

RESEARCH ARTICLE

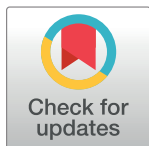
The prediction of fluctuation in the order-driven financial market

Fabin Shi[✉], Xiao-Qian Sun^{✉*}, Jinhua Gao[✉], Zidong Wang, Hua-Wei Shen, Xue-Qi Cheng

Data Intelligence System Research Center, Institute of computing technology, Chinese Academy of Sciences, Beijing, Haidian, China

[✉] These authors contributed equally to this work.

* sunxiaoqian@ict.ac.cn



OPEN ACCESS

Citation: Shi F, Sun X-Q, Gao J, Wang Z, Shen H-W, Cheng X-Q (2021) The prediction of fluctuation in the order-driven financial market. PLoS ONE 16(11): e0259598. <https://doi.org/10.1371/journal.pone.0259598>

Editor: Jie Zhang, Newcastle University, UNITED KINGDOM

Received: July 14, 2021

Accepted: October 21, 2021

Published: November 18, 2021

Copyright: © 2021 Shi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: <https://lightning.bitflyer.com/docs> <https://developers.coinbase.com/api/v2#introduction>.

Funding: This work was funded by the National Natural Science Foundation of China under grant number 91746301, 61802370. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Risk prediction is one of the important issues that draws much attention from academia and industry. And the fluctuation—absolute value of the change of price, is one of the indexes of risk. In this paper, we focus on the relationship between fluctuation and order volume. Based on the observation that the price would move when the volume of order changes, the prediction of price fluctuation can be converted into the prediction of order volume. Modelling the trader's behaviours—order placement and order cancellation, we propose an order-based fluctuation prediction model. And our model outperforms better than baseline in OKCoin and BTC-e datasets.

Introduction

Risk prediction is one of the important issues that draws much attention from academia and industry. It is the key element for financial decision making and plays an increasing role in the financial market. The fluctuation—the absolute value of the change of price, which is one of the indexes of risk. In this paper, we focus on the fluctuation prediction.

In finance, the traditional methods applied to predict the market include fundamental analysis and technical analysis. Fundamental analysis involves assessing a firm's equity value based on the analysis of related economic and financial factors [1]. Fundamental researchers study everything that can affect the security's value, from microeconomic factors like the effectiveness of the company's management to macroeconomic factors such as the state of the economy and industry conditions. Unlike fundamental analysis, technical analysis is an analysis methodology for forecasting the direction of prices by analyzing statistical trends gathered from trading activities, such as price movements and volumes. It focuses on patterns of price movements, trading signals, and various other analytical charting tools to evaluate a security's strength or weakness.

Besides, the financial time series models expressed by financial theories have provided classic forecasting methods. For the stationary time series, Yule [2] proposed the autoregressive model(AR) which assumes that the current state is linear to the previous state. Similarly, Walker [3] proposed the moving average model(MA). The difference is that MA concerns the

variance. The automatic regression moving average model (ARMA) is a mixed model combined with AR and MA, which combines the advantages of AR and MA. As for the non-stationary time series, the autoregressive conditional heteroscedasticity model (ARCH) adopted by Engle [4] is commonly used, which solves the problem brought by the hypothesis that the variance is constant. In ARCH, the unconditional variance of the model is a constant, but its conditional variance changes with time. Besides, there are many ways to improve [5–7], such as the generalized autoregressive conditional heteroscedasticity model (GARCH) [8], asymmetric power autoregressive conditional heteroscedasticity model (APGARCH) [9].

In recent years, with the development of artificial intelligence, many researchers have applied machine learning methods to the financial field [10–20], including genetic algorithm, convolutional neural network, and support vector machine. Hammad [21] found the artificial neural network has a strong predictive ability and high precision compared with econometric analysis methods. Besides, the researchers attempted to integrate artificial neural networks, moving average autoregressive models, maximum entropy models, particle swarm optimization algorithms, and genetic algorithms into a mixed model [22–26]. Hassan [27] combined the hidden Markov model, artificial neural networks, and genetic algorithms to predict the price of three stocks. Their results show that their methods have significantly improved over other models in accuracy.

However, these models ignore the relationship between order and price. There exist many studies that are concerned with this issue in theory [28–36]. Shi [37] found the change of price is inverse of the volume of best ask and best bid. Yura [38] found that the price change is a proportion of the change of order between ask orders and bid orders. All of these suggest that the order is key for the fluctuation prediction which can not be neglected. In this paper, we focus on the fluctuation prediction using order book data. We find that the volume of the order would change when the price has moved. Based on this observation, the prediction of price change can be converted into the prediction of order volume. Modelling the trader's behaviour—order placement and order cancellation, we conduct an order-based fluctuation prediction model. Our experimental results show that the proposed model outperforms other methods on OKCoin and BTC-e datasets.

Results

Data

We collect order book data from the Bitcoin transaction platform. Compared with the stock market or futures market, it is easy to collect order book data from the Bitcoin market. We use the API provided by the Bitcoin platforms to collect order book data. It includes

- OKCoin was the largest Bitcoin exchange platform in China which consisted of millions of users and billions of turnovers per day. We collected order book data at OKCoin from Nov. 3rd, 2016 to Jul. 28th, 2017.
- BTC-e is one of the largest Bitcoin trading platforms headquartered in Russia, whose data was collected from May 3rd, 2017 to Jul. 26th, 2017.

The data description is in Table 1.

Table 1. The basis information about the OKCoin and BTC-e dataset.

Platform	date	country	record
OKCoin	Nov. 3rd, 2016–Jul. 28th, 2017	China	3.8×10^6
BTC-e	May 3rd, 2017–Jul. 26th, 2017	Russia	3.7×10^6

