OPEN ACCESS
Fractals, Vol. 30, No. 2 (2022) 2240084 (12 pages)

DOI: 10.1142/S0218348X22400849

World Scientific
www.worldscientific.com

FINANCIAL TIME SERIES USING NONLINEAR DIFFERENTIAL EQUATION OF GAUSSIAN DISTRIBUTION PROBABILITY DENSITY

PENGBO WAN,* NAWAF ALHEBAISHI†,§ and QI LIU‡,¶
*Finance School
Hubei University of Economics
Wuhan, P. R. China
†Information System Department
Faculty of Computing and Information Technology
King Abdulaziz University
Jeddah, Saudi Arabia
‡School of Employment and Entrepreneurship
Hubei University of Technology
Wuhan, P. R. China
§nalhebaishi@kau.edu.sa
¶stephanie20058@163.com

Received February 6, 2021
Accepted July 11, 2021
Published March 7, 2022

¶Corresponding author.

Abstract

To further explore the research of financial time series prediction and broaden the application scope of Gaussian distribution probability density equation, based on the nonlinear differential equation of Gaussian distribution probability density, the semi-supervised Gaussian process model is taken as the research object to discuss the application of semi-supervised Gaussian process model in the stock market. Shanghai 180 Index, Shanghai (Securities) Composite Index and the yield of three stocks are studied in detail. The specific results are as follows. The prediction accuracy of Shanghai 180 Index is 83%, and the prediction accuracy of Shanghai (Securities) Composite Index is 78%. The stock yield series curves of the three stocks have the characteristics of peak and thick tail, which do not obey the normal distribution. Among the three models of semi-supervised Gaussian process model, initial Gaussian process model and SVM algorithm model, the prediction accuracy of semi-supervised Gaussian process model is the best, and the prediction accuracy of the yield of three stocks is 82.45%, 85.03% and 84.53%, respectively. The research results can fully prove that the semi-supervised Gaussian process model has good application effect in stock time series prediction. The research content can provide scientific and sufficient reference for the follow-up research of financial time series, and also has important significance for the research of Gaussian process model.

Keywords: Gaussian Distribution; Gaussian Process; Financial Time Series Research; Nonlinear.

1. INTRODUCTION

Due to the development of Internet technology and the gradual deepening of machine learning research, researchers fully analyze many mathematical models and apply them to solve practical problems, which achieve good application results.[1] Among them, Gaussian process (GP) is widely concerned as a very important model.[2] GP model fully combines the relevant theoretical basis of statistics and machine learning, which makes the model have the advantages of both statistical learning method and machine learning method.[3] GP model has a strict basis of statistical learning theory, which is a natural extension of the traditional multivariable Gaussian distribution from vector to function. GP method has strong applicability and generalization ability for high-dimensional nonlinear small sample data. Moreover, in GP, related concept of kernel function is used to make Gaussian function have obvious nonlinear characteristics.[4] In this way, it can ensure that Gaussian function can better deal with many complex and changeable problems, and the calculation cost and the calculation time can be further reduced.[5] With the extensive attention of researchers on GP model, good application results of GP model have been achieved in many fields. For example, there are GP application examples in the aspects of image classification technology, classification theory, target object recognition, fault anomaly detection and target state prediction.[6–8]

At present, with the deepening of the research, the theory of mathematical model is gradually popular. Researchers find that many scenes in modern life can be explained by mathematical models, among which time series model is one of the most common mathematical models.[9] The so-called time series data refers to the dataset arranged according to the time sequence of each observation record, which can also be a special kind of data.[10] Datasets in financial securities, e-commerce, biomedicine, DNA sequencing, atmospheric forecasting and many other fields can be regarded as a time series.[11–13] In the related research in the financial field, the effective storage, analysis, prediction and utilization of financial time series data had become a hot issue.[14] The financial industry has the characteristics of high risk and high income. Therefore, it is of great practical significance to understand the changing law of the financial market, carry out effective financial management and improve the efficiency of financial investment.[15] The use of financial time series prediction can provide sufficient reference for relevant national departments and financial enterprises to make relevant decisions, formulate reasonable capital plans and investment programs, and improve the level of marketing. Therefore, it is particularly important to explore a reasonable model to further improve the performance of financial time series prediction.[16]

Based on the theory and technology of GP, semi-supervised GP (SSGP) is taken as the research model and its mathematical derivation process and model principle is discussed in this exploration. Meanwhile, in the aspect of stock market research, the prediction results of the SSGP model on the Shanghai 180 Index, Shanghai (Securities) Composite Index and three Shanghai stocks are explored, and the prediction accuracy of the SSGP model is analyzed.

2. METHOD

2.1. Basic Concepts of GP

The original GP is also called normal stochastic process. In a random process, if the distribution of random variables at any time is Gaussian normal distribution, the random process is GP.[17] GP can be a finite dimensional or infinite dimensional Gaussian distribution, in which each subset of the infinite dimensional Gaussian distribution conforms to the Gaussian distribution. GP is to establish the prior probability of the model through probability method, in order to learn the kernel function, and finally gives the output function.[18]

With the rapid development of computer network, machine learning and deep learning, researchers have successfully applied the concept of Gaussian distribution to the field of statistics and machine learning. Moreover, with the deepening of the research, its application scope is constantly expanding, which has brought significant impact to various industries. Among them, Gaussian distribution has good application effect and research prospect in the research of time series.[19] The most famous is the application of Kriging method (GP regression). Kriging method is one of the most widely used prediction analysis methods in the field of geographic information, which has good application in remote sensing monitoring, mineral exploration, environmental protection and other aspects.[20] GP is a learning method based on Bayesian theory. Its output is a random vector which obeys joint Gaussian distribution. GP is uniquely determined by the covariance function and the mean function of the set. Equation (1) is its specific expression.

$$f = (f_1, f_2, \ldots, f_n) \sim N(\mu, N), \tag{1}$$

where f_i is a random variable.

