case he will make a decision according to the sign of I(t). To describe the heterogeneity, we introduce the sentiment of each agent to the news as a decision threshold $\theta_i(t)$ with the initial trading threshold set between 1 and 2 times the standard deviation of the news, $\theta_i(0) = \sigma_I [1 + U(0, 1)]$.

Let N_{noise} denote the number of noise agents. The trading rule of each noise agent $i, i \in N_{noise}$, is represented by

$$s_i^{noise}(t) = \mathbb{1}_{I(t) > \theta_i(t)} - \mathbb{1}_{I(t) < -\theta_i(t)}.$$

Market-related sentiments are formed differently for each agent and each responds differently to the information. To avoid the artificial ordering of agents as in sequential choice models, we follow [20] to generate an asynchronous strategy to update the agents' threshold. At each time step, each agent *i* will update his threshold $\theta_i(t)$ with probability $0 \le p_i \le 1$. The threshold is set equal to the absolute return at time t, $|r(t)| = |ln \frac{S(t)}{S(t-1)}|$. Introducing independent and identically distributed random variables $u_i(t)$, $i \in N_{noise}$, uniformly distributed on [0, 1], the updating scheme is represented as

$$\theta_i(t) = \begin{cases} |r(t)|, & \text{if } u_i(t) < p_i; \\ \theta_i(t-1), & \text{otherwise.} \end{cases}$$

In the next addition to the model, technical agents and fundamental agents are introduced.

3.3. Technical Agent

Unlike the noise agents, technical agents are traders who use the past prices to infer private information. They make decisions by calculating the technical indicators from historical prices [34]. The characteristic of technical analysis is to make decisions after identifying the trends at an early stage and reverse the decision when the trend reversal occurs [35]. In this model, the agents are assumed to use a technique in the real world called a moving average-oscillator (MAO) to predict future price moves [36]. The technique first involves taking two moving averages of the price, a short-term moving average *A* and a long-term moving average *B*, of different lengths $l_A < l_B$. The moving average crossover is calculated as the difference M(t) between these two moving averages. M(t) > 0 indicates that the price is trending upwards and the agents will make a buy. M(t) < 0 indicates that price is trending down and the agents will sell the stock.

In order to account for individual heterogeneity, each agent can use different lengths of historical data, so one moving average strategy may indicate an uptrend while another moving average strategy indicates a downtrend. However, in our model, the difference in window lengths, l_A and l_B , has no effect on agents' trading decision. When the historical stock price data has a strong uptrend or downtrend, all investors will make the same decision. Thus, we assume that all technical agents can use Bayesian optimization to estimate the optimal length for short and long windows.

We formulate an optimization problem with three main ingredients:

- Domain space: The range of each hyperparameter is selected. Following the studies in [37], our approach starts the domain from a wide range and then focuses on specific areas around the optimal parameters calculated from the previous run. Thus, each agent is assigned a random short window and long window derived from a random uniform distribution in [5, 40] and [50, 100], respectively.
- Objective function: We define a score function that indicates how well a set of long-short periods performs. Cumulative returns and Sharpe ratio are usually used to indicate the performance of a technical strategy [38]. In our model, the Sharpe ratio is used as the evaluation metric of choice. We put a minus sign before the objective function since Hyperopt by default is defining a function to minimize.
- Surrogate function and selection function: The surrogate function is used to propose sets of values that increase performance in minimizing the score of the objective function. Then, they are selected by applying a criterion. In this model, we follow [30] to use the tree Parzen estimator as the surrogate function. The expected improvement criterion is used as the selection function.

The optimal lengths for short and long windows with the largest Sharpe ratio are 35 and 51, respectively. The buy/sell signals are shown in Figure 1.



Figure 1. Moving average oscillator with buy/sell signals.

To describe the heterogeneity, we introduce $\psi_i(t)$ as the average profit a technical agent *i* gained in the previous n_i trading day, $\psi_i(t) = \sum_{j=1}^{n_i} \frac{1}{n_i} r(t-j)$, which is considered as a threshold to make a decision. Each agent *i* will update her threshold $\psi_i(t)$ with probability $0 \le p_i \le 1$ at each time step *t*. The updating scheme is represented as in the description of noise agents.