<https://doi.org/10.1371/journal.pone.0259598.t001>

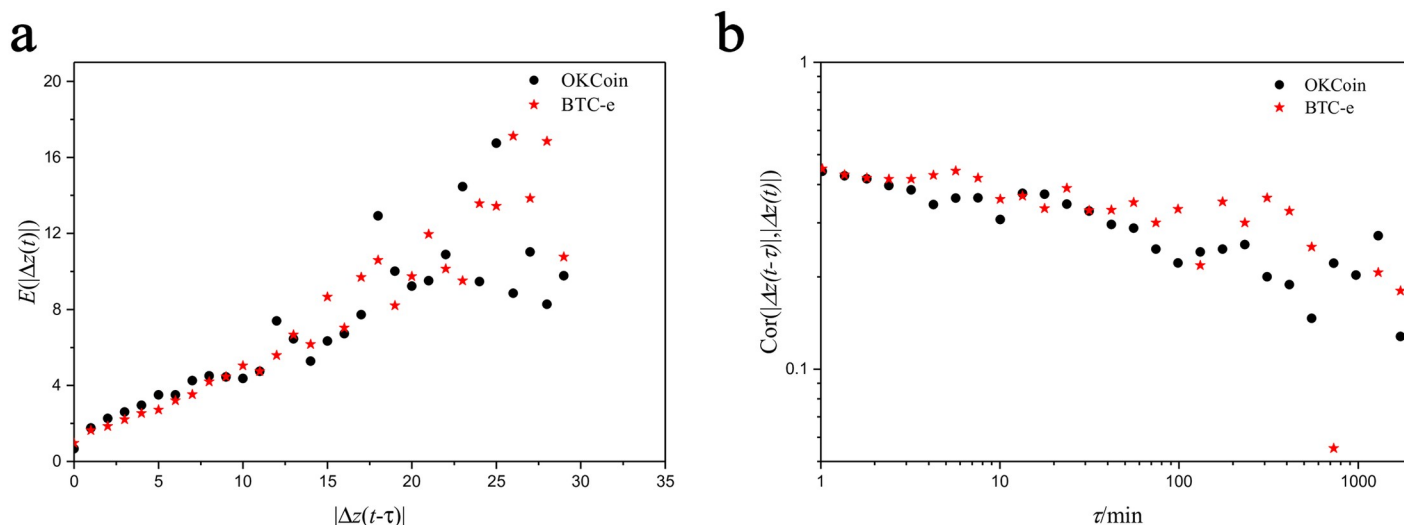


Fig 1. The predictability of the fluctuation in price in Bitcoin transaction platform. a. The relationship between the fluctuation of price in time $t - \tau$ and the fluctuation in time t . b. The autocorrelation of the price fluctuation $Cor(|\Delta z(t - \tau)|, |\Delta z(t)|)$ versus time interval τ .

<https://doi.org/10.1371/journal.pone.0259598.g001>

Price fluctuation

The most widely accepted models state that the variation of share prices is a random process. Investigating the time series of fluctuation on varying time scales $|\Delta t|$ is useful in probing the underlying nature of the stochastic process. The fluctuation $|\Delta z(t)|$ are defined as the difference between two successive price $X(t)$.

$$|\Delta z(t)| = |X(t + |\Delta t|) - X(t)| \quad (1)$$

The expected value of fluctuation $E(|\Delta z(t)|) = \sum_d P(|\Delta z(t)| = d)d$. Fig 1a shows the relationship between the expected value of fluctuation of price at time $t - \tau$ and the fluctuation at time t . In Fig 1a, we find that the expected value of fluctuation $E(|\Delta z(t)|)$ increases as the fluctuation $|\Delta z(t - \tau)|$ increases, where it is approximately linear. It suggests that if the fluctuation at time $t - \tau$ is large, the expected fluctuation would be large next time. From this aspect, the fluctuation of price has a strong correlation. Moreover, we measure the autocorrelation of the series of price fluctuation. The autocorrelation is calculated as

$$\begin{aligned} &Cor(|\Delta z(t - \tau)|, |\Delta z(t)|) \\ &= \frac{\langle [|\Delta z(t)| - \langle |\Delta z(t)| \rangle][|\Delta z(t - \tau)| - \langle |\Delta z(t - \tau)| \rangle] \rangle}{\sigma^2(|\Delta z(t)|)}, \end{aligned} \quad (2)$$

while $\langle |\Delta z(t)| \rangle$, $\sigma(|\Delta z(t)|)$ are the mean and the standard derivation of $|\Delta z(t)|$, respectively. In Fig 1b, we find there exists autocorrelation between the fluctuation of price when the time interval $\tau < 10min$. Besides, the autocorrelation decreases as time interval τ increases. When the time interval $\tau = 1$ min, the autocorrelation $Cor(|\Delta z(t - \tau)|, |\Delta z(t)|) \approx 0.5$; and when $\tau = 10$ min, the autocorrelation $Cor(|\Delta z(t - \tau)|, |\Delta z(t)|) \approx 0.2$.

Experiment

In this section, we evaluate the effectiveness of our model in two dataset and compare our model with several state-of-the-art methods for fluctuation prediction. we divide the entire

data into two parts. 80% of the data is used for in-sample training, which is used to determine the specifications of the model. And the other is considered as test data. For training, we use mean-square error to optimize our model.

Evaluation. To test the prediction of our model, we use the mean-square error and coefficient of determination to evaluate.

- MSE(mean-square error): The mean-square error is the average squared difference between the estimated values and the actual value. It is defined as

$$MSE = \sum_t (z(\hat{t}) - z(t))^2, \quad (3)$$

while $z(\hat{t})$, $z(t)$ are estimated values and the actual value, respectively.

- R^2 (coefficient of determination): The coefficient of determination is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_t (z(\hat{t}) - \bar{z})^2}{\sum_t (z(t) - \bar{z})^2}, \quad (4)$$

while $z(\hat{t})$, $z(t)$ are estimated values and the actual value, respectively. And \bar{z} is the average of the actual value.

Baseline. We compare our model with several commonly used methods, which are briefly described below.

- ARMA: ARMA(Autoregressive moving average) is one of the common methods for financial time series, which is a mixture of autoregressive model and moving average model. In this paper, we use the change of price series as input data, and apply the ARMA method for fluctuation prediction. The autoregressive polynomial degree and moving average polynomial degree are set into 10.
- SVR: The SVR(support vector regression) method is widely used for the stock prediction in the previous work [39–41]. In this paper, we use the change of price series as input data and apply the SVR method for fluctuation prediction.
- ANN: ANN(Artificial Neural Network) is one of the commonly used machine learning methods. There is much previous work that has been reported to used ANN to predict stock price [42, 43]. In this paper, we use the change of price series as input data and apply the ANN method for fluctuation prediction. The ANN model consists of three layers with 20 nodes. And two Relu are used in the first two layers.
- CNN: CNN(Convolutional Neural Network) is a class of deep neural networks, which is good at capturing the local features. Persio [44] used CNN to predict stock price. In this paper, we use the change of price series as input data and apply the CNN method for fluctuation prediction. The convolutional neural network consists of two convolutional layers. The parameters of these two convolutional layers are (kernel = 1×3 , filters = 4, stride = 1, active function = ReLU) and (kernel = 1×3 , filters = 8, stride = 1, active function = ReLU). And two max-pooling layers are used after convolutional layers respectively. The bottom of the network is built by two fully connected layers.

Experiment result. Table 2 shows the experiment result in OKCoin dataset. The MSE material is 0.761, 0.780, 0.749, 0.783 for four baseline methods when time interval $\Delta t = 10s$.

Table 2. The experiment result in OKCoin dataset.

		ANN	CNN	ARMA	SVR	Our model
$\tau = 10s$	MSE	0.761	0.780	0.749	0.783	0.712
	R^2	0.164	0.143	0.174	0.140	0.219
$\tau = 20s$	MSE	1.625	1.652	1.577	1.657	1.531
	R^2	0.147	0.132	0.172	0.130	0.196
$\tau = 40s$	MSE	3.622	3.674	3.589	3.667	3.398
	R^2	0.0981	0.0852	0.106	0.0871	0.154
$\tau = 80s$	MSE	7.437	7.257	6.944	7.353	6.887
	R^2	0.0397	0.0630	0.103	0.0506	0.111

<https://doi.org/10.1371/journal.pone.0259598.t002>

SVR performs the worst in MSE material. It is because the complexity of SVR is much higher than others when the train dataset is larger. Thus, the experiment result of SVR does not work well when the train time is limited. ANN and CNN perform better than SVR. And the performance of ARMA is the best in four baseline methods. However, the MSE metric in our model is 0.712. It is approximately 5% improvement compared with ARMA, which is the best method in the baseline. Similarly, our method outperforms when $\Delta t = 20s, 40s, 80s$. Table 3 shows the experiment results in the BTC-e dataset. And we find a similar experiment results in the BTC-e dataset except the experiment result when $\tau = 80s$. In summary, our method performs better than the baseline for MSE material.

