Received December 31, 2019, accepted January 9, 2020, date of publication January 14, 2020, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966553

Time Series Data Mining: A Case Study With Big Data Analytics Approach

FANG WANG[1,3], MENGGANG LI[2,3,4], (Member, IEEE), YIDUO MEI[5], AND WENRUI LI[1,3]

[1]School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China
[2]National Academy of Economic Security, Beijing Jiaotong University, Beijing 100044, China
[3]Beijing Laboratory of National Economic Security Early-Warning Engineering, Beijing Jiaotong University, Beijing 100044, China
[4]Beijing Center for Industrial Security and Development Research, Beijing Jiaotong University, Beijing 100044, China
[5]Postdoctoral Programme of China Centre for Industrial Security Research, Beijing Jiaotong University, Beijing 100044, China

Corresponding authors: Menggang Li (mgli1@bjtu.edu.cn) and Wenrui Li (liwenrui@bjtu.edu.cn)

This work was supported by the Program of the Co-Construction with the Beijing Municipal Commission of Education of China under Grant B18H100040.

ABSTRACT Time series data is common in data sets has become one of the focuses of current research. The prediction of time series can be realized through the mining of time series data, so that we can obtain the development process and regularity of social economic phenomena reflected by time series, and extrapolate to predict its development trend. More and more attention has been paid to time series prediction in the era of big data. It is the basic application of time series prediction to accurately predict the trend. In this paper, we introduce various time series autoregressive (AR) model, moving average (MA) model, and ARIMA model that is combined by AR and MA. As the time series prediction in general scenarios, the ARIMA is applied to the risk prediction of the National SME Stock Trading (New Third Board) in combination with specific scenarios. The case studies show that the results of our analysis are basically consistent with the actual situation, which has greatly helped the prediction of financial risks.

INDEX TERMS Data mining, time series, financial forecast, AR, MA, ARIMA, financial risk.

I. INTRODUCTION

Time series data mining comes from the need of people to visualize data models according to their abilities. People rely on complex methods to perform these tasks. In fact, we can ignore small fluctuations to get the conceptual model and distinguish different time models based on the similarity between models. The main time series related tasks include content-based querying, anomaly checking, pattern recognition, prediction, clustering, classification and segmentation. A large number of decision-making problems cannot be separated from prediction in various research fields of the natural sciences and social sciences, forecast is the basis of the decision-making [1]. Therefore, we mainly explored the time series data analysis and prediction.

Time series prediction methods are divided into traditional time series prediction methods and machine learning methods. The traditional time series forecasting method refers to predicting the trend development of future time series only based on the trend development of historical time series. This method fits the historical time trend curve by establishing an appropriate mathematical model and predicts the trend of future time series according to the established model Curves, our common models include ARMA [2], VAR [3], TAR [4], ARCH [5], etc. The traditional time series method can be applied to a variety of scenarios because it relies on relatively simple data and only needs the historical time series trend curve to build a model. However, the traditional time series prediction method often faces the problem of lag, which is that the predicted value is several time units later than the true value. In order to improve the accuracy of prediction, machine learning algorithms are introduced into time series prediction. The machine learning methods select features that may affect the predicted value according to the specific application scenario, then introduces these features into the model, finally applies machine learning classification models for prediction. Machine learning methods need to extract more features from data in multiple dimensions. The more complex the model, the more accurate the prediction. However, models are often not universal and features need to be re-extracted for different application scenarios to build models. In reality prediction, machine learning methods are

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

often combined with traditional time series prediction methods. We mainly explore the AR and MA prediction models, and then explore the combination of the two models, ARIMA, which has a good method for processing non-stationary time series [6].

At present, there are many methods for analyzing and predicting factors related to the relationship between supply and demand in the financial market, but the effect of this method is not obvious. We conducted a time series analysis of the financial and economic fields and used the ARIMA model to predict the risks of the National SME Stock Trading (New Third Board). The final results are basically consistent with the actual results, and good prediction results have been obtained.