Technical agents are also affected by the public information like the noise agents. The news reactions $Q_i(t)$ are represented by

$$Q_i(t) = \mathbb{1}_{I(t) > \theta_i(t)} - \mathbb{1}_{I(t) < -\theta_i(t)}$$

Define the sign function, sign(x), as

$$sign(x) = \begin{cases} 1, & \text{if } x > 0; \\ -1, & \text{if } x < 0; \\ 0, & \text{if } x = 0. \end{cases}$$

Let N_{ta} denote the total number of technical agents. The trading rule of the technical agent $i, i \in N_{ta}$, is represented by

$$s_i(t) = \operatorname{sign}\left[\psi_i(t)M(t) + Q_i(t)\right].$$

The amplification of trends and unlimited buying power of technical agents can cause problems because the price can grow to infinity or drop to extremely small values very quickly. Thus, fundamental agents are necessary to keep the market reasonably stable.

3.4. Fundamental Agent

Fundamental agents are traders who measure the intrinsic value by analyzing the accounting, finance, and economics of the stock [39]. They make decisions with the belief that the stock price will return to its fundamental value in the long-run. The fundamental value of the stock f_t is public to all agents and is assumed to follow a random walk process

$$f(t) = f(t-1) + \eta,$$

where the fundamental shocks $\eta \sim N(0, \sigma_{\eta}^2)$.

At each time step t, fundamental agents make a decision based on the difference between S(t) and f(t). Therefore, if the price is below (above) its fundamental value, the agents will make a buy (sell) order. This strategy shortens the difference between price and its fundamental value. They help to stabilize the market and cause the opposite effect on prices for technical agents. In order to account for individual heterogeneity, they are given heterogeneous beliefs about the fundamental price for more interesting dynamics [40].

Let N_{fa} denote the total number of fundamental agents. The trading rule of the fundamental agent $i, i \in N_{fa}$, is represented by

$$s_i^{fa}(t) = sign\left[\kappa(f(t) - S(t)) + \epsilon_i\right],$$

where κ is a positive coefficient that describes the speed of adjustment. The idiosyncratic term ϵ_i embodies a random term that accounts for each individual's own interpretation. We take it as being normally distributed around zero and with a standard deviation controlled by the user, $\epsilon_i \sim N(0, \sigma_{\epsilon}^2)$.

Finally, all agents are put together in the artificial financial market. This includes N_{noise} noise agents, N_{ta} technical agents, and N_{fa} fundamental agents. The total number of traders is N,

$$N = N_{noise} + N_{ta} + N_{fa}.$$

4. Bayesian Calibration

4.1. Data Description

As an empirical basis, we used a data set that covers 6 years of daily quotes of the DJIA Index, from of January 2013 to January 2019. This period contains 1508 daily returns, of which the main statistical properties are given in Table 1. The analysis focused on stylized facts of the distribution of stock returns and the volatility clustering in the market. Based on this data set, we estimated the proportion of each type of trader in our artificial stock market.

Table 1. Statistical properties of the actual returns.

Number of Observations	Mean	Standard Deviation	Kurtosis	Skewness	Min	Max
1508	0.03%	8.09%	4.31	-0.64	-5.07%	3.81%

4.2. Model Calibration

We constructed an artificial stock market of *N* traders. Each type of trader haD a minimum number of 10 and a maximum of 100. The trading period was considered as days, thus the results can be compared with properties of daily returns. We chose $\lambda \in [5, 20]$ which means the minimum value of daily returns was 5% and the maximum value of daily returns was 20%. We denote $1/p_i$ as the latency of an agent *i* when making his decision $\theta_i(t)$. *p* was chosen within [0.001, 0.1] to be interpreted as days. For example, q = 0.1 means the average updating period of agents was 100 days. The number of days each technical agent collected to calculate profit was chosen within [5,22], which describes the interest in the returns of traders in the range from 1 week to 1 month. The short and long windows for the technical strategy of all technical traders are given in Section 2, $l_A = 35$, $l_B = 51$. The amplitude σ_I of the noise of the news arrival process can be chosen in the range [0.001, 0.01] to reproduce a realistic range of values for the volatility. The standard deviation of the noise process in fundamental value, σ_n , was calculated from the standard deviation of the real price ($\sigma_{\eta}^{p} = 3255.43$), which is in $[0.5\sigma_{\eta}^{p}, 1.5\sigma_{\eta}^{p}]$. The value of the standard deviation in idiosyncratic term was chosen in the range [0.001, 1]. The results discussed in the next section are generic within these parameters' ranges.