For any GP, it can be determined by the mean function and covariance function of a random process $f(x)$.

$$f \sim GP(m(x), K(x, x')), \tag{2}$$

$$m(x) = E[f(x)], \tag{3}$$

$$K(x, x') = E[(f(x) - m(x))(f(x') - m(x'))], \tag{4}$$

where x_i is a random variable, $m(x)$ is a mean function and $K(x, x')$ is a covariance function.

GP is expressed as any combination of function variables and obeys the Gaussian joint distribution with zero mean value. The expression of one-dimensional GP can be expressed as follows:

$$\begin{aligned} &p(f(x)) \sim gp(m(x) = 0, \\ &K(x, x') = \exp(-0.5(x - x')^2)). \end{aligned} \tag{5}$$

In the above equation, the value of the mean function $m(x)$ is 0, and the covariance function is in the form of radial basis function.

GP is a nonparametric kernel method based on Bayesian probability framework, which is mainly used in supervised learning regression and classification problems.[21] At the same time, GP is a set of random variables, and any number of random variables in the set obey the joint Gaussian distribution. The main advantages of GP model are as follows. First, GP is a nonparametric probability model, which can not only predict the output of unknown inputs, but also give the accuracy parameters of the prediction. Then, GP can express the prior knowledge of the process in the form of prior probability, so as to improve the performance of GP model. Compared with neural network, support vector machine (SVM) and other methods, GP model parameters are significantly less, so parameter optimization is relatively easy, and it is easier to converge.[22,23]

2.2. Derivation of Probability Density Function of Gaussian Normal Distribution

Normal distribution is the most widely used continuous probability distribution because its analytical equation of probability density function is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(x-\mu)^2}{2\delta^2}}. \tag{6}$$

In the derivation of Gaussian normal distribution probability density function, the following assumptions need to be determined first. It is known that the function

$$L(\theta) = f(x_1 - \theta) f(x_2 - \theta) \cdots f(x_n - \theta), \tag{7}$$

$$\ln L(\theta) = \ln f(x_1 - \theta) \ln f(x_2 - \theta) \cdots \ln f(x_n - \theta). \tag{8}$$

The maximum value problem of Eq. (7) is equal to the maximum value problem of Eq. (8). Therefore, the left and right sides of Eq. (8) can be derived, and the result is 0. Then, the following expression can be obtained:

$$\frac{f'(x_1 - \theta)}{f(x_1 - \theta)} + \frac{f'(x_2 - \theta)}{f(x_2 - \theta)} + \cdots + \frac{f'(x_n - \theta)}{f(x_n - \theta)} = 0. \tag{9}$$

According to the fact that the arithmetic mean of the observed value is the most reasonable estimate of the true value, it can be obtained that

$$\hat{\theta} = \bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n}. \tag{10}$$

Equation (10) should be the value of the solution obtained in Eq. (9). After the two equations are combined, the following results are obtained:

$$\frac{f'(x_1 - \hat{\theta})}{f(x_1 - \hat{\theta})} + \frac{f'(x_2 - \hat{\theta})}{f(x_2 - \hat{\theta})} + \cdots + \frac{f'(x_n - \hat{\theta})}{f(x_n - \hat{\theta})} = 0. \tag{11}$$

At the same time, the auxiliary function is introduced.

$$g(x) = \frac{f'(x)}{f(x)}. \tag{12}$$

Therefore, the following equation can be obtained:

$$g(x_1 - \hat{\theta}) + g(x_2 - \hat{\theta}) + \cdots + g(x_n - \hat{\theta}) = 0. \tag{13}$$

Since $f(x)$ is an even function, if $n = m + 1$, and $x_1 = x_2 = \cdots = x_m = -x$, $\hat{\theta} = 0$. When the above conditions are brought into Eq. (13), the following equation can be obtained:

$$g(mx) = mg(x). \tag{14}$$

The equation holds for all natural numbers m and real numbers x. In addition, because it is known that the continuous function $g(x)$ is an odd function defined on R, for any natural number and real number, there is $g(mx) = mg(x)$. Then

$$g(x) = cx. \tag{15}$$

That is

$$\frac{f'(x)}{f(x)} = cx. \tag{16}$$

According to the related theorem, it can be obtained that

$$[\ln f(x)]' = \frac{f'(x)}{f(x)}, \tag{17}$$

$$\ln f(x) = \frac{1}{2} cx^2 + c_0. \tag{18}$$

Therefore, the following expression can be obtained:

$$f(x) = e^{\frac{1}{2}cx^2 + c_0} = Me^{\frac{1}{2}cx^2}, \tag{19}$$

where $M = e^{c0}$ and M is greater than 0. The function $f(x)$ decreases with the increase of the absolute value of x, so c must be less than 0. Therefore, it can make

$$c = -\frac{1}{\sigma^2} \ (\sigma > 0). \tag{20}$$

Therefore, the following equation can be obtained:

$$f(x) = Me^{\frac{x^2}{2\sigma^2}}. \tag{21}$$

According to the related theorem, the integral value of the density function on the whole real number is 1, so it can be obtained that

$$\int_{-\infty}^{+\infty} Me^{\frac{x^2}{2\sigma^2}} dx = 1. \tag{22}$$

At the same time, the following assumptions are set in the exploration:

$$t = \frac{x}{\sqrt{2}\sigma}, \tag{23}$$

$$dx = \sqrt{2}\sigma dt. \tag{24}$$

Therefore, the following equation can be obtained:

$$\int_{-\infty}^{+\infty} \sqrt{2}\sigma M e^{-t^2} dt = 1. \tag{25}$$

According to the relevant equation, it can be obtained that

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}. \tag{26}$$

Therefore, it can be obtained that

$$M = \frac{1}{\sqrt{2\pi}\sigma}. \tag{27}$$

In conclusion, the analytical expression of the density function of error normal distribution is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \tag{28}$$

2.3. Regression Prediction Based on GP Model

The basic definition of regression prediction is as follows. In the n pairs of continuous vectors x_i, and y_i of a given training dataset D, for the new input value x, its distribution function value or its output value is predicted, and the prediction probability function and parameter estimation are given. The targets to be studied are Gaussian noise with zero mean value, and the difference from the real input value is ε. Then, the relationship between the target value y to be observed and the potential function can be expressed as follows:

$$y_i = f(x_i) + \varepsilon_i. \tag{29}$$

The expression of likelihood function is as follows:

$$p(y \mid f) = N(f, \sigma^2 I). \tag{30}$$

In GP regression, if the sample data can get its GP prior probability, and the mean value of the multivariate Gaussian normal curve obeyed by several subsets in any range of the sample set is zero, the kernel function is

$$p(f|(x_1)(x_2), \ldots, (x_n)) = N(0, K), \tag{31}$$

where f is the vector and X is the covariance matrix.

