ATLANTIS PRESS
Advances in Economics, Business and Management Research, volume 110
5th International Conference on Economics, Management, Law and Education (EMLE 2019)

Research on Economic Forecast of High-Tech Park Based on Combination Model

Taking the Zhongguancun Demonstration Zone as an Example*

Xiaofeng Wu
School of Economics and Management
Beijing Jiaotong University
Beijing, China

Da Li
School of Economics and Management
Beijing Jiaotong University
Beijing, China

Yanan Jin
School of Economics and Management
Beijing Jiaotong University
Beijing, China

Yaotong Chen
School of Economics and Management
Beijing Information Science and Technology University
Beijing, China

***Abstract*—In recent years, the domestic high-tech park economy has achieved rapid development. The park economic forecast is the basis for formulating macroeconomic policies and improving risk management capabilities. The traditional time series forecasting model is more and more difficult in the increasingly complex national economic development to meet the needs of economic forecasting. Based on this, an EMD-SVM based on high-tech park total revenue error correction combined forecasting model is established. The data were selected from the 2009-2017 Zhongguancun High-tech Park total revenue monthly data as a sample to verify. The results show that the model can provide more effective predictions for the operational trends under the characteristics of economic fluctuations in China's high-tech parks.**

***Keywords*—high-tech park; economic forecast; combined model**

I. INTRODUCTION

In recent years, China's high-tech parks have been intellectually intensive, and have achieved rapid development with open environmental conditions. Their development speed is much higher than that of the national economic development in the same period [1]. The total assets of the national high-tech industry increased from 2,949.97 billion yuan in 2009 to 136,337.0 billion yuan in 2016, with an average annual growth rate of 26.8%. The high-tech park has gradually become a new growth point for China's national economic development and a base for promoting the development of China's high-tech industries, which has had a major impact on China's economic development. Therefore, in order to continue to promote the sustainable and healthy development of high-tech parks and correctly predict the economic operation of the park, it is particularly urgent to effectively predict the economic development of the park.

At present, domestic and foreign scholars have carried out a large number of researches on economic forecasting and have achieved fruitful results. The methods used can be divided into two categories: data-driven models and data mining models [2]. The data-driven model focuses on analyzing and simulating historical economic data and predicting future trends, including time series models such as ARIMA, ARCH, GARCH, and TGARCH. Cai Chengzhi used the ARIMA model to predict the world soybean yield [3]; Ling Zhenghua used the ARCH model to study the fluctuation characteristics and impact of China's egg futures price [4]; Shen Shichang used the GARCH model to conduct the Chinese health insurance market [5]. The risk measurement; Xie Chi applied the TGARCH model to the research of securities investment strategy and achieved effective results [6]. Data mining models, such as chaos theory, grey theory, neural networks, and support vector machines (SVM), can extract implicit and valuable information from a large number of fuzzy random data, and have been increasingly cited in time series. Forecasting to solve the non-stationary, non-linear characteristics implied in the data. Li Qing applied chaos theory in the study of China's economic system [7]; Tian Yuchen used the improved GM model to predict the running trend of GDP [8]; Huang Yang and Lu Shichang used neural networks respectively[9][10]. The CPI is predicted with the support vector machine. Among them, the support vector machine method based on statistical learning theory has the advantages of effectively reducing the error interval and reducing the structural risk of the model in time series prediction, and ensuring that the sample prediction error is small. In view of the strong noise characteristics of the economic development time series, in recent years, many scholars have introduced the SVM method into economic forecasting analysis. Zheng Yanqiu used SVM to evaluate the financing risk [11]; Wu Mu used SVM to predict the sales of cigarettes [12]; Ping Weiying

*Fund: This article is in the National Social Science Fund Major Project "Research on the Framework and Evolution Path of China's High-speed Rail Economy Theory" (NO. 17ZDA084).

used SVM to evaluate the development of China's provincial circular economy [13]. These SVM-based improved prediction methods have significantly improved the prediction accuracy compared with the traditional prediction methods. However, the existing methods still cannot effectively solve the shortcomings of the SVM method prediction error, which affects the prediction accuracy of the model.