For clarity, we summarize the parameters and calibration ranges used in our model in Table 2. We formulated an optimization problem using the *"hyperopt"* package of Python with the three main ingredients:

- Domain space: Each trader was assigned initial values derived from random uniform distributions within the range of Table 2, respectively.
- Objective function: KL divergence and KS test were used as the evaluation metrics of choice. For this
 run, the algorithm found the best value of hyperparameters (the one which minimized the loss) in
 just under 100 trials.
- Surrogate function and selection function: Tree Parzen estimator was used as the surrogate function.
 Then, the expected improvement criterion was used as the selection function.

Parameters	Description	Value
N _{noise}	Number of noise agents	[10, 100]
N _{ta}	Number of technical agents	[10, 100]
N_{fa}	Number of fundamental agents	[10, 100]
$\dot{\lambda}$	Market depth	[5,20]
р	Fraction of agents updating their news reaction	[0.001, 0.1]
п	Number of days each technical agent collected to calculate profit	[5,22]
κ	Adjustment speed	[0.001,1]
σ_{I}	Standard deviation of the noise representing the news arrival process	[0.001, 0.01]
σ_{η}	Standard deviation of the noise process in fundamental value	[1627.72, 4883.15]
σ_{ϵ}	Standard deviation of idiosyncratic term	[0.01, 1]

Table 2. Description of parameters.

Both KL divergence and KS test found the optimal set of parameters with small simulation error as shown in Table 3. The proportion of each type of trader with KL and KS tests were: N_{noise} : 12.2%, N_{ta} : 78.3%, N_{fa} : 9.6% and N_{noise} : 13.6%, N_{ta} : 75.6%, N_{fa} : 10.6%, respectively. Note that we did not compare values of KL and KS due to differences in measurement definitions. In this work, we picked out the better objective function by investigating their stability over many simulations. KL divergence represents the instability in many simulations; the minimum value ranged from 0.08 to 4.5. In contrast, the KS test showed stability with small fluctuations in the range of 0.03 to 0.06. The stylized facts which we found to be present in the simulated returns are discussed below.

Table 3. Optimized parameters. KL: Kullback-Leibler divergence; KS: Kolmogorov-Smirnov.

Method	Optimal	Parameters									
Wiethou	Value	N _{noise}	N _{ta}	N _{fa}	λ	p	n	к	σ_I	σ_{η}	σ_{ϵ}
KL	0.08	14	90	11	20	0.061	16	1	0.001	2700	0.004
KS	0.03	18	100	14	15	0.1	9	1	0.001	2000	0.003

Moreover, we considered many different sets, which contained fixed numbers of different types of traders in ABM. The results in Table 4 show that the KS test had the lowest value when there were three types of traders in the market. Thus, the KS test is an effective objective function to measure simulation errors in the Bayesian calibration of our model.

KS Test	Trader Set					
K5 lest	N _{noise}	N _{ta}	N_{fa}			
0.10	100	0	0			
0.22	0	100	0			
0.13	100	100	0			
0.25	0	100	100			
0.06	100	100	100			

Table 4. Comparing the KS test results of different fixed numbers of types of traders.

5. Simulation Results

In this section, we consider a stock market with three types of trader: noise, technical, and fundamental traders, with the KS test as the objective function for calibration. In the first stage, only the noise traders joined the market since the technical traders and fundamental traders need to collect the historical data. In the second stage, after a few days (here we chose the number of days equal to the optimal long window), technical traders and fundamental traders joined the market. The calibrated parameter set is shown in Table 3. Figure 2 shows the daily movement of actual return and simulated return.



Figure 2. Simulation of returns.

5.1. Distribution of Returns

The statistical analysis of the actual returns and simulated returns is given in Tables 5 and 6. The simulated returns had similar properties to the real ones. In financial data, the heavy tail of the distribution of returns is displayed by a positive kurtosis [41]. We found that the Shapiro–Francia test on actual returns and simulated returns rejected the normality at a significance level of 0.1% with *p*-values of 2.2×10^{-22} and 1.5×10^{-12} , respectively. The quantile–quantile (Q–Q) plots in Figure 3 are also show the same results. In Figure 4a, the distribution of actual and simulated returns do not display the heavy tail clearly on linear axes. Thus, we change the vertical axis to use logarithmic scaling in Figure 4b to zoom in on the tails. The returns produced by the model were found to be consistently leptokurtically distributed with a heavy tail.