As for R^2 , it is found that our method also outperforms all other baselines. When $\tau = 10s$, the R^2 in our method is 0.219, while it is 0.174 for ARMA. The improvement is more than 25%, which is significant. It is near 10% improvement when $\tau = 80s$.

On the other hand, we find that the experiment results on MSE increase as the time interval τ increase. Also, the experiment results on R^2 decrease when time interval τ increases. It is because that the variance increases when time interval τ increases. Fig 2 shows the relationship between the variance of the change of price and time interval τ . The variance of the change of price is linear to the time interval τ . Thus, the predictability of the fluctuation decrease when time interval τ increase.

Method

Fluctuation prediction model

In this section, we will introduce our model, including the motivation, the framework of our model, the influence of order, the influence of the fluctuation at last moment and parameters update.

Table 3. The experiment result in BTC-e dataset.

		ANN	CNN	ARMA	SVR	Our model
$\tau = 10s$	MSE	1.461	1.451	1.469	1.515	1.413
	R^2	0.0706	0.0770	0.0655	0.0368	0.101
$\tau = 20s$	MSE	2.201	2.181	2.176	2.331	2.143
	R^2	0.0806	0.0890	0.0911	0.0265	0.105
$\tau = 40s$	MSE	3.429	3.468	3.360	3.518	3.342
	R^2	0.0871	0.0767	0.105	0.0634	0.110
$\tau = 80s$	MSE	6.197	7.257	6.012	6.269	6.238
	R^2	0.101	0.0630	0.127	0.0904	0.0948

<https://doi.org/10.1371/journal.pone.0259598.t003>

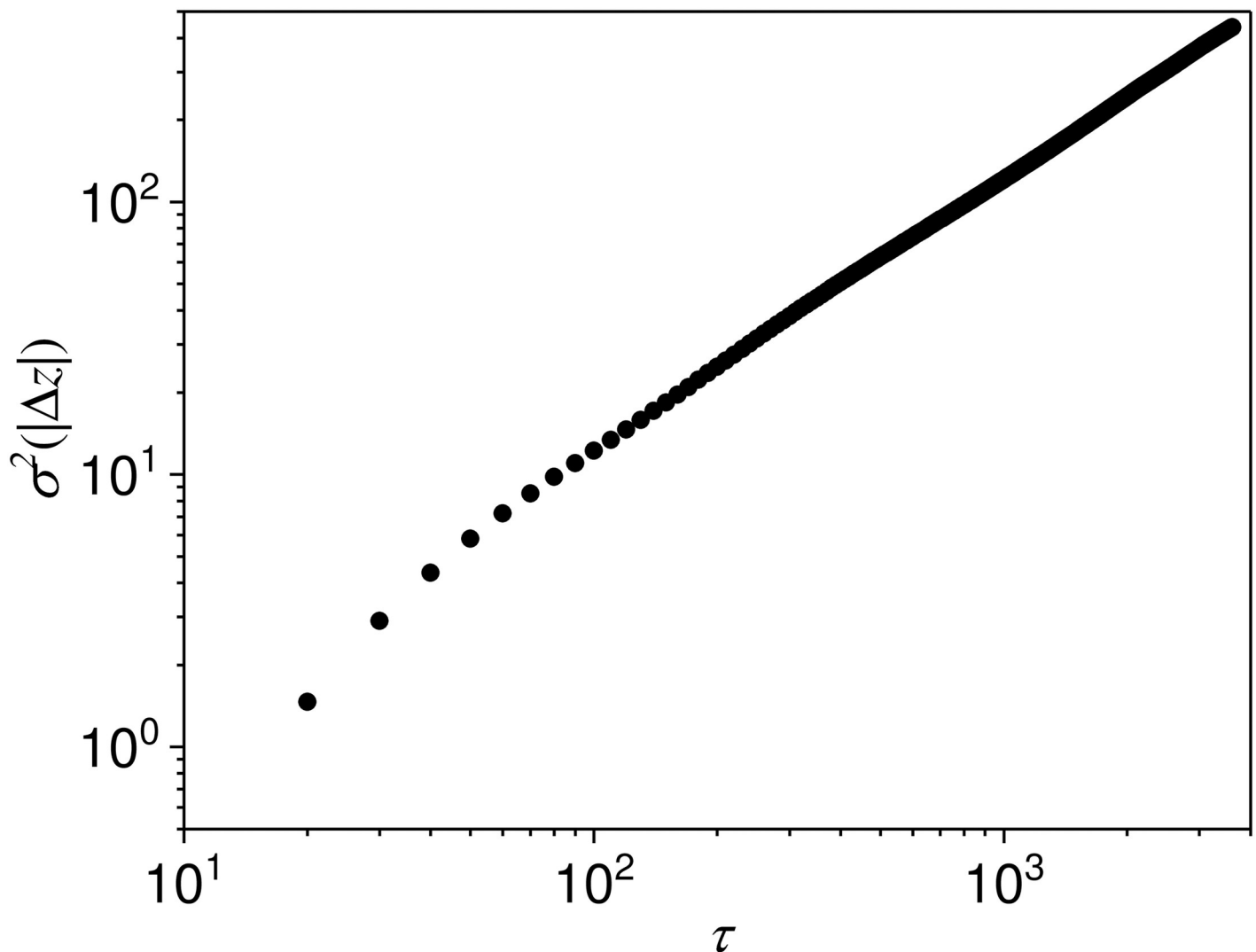


Fig 2. The variance of the change of price versus time interval τ .

<https://doi.org/10.1371/journal.pone.0259598.g002>

Basic definition

There are some basis terms which need to be introduced.

- **Limit Order.** Limit order definition: <https://www.investopedia.com/terms/l/limitorder.asp>. A limit order is a type of order to purchase or sell a security at a specified price or better. For buying limit orders, the order will be executed only at the limit price or a lower one, while for selling limit orders, the order will be executed only at the limit price or a higher one.
Note: the order mentioned in this paper is the limit order.
- **Order book.** Order book definition: <https://www.investopedia.com/terms/o/order-book.asp>. An order book is an electronic list of buying and selling orders for a specific security or financial instrument organized at a price level. An order book lists the number of shares being bid or offered at each price point.

Fig 3 demonstrates typical order book profiles at time $t - \tau$ and time t . The best ask $a(t)$ (or best bid $b(t)$) is defined to be the lowest ask price (or the highest bid price) at time t . The

midprice $z(t)$ is calculated as $z(t) = \frac{a(t)+b(t)}{2}$. And the spread $s(t) = a(t) - b(t)$, which represents the gap between best ask and best bid.

The blue orders and red orders represent bid order and ask order, respectively. And $n_x^a(t)$ and $n_x^b(t)$ represent the volume of ask order and bid order at time t , respectively. The relative position $x = \lceil \frac{|z_x - z(t)|}{\bar{z}} \rceil$ is the distances from price z_x to the midprice $z(t)$, where z_x is the price of order and \bar{z} is the price unit. Thus, the value $n_x^a(t)$ is the total volume of ask order from $z(t) + x * \bar{z}$ to $z(t) + (x + 1) * \bar{z}$. The value $n_x^b(t)$ is the total volume of ask order from $z(t) - x * \bar{z}$ to $z(t) - (x + 1) * \bar{z}$.