Aiming at the above problems, this paper establishes an error correction prediction model based on EMD-SVM. That is to say, based on the prediction of SVM method, the EMD method is introduced, the error obtained by SVM prediction is decomposed into modal components of different frequencies, and these components are trained by SVM method, and the predicted values of errors are superimposed to use Correct the preliminary SVM predictions. The model uses the error correction method to solve the problem of increasing error, which improves the prediction accuracy. The monthly data of the accumulated total revenue of Zhongguancun High-tech Park in 2009-2017 was selected for empirical simulation. The EMD-SVM model was used to reconstruct the total revenue forecast of Zhongguancun High-tech Park, and it was combined with SVM, EMD-BP and BP methods. Comparing the prediction results, the root mean square error (RMSE) and the mean absolute error (MAE) are smaller than the other three models, and the prediction effect is obviously better.

II. THEORETICAL MODEL DESIGN

A. *Empirical Mode Decomposition (EMD)*

In 1998, American Chinese scientists N.E.Huang et al. proposed an EMD (Empirical Mode Decomposition) method based on instantaneous frequency signal processing [14]. In essence, the method smoothes a signal, decomposes the fluctuations or trends of different scales existing in the original data step by step, and generates a series of sequences with different wave frequencies. Each sequence is called an eigenmode function (IMF), and the lowest frequency IMF is called the residual term R(n), which contains information about the trend or average of the original data. Each IMF component highlights the original data. The local features of the sequence have obvious physical meaning.

Each IMF decomposed by the EMD method can be linear or non-linear, but must meet the following conditions: first, the extreme points of the original sequence curve (maximum point, minimum point) and the intersection with the abscissa The number is the same or at most one difference; second, at any time, the upper envelope defined by the maximum value point and the minimum envelope defined by the minimum value point are zero, and the upper and lower envelopes of the signal surround the time. Axisymmetric. The process of decomposing a data sequence into several IMFs using the EMD method is also referred to as a screening process. Specific steps are as follows:

First, determine the original signal: all local extremum points in x(t). All the local maximum points and minimum points are fitted by the cubic spline function to obtain the upper and lower envelopes xup(t) and xlow(t) of the original signal.

Second, calculate the mean m(t) of the upper and lower envelopes

$$\mathrm{m}(t)=[\chi_{up}(t)+x_{low}(t)]/2 \quad (1)$$

Third, calculate the new series h(t)

$$h(t)=x(t)-m(t) \quad (2)$$

If h(t) meets the condition that the IMF component is true, then h(t) is an IMF component of the original signal, denoted as c1(t). If h(t) does not meet the condition that the IMF component is true, h(t) is used. Repeat the above process instead of the original signal x(t) until the c1(t) that satisfies the component condition is selected

Fourth, for the remainder:

$$r_1(t)=x(t)-c_1(t) \quad (3)$$

Using r1(t) as the original signal, the steps of (1), (2), and (3) are repeated to obtain the second IMF component c2(t), which is repeated n times, thereby obtaining all n IMF components. This has:

$$\left.\begin{array}{l} r_1(t)-c_2(t)=r_2(t) \\ r_{n-1}(t)-c_n(t)=r_n(t) \end{array}\right\} \quad (4)$$

Fifth, the above decomposition process is stopped when the component cn(t) or the residual component rn(t) is smaller than a predetermined value, or when the remaining component rn(t) is not screened for the eligible IMF component.

Thus, the original sequence x(t) is decomposed into n IMFs and one residual component rn(t).

$$x(t)=\sum_{i=1}^{n}c_i(t)+r_n(t) \quad (5)$$

B. *Support Vector Machine (SVM) Principle*

SVM (Support Vector Machines) is a statistical theory-based method of instructor learning proposed by Cortes and Vapnik that can solve classification problems and solve regression problems [15]. Its mechanism is to construct a classification hyperplane to achieve the division between samples, and the core of SVM is to find an optimal classification hyperplane, so that the shortest distance between the two types of samples and the hyperplane is maximized. According to the training set {xi, yi}, where xi ∈RD (xi contains D feature attributes), i=1, 2...n, is n D-dimensional vectors, yi ∈ R, F={f｜f : RD→ R}. For nonlinear classification problems that cannot satisfy linear separability, SVM can map the linear inseparable problem of the original input training samples in a low-dimensional space to a high-dimensional space to achieve linear separability.