Table 5. Comparing statistical properties of the actual returns and simulated returns.

Data	Mean	Standard Deviation	Kurtosis	Skewness
Real returns	0.03%	8.09%	4.31	-0.64
Simulation	0.02%	8.09%	2.32	-0.08



Figure 3. Quantile-quantile plot of returns.



Table 6. Comparing statistical properties of the actual returns and simulated returns (cont.).

Figure 4. Distribution of actual and simulated returns.

5.2. Absence of Autocorrelations in Returns

Figure 2 shows the plots of stock returns for both actual data and simulated data. We studied the stationarity of the stock returns by using an augmented Dickey–Fuller (ADF) unit root test. The results are shown in Table 7. The *p*-values of the ADF test statistic for the real returns and the simulated returns are 2.31×10^{-30} and 3.44×10^{-30} , respectively. Since both values are smaller than 0.01, we rejected the null hypothesis that they are non-stationary. In addition, the autocorrelation function (ACF) plots in Figure 5 show the same results, and so we can conclude that there was no correlation between the returns and its lags.



(a) Autocorrelations of actual returns.

(b) Autocorrelations of simulated returns.

Figure 5. Autocorrelations of actual and simulated returns.

Table 7. ADF unit root test for real returns and simulated returns.

Data	ADF	<i>n</i> -Value	Critical Values			
	Statistic	<i>p</i>	1%	5%	10%	
Real Returns Simulation	-18.28 -17.73	$\begin{array}{c} 2.31 \times 10^{-30} \\ 3.44 \times 10^{-30} \end{array}$	$-3.44 \\ -3.44$	-2.86 -2.86	-2.57 -2.57	

5.3. Volatility Clustering

The next stylized fact we studied was volatility clustering, which describes the the tendency of large changes in prices to cluster together. The volatility clustering can be seen in Figure 6, where the ACFs of the absolute value of returns decay slowly. Notice that for both cases, the autocorrelations of absolute returns remained significantly positive over many time lags. This is clear evidence of volatility clustering.



(a) Autocorrelations of absolute actual returns.



Figure 6. Autocorrelation of absolute returns.

6. Discussion and Concluding Remarks

In this work we proposed a novel approach to calibrate hyperparameters in an agent-based model with different types of traders. The Bayesian calibration with Kolmogorov–Smirnov test proposed an optimal set of parameters. The model was able to highlight some of the important stylized facts of stock returns, such as the leptokurtosis, the heavy tail of the returns, and volatility clustering.

Our Bayesian approach is different from the recent studies on the calibration of ABMs applied to financial market [12,15,16]. The calibration considers large parameter sets which can become quite a complex and costly operation. In particular, in [15], the authors generated 400,000 initial points to

calibrate their model. In [12], the amount of data was reduced due to the high computational complexity of the Island model. Our approach is based on Bayesian rules and looks at previous experience to scale down the search space. Thus, the calibration requires less time and iterations to get the optimal hyperparameters. Moreover, the calibrated results showed that fundamental traders accounted for 9%–11% of all traders in the United States stock market. The proportion of traders in the stock market provides an additional indicator so that traders can come up with effective strategies.

Although Bayesian approaches provide several advantages, the complexity of the objective function could be a problem since it still has its own costs. In our model, the Sharpe ratio used as the objective function is very simple to calculate. However, if we want to consider a complex model with more types of investors and interactions between them, the dimension of domain space will increase. Optimal models require different settings between datasets, and with a high-dimensional space, the correlation between hyperparameters is difficult to study.

This work is only the first step toward fully assessing the properties of agent-based models in the United States stock market. In future research, the model in this research can be applied to other exchange such as the New York Stock Exchange or emerging markets in Asia. The different characteristics of the market and investors provide many interesting results for managers and investors. Moreover, deep learning is a good methodology for dealing with high-dimensional problems. Reinforcement learning can be considered to address Bayesian optimization issues [42], in which our Bayesian approach can be used as a benchmark method for the comparison.