Motivation

Fig 4 shows the relationship between the fluctuation $|\Delta z(t)|$ at t and the fluctuation $|\Delta z(t - \tau)|$ at $t - \tau$. The red line is the expected value of prediction fluctuation $|\Delta z(t)|$ without considering the the volume of order. And the blue line and black line are the expected value of prediction fluctuation when $n_0 < 250$ and $n_0 > 800$ (n_0 is the sum of order volume at best bid and best ask). The expected value of prediction fluctuation $E(|\Delta z(t)|)$ is obviously different when the fluctuation at last moment $|\Delta z(t - \tau)|$ is the same. When $n_0 < 250$, the fluctuation prediction is bigger than that $n_0 > 800$. The redline is situated between the two. For example, when $\Delta z(t - \tau) = 3$, the expected value of fluctuation at next time $E(|\Delta z(t)|) = 1.3$ for $n_0 > 800$ (C point) and $E(|\Delta z(t)|) = 2.4$ for $n_0 < 250$ (A point). It seems that the smaller the order volume n_0 is, the larger the fluctuation is. The fluctuation is related to the order volume n_0 .

Moreover, we consider the prediction value and the fluctuation $|\Delta z(t - \tau)|$. When the order information is given, the prediction value of fluctuation is

$$f_1 = \sum_d P(|\Delta z'(t)| = d | \Delta z(t - \tau), n_0) d. \quad (5)$$

Only considering the price information, the prediction value is

$$\begin{aligned} f_2 &= \sum_d \sum_{n_0} P(|\Delta z'(t)| = d | \Delta z(t - \tau), n_0) d \\ &= \sum_{n_0} \sum_d P(|\Delta z'(t)| = d | \Delta z(t - \tau), n_0) d \end{aligned} \quad (6)$$

Thus, this prediction value f_1 is the average value of f_2 for different order volume. Obviously, the prediction value f_2 would be more precision when the order information is given. Thus, the order information is conducive to fluctuation prediction.

To summarize, the order information is important and helpful for fluctuation prediction.

Framework of model

In our model, the prediction of fluctuation aims to minimum the average squared difference between the prediction value and the actual value. Considering the order information, the loss function of fluctuation prediction L can be formation as:

$$\begin{aligned} L &= \min \sum_t [|\Delta z'(t)| - |\Delta z(t)|]^2 \\ &= \min \sum_t \left[\sum_d^{D_{MAX}} P(|\Delta z'(t)| = d | H(t - \tau)) d - |\Delta z(t)| \right]^2, \end{aligned} \quad (7)$$

while $|\Delta z'(t)|$ is the value that our model predicts. The history information $H(t - \tau) = \{\Delta z(t -$

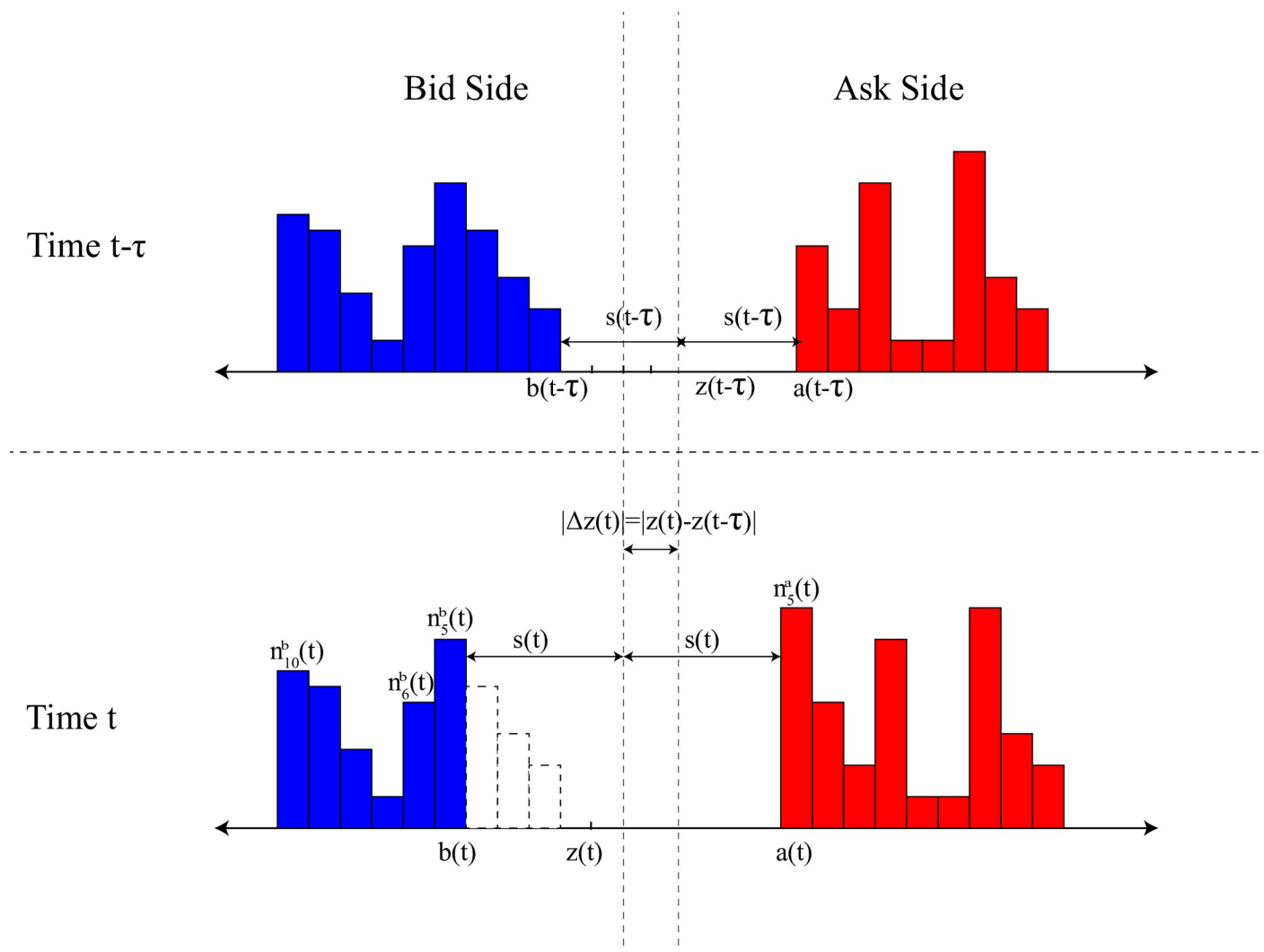


Fig 3. The change in order book from time $t - \tau$ to time t . The blue orders and red orders represent bid orders and ask orders, respectively.

<https://doi.org/10.1371/journal.pone.0259598.g003>

τ), $n(t - \tau)$ }, which represents the history data of order book and price. The value D_{MAX} is the maximum value of price fluctuation, which is hyperparameters in our model. From Eq (7), it is significantly that the key for this problem is to calculate $P(|\Delta z'(t)| = d|H(t - \tau))$.