The edge distribution of $p(y)$ can be obtained by synthesizing the above equation. The specific equation is as follows:

$$p(y) = \int p(y \mid f) p(f) df = N(y \mid 0, C), \tag{32}$$

where $p(y)$ is subject to the GP regression model, and the elements in the covariance matrix C are denoted as

$$C(x_n, x_m) = K(x_n, x_m) + \sigma^2 \delta_{nm}. \tag{33}$$

For the input x, to predict its specific value y, that is, to predict the probability distribution of a given sample, X_N is recorded as the matrix composed of the input vector of the observed sample, and y_N is the corresponding output value vector of the observed data. The specific expression is as follows:

$$X_N = [x_1, x_2, \ldots, x_n], \tag{34}$$

$$Y_N = [y_1, y_2, \ldots, y_n]^T, \tag{35}$$

$$Y_{N+1} = [y_1, y_2, \ldots, y_n, y_*]^T. \tag{36}$$

Equation (33) is the assumption of GP regression prediction, and the joint probability distribution of the sample data can be obtained as follows:

$$p(y_{N+1}) = N(y_{N+1}|0, C_{N+1}), \tag{37}$$

$$C_{N+1} = \begin{pmatrix} C_N & K \\ K_T & c \end{pmatrix}, \tag{38}$$

$$K = \{k(x_1, x_{N+1}) \cdots (x_n, x_{N+1})\}|n = \{1, 2, \ldots, n\}, \tag{39}$$

$$c = k(x_{N+1}, x_{N+1}) + \sigma_n^2, \tag{40}$$

where C_N represents the covariance matrix of $N*N$, the quantity K is composed of N related elements corresponding to the covariance matrix, and c is a scalar.

For N training samples including input x and output y, the joint probability distribution is as follows:

$$\begin{bmatrix} y \\ f_0 \end{bmatrix} \sim N\left(\begin{bmatrix} K(X,X) + \sigma^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right). \tag{41}$$

Among them, $K(X, X)$ is the $N * N$ order symmetric positive definite covariance matrix; $K(X, X_*)$ is the $N * 1$ order covariance matrix of sample test data points and all observation data X of training set; $K(X_*, X_*)$ is the covariance of test data itself.

Therefore, when the test point and training set D are given, according to the Bayesian posterior probability, the conditional distribution can be deduced as follows:

$$p(f_*|x_*, X, y_N, k) \sim N(\bar{f}_*, \mathrm{cov}(f_*)). \tag{42}$$

The expressions of mean and variance are shown in the following equation:

$$\bar{f} = E[f_*|X, y, X_*], \tag{43}$$

$$\mathrm{cov}(f_*) = K(X_*, X_*) - K(X_*, X) \times [K(X,X) + \sigma^2 I]^{-1} K(X, X_*). \tag{44}$$

2.4. Nonlinear Analysis in Financial Stock Market

Among the core analysis methods of modern financial market, there are many traditional linear analysis methods such as efficient market theory and CAPM model.[24] These techniques are to use the method of linear mathematical statistics to guide financial investment, and have a certain application effect in financial market research. However, with the deepening of research, the traditional linear analysis method is unable to meet the existing complex situation, and cannot explain problems such as the US stock market crash.[25] Compared with linear system, nonlinear system is more complex. In the related research in the field of financial market, the introduction of nonlinear model can not only deeply understand and analyze the relevant theoretical knowledge of financial market, but also interpret and analyze more abstract economic problems through model prediction, parameter comparison, system simulation and other means.[26] At the same time, prediction of the corresponding problems can be put forward and the corresponding solutions can be given.

Many studies show that China's stock market has a very significant volatility.[27] Specifically, the volatility of China's stock market is mainly reflected in three aspects: state persistence, correlation and volatility periodicity. The main performance is as follows. First, in China's stock market, it often shows a continuous upward or downward trend. If the sequence is up in the previous interval, it is also up or down in the next interval. Therefore, in practice, the stock market often presents a continuous upward or downward trend.[28] Then, in practical application, it is found that the correlation scale of Shanghai and Shenzhen stock markets is usually not equal to zero, which proves the correlation of stock price volatility. Moreover, the fluctuation of stock price is affected by historical information to a certain extent. Therefore, investors tend to pay more attention to stock history information when they invest.[29] Finally, the current event information can affect the future stock trend for a long time, and the measurement of cycle length refers to how long it takes for the impact of a single period to be reduced to a degree that cannot be measured.[30]

In the research of stock market in China, the most commonly used theory is chaos theory. The so-called chaos refers to a seemingly irregular and random phenomenon in the deterministic system.[31] Chaos is not a simple disorder, but has no obvious period and symmetry. It has an ordered structure with rich internal levels. It is a new form of existence in nonlinear systems. The main definitions of chaos theory are as follows.

First, V is set as a compact metric space. In this case, it is assumed that V satisfies the following conditions:

(1) Sensitive dependence on initial value: If there is σ, any $\varepsilon > 0$ and $x \in V$, there is y and natural logarithm n in the field of x, which make:

$$d(f^n(x), f^n(y)) > \delta. \tag{45}$$

(2) Topological transitivity: for any pair of open sets X and Y on V, there is $k > 0$, which make:

$$f^k(X) \cap Y = \varnothing. \tag{46}$$

According to the above conditions, the sensitive dependence on the initial value means that no matter how close x and y are, their orbits may be far apart under the action of f. In the vicinity of each point x, the point y which is very close to it but finally separated can be found. Based on the above situation, if the orbit of f is calculated by computer, any small initial error will lead to the failure of the calculation results after several iterations.