The training sample xi satisfies the following conditions:

$$y_i(x_i\cdot\omega+b)-1+\xi_i\geq 0,\xi_i\geq 0,i=1,\cdots n \quad (6)$$

ξ i is called the loose her variable. When 0< ξ i<1, the sample xi is correctly classified; when ξ i≥1, the sample xi is misclassified. To do this, add the penalty C $\sum_{i=1}^{n}\xi_i$ to the minimized target $\frac{1}{2}\|\omega\|^2$ and introduce the following objective function:

$$\Phi(\omega,\xi)=\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}\xi_i \quad (7)$$

C denotes a penalty factor that balances the magnitude of the slack variable and the classification interval and controls the tolerance to the abnormal sample. The larger C, the more attention is paid to the item, and the lower the tolerance for classification errors. In this way, the problem of maximizing the shortest distance between the classified sample and the hyperplane is transformed into solving the following optimization problem:

$$\begin{cases} \min\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}\xi_i \\ s.t. y_i(x_i\cdot\omega+b)-1+\xi_i\geq 0, \xi_i\geq 0, i=1,\cdots n \end{cases} \quad (8)$$

The above formula is the optimal final discriminant function:

$$f(x)=\text{sgn}(\sum_{i=1}^{n}a^* y_i K(x_i,x)+b^*) \quad (9)$$

K(xi,x) is a kernel function, the role of the kernel function: (1) maps the linear indivisible problem in low-dimensional space to a high-dimensional space to achieve linear separability; (2) mapping to high-dimensional space will make it between samples The complexity of the product calculation increases, and another function of the kernel function is to make the calculation amount in a low-dimensional capacity, but the output is in a high-dimensional space. This paper chooses the radial basis function as its kernel function:

$$K(x_i,x)=\exp(\frac{-\|x_i-x\|^2}{\sigma^2}), \quad i=1,2\cdots n \ (10)$$

σ is the nuclear density.

C. *EMD-SVM Integrated Prediction Model*

The EMD-SVM integrated model algorithm is divided into four parts: total revenue forecasting, error decomposition, quantity forecasting and integration, total revenue forecasting and error forecasting reorganization. The prediction algorithm design flow of the EMD-SVM integration model is shown in "Fig. 1".

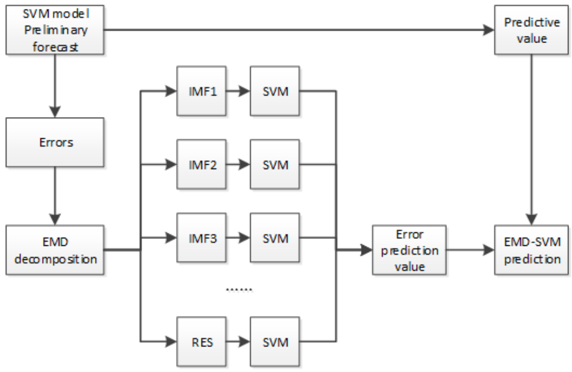

Fig. 1. EMD-SVM integrated model prediction algorithm design flow.

In the process of predicting and analyzing the high-tech park economy, the SVM method is used to make a preliminary prediction of the original sequence. The obtained prediction results are compared with the original sequence because of the error between the prediction result and the actual value, the impact prediction Precision. In order to improve the prediction effect of the model, this paper uses the EMD method to decompose the error value and simulate and predict each component, and then superimpose the result to obtain the predicted value of the error. Finally, the error prediction value is fed back to the SVM. The results of the preliminary forecast are used to obtain the predicted value of the final high-tech park economy. The specific steps of the model are described as follows:

First is data preprocessing. First, the selected high-tech park sample data sequence {x(t), t=1, 2,. . . , n} is transformed into a matrix form, and the sample data (X(t), Y(t)) is constructed, where X(t)={x(t-m), x(t-m+1),. . . , x(t-1)}, Y(t)=x(t), where m represents the (m+1) value predicted by the first m values. The sample data sequence is segmented into training data set I1 and test data set I2, and spatial reconstruction is performed on the two data sets according to the spatial reconstruction principle.