Author Contributions: Conceptualization, M.T.; formal analysis, M.T.; investigation, M.T.; methodology, M.T.; supervision, D.P.H. and M.B.; validation, T.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Vietnam National University Ho Chi Minh City under grant number C2018-42-03.

Acknowledgments: We thank Man Ngo (John von Neumann Institute) for useful discussions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- 1. Gazit, E.; Yair, Y.; Chen, D. Emerging conceptual understanding of complex astronomical phenomena by using a virtual solar system. *J. Sci. Educ. Technol.* **2005**, *14*, 459–470. [CrossRef]
- Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabási, A.L. The large-scale organization of metabolic networks. *Nature* 2000, 407, 651. [CrossRef] [PubMed]
- 3. Faloutsos, M.; Faloutsos, P.; Faloutsos, C. On power-law relationships of the internet topology. *ACM SIGCOMM Comput. Commun. Rev.* **1999**, *29*, 251–262. [CrossRef]
- 4. Pagani, G.A.; Aiello, M. The power grid as a complex network: a survey. *Phys. A Stat. Mech. Appl.* **2013**, 392, 2688–2700. [CrossRef]
- Farmer, J.D.; Gallegati, M.; Hommes, C.; Kirman, A.; Ormerod, P.; Cincotti, S.; Sanchez, A.; Helbing, D. A complex systems approach to constructing better models for managing financial markets and the economy. *Eur. Phys. J. Spec. Top.* 2012, 214, 295–324. [CrossRef]
- 6. Anderson, P.W. The Economy as an Evolving Complex System; CRC Press: Boca Raton, FL, USA, 2018.
- 7. Arthur, W.B. The Economy as an Evolving Complex System II; CRC Press: Boca Raton, FL, USA, 2018.
- 8. Kirman, A.P. Whom or what does the representative individual represent? *J. Econ. Perspect.* **1992**, *6*, 117–136. [CrossRef]
- 9. Campbell, J.Y.; Cochrane, J.H. By force of habit: A consumption-based explanation of aggregate stock market behavior. *J. Polit. Econ.* **1999**, 107, 205–251. [CrossRef]
- Hommes, C.H. Heterogeneous agent models in economics and finance. *Handb. Comput. Econ.* 2006, 2, 1109–1186.
 [CrossRef]