2.5. Theoretical Analysis and Model Discussion of Financial Time Series Prediction

According to the different methods and theories of models, financial time series prediction methods can be divided into linear and nonlinear methods. At present, the widely used linear methods include moving average method, exponential smoothing method, multiple regression method and autoregressive conditional heteroscedasticity method; the commonly used nonlinear prediction methods include nonlinear prediction method based on phase space reconstruction, artificial neural network method, wavelet method, SVM prediction method and nonlinear tracking method.[32]

There are three most important characteristics of linear financial time series: stationarity, autocorrelation and volatility aggregation test. Stationarity test is the basis of time series analysis. Different modeling methods are selected according to whether the time series are stationary or not.[33] Time series cannot be described only through one or more random variables, and its dynamic change process must be investigated. The dynamic change process of this kind of random phenomenon is called random process, and the financial time series is a random process. Autocorrelation mainly refers to dynamic correlation, that is, the correlation of the same variable in different time domains. Volatility aggregation means that volatility is larger in a certain period of time and smaller in another period of time. Although the volatility of financial time series cannot be directly observed, it has generally accepted characteristics. For example, volatility aggregation, leverage effect, continuous changes of volatility over time and stable volatility.

The nonlinear characteristics of financial time series are mainly reflected in the following aspects. The most important thing is that the change of financial time series data flow has strong positive correlation, long memory and self-similarity, which indicates that the trend change of financial time series has great correlation with its own historical data.[34] Then, information is usually presented in a nonlinear way, and financial investors usually process the information in a nonlinear way. At the same time, the change of financial market will be reflected in the price of financial products through the trading activities of financial market, which makes the price change of financial products nonlinear. In nonlinear time series theory, there are many analysis models, including SVM model.[35] SVM is developed based on statistical theory. The core idea is to map the nonlinear sample data into high-dimensional space, so that it can be transformed into classification or regression problem which can be dealt by linear method. That is to say, the nonlinear problem in low-dimensional state is transformed into linear problem in high-dimensional state by nonlinear transformation. Figure 1 is the specific feature form. This exploration will also focus on the application effect of SVM model.

SVM function mainly uses kernel function as l stock sample data with independently identically distribution, so that the main expressions for predicting the expected risk are as follows:

$$Q(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j (x_i, y_i), \quad (47)$$

$$f(x) = \mathrm{sgn}\{(w \cdot x) + b\}, \quad (48)$$

where $f(x)$ is the decision function; $Q(a)$ is the loss function; a is the weight vector.

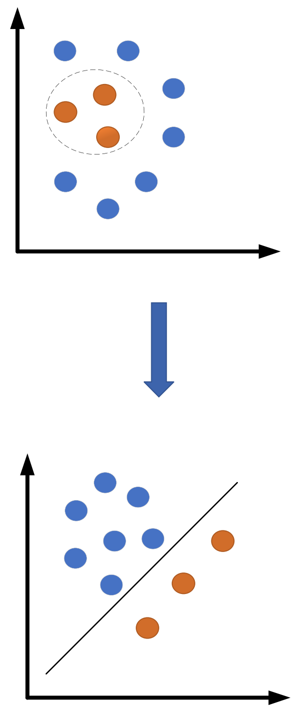

Fig. 1 Solution idea of SVM model.

SVM is composed of three layers, in which the input layer is used to obtain the original stock sample data; the middle layer is used to map the nonlinear sample data in the low-dimensional state to the high-dimensional feature space by kernel function, so that it is linearly separable; the top layer is used to construct the decision function $f(x)$. Figure 2 shows the structure of SVM.

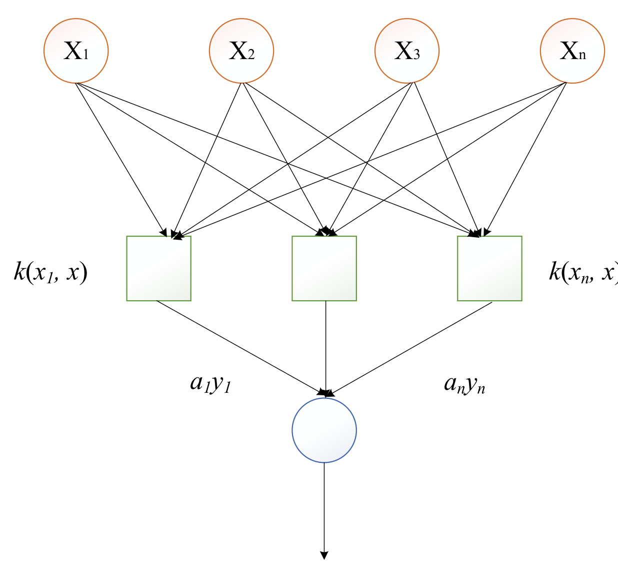

Fig. 2 SVW structure diagram.

2.6. Financial Time Series Prediction Model Based on Semi-Supervised Gaussian Distribution

Common learning methods include supervised learning, semi-supervised learning and unsupervised learning. Semi-supervised learning is not only learning from training data with labels, but also developing structural information from other unlabeled data.[36] It improves the performance of the classifier through the joint probability distribution of labeled data and unlabeled data. Semi-supervised learning has the following advantages. First, it can effectively use domain specific knowledge to assist analysis; then, it can improve the effect of data classification; at the same time, it can classify special shapes or standards according to different tag information.[37] Users can use less known information to guide the classification process, which can better reflect their needs. Figure 3 shows the relationship between the three learning methods and the sample dataset.