Second is preliminary prediction of sample data. The training data set I1 is simulated by the SVM method, and the preliminary predicted value P(I2) of the data set I2 is obtained.

Third is EMD decomposition and prediction of error terms. Using the training data set I1 to train the simulated error term as a sample, use the EMD method to decompose it and use the SVM method to simulate and predict each IMF component of different frequencies, and then superimpose the prediction result to obtain the data set. The predicted value of the error term of I2 is EP(I2).

Fourth is error correction prediction of sample data. The error prediction value EP(I2) obtained by the step (3) is fed back to the result P(I2) which is initially predicted using the SVM to obtain the predicted value P(I2)* of the final high-tech park economy.

In this paper, two indicators, root mean square error (RMSE) and mean absolute error (MAE), are used as the evaluation criteria for model prediction performance. RMSE reflects the statistical characteristics of the error, that is, the degree of dispersion, the smaller the value of RMSE, the higher the prediction accuracy of the model, the better the effect; since the MAE considers the absolute value, the error produced by the model prediction has no positive or negative phase. Offset, which can better describe the actual difference in the final prediction error.

Let yt be the true value and $\hat{y}$t be the predicted value. The expressions of RMSE and MAE are:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2} \quad (11)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_t - \hat{y}_t\right| \quad (12)$$

III. Empirical Analysis

A. Data Source and Description

The monthly data of the cumulative total revenue of Zhongguancun High-tech Park in 2009-2017 was selected as the analysis sample. By the end of 2017, the total revenue of Zhongguancun High-tech Park has reached 5,302.58 billion yuan, a year-on-year growth rate of 15.2%. It can be seen from "Fig. 2" that the total income of Zhongguancun High-tech Park has shown a trend of increasing year by year and growing well. When making predictions, the sample data is a one-dimensional sequence, but the input data in the SVM algorithm requires multi-dimensional sequences to learn, so it is necessary to preprocess the original sample data, construct a prediction matrix, and consider the existence of periodic factors. The total income data of the 12-period period is selected as an input to predict the total income of the month, the data of 2009-2016 is used as the training sample, and the data of 2017 is used as the test sample.

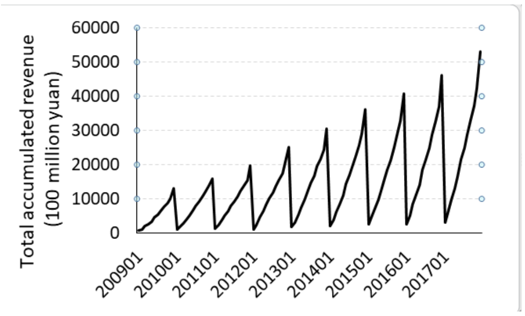

Fig. 2. Monthly data of total accumulated income of Zhongguancun High-tech Park.

B. SVM Preliminary Forecast

The SVM model is established by using the training set of the total income sample data of Zhongguancun High-tech Park and the preliminary prediction is made. The comparison curve between the predicted value and the actual value is shown in "Fig. 3". The predicted value curve deviates significantly from the actual value curve, and there is the phenomenon that the error gradually becomes larger, which affects the prediction accuracy of the model.

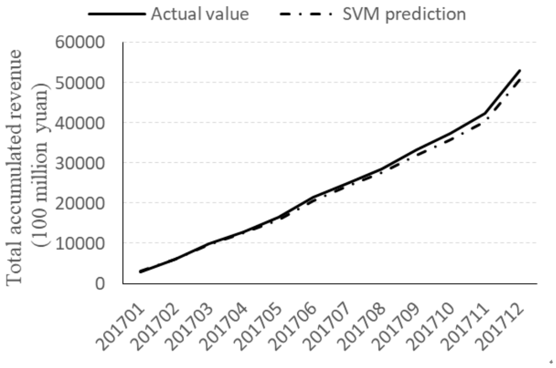

Fig. 3. SVM model prediction results.