The value $P(|\Delta z'(t)| = d|H(t - \tau))$ is the probability of the change of price, which can be converted to the probability of the change of order. From the price perspective, the price changes from $z(t - \tau) = \frac{a(t - \tau) + b(t - \tau)}{2}$ to $z(t) = \frac{a(t) + b(t)}{2}$ (Fig 3). And the best bid(ask) price changes from $b(t - \tau)(a(t - \tau))$ to $b(t)(a(t))$. From the order perspective, the order with the price $z_x \in [b(t - \tau), b(t)]$ disappear when the best bid change from $b(t - \tau)$ to $b(t)$. Thus, the change in price is actually equivalent to the change in the order volume.

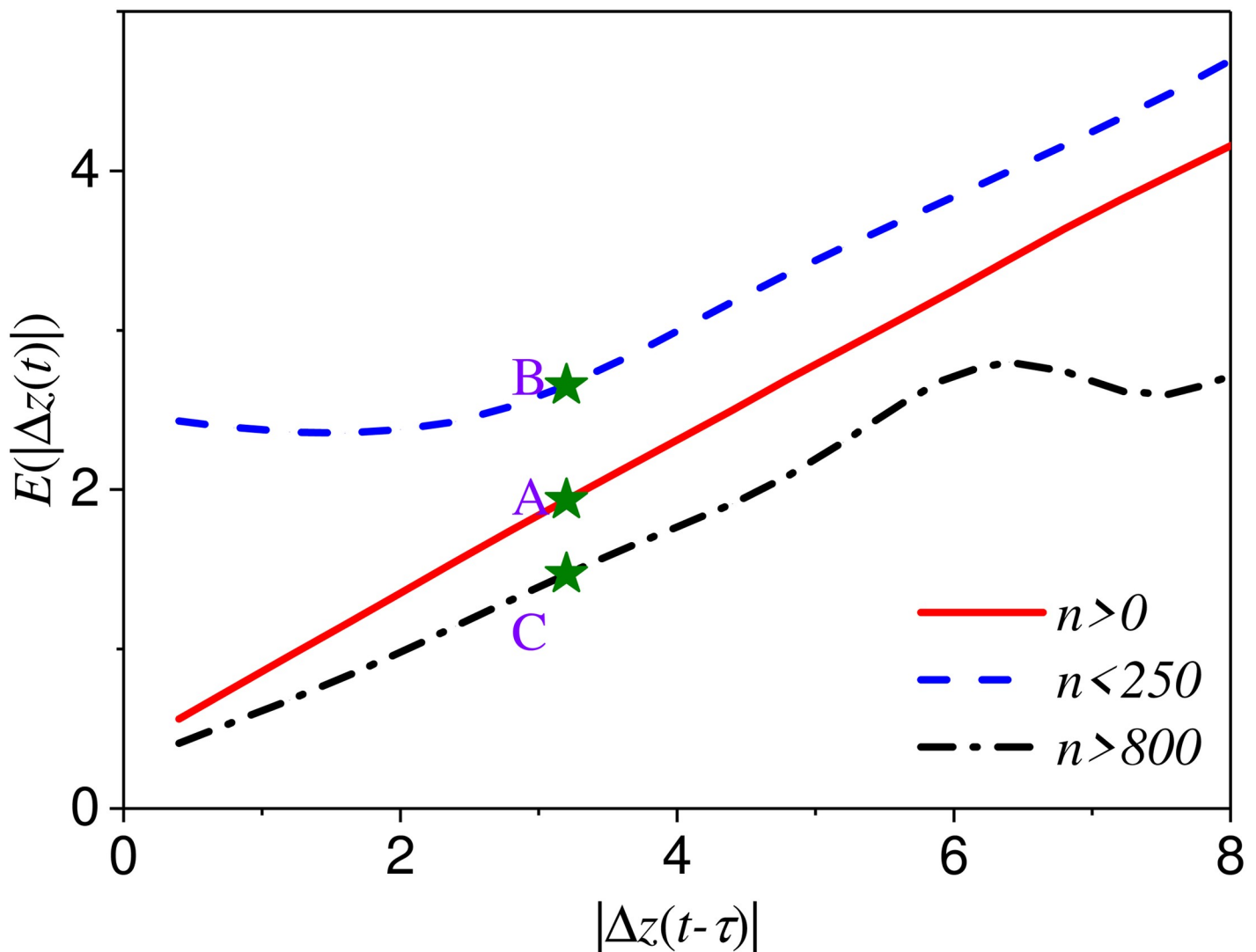


Fig 4. The relationship between the fluctuation of price in time $t - \tau$ and the fluctuation in time t for different order volumes.

<https://doi.org/10.1371/journal.pone.0259598.g004>

Based on this observation, the prediction of fluctuation can be converted into the prediction of the order volume. The conditional probability $P(|\Delta z'(t)| = d | H(t - \tau))$ is calculated as

$$\begin{aligned}
 & P(|\Delta z'(t)| = d | H(t - \tau)) \\
 &= P(|a(t) - a(t - \tau) + b(t) - b(t - \tau)| = 2d | H(t - \tau)) \\
 &= \sum_{|i+j|=2d} P(a(t) - a(t - \tau) = i, b(t) - b(t - \tau) = j | H(t - \tau)) \\
 &= \sum_{|i+j|=2d} P(n_1^a(t) = 0, \dots, n_{s(t-\tau)+i-1}^a(t) = 0, n_{s(t-\tau)+i}^a(t) > 0, \\
 &\quad n_1^b(t) = 0, \dots, n_{s(t-\tau)+j-1}^b(t) = 0, n_{s(t-\tau)+j}^b(t) > 0 | H(t - \tau))
 \end{aligned} \tag{8}$$

To simplify our model, we assume that the changes of order at different positions are independent. Thus,

$$\begin{aligned}
 P(|\Delta z'(t)| = d|H(t - \tau)) \\
 &= \sum_{|i+j|=2d} \prod_{k=0}^{s(t-\tau)+i-1} P(n_k^a(t) = 0|H(t - \tau))P(n_{s(t-\tau)+i}^a(t) > 0|H(t - \tau)) \\
 &\quad \prod_{k=0}^{s(t-\tau)+j-1} P(n_k^b(t) = 0|H(t - \tau))P(n_{s(t-\tau)+j}^b(t) > 0|H(t - \tau))
 \end{aligned} \quad (9)$$

Thus, how to predict the volume of order at next time is key for the prediction of fluctuation.

Influence of orders. The change of an order volume is resulted from two operation: order placement and order cancellation.

- **Order placement.** The order placement at time t denotes as $\Delta n_i^{OP}(t)$. The previous study shows that the order placement is related the order price.
- **Order cancellation.** The order cancelled at time t denotes as $\Delta n_i^{OC}(t)$. We assume that every unit order is cancelled with the same probability q_i . Based on this assumption, the process of cancellation is similar to the box picking problem: there are $\frac{n_i}{c}$ ball in the box at time $t - \tau$, while c is the minimum unit of order. And the balls are picked with the same probability q_i independently. Therefore the probability $P(\Delta n_i^{OC}(t - \tau) = m|H(t - \tau))$ can be calculated as

$$\begin{aligned}
 &P(\Delta n_i^{OC}(t - \tau) = m|H(t - \tau)) \\
 &= C\left(\frac{n_i(t - \tau)}{c}, \frac{n_i(t - \tau) - m}{c}\right)(1 - q_i)^{\frac{n_i(t - \tau) - m}{c}} q_i^m \\
 &\approx \frac{\left[n_i(t - \tau)\left(1 - \frac{q_i}{c}\right)\right]^{\frac{n_i(t - \tau) - m}{c}}}{\left(\frac{n_i(t - \tau) - m}{c}\right)!} e^{-n_i(t - \tau)\left(1 - \frac{q_i}{c}\right)}
 \end{aligned} \quad (10)$$

The approximation in Eq (10) is because $\lim_{n \rightarrow \infty, p \rightarrow 0} C_n^k p^k (1 - p)^{n-k} \approx \frac{(np)^k}{k!} e^{-np}$. When $m = n_i(t - \tau)$, the conditional probability $P(\Delta n_i^{OC}(t - \tau) = m|H(t - \tau)) \approx e^{-n_i(t - \tau)\left(1 - \frac{q_i}{c}\right)}$.