The main time series prediction method in this exploration is SSGP time series classification algorithm. The calculation form is as follows:

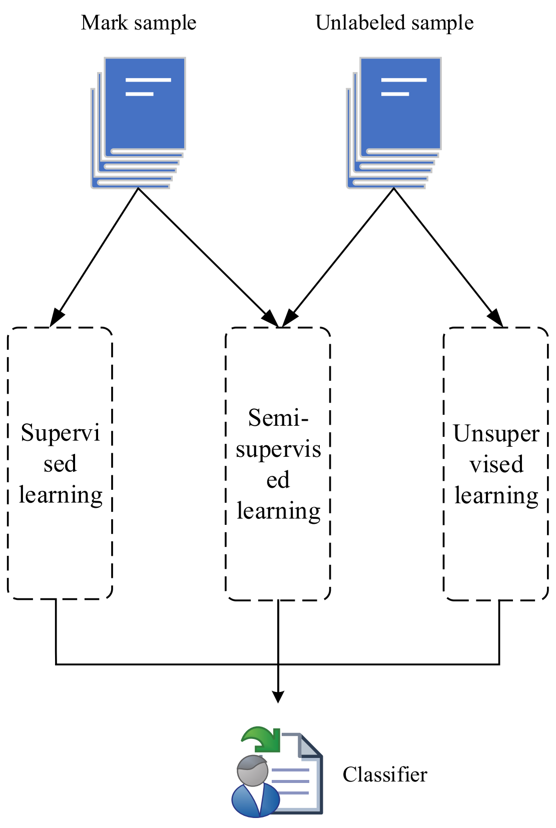

Fig. 3 Relationship between data sample set and learning style.

First, dataset $X = \{X_l, X_u\}$ is given, where X_l is the labeled data sample and X_u is the unlabeled data sample. There is the following expression:

$$f = \arg\min \left\{ \frac{1}{l} \sum_{i}^{l} C(f, xi, yi) + \Omega(\|f\|) \right\}. \quad (49)$$

Among them, the C-cost function represents the matching degree between the function f and the given data sample.

For semi-supervised learning, it is necessary to consider both labeled data and unlabeled data. Therefore, it can be obtained that

$$f = \arg\min \left\{ \frac{1}{l} \sum_{i=1}^{l+u} C(f, xi, yi) + \Omega(\|f\|) \right\}. \quad (50)$$

At the same time, for SSGP, there is also the problem of hyperparameter learning. First, the logarithmic marginal likelihood function needs to be Laplacian approximated. The specific equation is as follows:

$$L = \log p(y \mid \hat{f}, \theta) - \frac{1}{2}\hat{f}K^{-1}\hat{f} - \frac{1}{2}\log|B|. \quad (51)$$

The maximum value of θ can be obtained by calculation

$$\theta^* = \arg\max_{\theta} \log q(y|X, \theta). \quad (52)$$

In conclusion, it can be concluded that the functional dependence of the approximate marginal likelihood function on the hyperparameter θ consists of two parts.

$$\frac{\partial \log p(y|X,\theta)}{\partial \theta_j} = \frac{\partial \log p(y|X,\theta)}{\partial \theta_j} + \sum_{i}^{n} \frac{\partial \log q(y|X,\theta)}{\partial \theta_j} \frac{\partial \hat{f}_i}{\partial \theta_j}. \quad (53)$$

The classification principle of SSGP algorithm is to use semi-supervised information to inject the corresponding class label information into unlabeled data, and continuously expand the set of labeled data samples to construct a reasonable classifier. The prediction results of unlabeled data samples can be used to guide the next data classification and prediction. Figure 4 is the main schematic diagram.

2.7. Research Data Sources and Test Environment

In this exploration, Shanghai 180 Index prediction analysis and Shanghai (Securities) Composite Index

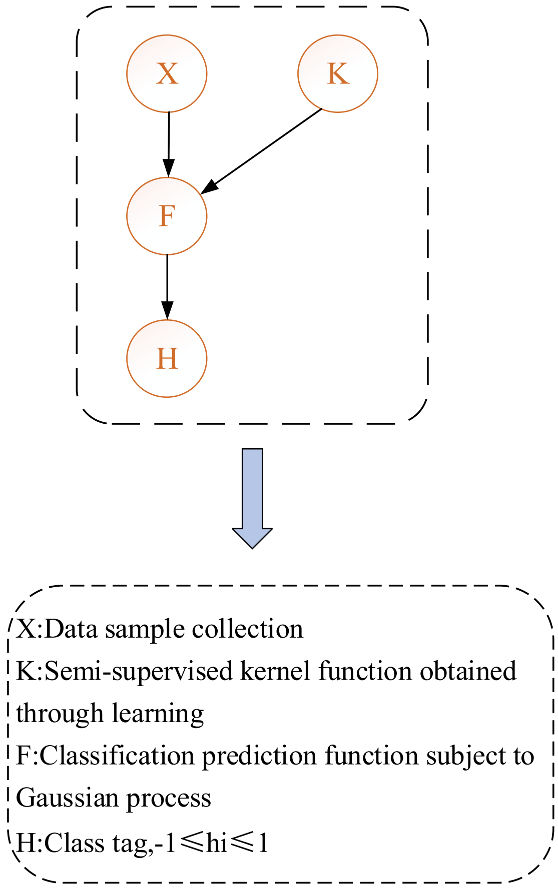

Fig. 4 Principle of SSGP classification and prediction.

prediction analysis are mainly carried out. At the same time, three stocks in Shanghai Stock Exchange are selected to study and analyze their stock yield. The Shanghai 180 Index, also known as the SSE Constituent index, is a result of the Shanghai Stock Exchange's adjustment and renaming of the original Shanghai index. Its sample stock is the most representative sample stock selected from all the stocks in the market. It has been officially released since July 1, 2007. The Shanghai (Securities) Composite Index is a weighted composite stock price index compiled by the Shanghai Stock Exchange, which takes all the stocks listed on the Shanghai Stock Exchange as the calculation range and takes the issuance as the weight.