II. RESEARCH BACKGROUND

Time series data is encountered in every aspect of the scientific field [7]. A time series is a series of observations taken in chronological order. For examples, a time series can be constituted by the closing price of a stock A on each trading day from June 1, 2015 to June 1, 2016; a time series can be constituted by the daily maximum temperature in a certain place; The station's environmental detection data records consist of a time series and so on. With the rapid development of big data, more and more time series data are stored in computers, so that we have a huge amount of time series data. Faced with these time series data, people want to reveal the information existing in these series data sets through effective methods or techniques. Today, the study of time series data has been rapidly developed and has become an important research direction in data mining. We can discover the inherent rules of things change and provide a reference for relevant people through the study of time series data.

The basis of time series analysis is to believe that prices follow trends and that past price information is useful for predicting future prices. Analysis of the dynamic changes in stock prices is one of the most difficult challenges for human intelligence. There are many prediction methods for predicting stock price changes, such as the Box Jenkins method [8], the Black-Scholes model [9], and the binomial model [10]. The Box-Jenkins method is a five-step process of identifying, selecting, and evaluating conditional averaging models. The ARIMA method is popularized by Box and Jenkins, and the ARIMA model is often called the Box-Jenkins model. Box and Tiao discuss the general transfer function model used by the ARIMA program. The Black-Scholes model is XXX. The binomial model is YYY. The Autoregressive Moving Average (ARMA) method is one of the most popular linear models in time series prediction because of its good statistical characteristics and great flexibility. The current stock forecast is based on market demand, the effect of this method is not very satisfactory due to the lack of time series. This paper uses the processing of time series data to obtain future stock conditions and produces good results.

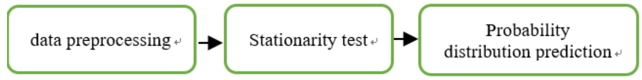


FIGURE 1. Three-step analysis method.

III. TIME SERIES DATA ANALYSIS AND FORECAST

In this paper, a three-step analysis method for time series data analysis is proposed (see Figure 1): Firstly, the data is pre-processed, which includes stationary processing of time series that are in an unstable state. Secondly, the pre-processed data is tested for stationarity. Finally, the prediction model is used to predict the probability distribution in the same time period in the future.

A. STATIONARITY DETERMINATION AND PROCESSING

A time series can be considered stable when it has no systematic changes in the mean (no trend), no systematic changes in the variance and periodic changes strictly eliminated. The time series can be further subdivided into strict stationary and weak stationary. For all time t, any positive integer k and any k positive integers $(t_1, t_2, \cdots, t_k)$, the joint distribution of $(r_{t1}, r_{t2}, \cdots, r_{tk})$ is the same as the joint distribution of $(r_{t1+t}, r_{t2+t}, \cdots, r_{tk+t})$, we call the time series $\{r_t\}$ to be strictly stable and the joint distribution of $(r_{t1}, r_{t2}, \cdots, r_{tk})$ remains unchanged under the translational transformation of time. The above time series are strong stationary time series, but the time series we use are generally weak stationary sequence.

A weakly stationary sequence $\{r_t\}$ must satisfy the following two conditions: $E(r_t) = \mu$ (μ is constant). Variance $Cov(r_t, r_{t-1}) = \gamma_l, \gamma_l$ only depends on l (l is any integer). For weakly stationary time series, the mean and the covariance of r_t and r_{t-1} do not change with time. We usually call a stationary sequence is weakly stationary in financial data.

Differential operation is usually used to achieve the stable condition when the time series is not stable. The difference (forward here) is to find the difference between the value r_t of the time series $\{r_t\}$ at time t and the value r_{t-1} at time t-1. Let us consider it as d_t, it is a first-order difference. If the same operation is performed on the new sequence $\{d_t\}$, it is a second-order difference. Generally, non-stationary time series can be processed through d-time difference to be as stationary or as approximate as stationary time series.

B. TIME SERIES PREDICTION MODEL

1) CORRELATION COEFFICIENT AND AUTOCORRELATION FUNCTION

The correlation coefficient is actually the angle between the two vectors in the vector space and the covariance is the expected value (or mean) of the product of their deviations from their individual expected values. The correlation coefficient is equal to 1 or -1 when the two vectors are parallel (In particular, 1 means the same direction, -1 means the reverse). If the two vectors are perpendicular and the cosine of the included angle is equal to 0, it means that the two vectors are uncorrelated. The smaller the angle between the two vectors, the closer the absolute value of the correlation coefficient is to 1, and the higher the correlation between the two vectors.