C. EMD-SVM Error Correction Prediction Model

The error correction of the total economic income of Zhongguancun High-tech Park is firstly solved by using the EMD method to decompose the error sequence of total income to obtain 5 IMF components and one residual component (See "Fig. 4"). Different frequencies of IMF components and residual terms imply strong information, which can be used to explain the inherent characteristics implied in the time series of total income of Zhongguancun High-tech Park. The residual term is the general trend of the error data and can be used to describe the long-term trend of the SVM prediction error term of the total income of Zhongguancun High-tech Park. The rising trend of the trend item indicates that the SVM forecasting model has a drawback of gradually increasing the total revenue forecast error of Zhongguancun High-tech Park.

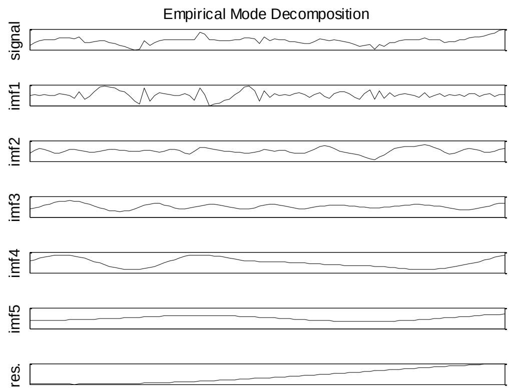

Fig. 4. Error component of EMD decomposition.

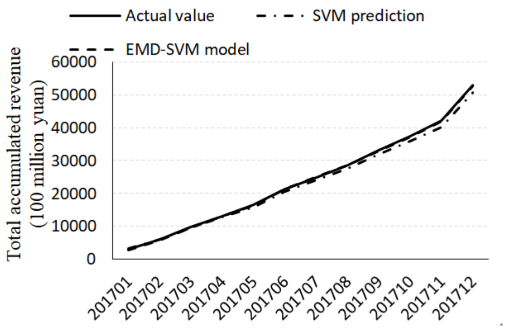

Fig. 5. EMD-SVM predictor.

The upper and lower oscillating transitions of the low-frequency part indicate the impact of major events on the total revenue forecast error of Zhongguancun High-tech Park. The period of the low-frequency curve indicates the length of the impact of major events on the total revenue forecast error of Zhongguancun High-tech Park. The longer the period, the longer the impact of major events on the park economy, and the more difficult it is to eliminate; the amplitude indicates the impact on the total income of Zhongguancun High-tech Park. The larger the amplitude, the greater the impact of major events on total income, the greater the impact. The high-frequency part reflects the imbalance of the model prediction in the short-term of the total income of Zhongguancun High-tech Park. Usually, the imbalance of prediction error exists objectively.

The SVM method is used to simulate and predict each IMF component, and the predicted values are superimposed to obtain the error prediction result, and then fed back to the initial prediction sequence, and finally the predicted value of the total income of the Zhongguancun High-tech Park after error correction is obtained (See "Fig. 5").

D. Analysis of Results

According to the final prediction result after the error correction shown in "Fig. 5", the corrected prediction value is highly consistent with the sample data of the total income sequence of Zhongguancun High-tech Park, and the problem of gradually increasing the error is well solved. It can be seen that the prediction effect of the error correction model based on EMD-SVM is relatively good, which also indicates that the model is feasible in the economic income forecast of the park. In order to reflect the EMD-SVM model has better predictive ability, this paper also uses BP and EMD-BP methods to train and predict the total income of Zhongguancun High-tech Park, and compare and analyze the final error values of each model (See "Table I").

TABLE I. COMPARISON OF PREDICTION RESULTS OF VARIOUS MODELS

Model	RMSE	MAE
EMD-SVM	189.9139684	162.0399606
SVM	1169.2148624	943.5088476
EMD-BP	3492.8006095	2156.1178726
BP	3891.1788127	2213.6934129

The RMSE and MAE values of EMD-SVM are smaller than SVM model, EMD-BP model and BP model. The prediction accuracy is higher than the other three models, indicating that the model can better grasp the change of the total income of Zhongguancun High-tech Park trend.