The Shanghai 180 Index selected in this exploration is from the data research of Shanghai Stock Exchange (http://www.sse.com.cn/) from July 15, 2014 to December 31, 2015, with a total of 100 data between 171 and 270. The Shanghai (Securities) Composite Index data selected in this exploration are also the data of Shanghai Stock Exchange (http://www.sse.com.cn/) from July 15, 2014 to December 31, 2015. A total of 100 data of Shanghai (Securities) Composite Index between 431 and 530 are predicted. The data of three stocks of Shanghai Stock Exchange analyzed by in this exploration is from July 15, 2015 to December 31, 2015.

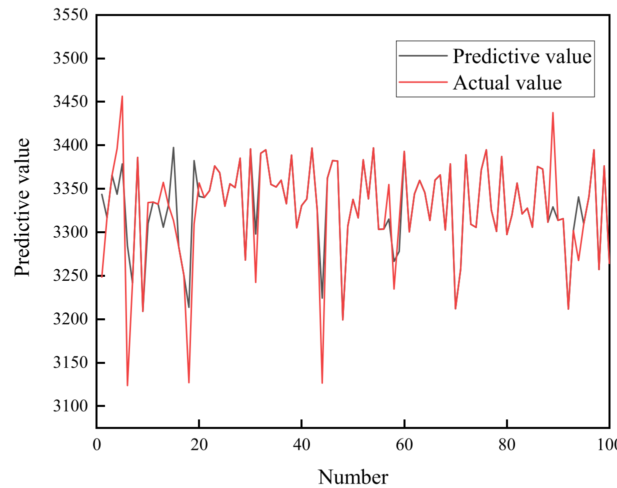

Fig. 5 Prediction results of Shanghai 180 Index.

3. RESULTS AND DISCUSSION

3.1. Prediction Analysis of Shanghai Stock Index

(1) Prediction analysis of Shanghai 180 Index

Figure 5 shows the prediction results of Shanghai 180 Index. The black line represents the predicted value of 100 data, and the red line represents the actual value. The comparison with the actual value shows that the prediction accuracy of Shanghai 180 Index can reach 83% by using Gaussian distribution probability density equation. Experiments show that the prediction effect is good.

(2) Prediction analysis of Shanghai (Securities) Composite Index

Figure 6 shows the prediction results of the Shanghai (Securities) Composite Index. Similarly, the black line represents the predicted value of the Shanghai (Securities) Composite Index data, while the red line represents the actual value of the Shanghai (Securities) Composite Index. In the prediction of Shanghai (Securities) Composite Index, after Gaussian distribution probability density equation is used, the prediction accuracy of Shanghai (Securities) Composite Index is slightly lower than that of Shanghai 180 Index, which is 78%.

Moreover, the research results reveal that the selection of input variables has a very important impact on the prediction results. In fact, the input

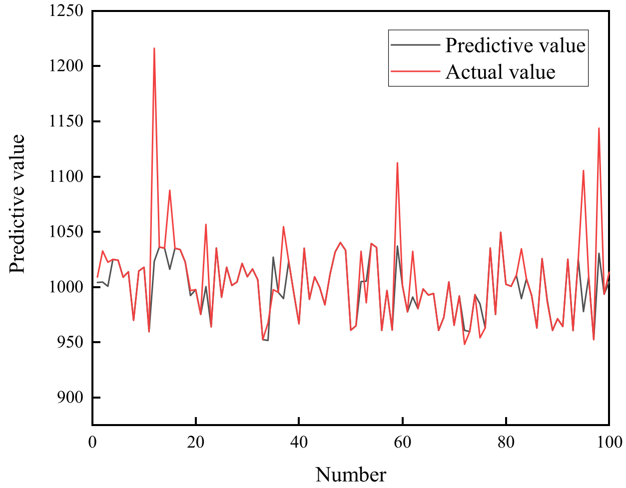

Fig. 6 Prediction results of Shanghai (Securities) Composite Index.

variables refer to the basic factors that affect the change of stock price. The choice of input variables reflects the choice of influencing factors of stock price. Different input variables will lead to different prediction results.

3.2. Stock Yield Analysis

In this exploration, the distribution of the yield of three stocks is mainly tested. Figure 7 shows the specific results. Among them, Fig. 7a shows the distribution and fitting curve of A stock yield; Fig. 7b shows the distribution and fitting curve of B stock yield; Fig. 7c shows the distribution and fitting curve of C stock yield.

Figure 7 shows that the prediction effect is good when the data sample size is 100. At the same time,