The linear correlation between the two vectors is measured by correlation coefficient. In the stable time series $\{r_t\}$, the linear correlation between r_t and its past value r_{t-i} is measured by autocorrelation coefficient. The correlation coefficient between r_t *and* r_{t-i} is called the autocorrelation coefficient of spacing l of r_t, which is usually recorded as p_1. specific:

$$\rho_1 = \frac{Cov\left(r_t, r_{t-1}\right)}{\sqrt{Var\left(r_t\right) Var\left(r_{t-1}\right)}} = \frac{Cov\left(r_t, r_{t-1}\right)}{Var\left(r_t\right)}$$

The above formula uses the property of weak stationary: $Var\left(r_t\right) = Var\left(r_{t-1}\right)$. For $\{r_t\}$ samples of stationary time series, then the autocorrelation coefficient of the samples with an interval of 1 is estimated as:

$$\hat{\rho}_1 = \frac{\sum_{t=l+1}^{T}\left(r_t - \overline{r}\right)\left(r_{t-1} - \overline{r}\right)}{\sum_{t=1}^{T}\left(r_t - \overline{r}\right)^2}$$

A series of autocorrelation sequences $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3 \cdots$ is called the sample autocorrelation function of r_t. We consider that the time series is completely uncorrelated when all the values in the autocorrelation function are 0. Therefore, we often need to check whether multiple autocorrelation coefficients are 0.

2) AUTOREGRESSIVE (AR) MODEL

The data r_{t-1} at time t-1 may be useful in predicting r_t at time t when the time series data interval is 1 and the autocorrelation coefficient ACF is significant. We can build the following model according to the above principles:$r_t = \emptyset_0 + \emptyset_1 r_{t-1} + a_t$, a_t is a white noise sequence, this model is called a first-order autoregressive (AR) model. We can introduce an AR (p) model from AR model: $r_t = \emptyset_0 + \emptyset_1 r_{t-1} + \emptyset_2 r_{t-2} + \cdots + \emptyset_p r_{t-p} + a_t$. We generally choose partial correlation function and information criterion function to determine the order. The information criterion usually uses the AIC rule. The following methods are proposed for the test of AR (p) stationarity. We first assume that the sequence is weakly stationary, then $E\left(r_t\right) = \mu, Cov\left(r_t\right) = \gamma_0, Cov\left(r_t, r_{t-j}\right) = \gamma_j, (\mu, \gamma_0)$ are constants. Because a_t is a white noise sequence, there are: $E\left(a_t\right) = 0, Var\left(a_t\right) = \sigma_a^2$, so there are: $E(r_t) = \emptyset_0 + \emptyset_1 E(r_{t-1}) + \emptyset_2 E(r_{t-2}) + \cdots + \emptyset_p E(r_{t-p})$. According to the nature of stationary, $E\left(r_t\right) = E\left(r_{t-1}\right) = E\left(r_{t-2}\right) = \cdots = \mu$, which has: $\mu = \emptyset_0 + \emptyset_1 \mu + \emptyset_2 \mu + \cdots + \emptyset_p \mu, E\left(r_t\right) = \mu = \frac{\theta_0}{1-\emptyset_1-\emptyset_2-\cdots-\emptyset_p}$. We have the equation $1 - \emptyset_1 x - \emptyset_2 x - \cdots - \emptyset_p x = 0$ as the characteristic equation when the denominator is not 0. The inverses of all the solutions of the equation are the characteristic roots of the model. The AR (p) sequence is stationary when all the characteristic roots are less than 1.

3) MOVING AVERAGE (MA) PREDICTION MODEL

We directly give the form of the MA (q) model: $r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_q a_{t-q}^{[11]}$, c_0 is a constant term. The a_t is the perturbation or information of the AR model at time t, then it can be found that the model uses random interference or prediction error in the past q periods to linearly express the current prediction value. The autocorrelation function is always q-step truncated for q-order MA models. Therefore, the MA (q) sequence is only linearly related to its first q delay values, so it is a "limited memory" model. This feature can be used to determine the order of the model. MA models are always weakly stationary because they are a finite linear combination of white noise sequences. Therefore, MA model has the properties of weak stationarity: stationarity, finality and reversibility.