Based on the analysis above, the probability $P(n_i(t) = 0|H(t - \tau))$ can be calculated as:

$$\begin{aligned}
 &P(n_i(t) = 0|H(t - \tau)) \\
 &= P(n_i(t - \tau) + \Delta n_i^{OP}(t - \tau) - \Delta n_i^{OC}(t - \tau) = 0|H(t - \tau)) \\
 &= P(\Delta n_i^{OP}(t - \tau) = 0|H(t - \tau))P(\Delta n_i^{OC}(t - \tau) = n_i(t - \tau)|H(t - \tau)) \\
 &= P(\Delta n_i^{OP}(t - \tau) = 0|H(t - \tau))e^{-n_i(t - \tau)\left(1 - \frac{q_i}{c}\right)}
 \end{aligned} \quad (11)$$

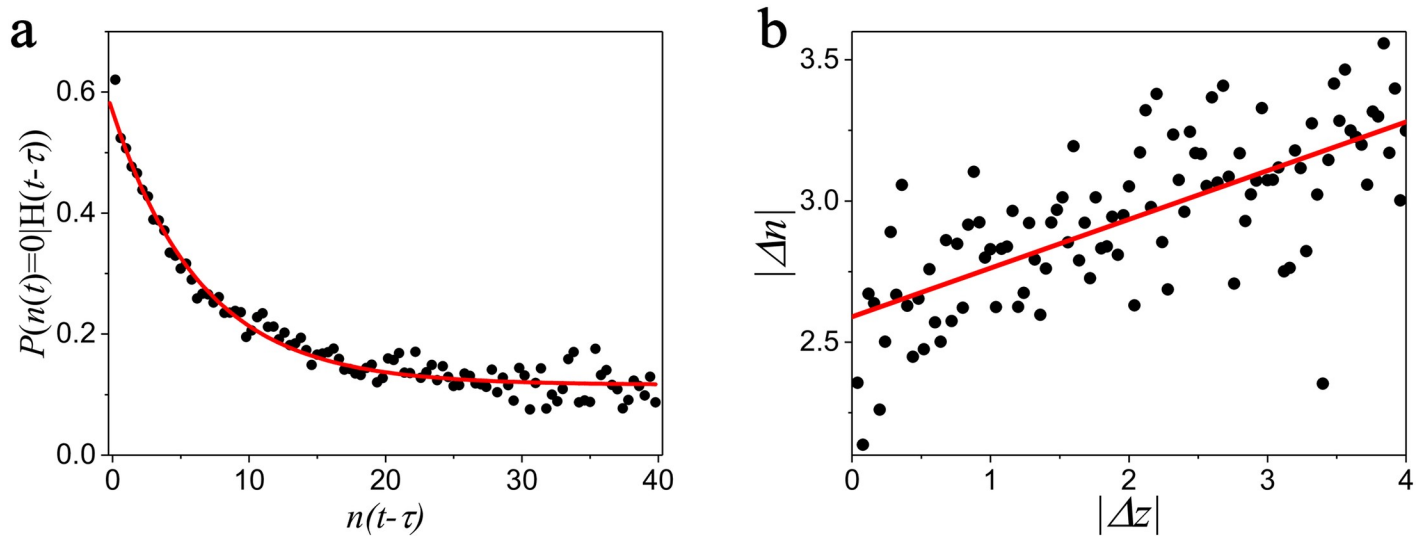


Fig 5. a. The relationship between the conditional probability $P(n(t) = 0|H(t - \tau))$ and the order volume $n(t - \tau)$. The red line is the fitting line. b. The relationship between the absolute value of the change of order volume $|\Delta n|$ versus the fluctuation $|\Delta z|$.

<https://doi.org/10.1371/journal.pone.0259598.g005>

Eq (11) show the relationship between the probability $P(n_i(t) = 0|H(t - \tau))$ and the order volume $n_i(t - \tau)$. Thus, we set the formation of the probability $P(n_i(t) = 0|H(t - \tau))$ as

$$\begin{aligned} P(n_i(t) = 0|H(t - \tau)) \\ = w_i(1)e^{w_i(2)n_i(t-\tau)+w_i(3)n_i(t-\tau)q_i} + w_i(4). \end{aligned} \quad (12)$$

To verify whether our model is correct, we calculate the conditional probability $P(n(t) = 0|H(t - \tau))$ for different order volume $n(t - \tau)$ in Bitcoin transaction platform. We find the conditional probability $P(n(t) = 0|H(t - \tau))$ decreases as the volume of order $n(t - \tau)$ increases in Fig 5. It is found that $P(n(t) = 0|H(t - \tau)) = w_i(1)e^{w_i(2)n(t-\tau)} + w_i(3)$ fit the data well, which means that our model capture the relationship between $P(n_i(t) = 0|H(t - \tau))$ and $n_i(t - \tau)$.

Influence of the fluctuation at last moment. In addition, we need to determine the the relationship between rate of cancellation q_i and the fluctuation $|\Delta z(t)|$. Fig 5b shows the relationship between $|\Delta n|$ and $|\Delta z|$. It is found that the change of order volume $|\Delta n|$ increases as the fluctuation $|\Delta z|$ increases, implying that q_i is related to $\Delta z(t)$. To simplify our model, we set $q_i \approx k|\Delta z(t)|$. However, $|\Delta z(t)|$ is unknown when we predict the fluctuation at time $t - \tau$. So we replace $|\Delta z(t)|$ with the average fluctuation before $\sum_j |\Delta z(t - j)|$. Thus, Eq (12) can be rewrite as

$$\begin{aligned} P(n_i(t) = 0|H(t - \tau)) \\ = w_i(1)e^{w_i(2)n_i(t-\tau)+w_i(3)n_i(t-\tau)\sum_j |\Delta z(t-j)|} + w_i(4). \end{aligned} \quad (13)$$

Parameters update. In our model the parameters $w_i(1)$, $w_i(2)$, $w_i(3)$, $w_i(4)$, $w_i(5)$ are need to form the train data. Using the stochastic gradient descent method, the parameter $w_i(2)$ is

updated as

$$\begin{aligned} w_i(2) &= w_i(2) + \eta \frac{\partial L}{\partial w_i(2)} \\ &= w_i(2) + \eta \frac{\partial L}{\partial P(n_i(t) = 0|H(t - \tau))} \frac{\partial P(n_i(t) = 0|H(t - \tau))}{\partial w_i(2)}. \end{aligned} \quad (14)$$

But we find that the distribution of $n_i(t - \tau)$ is fat-tail, which means that the maximum value of $n_i(t)$ is many times of the minimum value of $n_i(t)$. In the Bitcoin transaction platform, we find that the maximum value of $n_i(t)$ is more than 100 times the average value of $n_i(t)$. The extreme value of order volume would result in the parameter $w_i(2)$ fluctuate violently.

