Herding model

Zhuoyue WAN

1 Introduction

Herd behavior, in which many investors adopt the same trading strategy or have the same preferences for a particular asset over a certain period, is a good explanation of "bounded rationality" in the real world. This project sets out over 80 rules containing three types of investors and three types of market news to simulate herd behavior in the real-world stock market. Based on the comparison of stock price, returns time series and investor behavior chart per round, we find that Herd Behavior will begin to highlight in the period of stock price acute fluctuations. A direct comparison is made with the daily closures of the Hang Seng index. Our stock sequences exhibit properties like actual market data.

2 Modelling method:

In the present work we simulate the financial market dynamics via a stochastic cellular automata model. The agents of the market are represented by cells on a two-dimensional grid, 100 × 100. The *i*-th agent at the discrete time step t is characterized by three possible states or spin orientations, $\sigma_i(t)$ = 0, ±100. The value $\sigma_i(t)$ = +100 is associated with the purchase of a stock, while $\sigma_i(t)$ = −100 with selling. The former states are called active. The cells with spin value $\sigma_i(t)$ = 0 are inactive investors which means holding stock. The active investors herd in networks or clusters via a direct percolation method related to a *forest fir*e model [1]. The information carried by the active investors and market message, that is their spin state, is shared with the other investors. The percolation dynamics allow a time-dependent herding behavior, and the market can be interpreted as an open system not bounded by conservation laws [2]. We divide investors into three categories and message into three different categories. Meanwhile we set different information values and acceptance thresholds for different investors. Furthermore, we divide investor behavior into two types: follow and don't follow. Investors update our simulated markets by following the behavior of those with the greatest returns within the appropriate distance. The skeleton of model is shown in Session 6.1.

3 Analysis:

After modelling by the above method, we conducted 500 rounds of simulation, producing 500 days of stock price data in Session 6.2 and market behavior chart (See Fig 1). We take the market behavior charts of the ten days in which the stock price fluctuates the most in Session 6.3.

3.1 Do Markets Follow Random Walks?

The steady-state distribution of a random walk is a Gaussian (normal) distribution:

$$P(r) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{r^2}{2\sigma^2})$$

We compare simulated and real market returns to the standard normal distribution by KS-test in Session 6.4. The return distribution of simulated data and real data looks very different from the Gaussian distribution. More importantly, they all exhibit fat-tailed traits.

3.2 Fat Tails

While $F(r)$ is the cumulative distribution, large returns asymptotically follow a power law $F(r) \sim |r|^{-\alpha}$, with $\alpha > 2$. After judging that the data did not conform to a normal distribution, we examined their fat-tails in Session 6.5 (See Fig 4).

3.3 Power law behavior

Firstly, we divide the simulation return data into positive part and negative part. Then we check whether they obey this relationship:

$$P_{Levy}(|x|) \sim |x|^{-(1+\alpha)}$$

Both the simulated data and the real data we generated showed the power law behavior in Session 6.6 (See Fig 5&6). Then we use TP and TE test to check whether the tail distribution is Pareto-like (power law) or exponential (See Fig 7&8&9&10).

3.4 Volatility Clustering

Volatility is measured by the absolute value of the return. Return (See Fig 11) and volatility look very different. We first compare the autocorrelation function between simulated data and real data in Session 6.7 (See Fig 12), and detrended the fluctuation of the data through the DFA (Detrended Fluctuation Analysis [5]) method (See Fig 13). Then we analysis PSD (Power Spectral Density from the Power laws slide) (See Fig 14).

4 Result and Conclusion:

We observe that the market shows obvious herd behavior in most of these ten days. So, we draw a conclusion that Herd Behavior highlight in the period of stock price acute fluctuations. A direct comparison is made with the daily closures of the Hang Seng index. Our stock sequences exhibit properties such as Fat tails and Volatility clustering (power law behavior) like actual market data (HS index data). However, it is worth noting that the convergence of autocorrelation function generated by our model is different from that of real data. At the same time, the data generated by our model failed during PSD analysis.

5 References:

1. D. Stauffer, *Introduction to Percolation Theory*, (Taylor & Francis, London,1985).
2. Bartolozzi, M., & Thomas, A. W. (2004). Stochastic cellular automata model for stock market dynamics. *Physical review E*, *69*(4), 046112.
3. S.Sinha, A.Chatterjee, A.Chakraborti,B.K.Chakrabarti (2010) Econophysics : An Introduc-tion JohnWiley&Sons
4. A. Krawiecki *et al*., Phys. Rev. Lett.89, 158701 (2002); A. Krawiecki *et al*., Physica A 317,597 (2003).
5. Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley, Statistical properties of the volatility of price fluctuations, Phys. Rev. E 60, 1390-1400 (1999).

6 Appendix: Theory and numerical implementation

6.1 The skeleton of model

- The two-dimensional grid is 100×100. which means there are up to 10,000 investors (See the right figure in Fig 1).

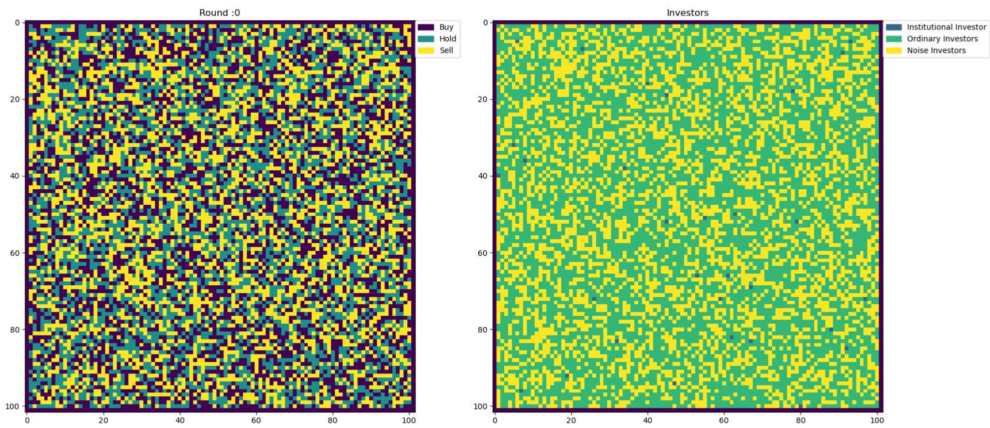
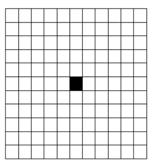
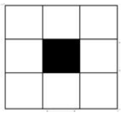