TABLE 1. Unit root inspection table.

	value
Test Statistic Value	-2.30472
p-value	0.170449
Lags Used	1
Number of Observations Used	379
Critical Value(1%)	-3.44772
Critical Value(5%)	-2.8692
Critical Value(10%)	-2.57085

C. ARIMA PREDICTION MODEL

So far, we have focused on stationary sequences. We can consider using the ARIMA model if the sequence is non-stationary. The ARIMA can be used for statistics and artificial intelligence [12]. ARIMA has only one more letter "I" than ARMA, which means that it has one more level of connotation than ARMA. A non-stationary sequence can be transformed into a stationary time series after d times of difference. For the specific value of d, we first perform a stationary test on the sequence after the first difference. Then we will continue to make the difference if it is still non-stationary until the test is stationary after d times. Finally, the specific value of d is calculated.

1) UNIT ROOT TEST

ADF is a common unit root test method [13]. Its original hypothesis is that the sequence has a unit root, and the sequence is non-stationary. It is necessary to be significant at a given confidence level and reject the original hypothesis for a stable time series data.

According to Table 1 and Figure 2 above, we assume the original hypothesis that the sequence has a unit root. The original hypothesis cannot be rejected because we can see that the value of p-value is 0.1704489, which is much larger than the significant level. Therefore, the daily index series of the Shanghai Stock Index is non-stationary. We make a difference to the sequence as shown in Figure 3:

We can know from the figure 3 that the sequence is approximately stationary. Let's perform ADF test, p-value: 2.31245750144e-30. We can think the sequence is stationary

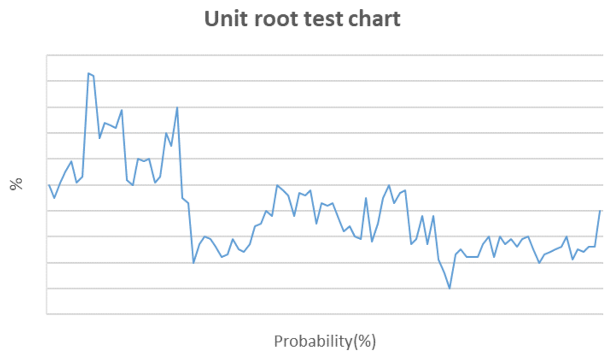


FIGURE 2. Unit root test chart.

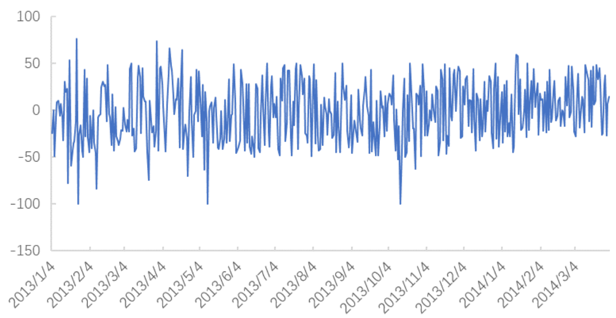


FIGURE 3. Sub-difference.

Because it can be seen that the p-value is very close to 0 and the original hypothesis is rejected. The value of d for the original sequence can be 1 because the sequence is stable after one difference. An ARMA model can be built from the differential sequence after the value of d is determined. At present, ARIMA has been widely used in various fields [14]. Next, we will use the ARIMA model to analyze example in the financial field.

IV. NEW THIRD BOARD RISK FORECAST

The unit root analysis of the rise based on the stock price within a certain period of time can determine the stability of the rise series. The probability distribution of different rises and falls in the same period in the future can be inferred based on the historical distribution of the rise when the sequence is stable, so that the interested parties prepare plans for extreme situations that seriously affect the level of net worth of funds. In recent years, the OTC New Third Board has developed rapidly in China. We can find that the New Third Board market has two characteristics after careful observation. Firstly, the overall market price volatility is significantly higher than that of the Shanghai and Shenzhen markets. Secondly, the volatility distribution is severely rightward. The fluctuation risk of individual stocks is often released quickly and violently because there is no limit of the daily limit system. In the following, we focus on the practical problems in the Chinese NEEQ stock market. The time series analysis method was used to estimate the distribution probability of future rise and fall based on the differential time series of daily rise and fall of stock prices.