To solve this problem, we proposed a two-step method to update parameters (algorithm 1). At first, we calculate the conditional probability $P(n_i(t) = 0|H(t - \tau))$ using the stochastic gradient descent method [45]. Secondly, we use Eq (13) to fit $P(n_i(t) = 0|H(t - \tau))$. The difference between our method and the stochastic gradient descent method is that we replace derivation with fitting.

Algorithm 1: Learning the conditional probability $P(n(t) = 0|H(t - \tau))$

```

Input: The train dataset;
Output: The probability  $P(n_i(t) = 0|H(t - \tau))$ ;
1 //ITER: Number of iterations; T: size of train dataset; DMAX: maximum price movement; LEN: depth of order book
2 for  $i = 1$  to  $ITER$  do
3   // Train
4   for  $t = 1$  to  $T$  do
5     for  $d = 1$  to  $DMAX$  do
6       cal  $P(|\Delta z'(t)| = d|H(t - \tau))$  using Eq (9)
7     end
8     1. calculate prediction value  $|\Delta z'(t)|$ 
9     2. calculate mean-square-error  $MSE = (|\Delta z'(t)| - |\Delta z(t)|)^2$ 
10    3. update conditional probability  $P(n(t) = 0|H(t - \tau))$ 
11  end
12  // Fit probability
13  for  $k = 1$  to  $LEN$  do
14    Fit  $P(n_k(t) = 0|H(t - \tau))$  using Eq (12)
15  end
16 end
17 Return  $P(n_i(t) = 0|H(t - \tau))$ 

```

Discussion

In this paper, there are five fluctuation prediction methods. The baseline models are the end-to-end models, which ignore the relationship between fluctuation and order. Comparing with the baseline models, we convert the prediction of price fluctuation into the prediction of order volume. Our model explicitly builds the relationship between the fluctuation and the order book. It uses more prior knowledge about order books and price. This may be the reason why our model performs better than the four baseline models.

High-frequency trading is the operation of seeking profits in the short-term market changes, which is usually carried out by computer programs. Dozens of orders can be placed in a second, and these orders will have an impact on the price. The study of the impact of order volume changes on short-term price, allows us to monitor high-frequency trading behaviour. However, the prediction of price direction is not within the scope of our study,

because the price rise and fall direction of the actual market is not random distribution. The accuracy of predicting price trends (rise or fall) is out of this research scope. In the future, the prediction of price trends will be future work. Our method is not only suitable for the bitcoin market, but also for the order-driven market, for example the stock market which has a lot of high-frequency transactions.

Conclusion

In this paper, we focus on the fluctuation prediction. We find that the price would move when the volume of order changes. Based on this observation, the prediction of price fluctuation can be converted into the order volume prediction. Modeling the trader's behaviors—order placement and order cancellation, we conduct an order book based fluctuation prediction model. The developed model performs better on the OKCoin and BTC-e datasets. But there is a limitation on our model—the time lag that our model can predict is small. We try to solve this problem in the future work.

Author Contributions

Data curation: Fabian Shi, Zidong Wang.

Formal analysis: Fabian Shi, Zidong Wang.

Investigation: Fabian Shi, Xiao-Qian Sun, Jinhua Gao, Hua-Wei Shen.

Methodology: Fabian Shi, Xiao-Qian Sun, Jinhua Gao, Hua-Wei Shen, Xue-Qi Cheng.

Project administration: Xue-Qi Cheng.

Resources: Hua-Wei Shen, Xue-Qi Cheng.

Writing – original draft: Fabian Shi, Xiao-Qian Sun.

Writing – review & editing: Xiao-Qian Sun.

References

1. Bauman Mark P. A review of fundamental analysis research in accounting, *Journal of Accounting Literature*. 15, 1 (1996).
2. Yule George Udny. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers, *Philosophical Transactions of the Royal Society of London Series A*, 226, 267–298, (1927). <https://doi.org/10.1098/rsta.1927.0007>
3. Walker Gilbert Thomas. On periodicity in series of related terms, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 131(818), 518–532, (1931).
4. Engle Robert F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica: Journal of the Econometric Society*. 987–1007, (1982). <https://doi.org/10.2307/1912773>
5. Li Dong and Zhang Xingfa and Zhu Ke and Ling Shiqing. The ZD-GARCH model: A new way to study heteroscedasticity, *Journal of Econometrics*, 202(1), 1–17, (2018). <https://doi.org/10.1016/j.jeconom.2017.09.003>
6. Posedel Petra Analysis of the exchange rate and pricing foreign currency options on the croatian market: the ngarch model as an alternative to the black-scholes model, *Financial theory and practice*, 30 (4), 347–368, (2006).
7. Engle Robert F and Ng Victor K. Measuring and testing the impact of news on volatility, *The journal of finance*, 48(5), 1749–1778, (1993). <https://doi.org/10.1111/j.1540-6261.1993.tb05127.x>
8. Bollerslev Tim. Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics*, 31 (3), 307–327, (1986). [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)