Fig 1 The two-dimensional 100×100 grid

- Investors are divided into 3 categories, namely, institutional investors (30), ordinary investors (6000), and noise investors (3970).
- Different range of "visibility":
  - Institutional investors: we set they can make use of those investors who five squares away from them:



  - Ordinary investors and noise investors: we set they can make use of those investors who one square away from them:



- There are 3 different categories of message in the market, obeying $N(0,1)$. If greater than 0.75, it is called "good news". If smaller than 0.25, it is called "bad news". Otherwise, it is called "No news".
- There are three possible states, $\sigma_i(t)$ = 0, ±100.
  - $\sigma_i(t)$ = 0, holding stock.
  - $\sigma_i(t)$ = +100, buying stock.
  - $\sigma_i(t)$ = -100, selling stock.
- Produce the stock price and return:
  - The initial price of the stock is p_0= 10
  - $p_{t+1} = [1 + \frac{k}{100^2}(n_{buy} - n_{sell})]\, p_t$, where $k \sim N(0,1)$, n_{buy} is the number of “buyers” and n_{sell} is the number of “sellers” at time t.
  - Return is $R_{t+1} = 100 \times ln\frac{p_{t+1}}{p_t}$ where R_0=0
- Update rule: Find the person with the highest return within sight range and follow their choices (Buy,Sell,Hold).

Based on three kinds of investors and three kinds of market message, there are up to 80 imitation rules. Herding behavior is the tendency of people involved in the market to aggregate in networks or clusters of influence. The investors then use the information obtained by their network in order to formulate a market strategy. Even if the topological structure of these networks of information is optional, since several kinds of long-range interaction are available nowadays [4], the number of connections for each trader must be, in any case, finite and not extended over the whole market. In the model, we set a different number of correlations depending on the type of investors because institutional investors are getting more market news than ordinary investors and noise investors. See simulation program for the specific implementation process.

6.2 Results of simulation

The stock price data generated by our model shows rapid increases and rapid decreases in most cases. This reflects from the side that our model has well simulated the irrational market behavior of investors in the real market, which is influenced by the herd behavior. It shows that our modeling method is reasonable and herd effect is considered in the stock pricing process.

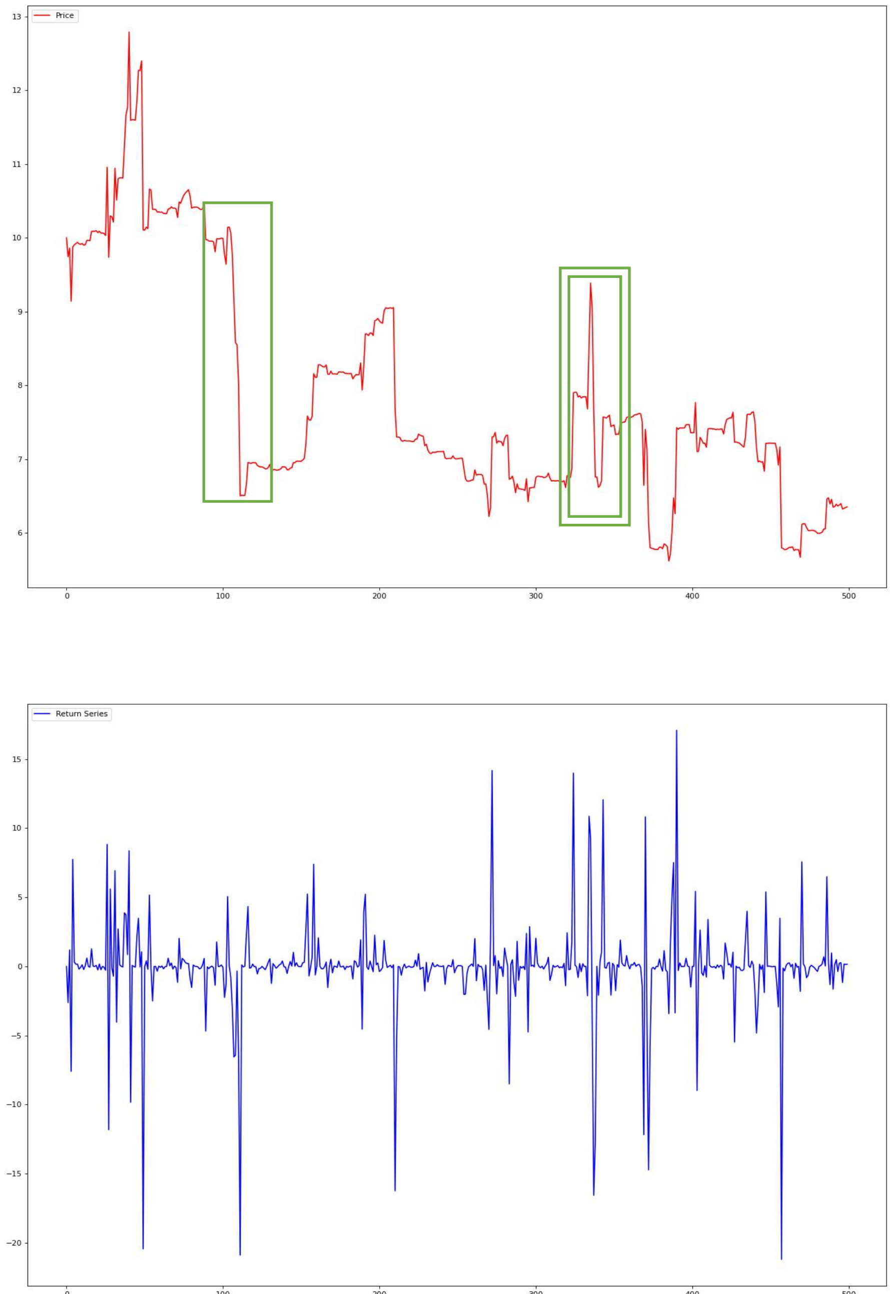
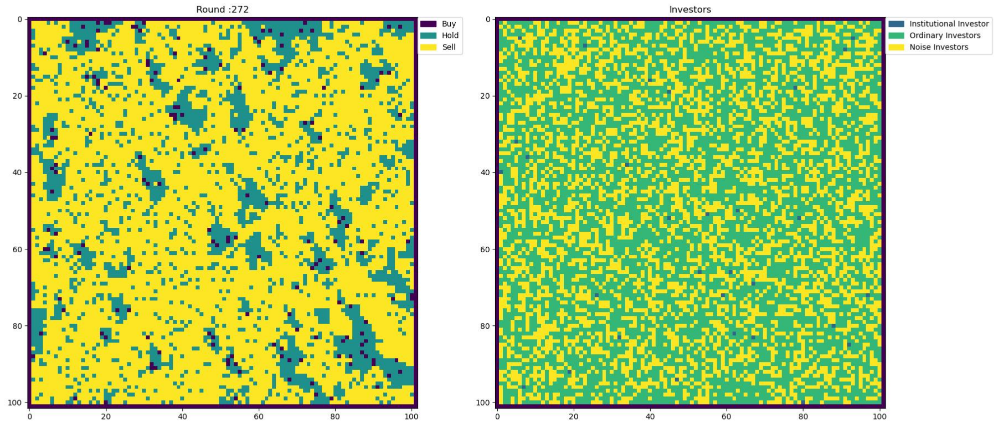
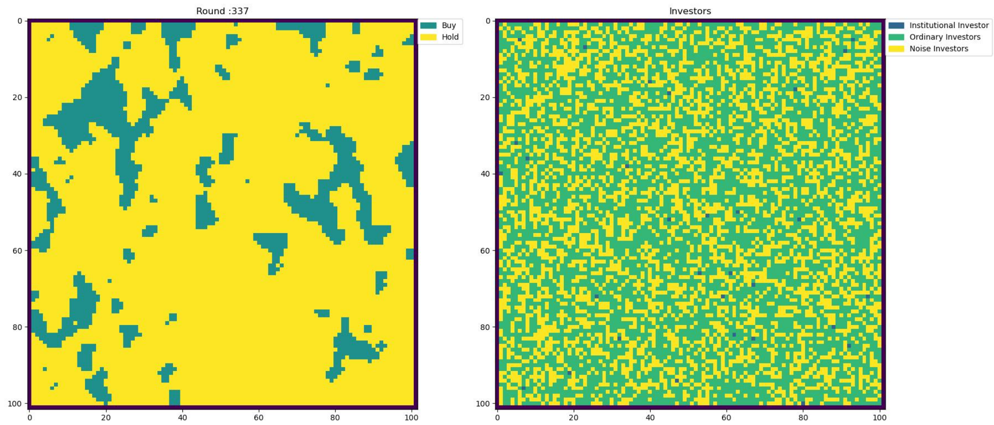