A. ANALYTICAL METHOD

We used the three-step analysis method proposed above for analysis: In the first stage, a non-stationary sequence is transformed into a stationary time series by differential processing. In the second stage, we use the ADF unit root test to check whether the time series is stable. In the third stage, the probability distribution of different rises and falls within the same time period based on the historical distribution of rises and prepare is inferred for extreme situations that may seriously affect the level of NAV.

STEP 1: data preparation (data preprocessing). The time series is defined as a series of quantitative observations at consecutive times. In the analysis of financial time series, the price time series itself is generally unstable, not completely random distribution, and has obvious autocorrelation. At the same time, the law of price distribution may also change abruptly due to a variety of factors, so that the law established in the past stage may not still hold in the future. Therefore, it is generally invalid to analyze the price time series directly in an attempt to find the law or regression formula. We pre-process the time series before applying the ARMA model if the sequence is non-stationary. Generally, the method for dealing with unstable time series is to make first order difference of the time series [15]. Generally, two methods can be used:

The first is to find the difference between adjacent variables to build a first-order difference sequence, we can build a new sequence y_t:

$$y_t = x_t - x_{t-1}$$

The second is to find the ratio of adjacent variables to build a first-order difference sequence, we can build a new sequence y_t:

$$y_t = \frac{x_t}{x_{t-1}}$$

A non-stationary sequence can be transformed into a stationary sequence after d times of difference. The specific value of d depends on the structure of the stationarity test after the time series difference. we will continue to make the difference if it is still non-stationary until the test is stationary after d times.

The relative ratio of the stock prices (the relative increase) is more concerned about the absolute value of stock price changes, so that the ratio method is generally used in the analysis of financial product price time series [16]. At the same time, the stock price difference will continue to increase or decrease accordingly after the stock price continues to rise or fall. Therefore, it is proposed to use the natural logarithm of the ratio of adjacent variables in the time series of stock prices to perform first-order difference processing, we need to construct a new series y_t:

$$y_t = \ln(\frac{x_t}{x_{t-1}})$$

A prominent advantage of this method is that the first-order difference sequence y_t obtained from this method is approximately equal to the stock price increase, which can be directly used for the probability prediction of the future stock price distribution. In this paper, we use the ratio method to deal with time series.

STEP 2: Stationarity check. We apply the unit root test to the logarithmic rise series. Our goal is to investigate the stationarity of the residuals to determine if the ARMA model is a good model for them. The original hypothesis of the unit root test is to test whether the sequence is stationary. Then, negating the original hypothesis means that the series (or the differential sequence in this example) is stationary. Specifically, we use ADF (Augmented Dicky-Puller) to check whether the time series is stable.

STEP 3: Prediction of the probability distribution of risk. We can prepare for extreme situations that severely affect asset benefit levels with the probability distribution of price the next day. The basic idea of the ARMA model is to combine the AR and MA models so that the number of parameters used is kept small. The form of the model is:

$$r_t = \varphi_0 + \sum_{i=1}^{i=p} \varphi_i r_{t-1} + a_t + \sum_{i=1}^{i=q} \theta_i a_{t-1}$$

Among them, $\{a_t\}$ is a white noise sequence, p and q are both non-negative integers. We use the sequence of moving operator B backward, the previous moment, the above model can be written as: $\left(1 - \emptyset_1 B - \cdots - \emptyset_p B^p\right) r_t = \emptyset_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t$

At the time we get the expectation of r_t:

$$E\left(r_t\right) = \frac{\emptyset_0}{1 - \emptyset_1 - \cdots - \emptyset_p}$$

The inverse of all solutions of the equation $1 - \emptyset_1 x - \emptyset_2 x^2 - \cdots - \emptyset_p x^p = 0$ is called the characteristic root of the model. We can think the ARMA model is stable if the modulus of all characteristic roots is less than 1. We limit the maximum order of AR to less than 6 and the maximum order of MA not to exceed 4 in order to control the amount of calculation. Then we established the ARMA model based on the (3,3) order model solved by the AIC criterion.