The error of the results of EMD-SVM model and EMD-BP model is compared with the error of SVM model and BP model. It can be seen that the prediction effect of the model after EMD error correction is better than that of the model without error correction. Predictive effects, this result also fully reflects the advantages of integrated thinking.

Comparing the error of EMD-SVM model and EMD-BP model, and comparing the error of SVM model and BP model, it can be concluded that SVM is more suitable for small sample data.

By comparing the prediction results and error indicators of each model, it can be seen that in most cases, the predicted value of the EMD-SVM integrated model is closer to the actual value of the monthly total income of the Zhongguancun High-tech Park, indicating that the prediction model is practical, effective, more suitable for the forecast of the total economic income of domestic high-tech parks.

IV. CONCLUSION

The EMD-SVM integration model of this paper fully considers the randomness, periodicity and trend characteristics of the economic data of Zhongguancun High-tech Park, explains the inherent meaning of the fluctuation of total income error, and compares it with SVM, EMD-BP and BP models. It shows high prediction accuracy, and can better grasp the running trend of the total income of Zhongguancun High-tech Park, and also provides new ideas and methods for China's future high-tech park economic forecast.

REFERENCES

[1] Bai Lulu, Xie Ya, Guan Chiming. Research on the Development and Planning of International Cooperation Parks in China. International Urban Planning, 2018, (02): 23-32.

[2] Gao Yang, Li Jian. International Carbon Finance Market Price Forecast Based on EMD-PSO-SVM Error Correction Model. China Population, Resources and Environment, 2014,(06):163-170.

[3] Cai Chengzhi, Zhang Jianwei, Liang Ying. Prediction of World Soybean Yield Based on ARIMA Model. Soybean Science, 2018, (03): 452-457.

[4] Ling Zhenghua. Research on the Fluctuation Characteristics of China's Egg Futures Price and Its Influencing Factors — Based on the Analysis of ARCH Models. Price Theory and Practice, 2018, (02): 91-94.

[5] Shen Shichang, Shao Sheng. Risk Measurement of China's Health Insurance Market by GARCH Family Model. Systems Engineering, 2017, (12): 59-64.

[6] Xie Chi,Xi Xiang,Chu Huibin. An Empirical Study on the Inverse Trading Strategy of Securities Investment Based on TGARCH Model. Management Science,2007,(03):68-75.

[7] Li Qing, Zhao Xinquan. Research on the Operation of China's Macroeconomic System Based on Chaos Theory. Studies in Dialectics of Nature, 2013, (12): 66-72.

[8] Tian Yuchen, Liu Wei. Empirical Prediction of GDP Based on Improved Gray GM(1,1) Model. Statistics & Decision Making,2018,(11):83-85.

[9] Huang Yang, Lu Haiyan, Cheng Bizhen. Prediction Model of Consumer Price Index Based on New Dimensional Unbiased Grey RBF Neural Network. Statistics & Decision,2018,(10):74-76.

[10] Lu Shichang, Zhao Boqi, Bi Jianwu. CPI Prediction Based on Fuzzy Information Granular SVM Time Series Regression. Statistics & Decision,2015,(14):82-84.

[11] Zheng Yanqiu, Jiang Tao. Application of Mixed Kernel Function SVM in Financing Risk Assessment. Statistics and Decision, 2018, (05): 74-77.

[12] Wu Mu, Lin Huiping, Li Sike. A Method for Predicting Cigarette Sales Based on Support Vector Machine. Tobacco Science & Technology, 2016, (02): 87-91.

[13] Ping Weiying, Zhong Chunlan. Research on China's Provincial Circular Economy Evaluation Based on Support Vector Machine. Journal of Beijing Technology and Business University, 2014, (06): 84-93.

[14] Huang NE, Shen Z, Long SR. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1998, (1971): 903-995.

[15] Cortes C, Vapnik V. Support-vector Networks. Machine Learning, 1995, (3): 273-297.