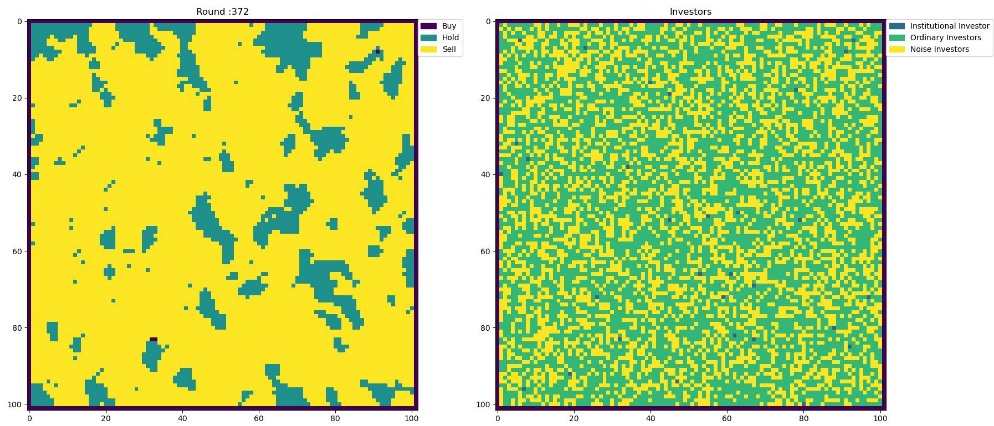
Fig 2 The simulation price and return

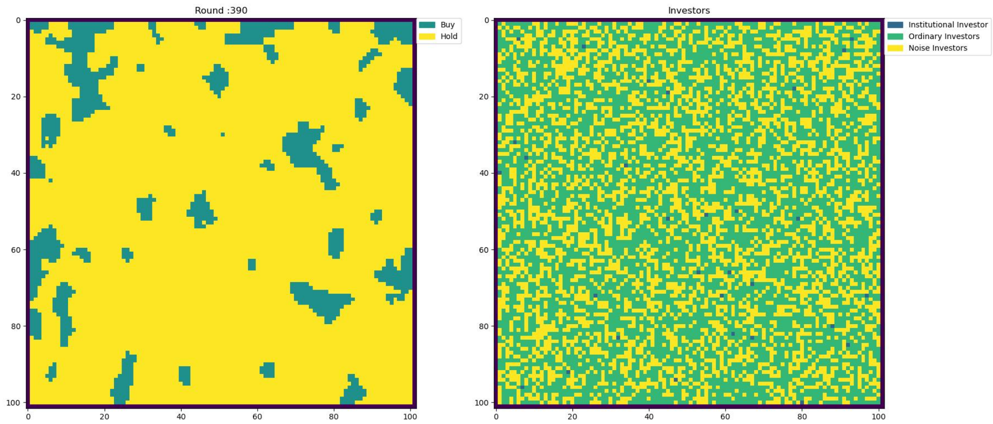
6.3 The ten days in which the stock price fluctuates the most