We take Jindalai of the New Third Board as an example and predict the probability distribution of future fluctuations through the analysis of the closing price increase. Since Jindalai was changed to a market-making transfer from November 25, 2014, historical data was selected as daily closing price data for 421 transfer days from January 5, 2015 to September 24, 2016. A logarithmic gain sequence can be obtained after differential processing of the closing price sequence. According to our analysis framework (see Figure 1), we next perform a unit root test on the above two time series and use ADF on Eviews 8 (Augmented Dicky-Fuller) test can be obtained:

It can be seen from table 2 that the existence of unit root in the closing price time series cannot be denied even at the confidence level of 10%, so that the closing price time series is basically non-stationary. The logarithmic rise and fall time series can be rejected to have a unit root when the confidence level is significantly higher than 1%, which is that the time series of logarithmic increase and decrease is basically stable. Therefore, we can use the time series analysis to predict the future probability of distribution based on the stock's past gain information. The following is a forecast of the distribution of the gain in the next 1 transfer day based on historical gain information:

TABLE 2. Unit root test results.

DF test results	Thresholds of different significance levels		
-2.456a	1%	5%	10%
-18.961b	-3.4457	-2.8682	-2.5703

TABLE 3. Daily increase probability distribution table.

probability	Probability distribution of daily increase					
interval (%)	<-5	<-3	<-1	>1	>3	>5
value (%)	2.14	6.9	23.33	23.81	10.24	4.76

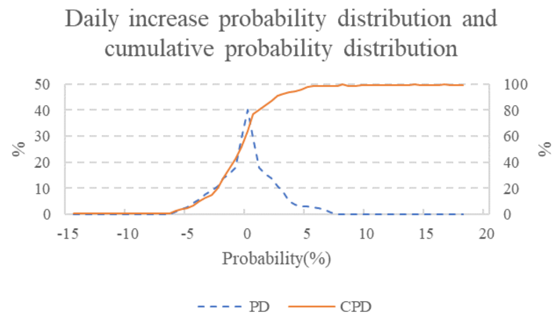


FIGURE 4. Daily increase probability distribution and cumulative probability distribution.

As can be seen from Table 3, the daily probability distribution of closing prices has a characteristic of peaks and long tails. When applied to intraday T+0 trading, we should pay attention to setting reasonable stop loss prices to avoid large losses caused by small probability events. As shown in Figure 4:

In addition to the daily increase distribution probability, we cannot directly use the cumulative distribution of the smaller period increase level probability when we want to obtain the increase distribution probability of different time periods. For example, it is not possible to obtain a weekly increase horizontal distribution probability or a monthly increase horizontal distribution probability from the daily increase horizontal distribution probability by a superimposed manner. The correct processing method is to directly find the logarithmic first-order difference of the weekly closing price sequence to obtain the time series of the logarithmic rise of the weekly closing price and process the sequence. For example, statistics on the weekly logarithmic rise time series of Jindalai are shown in Figure 5:

They can be obtained other distribution probabilities by interpolation in the distribution probability table or by taking intersection points on the probability distribution curve.

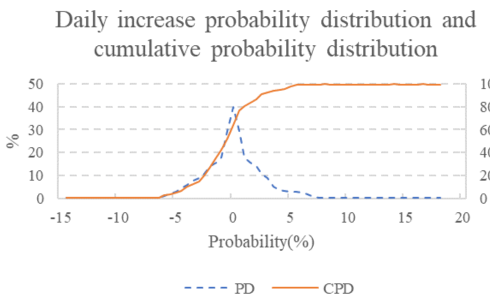


FIGURE 5. **Weekly probability distribution curve and weekly cumulative probability curve.**

Assuming that the probability of occurrence is less than 5% as the trading condition, by making two horizontal lines p = 5% and p = 95% on the ordinate to intersect the distribution curve at two points(Representing the probability distribution points of excessive decline and excessive increase, respectively), taking two points on the abscissa can get an excessive drop The corresponding abscissa is −4.5%, and the corresponding abscissa is 4. 35%. In other words, there is a 84% probability that the stock's individual stocks will rise or fall between −4.5% and 4.35%. Therefore, It can be bought and then sold intraday T+0 trading when the stock falls by more than 4.5% on the same day; It can be sold and then bought intraday T+0 trading when a stock rises more than 4.35% on the same day.