- 11. Cristelli, M.; Pietronero, L.; Zaccaria, A. Critical overview of agent-based models for economics. *arXiv* 2011, arXiv:1101.1847.
- 12. Lamperti, F.; Roventini, A.; Sani, A. Agent-based model calibration using machine learning surrogates. *J. Econ. Dyn. Control* **2018**, *90*, 366–389. [CrossRef]
- 13. Hessary, Y.K.; Hadzikadic, M. Role of behavioral heterogeneity in aggregate financial market behavior: An agent-based approach. *Procedia Comput. Sci.* **2017**, *108*, 978–987. [CrossRef]
- 14. Fenton-O'Creevy, M.; Soane, E.; Nicholson, N.; Willman, P. Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders. *J. Organ. Behav.* **2011**, *32*, 1044–1061. [CrossRef]
- 15. Recchioni, M.C.; Tedeschi, G.; Gallegati, M. A calibration procedure for analyzing stock price dynamics in an agent-based framework. *J. Econ. Dyn. Control* **2015**, *60*, 1–25. [CrossRef]
- Akram, W.; Niazi, M.A.; Iantovics, L.B.; Vassilakos, A.V. Towards Agent-Based Model Specification of Smart Grid: A Cognitive Agent-Based Computing Approach. *Interdiscip. Descr. Complex Syst.: INDECS* 2019, 17, 546–585. [CrossRef]
- Cheng, E. Just 10% of Trading is Regular Stock Picking, JPMorgan Estimates. Available online: https://www.cnbc.com/2017/06/13/death-of-the-human-investor-just-10-percent-of-trading-is-regularstock-picking-jpmorgan-estimates.html (accessed on 2 February 2019).
- Garifullin, M.; Borshchev, A.; Popkov, T. Using AnyLogic and agent-based approach to model consumer market. In Proceedings of the 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia, 9–13 September 2007; pp. 1–5.
- 19. Challet, D.; Marsili, M.; Zhang, Y.C. *Minority Games: Interacting Agents in Financial Markets*; Oxford University Press: New York, NY, USA, 2005; pp. 27–62.
- 20. Cont, R. Volatility clustering in financial markets: Empirical facts and agent-based models. In *Long Memory in Economics*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 289–309.
- 21. Gode, D.K.; Sunder, S. Human and artificially intelligent traders in computer double auctions. In *Computational Organization Theory*; L. Erlbaum Associates Inc.: Hillsdale, NJ, USA, 1994; pp. 241–262.
- 22. Frankel, J.A.; Froot, K.A. Understanding the US dollar in the eighties: the expectations of chartists and fundamentalists. *Econ. Record* **1986**, *62*, 24–38.
- 23. Murphy, J.J. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications;* New York Institute of Finance: Paramus, NJ, USA, 1999; pp. 1–22.
- 24. Graham, B.; Dodd, D.L.F.; Cottle, S. Security Analysis; McGraw-Hill: New York, NY, USA, 1934.
- 25. Plott, C.R.; Sunder, S. Efficiency of experimental security markets with insider information: An application of rational-expectations models. *J. Polit. Econ.* **1982**, *90*, 663–698. [CrossRef]
- 26. Lucey, B.M.; Dowling, M. The role of feelings in investor decision-making. *J. Econ. Surveys* **2005**, *19*, 211–237. [CrossRef]
- 27. Kirman, A. Epidemics of opinion and speculative bubbles in financial markets. *Money Financ. Mark.* **1991**, 17, 354–368.
- 28. Westerhoff, F. A simple agent-based financial market model: direct interactions and comparisons of trading profits. In *Nonlinear Dynamics in Economics, Finance and Social Sciences;* Springer: Berlin/Heidelberg, Germany, 2010; pp. 313–332.
- 29. Tauchen, G.E.; Pitts, M. The price variability-volume relationship on speculative markets. *Econom. J. Econom. Soc.* **1983**, *51*, 485–505. [CrossRef]
- Snoek, J.; Larochelle, H.; Adams, R.P. Practical bayesian optimization of machine learning algorithms. In Advances in Neural Information Processing Systems 25, Proceedings of the 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 2951–2959.
- 31. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 2012, 13, 281–305.
- 32. Tumminello, M.; Lillo, F.; Piilo, J.; Mantegna, R.N. Identification of clusters of investors from their real trading activity in a financial market. *New J. Phys.* **2012**, *14*, 013041. [CrossRef]
- 33. Teall, J.L. Financial Trading and Investing; Academic Press: Cambridge, MA, USA, 2018.

- 34. Brown, D.P.; Jennings, R.H. On technical analysis. Rev. Financ. Stud. 1989, 2, 527–551. [CrossRef]
- 35. Pring, M.J. Technical Analysis Explained, 4th ed.; McGraw-Hill: NewYork, NY, USA, 2002.
- 36. Brock, W.; Lakonishok, J.; LeBaron, B. Simple technical trading rules and the stochastic properties of stock returns. *J. Finance* **1992**, *47*, 1731–1764. [CrossRef]
- 37. Schwager, J.D. A Complete Guide to the Futures Markets: Fundamental Analysis, Technical Analysis, Trading, Spreads, and Options; John Wiley & Sons: New York, NY, USA, 1984.
- 38. Martin, A.D. Technical trading rules in the spot foreign exchange markets of developing countries. *J. Multinatl. Financ. Manag.* 2001, *11*, 59–68. [CrossRef]
- 39. Walraven, J.C. Fundamental analysis of aggregate interlock. J. Struct. Division 1981, 107, 2245–2270.
- 40. Ferreira, F.F.; de Oliveira, V.M.; Crepaldi, A.F.; Campos, P.R. Agent-based model with heterogeneous fundamental prices. *Physica A: Stat. Mech. Its Appl.* **2005**, *357*, 534–542. [CrossRef]
- 41. Lux, T.; Alfarano, S. Financial power laws: Empirical evidence, models, and mechanisms. *Chaos Solitons Fractals* **2016**, *88*, 3–18. [CrossRef]
- 42. Brochu, E.; Cora, V.M.; De Freitas, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv* **2010**, arXiv:1012.2599.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).