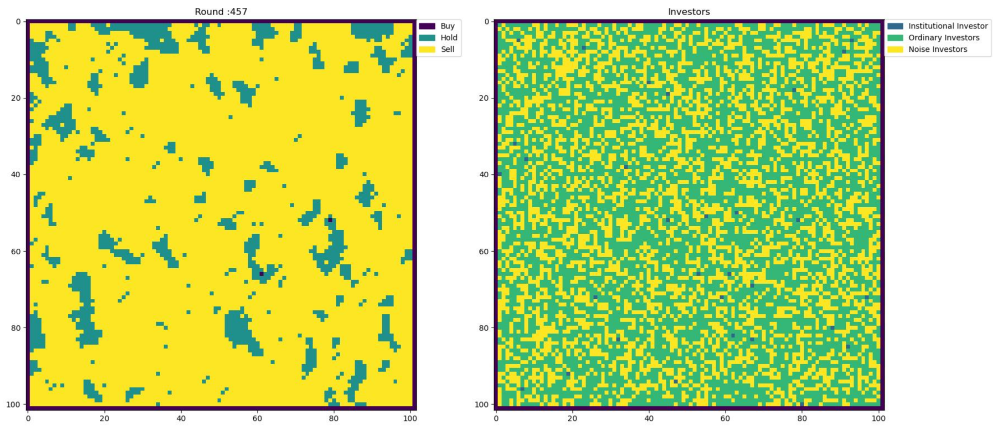
Seven figures show the herd behavior:

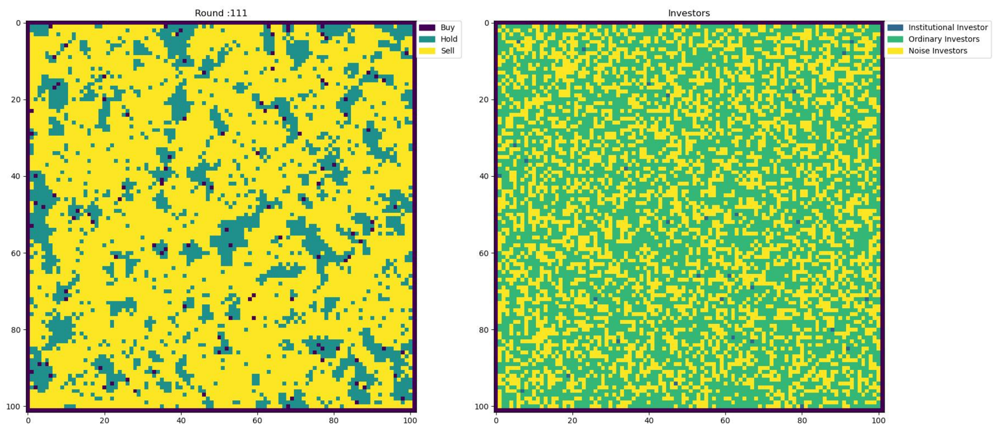


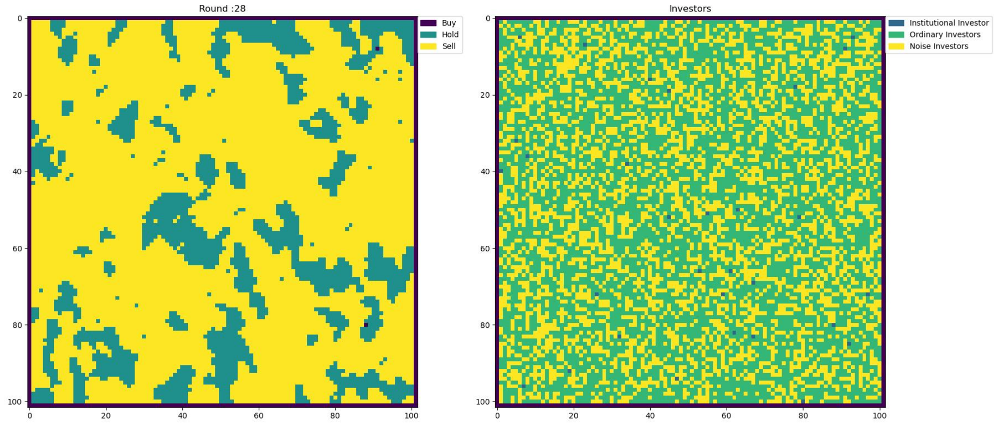




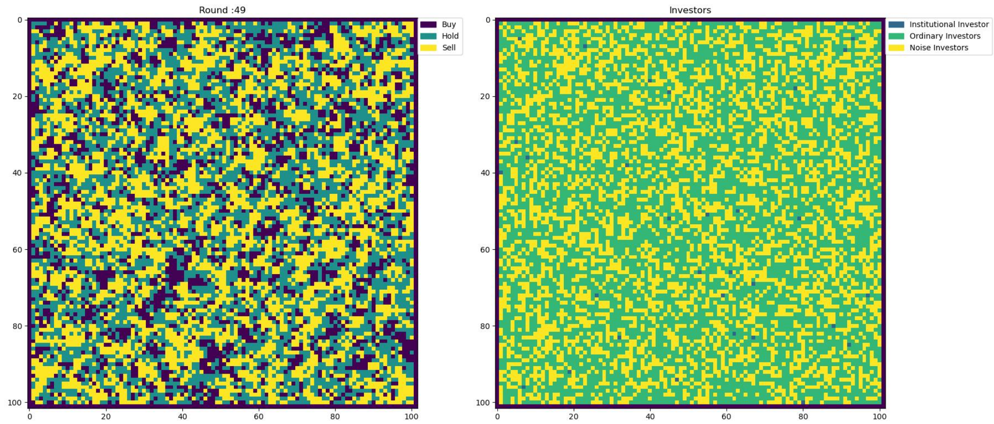

Round :390
Buy
Hold
Investors
Institutional Investor
Ordinary Investors
Noise Investors