B. CONCLUSION AND SUGGESTION

The time series analysis proposes a method of estimating the distribution probability of future stock price fluctuations based on historical price information, thereby we can avoid the risk of huge losses caused by intraday trading. The method of directly judging the future price trend based on the historical price often lacks sufficient reliability because price series are often unstable. We first perform a first-order difference processing on the stock price series, which is equivalent to changing the research object from the stock price itself to the change value of the stock price. Then check the stationarity and determine the stationarity of the increase series by unit root analysis of the logarithmic increase of the stock price within a certain period of time. Finally, we apply ARMA (3,3) to the stable differential sequence to obtain the historical distribution rule. We use this rule to obtain the probability distribution of different rises and falls in the same period of time in the future, so that we can prepare for extreme situations that can seriously affect the level of net asset value.

V. SUMMARY AND OUTLOOK

With the continuous progress on time series data mining technology, its application has been extended to financial analysis and it can well predict the risks in the financial field in the future. We analyzed the time series data and its various prediction models, then applied the ARIMA prediction model to the analysis of financial. The paper proposed a time series analysis method to predict the future rise through the historical rise and fall probability distribution curve. The probability of falling distribution makes a good prediction for extreme situations that may seriously affect the level of net assets in real life so that we can infer what will happen in the future through time series data sequences.

As the research continues, more technologies will be considered to expand the scope of prediction and improve the accuracy [17]–[19]. We expect the time-series predictions will have more applications in financial [20], [21]. By accurate predictions, we can improve response measures to discover possible emergencies; We can improve the prediction by adding the spatial dimension under the combination of time and space. For example, the better resource utilization for users and taxis can be generated by predicting the number of taxi rides in a certain area of Didi taxi [22]–[24]; This also can accurate financial predictions, such as, it can help managers Reasonably specify strategy by predicting the amount of money bought and sold, etc. The data mining of time series data whose guidance and help [25] to actual production and life will become more and more important.

REFERENCES

[1] J. F. Li and Q. Zong, "The forecasting of the elevator traffic flow time series based on ARIMA and GP," *Adv. Mater. Res.*, vols. 588–589, pp. 1466–1471, Nov. 2012.

[2] M. Rout, B. Majhi, R. Majhi, and G. Panda, "Forecasting of currency exchange rates using an adaptive ARMA model with differential evolution based training," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 26, no. 1, pp. 7–18, Jan. 2014.

[3] N. Groenewold, L. Guoping, and C. Anping, "Regional output spillovers in China: Estimates from a VAR model," *Papers Regional Sci.*, vol. 86, no. 1, pp. 101–122, Mar. 2007.

[4] C. Altavilla and P. De Grauwe, "Non-linearities in the relation between the exchange rate and its fundamentals," *Int. J. Fin. Econ.*, vol. 15, no. 1, p. n/a, 2008.

[5] B. Baldauf and G. J. Santoni, "Stock price volatility: Some evidence from an ARCH model," *J. Futures Markets*, vol. 11, no. 2, pp. 191–200, Apr. 1991.

[6] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *Proc. UKSim-AMSS 16th Int. Conf. Comput. Modelling Simulation*, Mar. 2014.

[7] C. C. Aggarwal, *Mining Time Series Data*. Springer, 2015.

[8] Y. Lu and S. AbouRizk, "Automated boxjenkins forecasting modelling," *Autom. Construction*, vol. 18, no. 5, pp. 547–558.

[9] J.-P. Bouchaud and D. Sornette, "The black-scholes option pricing problem in mathematical finance: Generalization and extensions for a large class of stochastic processes," *Sci. Finance Work. Paper Arch.*, vol. 4, no. 6, pp. 863–881, Jun. 1994.

[10] S. Muzzioli and C. Torricelli, "A multiperiod binomial model for pricing options in a vague world," *J. Econ. Dyn. Control*, vol. 28, no. 5, pp. 861–887, Feb. 2004.

[11] J. Che and J. Wang, "Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling," *Energy Convers. Manage.*, vol. 51, no. 10, pp. 1911–1917, Oct. 2010.

[12] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, Dec. 2012.