9. Ding Zhuanxin and Granger Clive WJ and Engle Robert F. A long memory property of stock market returns and a new model, *Journal of empirical finance*, 1(1), 83–106, (1993). [https://doi.org/10.1016/0927-5398\(93\)90006-D](https://doi.org/10.1016/0927-5398(93)90006-D)
10. Chang Pei-Chann and Liu Chen-Hao and Lin Jun-Lin and Fan Chin-Yuan and Ng Celeste SP. A neural network with a case based dynamic window for stock trading prediction *Expert Systems with Applications*, 36(3), 6889–6898, (2009). <https://doi.org/10.1016/j.eswa.2008.08.077>
11. Chen Miao-Sheng and Ying Li-Chih and Pan Mei-Chiu. Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system, *Expert Systems with Applications*, 37(2), 1185–1191, (2010). <https://doi.org/10.1016/j.eswa.2009.06.032> PMID: 32288334
12. Leu Yungho and Lee Chien-Pang and Jou Yie-Zu. A distance-based fuzzy time series model for exchange rates forecasting, *Expert Systems with Applications*, 36(4), 8107–8114, (2009). <https://doi.org/10.1016/j.eswa.2008.10.034>
13. Norio Baba and Motokazu Kozaki, An intelligent forecasting system of stock price using neural networks, [Proceedings 1992] IJCNN International Joint Conference on Neural Networks, 1, 371–377 (1992).
14. Chenoweth Tim and Obradović Zoran. A multi-component nonlinear prediction system for the S&P 500 index, *Neurocomputing*, 10(3), 275–290, (1996). [https://doi.org/10.1016/0925-2312\(95\)00109-3](https://doi.org/10.1016/0925-2312(95)00109-3)
15. Fernandez-Rodriguez Fernando and Gonzalez-Martel Christian and Sosvilla-Rivero Simon. On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market, *Economics letters*, 69(1), 89–94, (2000). [https://doi.org/10.1016/S0165-1765\(00\)00270-6](https://doi.org/10.1016/S0165-1765(00)00270-6)
16. Ghiassi M and Saidane H and Zimbra DK. A dynamic artificial neural network model for forecasting time series events, *International Journal of Forecasting*, 21(2), 341–362, (2005). <https://doi.org/10.1016/j.ijforecast.2004.10.008>
17. Hamzaçebi Coşkun and Akay Diyar and Kutay Fevzi. Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting, *Expert Systems with Applications*, 36(2), 3839–3844, (2009).
18. Lendasse Amaury and de Bodt Eric and Wertz Vincent and Verleysen Michel. Financial time series forecasting using support vector machines, *European Journal of Economic and Social Systems*, 14(1), 81–91, (2000). <https://doi.org/10.1051/ejess:2000110>
19. Roh Tae Hyup. Forecasting the volatility of stock price index, *Expert Systems with Applications*, 33(4), 916–922, (2007). <https://doi.org/10.1016/j.eswa.2006.08.001>
20. Zheng Linyu and He Hongmei. Share price prediction of aerospace relevant companies with recurrent neural networks based on pca. *Expert Systems with Applications*, page 115384, 2021. <https://doi.org/10.1016/j.eswa.2021.115384>
21. Hammad A and Ali Souma M Alhaj and Hall Ernest L. Forecasting the Jordanian stock price using artificial neural network, *Intelligent Engineering Systems through Artificial Neural Networks*, 17, (2007).
22. Egrigöglu Erol and Aladag Cagdas Hakan and Yolcu Ufuk and Uslu Vedide R and Basaran Murat A. A new approach based on artificial neural networks for high order multivariate fuzzy time series, *Expert Systems with Applications*, 36(7), 10589–10594, (2009). <https://doi.org/10.1016/j.eswa.2009.02.057>
23. Ghiassi MDKZ and Zimbra David K and Saidane H. Medium term system load forecasting with a dynamic artificial neural network model, *Electric power systems research*, 76(5), 302–316, (2006). <https://doi.org/10.1016/j.epsr.2005.06.010>
24. Güreşen, Erkam and Kayakutlu, Gülgün. Forecasting stock exchange movements using artificial neural network models and hybrid models, *International Conference on Intelligent Information Processing*, 129–137, Springer, (2008).
25. Khashei Mehdi and Bijari Mehdi. An artificial neural network (p, d, q) model for timeseries forecasting, *Expert Systems with applications*, 37(1), 479–489, (2010). <https://doi.org/10.1016/j.eswa.2009.05.044>
26. Liao Zhe and Wang Jun. Forecasting model of global stock index by stochastic time effective neural network, *Expert Systems with Applications*, 37(1), 834–841, (2010). <https://doi.org/10.1016/j.eswa.2009.05.086>
27. Hassan Md Rafiul and Nath Baikunth and Kirley Michael. A fusion model of HMM, ANN and GA for stock market forecasting, *Expert systems with Applications*, 33(1), 171–180, (2007). <https://doi.org/10.1016/j.eswa.2006.04.007>
28. Maslov Sergei. Simple model of a limit order-driven market. *Physica A: Statistical Mechanics and its Applications*, 278(3-4):571–578, 2000. [https://doi.org/10.1016/S0378-4371\(00\)00067-4](https://doi.org/10.1016/S0378-4371(00)00067-4)
29. Cont Rama, Stoikov Sasha, and Talreja Rishi. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010. <https://doi.org/10.1287/opre.1090.0780>

30. Bouchaud Jean-Philippe, Mézard Marc, and Potters Marc. Statistical properties of stock order books: empirical results and models. *Quantitative finance*, 2(4):251, 2002. <https://doi.org/10.1088/1469-7688/2/4/301>
31. Mike Szabolcs and Farmer J Doyne. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1):200–234, 2008. <https://doi.org/10.1016/j.jedc.2007.01.025>
32. Smith Eric, Farmer J Doyne, Gillemot László, and Krishnamurthy Supriya. Statistical theory of the continuous double auction. *Quantitative finance*, 3(6):481, 2003. <https://doi.org/10.1088/1469-7688/3/6/307>
33. Wang Jiahua, Zhu Hongliang, and Li Dongxin. Price dynamics in an order-driven market with bayesian learning. *Complexity*, 2018, 2018. <https://doi.org/10.1155/2018/8254068>
34. Kanazawa Kiyoshi, Sueshige Takumi, Takayasu Hideki, and Takayasu Misako. Derivation of the boltzmann equation for financial brownian motion: Direct observation of the collective motion of high-frequency traders. *Physical review letters*, 120(13):138301, 2018. <https://doi.org/10.1103/PhysRevLett.120.138301> PMID: 29694225
36. Sun Xiao-Qian, Shen Hua-Wei, Cheng Xue-Qi, Zhang Yuqing, Market confidence predicts stock price: beyond supply and Demand. *Plos One*, 11(7):e0158742, (2016). <https://doi.org/10.1371/journal.pone.0158742> PMID: 27391816
35. Sun Xiao-Qian, Shen Hua-Wei, Cheng Xue-Qi Trading network predicts stock price. *Scientific reports*, 4,3711, (2013). <https://doi.org/10.1038/srep03711>
37. Fabin Shi, Nathan Aden, Shengda Huang, Neil Johnson, Xiaoqian Sun, Jinhua Gao, et al. Modelling universal order book dynamics in bitcoin market. *arXiv preprint arXiv:2101.06236*, 2021.
38. Yura Yoshihiro, Takayasu Hideki, Sornette Didier, and Takayasu Misako. Financial brownian particle in the layered order-book fluid and fluctuation-dissipation relations. *Physical review letters*, 112(9):098703, 2014. <https://doi.org/10.1103/PhysRevLett.112.098703> PMID: 24655287
39. Kim Kyoung-jae. Financial time series forecasting using support vector machines, *Neurocomputing*, 55(1–2), 307–319, (2003). [https://doi.org/10.1016/S0925-2312\(03\)00372-2](https://doi.org/10.1016/S0925-2312(03)00372-2)
40. Kao Ling-Jing and Chiu Chih-Chou and Lu Chi-Jie and Yang Jung-Li. Integration of nonlinear independent component analysis and support vector regression for stock price forecasting, *Neurocomputing*, 99, 534–542, (2013). <https://doi.org/10.1016/j.neucom.2012.06.037>
41. Kumar Deepak and Meghwani Suraj S and Thakur Manoj. Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets, *Journal of Computational Science*, 17, 1–13, (2016). <https://doi.org/10.1016/j.jocs.2016.07.006>
42. Moghaddam Amin Hedayati and Moghaddam Moein Hedayati and Esfandiyari Morteza. Stock market index prediction using artificial neural network, *Journal of Economics, Finance and Administrative Science*, 21(41), 89–93, (2016). <https://doi.org/10.1016/j.jefas.2016.07.002>
43. Guresen Erkam and Kayakutlu Gulgun and Daim Tugrul U. Using artificial neural network models in stock market index prediction, *Expert Systems with Applications*, 38(8), 10389–10397, (2011). <https://doi.org/10.1016/j.eswa.2011.02.068>
44. Di Persio Luca and Honchar Oleksandr. Artificial neural networks architectures for stock price prediction: Comparisons and applications, *International journal of circuits, systems and signal processing*, 10, 403–413, (2016).
45. Robbins Herbert and Monro Sutton. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951. <https://doi.org/10.1214/aoms/1177729586>