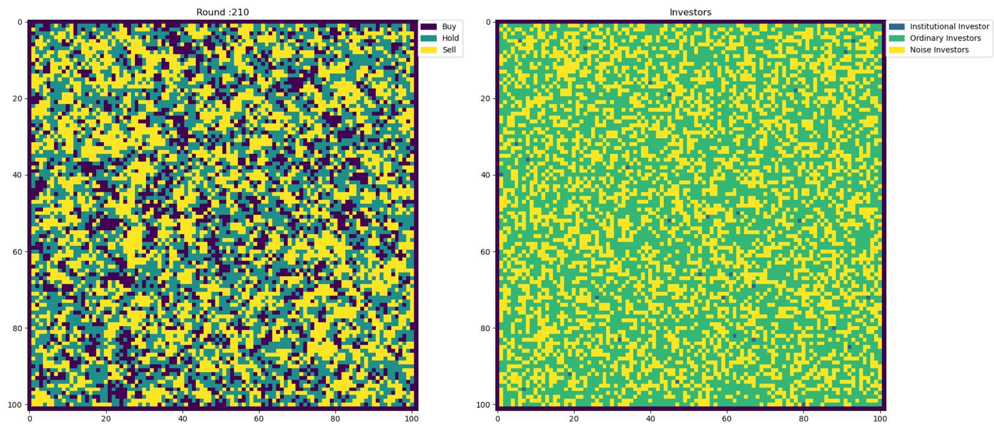

Round :457
Buy
Hold
Sell
Investors
Institutional Investor
Ordinary Investors
Noise Investors

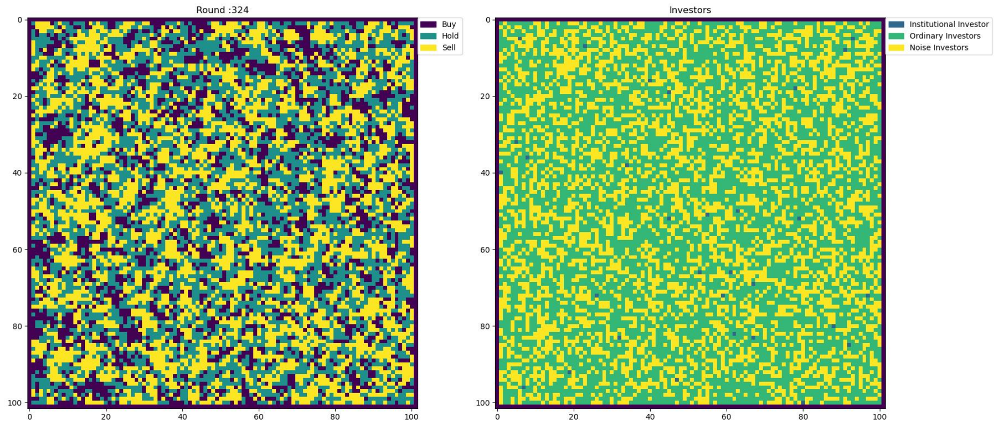

Round :111
Buy
Hold
Sell
Investors
Institutional Investor
Ordinary Investors
Noise Investors



Three figures do not show herd behavior:







6.4 Compared with normal distribution

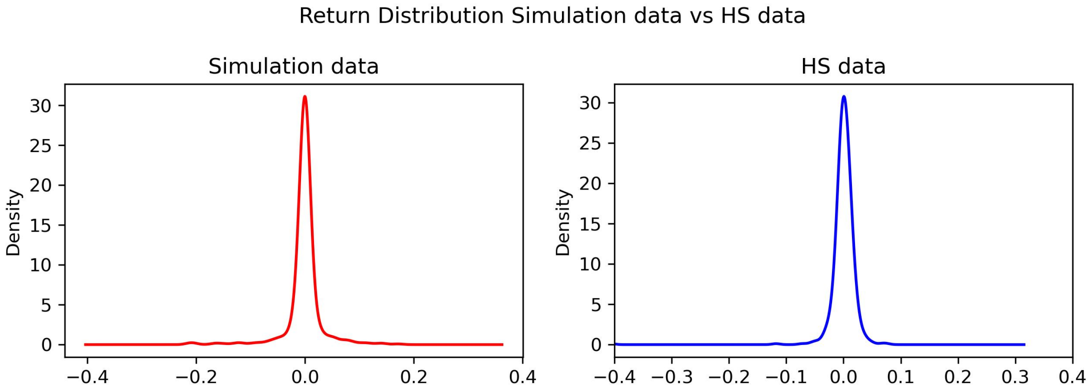


Fig 3 Return Distribution Simulation data vs HS data

Here, we use the Kolmogorov–Smirnov test which is a nonparametric test of the equality of continuous. The result as follows:

Dataset	Statistic value	P-value	Conclusion
Simulation data	0.281920386	1.1997606394346914e-35	No Gaussian
HS data	0.261837625	9.807279787167656e-31	No Gaussian

Since P-value is nearly zero, we draw a conclusion they reject null hypothesis which means there is no enough evidence to show they follow Gaussian distribution. Then we try to check the fat tail.

6.5 Fat tail test

We can see from the figure that the actual data and our simulated data have fat tail compared with the Gaussian Distribution.

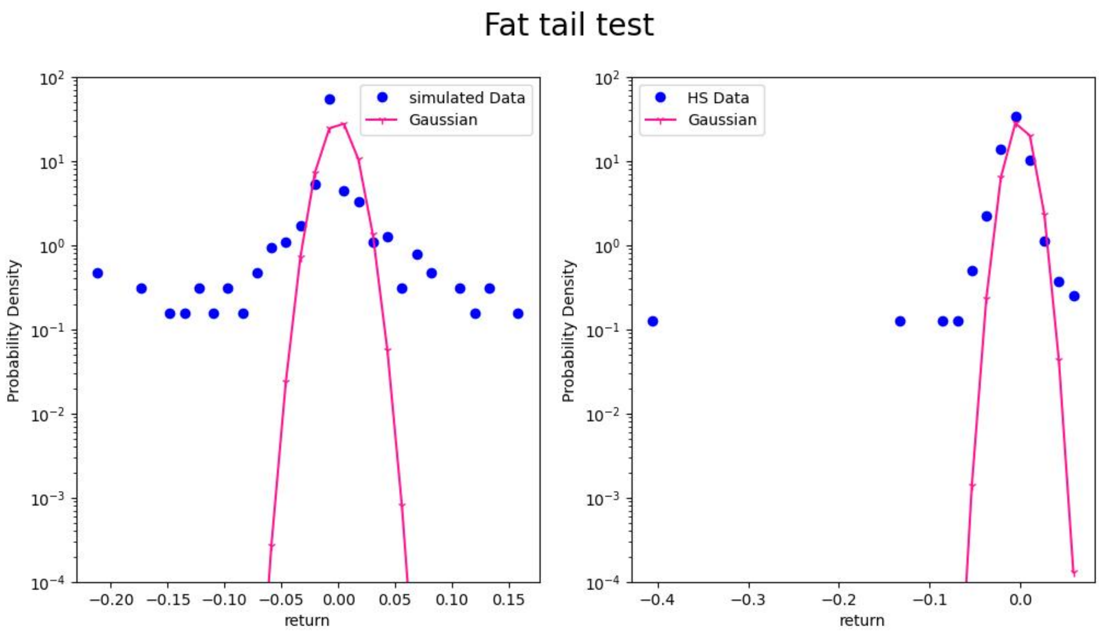


Fig 4 Fat tail test

6.6 Power law behavior

We use the method same as the HW6A, we get a great result in simulation and real market data as follows (More details in power law.xlsx):