[13] Q. A. Nie and X. Zhang, "The analysis of united test of statistics in adf unit root test," *Stat. Res.*, 2007.

[14] J. Contreras, R. Espinola, F. Nogales, and A. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003.

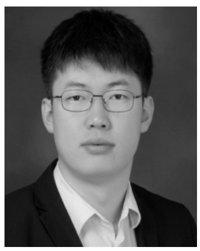
[15] F. Islam, M. Shahbaz, A. U. Ahmed, and M. M. Alam, "Financial development and energy consumption nexus in Malaysia: A multivariate time series analysis," *Econ. Model.*, vol. 30, pp. 435–441, Jan. 2013.

[16] F. E. Tay and L. Cao, "Modified support vector machines in financial time series forecasting," *Neurocomputing*, vol. 48, nos. 1–4, pp. 847–861, Oct. 2002.
[17] D. Zhang, "High-speed train control system big data analysis based on fuzzy RDF model and uncertain reasoning," *Int. J. Comput., Commun. Control*, vol. 12, no. 4, p. 577, Jun. 2017.
[18] W. Xu, L. Liu, Q. Zhang, and P. Liu, "Location decision-making of equipment manufacturing enterprise under dual-channel purchase and sale mode," *Complexity*, vol. 2018, Dec. 2018, Art. no. 3797131.
[19] D. Zhang, J. Sui, and Y. Gong, "Large scale software test data generation based on collective constraint and weighted combination method," *Tehnicki Vjesnik-Tech. Gazette*, vol. 24, no. 4, pp. 1041–1049, Jul. 2017.
[20] W. Xu and Y. Yin, "Functional objectives decision-making of discrete manufacturing system based on integrated ant colony optimization and particle swarm optimization approach," *Adv. Prod. Eng. Manage.*, vol. 13, no. 4, pp. 389–404, Dec. 2018.
[21] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.
[22] J. Y. Chen, "Thrown under the bus and outrunning it! The logic of DiDi and taxi drivers' labour and activism in the on-demand economy," *New Media Soc.*, vol. 20, no. 8, pp. 2691–2711, Aug. 2018.
[23] L. Zhang, J. Lu, J. Zhou, J. Zhu, Y. Li, and Q. Wan, "Complexities' day-to-day dynamic evolution analysis and prediction for a Didi taxi trip network based on complex network theory," *Mod. Phys. Lett. B*, vol. 32, no. 9, Mar. 2018, Art. no. 1850062.
[24] Y. Lu and X. Xiong, "Topic analysis of microblog about 'didi taxi' based on K-means algorithm," *Amer. J. Inf. Sci. Technol.*, vol. 3, no. 3, p. 72, 2019.
[25] J.-C. Kim and K. Chung, "Mining based time-series sleeping pattern analysis for life big-data," *Wireless Pers. Commun.*, vol. 105, no. 2, pp. 475–489, Mar. 2019.

FANG WANG received the Ph.D. degree in statistics from Beijing Jiaotong University, Beijing, China. She is currently a Postdoctoral Researcher with Beijing Jiaotong University. Her main research involves in game theory, machine learning, industrial economics, and industrial security.

MENGGANG LI (Member, IEEE) received the Ph.D. degree in applied economics from Beijing Jiaotong University, Beijing, China. He is currently the Dean of the National Academy of Economic Security, Beijing Jiaotong University, the Director of the Beijing Laboratory of National Economic Security Early-Warning Engineering and the Beijing Philosophy and Social Science Beijing Industrial Security and Development Research Base, and the Chairman of the IEEE Professional Committee in Logistics, Informatics, and Industrial Security System. His current research concerns national economic security, industrial economics, and industrial security.

YIDUO MEI received the B.E. and Ph.D. degrees in computer science and technology from Xi'an Jiaotong University, in 2004 and 2011, respectively. He is currently a Postdoctoral Researcher with the National Academy of Economic Security, Beijing Jiaotong University. His main research interests include blockchain, AI, cloud computing, grid computing, trust management, and big data.



WENRUI LI received the Ph.D. degree in applied economics from Beijing Jiaotong University, Beijing, China. He is currently a Lecturer of industrial economics with Beijing Jiaotong University. His current research involves in industrial economics and industrial security.