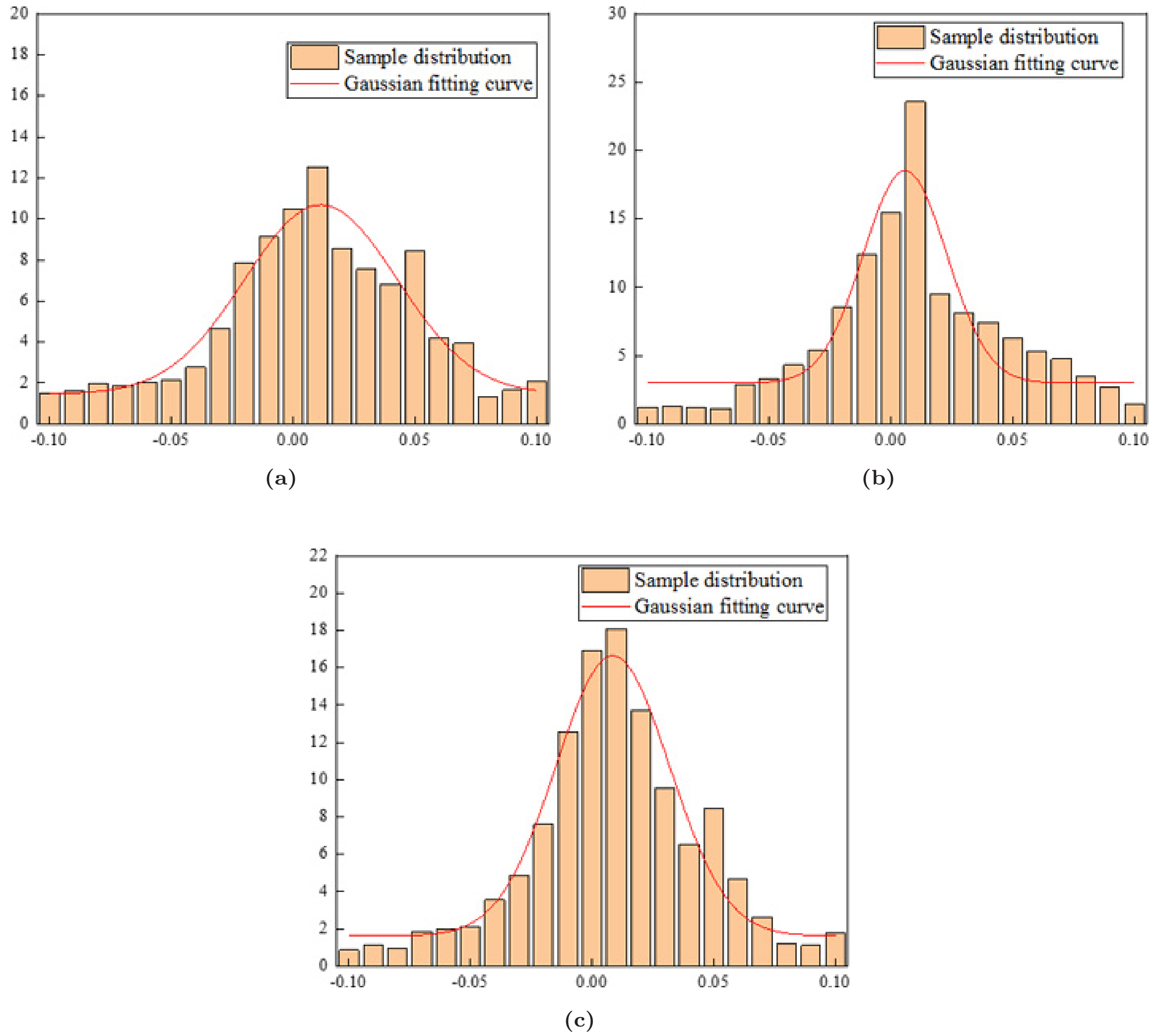

Fig. 7 The fitting of the stock yield of three stocks.

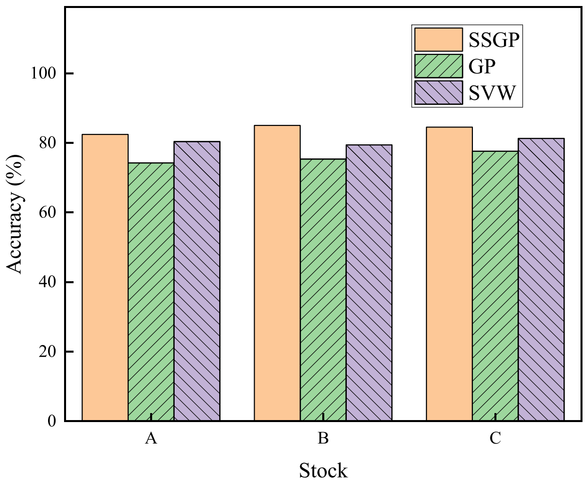

Fig. 8 Comparative analysis of stock prediction accuracy using different methods.

after Gaussian fitting, the prediction results of A, B and C stocks show the characteristics of peak, and the stock yield has the characteristics of right deviation. To sum up, the stock yield series has the characteristics of peak and thick tail, and does not follow the normal distribution.

3.3. Comparative Analysis of SSGP Model and Other Methods

In this exploration, SSGP model, original GP model and SVM algorithm model are studied respectively, and the accuracy results of yield of three stocks are obtained and compared. Figure 8 shows the specific results.

Figure 8 shows that three methods are used to predict the stock yield. Among them, the prediction accuracy of SSGP model for three stocks is 82.45%, 85.03% and 84.53%, respectively; the prediction accuracy of original GP model for three stocks is only 74.28%, 75.37% and 77.59%, respectively; the prediction accuracy of SVM model for three stocks is 80.39%, 79.43% and 81.31%, respectively. According to the above results, the prediction accuracy of SSGP model is the highest, which is significantly better than the original GP model and SVM model. It proves that SSGP model has good application effect.

4. CONCLUSION

In this exploration, the relevant prediction results of the financial time series by using the Gaussian probability density nonlinear differential equation (SSGP equation) are fully discussed. The specific research conclusions are as follows. First, the Shanghai 180 Index and the Shanghai (Securities) Composite Index are predicted. The results show that the supervised GP equation can effectively predict the two indexes. The prediction accuracy of Shanghai 180 Index is 83%, and that of Shanghai (Securities) Composite Index is 78%. It is considered that the selection of input variables can directly affect the prediction results. Then, the yield of three stocks is fitted. The results show that the curve of yield series has the characteristics of peak and thick tail, so it does not follow the normal distribution. Finally, the accuracy of SSGP model, original GP model and SVM algorithm model for the prediction of the yield of three stocks is analyzed, respectively. The experimental results show that the prediction accuracy of SSGP model is better than the original GP model and SVM model. Under the SSGP model, the prediction accuracy of the yield of three stocks is 82.45%, 85.03% and 84.53%, respectively. It is proved that SSGP model has good application effect in stock time series prediction. Meanwhile, there are still some problems need to be improved in this exploration. In today's big data era, the demand for the data amount in data analysis increases, which puts forward higher requirements for related technical methods. Therefore, how to continue to improve the accuracy of the algorithm in the case of large sample size will be further explored, which is also a key issue in the follow-up research.