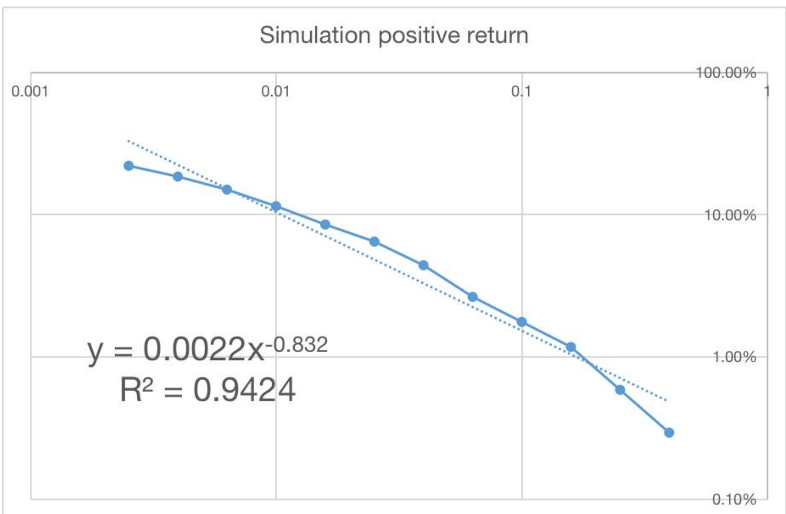


Fig 5 Positive simulation data

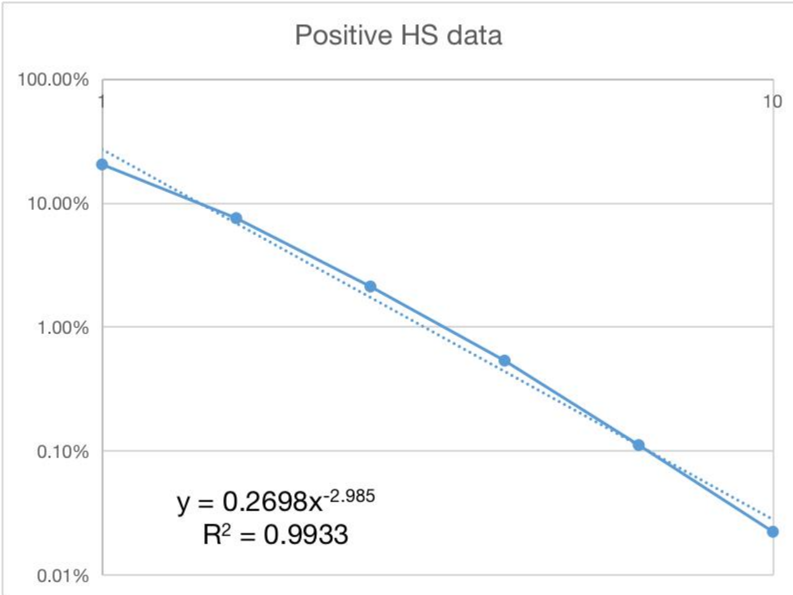


Fig 6 Positive HS data

Then we use TP [3] and TE[3] statistics method to check the return distribution.

TP test algorithm:

$$F_{TP}(X) = 1 - P_c(x) = 1 - (u/x)^{\alpha}; \quad for \;\; x \geq u$$

$$E_{1,TP} = E[log\frac{X}{u}]$$

$$E_{2,TP} = E[log^2\frac{X}{u}]$$

$$TP = [\frac{1}{n}\sum_{k=1}^{n} log(\frac{x_k}{u})]^2 - \frac{1}{2n}\sum_{k=1}^{n} log^2(\frac{x_k}{u})$$

$$SE_{TP} = (\frac{E_{2,TP}}{2} - 2E_{1,TP}^2) + \frac{1}{n}\sum_{k=1}^{n}(2E_{1,TP}log\frac{x_k}{u} - \frac{1}{2}log^2\frac{x_k}{u})$$

TE test algorithm:

$$F_{TE}(X) = 1 - P_c(x) = 1 - exp(-(u-x)/d); \quad for \;\; x \geq u$$

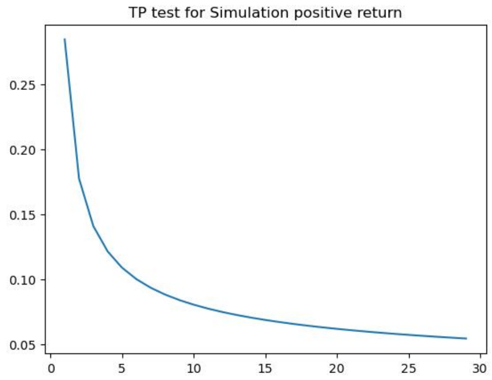
$$E_{1,TE} = E[log\frac{X}{u} - 1]$$

$$E_{2,TE} = E[log^2\frac{X}{u} - 1]$$

$$TE = \frac{1}{n}\sum_{k=1}^{n} log^2(\frac{x_k}{u} - 1) - (\frac{1}{n}\sum_{k=1}^{n} log(\frac{x_k}{u} - 1))^2 - \frac{\pi^2}{6}$$

$$SE_{TE} = \frac{1}{n}(\sum_{k=1}^{n} log((\frac{x_k}{u} - 1) - E_{1,TE})^2$$

The result as follows:



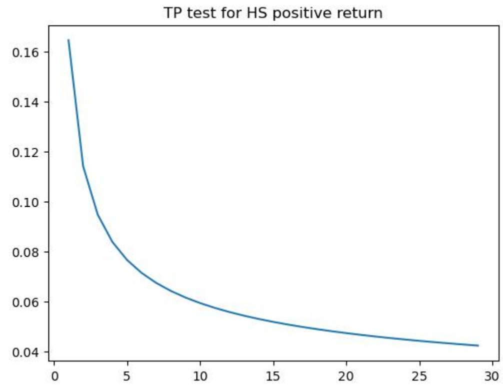


Fig 7 TP test for simulation data

Fig 8 TP test for HS data

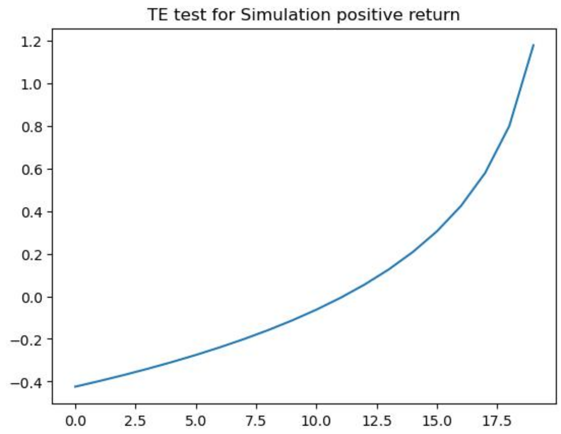


Fig 9 TE test for simulation data

Here we use the graph that 2021fall TSE, CHI HO got.

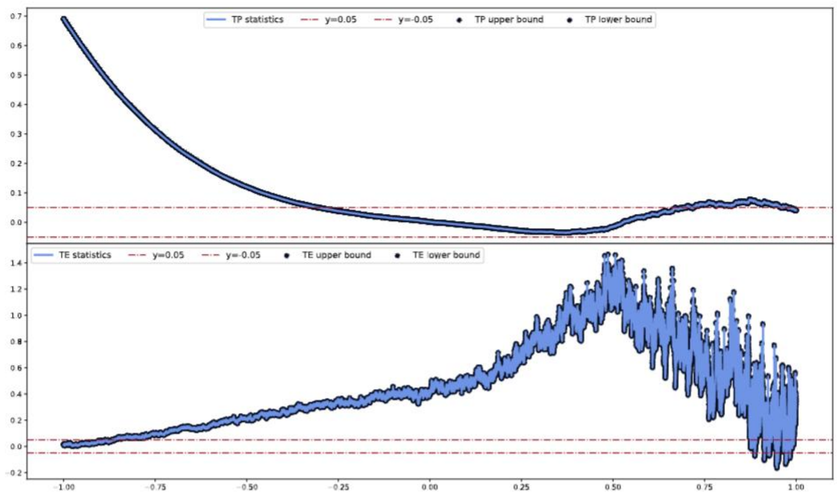


Fig 10 TSE, CHI HO showed in 2021 fall in HS data

TP test show simulation data and HS data convergence to power law. TE test show

Simulation data and HS data exponential law unlikely.

6.7 DFA and PSD

Autocorrelation:

Volatility Clustering as follows:

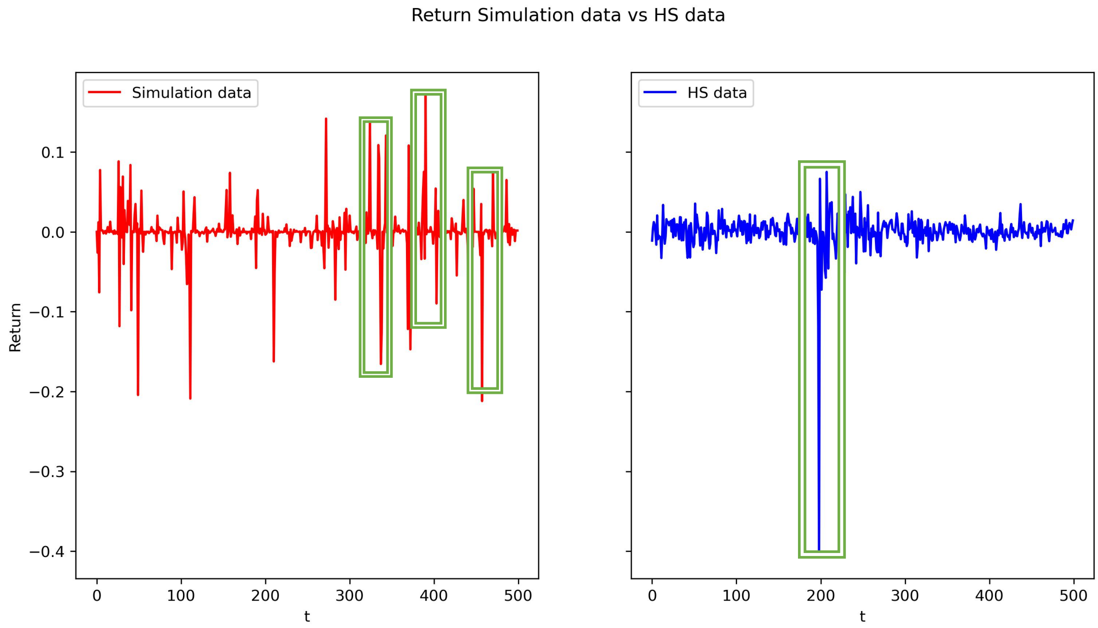


Fig 11 Return of simulation and HS data

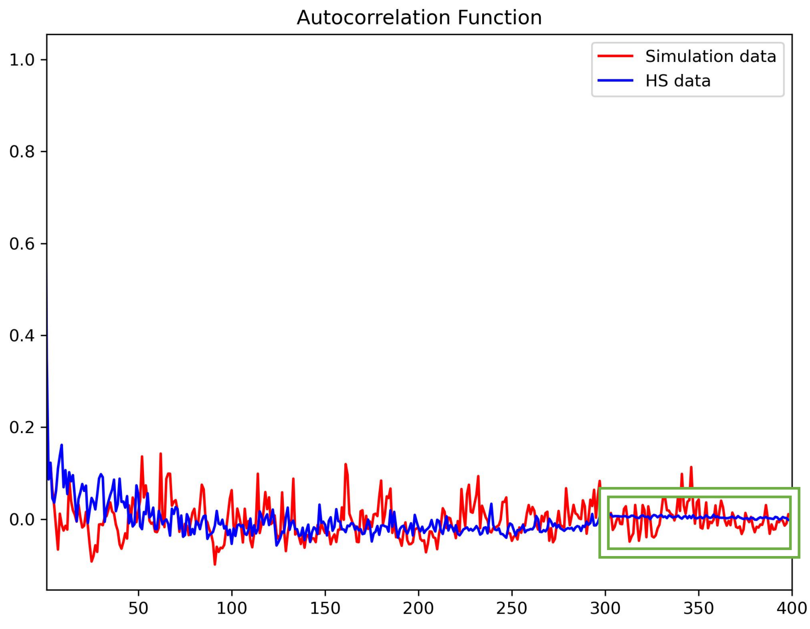


Fig 12 Autocorrelation function in simulation and real data

The autocorrelation function of the simulation data showed no convergence.

Detrended Fluctuation Analysis

Consider a time series ξ_i of length T.

Integrate the time series to an unbounded process (like a random walk)

$$X_t = \sum_{i=1}^{t} (\xi_i - \langle \xi \rangle)$$

Divide the length into N non-overlapping periods of length τ, such that $N\tau = T$.

In each period $i = 1, 2, \ldots, N$, fit the cumulative sum Xt with $z_{it} = a_{it} + b_i$, called the local trend. Detrend by subtracting the local trend to compute the fluctuation function:

$$F(\tau) = [\frac{1}{\tau} \sum_{t=(i-1)\tau+1}^{iT} (Xt - z\_it)^2]^{\frac{1}{2}}$$

Calculate the correlation exponent α from the power law relation:

$$F(\tau) \sim \tau^{\alpha}$$

If α=0.5, the signal is uncorrelated (white noise).

If α>0.5, there are persistent long-range correlations in the time series.

If α<0.5, anti-correlations are present. Large values are likely to be followed by small values and vice versa. The result as follows:

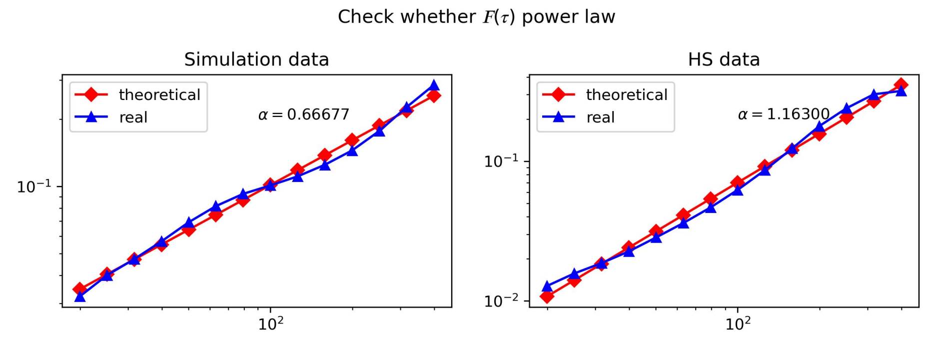


Fig 13 Check simulation and real data $F(\tau)$ power law

The α of simulation data and HS data greater than 0.5. There are persistent long-range correlations in the time series.

6.8 Power Spectral Density Analysis

The autocorrelation function $C(t)$ can be considered as the sum (or integral) of Fourier components $S(f)$ of frequency f. $S(f)$ is called the power spectrum. The low frequency components describes the long-time behavior, and the high frequency components describe the short-time behavior. It is calculated by

$$S(f) = \int_{-\infty}^{\infty} dt\, C(t) e^{-2\pi i f t}$$

Often in financial applications, only the amplitude of the power spectrum is studied.

$$|S(f)| = \sqrt{ReS(f)^2 + ImS(f)^2},$$

$$ReS(f) = \int_{-\infty}^{\infty} dt\, C(t) \cos(2\pi f t),$$

$$ImS(f) = -\int_{-\infty}^{\infty} dt\, C(t) \sin(2\pi f t)$$

The power spectral density of the absolute returns $|r_i|$ follows a power law

$$S(f) \sim \frac{1}{f^{\beta}}$$

The result as follows:

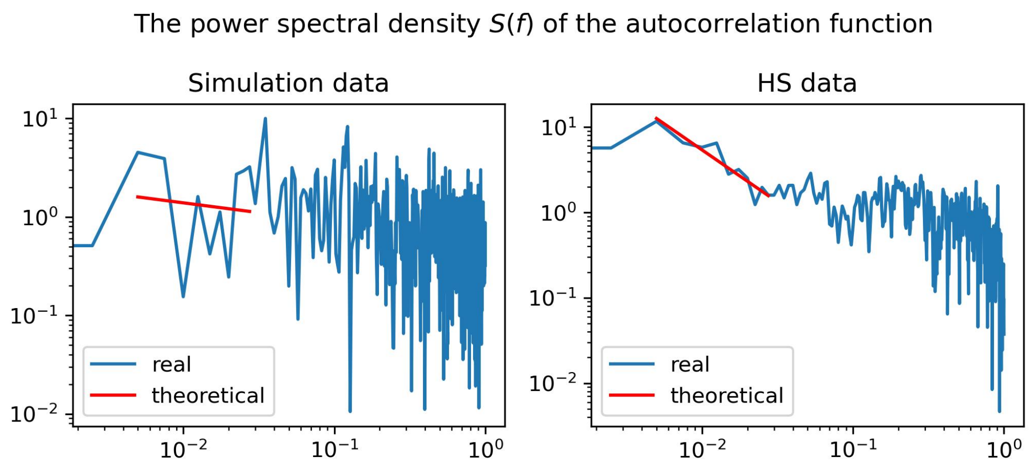


Fig 14 The power spectral density of the autocorrelation function

From the figure above, we can know that our model performs not well in the power spectral density, but the real market data is consistent with the result of DFA. There is no obvious linear trend in simulation data. But this does not mean that our model is not good in this aspect, because conclusions from graphical fitting may not be accurate. We need more reasonable and accurate methods.