REFERENCES

1. G. De *et al.*, GPflow: A Gaussian process library using tensor flow, *J. Mach. Learning Res.* **18**(1) (2017) 1299–1304.
2. Z. S. Ye and N. Chen, The inverse Gaussian process as a degradation model, *Technometrics* **56**(3) (2014) 302–311.
3. M. F. Huber, Recursive Gaussian process: On-line regression and learning, *Pattern Recognit. Lett.* **45** (2014) 85–91.
4. A. Banerjee, D. B. Dunson and S. T. Tokdar, Efficient Gaussian process regression for large datasets, *Biometrika* **100**(1) (2013) 75–89.
5. R. Bostanabad *et al.*, Leveraging the nugget parameter for efficient Gaussian process modeling, *Int. J. Numer. Methods Eng.* **114**(5) (2018) 501–516.
6. H. Liu, J. Cai and Y. S. Ong, Remarks on multi-output Gaussian process regression, *Knowl. Based Syst.* **144** (2018) 102–121.
7. S. Wu, Internet public information text data mining and intelligence influence analysis for user intent

understanding, *J. Intell. Fuzzy Syst.* **38** (2020) 487–494.
8. A. Denzel and J. Kästner, Gaussian process regression for geometry optimization, *J. Chem. Phys.* **148**(9) (2018) 094114.
9. V. Smirnov and D. Volchenkov, Five years of phase space dynamics of the Standard & Poor's 500, *Appl. Math. Nonlinear Sci.* **4**(1) (2019) 203–216.
10. J. Xu *et al.*, Sequentially site-specific delivery of thrombolytics and neuroprotectant for enhanced treatment of ischemic stroke, *ACS Nano* **13**(8) (2019) 8577–8588.
11. J. Chen *et al.*, Sequentially responsive shell — stacked nanoparticles for deep penetration into solid tumors, *Adv Mater.* **29**(32) (2017) 1701170.
12. M. A. Balc, Fractional interaction of financial agents in a stock market network, *Appl. Math. Nonlinear Sci.* **5**(1) (2020) 317–336.
13. N. Nce and A. Shamilov, An application of new method to obtain probability density function of solution of stochastic differential equations, *Appl. Math. Nonlinear Sci.* **5**(1) (2020) 337–348.
14. K. Nakagawa *et al.*, Sequentially timed all-optical mapping photography (STAMP), *Nat. Photonics* **8**(9) (2014) 695–700.
15. K. Gao *et al.*, Over 12% efficiency nonfullerene all – small – molecule organic solar cells with sequentially evolved multilength scale morphologies, *Adv. Mater.* **31**(12) (2019) 1807842.
16. S. Wu, Research on the application of spatial partial differential equation in user oriented information mining, *Alexandria Eng. J.* **59**(4) (2020) 2193–2199.
17. L. Lin and D. B. Dunson, Bayesian monotone regression using Gaussian process projection, *Biometrika* **101**(2) (2014) 303–317.
18. H. Rong, A. P. Teixeira and C. G. Soares, Ship trajectory uncertainty prediction based on a Gaussian Process model, *Ocean Eng.* **182** (2019) 499–511.
19. P. Li and S. Chen, A review on Gaussian process latent variable models, *CAAI Trans. Intell. Technol.* **1**(4) (2016) 366–376.
20. C. Park and D. Apley, Patchwork kriging for large-scale gaussian process regression, *J. Mach. Learning Res.* **19**(1) (2018) 269–311.
21. R. Frigola *et al.*, Bayesian inference and learning in Gaussian process state-space models with particle MCMC, *Adv. Neural Inf. Processing Syst.* **26** (2013) 3156–3164.
22. A. Stuart and A. Teckentrup, Posterior consistency for Gaussian process approximations of Bayesian posterior distributions, *Math. Comput.* **87**(310) (2018) 721–753.
23. S. Yiu and K. Yang, Gaussian process assisted fingerprinting localization, *IEEE Internet of Things J.* **3**(5) (2015) 683–690.
24. A. N. Ajmi *et al.*, How strong are the causal relationships between Islamic stock markets and conventional financial systems? Evidence from linear and nonlinear tests, *J. Int. Finan. Markets Inst. Money* **28** (2014) 213–227.
25. L. J. Kao *et al.*, Integration of nonlinear independent component analysis and support vector regression for stock price forecasting, *Neurocomputing* **99** (2013) 534–542.
26. S. Wu, An user intention mining model based on fractal time series pattern, *Fractals* **28**(8) (2020) 2040017.
27. E. Bouri *et al.*, Cointegration and nonlinear causality amongst gold, oil, and the Indian stock market: Evidence from implied volatility indices, *Resour. Policy* **52** (2017) 201–206.
28. Z. X. Ni, D. Z. Wang and W. J. Xue, Investor sentiment and its nonlinear effect on stock returns — New evidence from the Chinese stock market based on panel quantile regression model, *Econ. Modelling* **50** (2015) 266–274.
29. T. Choudhry, F. I. Papadimitriou and S. Shabi, Stock market volatility and business cycle: Evidence from linear and nonlinear causality tests, *J. Banking Finan.* **66** (2016) 89–101.
30. N. Naifar and M. S. Al Dohaiman, Nonlinear analysis among crude oil prices, stock markets' return and macroeconomic variables, *Int. Rev. Econ. Finan.* **27** (2013) 416–431.
31. S. Ghosh and K. Kanjilal, Co-movement of international crude oil price and Indian stock market: Evidences from nonlinear cointegration tests, *Energy Econ.* **53** (2016) 111–117.
32. S. Wu, Design of interactive digital media course teaching information query system, *Inf. Syst. e-Business Manage.* **18**(4) (2020) 793–807.
33. Z. Guo *et al.*, A feature fusion based forecasting model for financial time series, *PloS One* **9**(6) (2014) e101113.
34. S. A. Hamid and A. Habib, Financial forecasting with neural networks, *Acad. Accounting Finan. Stud. J.* **18**(4) (2014) 37.
35. G. Zhiqiang, W. Huaiqing and L. Quan, Financial time series forecasting using LPP and SVM optimized by PSO, *Soft Comput.* **17**(5) (2013) 805–818.
36. M. Kim, D. Lee and H. Shin, Semi-supervised learning for hierarchically structured networks, *Pattern Recognit.* **95** (2019) 191–200.
37. A. Shvets and A. Makaseyev, Deterministic chaos in pendulum systems with delay, *Appl. Math. Nonlinear Sci.* **4**(1) (2019) 1–8.