buildings

MDPI

Article

Application of GSM-SVM for Forecasting Construction Output: A Case Study of Hubei Province

Ming Lei [1], Yuejie He [1,*], Dandan Wang [2], Debin He [3], Yuhao Feng [1], Lianhuan Cheng [4] and Zihao Qin [5]

[1] School of Urban Construction, Yangtze University, Jingzhou 434023, China
[2] School of Economics and Management, Yangtze University, Jingzhou 434023, China
[3] Hunan Construction Engineering Real Estate Investment Co., Ltd., Changsha 410026, China
[4] School of Architecture and Planning, Hunan University, Changsha 410082, China
[5] PowerChina Sichuan Electric Power Engineering Co., Ltd., Chengdu 610041, China
* Correspondence: 202071573@yangtzeu.edu.cn

Abstract: It is significant to achieve the scientific forecast and quantitative analysis of construction output. In most existing construction economic forecasting methods, both time series models and BP neural network fail to consider the change in relevant influencing factors. This paper introduced the support vector machine (SVM) to solve the above problems based on the grid search method (GSM) optimization model. First, based on constructing an index system of influencing factors of the gross industrial output, a grey relational method is adopted to verify the correlation between the eight factors and output. Furthermore, a SVM forecast model of the gross output is constructed with the relative datasets and influencing factors of the construction industry in Hubei from 2001 to 2016 as a training sample, while the parameters are optimized using the GSM. Then, the model is used to forecast and analyze the gross output from 2017 to 2020 while checking errors. Finally, according to systematic comparison analyses among three forecast models, including the GSM-SVM model, BP neural network, and grey GM (1,1), the results showed that the GSM-SVM forecast model processed the higher solution accuracy and generalization ability. The effectiveness and reliability of our proposed model in the field of construction output forecasting are verified. It can provide a more effective modeling and forecasting method for the gross output value of the construction industry.

Keywords: construction output; support vector machine; economic growth; forecast model

check for updates

Citation: Lei, M.; He, Y.; Wang, D.; He, D.; Feng, Y.; Cheng, L.; Qin, Z. Application of GSM-SVM for Forecasting Construction Output: A Case Study of Hubei Province. *Buildings* **2023**, *13*, 48. https://doi.org/10.3390/buildings13010048

Academic Editor: Ahmed Senouci

Received: 24 October 2022
Revised: 20 November 2022
Accepted: 21 December 2022
Published: 25 December 2022

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Construction output denotes the sum of construction industry products and services generated in a country or a region within a certain period. It is a specific manifestation of production scale, development speed, and operating performance in the construction industry [1,2]; it is essential for governments or companies to avoid risks, position the industry, and formulate rules and regulations. It is necessary to forecast the future economic development of the construction industry based on the intrinsic factors affecting the gross construction output. The high-quality development of the construction industry will be achieved by systematically evaluating and measuring the economic growth of the construction industry [3].

The study of traditional forecasting methods based on construction output are generally achieved by establishing mathematical or physical models. Mainly includes linear regression [4,5], time series [6,7], and other linear analysis methods. These forecasting methods utilized the past construction output historical data to speculate on the future development trend of the construction economy. The research objects of these time series analysis methods require a continuous regular growth trend. However, the forecasting of gross construction output is a non-linear problem affected by many factors, so the above methods do not align with the objective development law of construction output. With the rapid development of artificial intelligence (AI) technology, economic development

forecasting methods based on AI algorithms have received widespread attention from researchers, mainly including grey forecasting models [8,9], BP neural network [10,11], and other integrated models of different algorithms. However, the grey model (1,1) has low accuracy on irregular, and unstable sample data. BP neural network requires large data samples during training and has poor learning ability and fault tolerance on the small sample time series. Generally, current construction output forecasting algorithms have the disadvantage of not considering the impact of relevant factors, slow convergence speed, and low fault tolerance.

According to the above discussion, a suitable construction output forecasting model for predictive economic development is urgently required. Firstly, it is necessary to introduce the influencing factors of construction output into the forecasting model. Most existing indicators of construction output influencing factors direct use macroeconomic indicators to forecast construction output development, such as gross regional product and fixed asset investment, which failed to summarize the impact of other factors. Therefore, a system of indicators with a more distinct hierarchical structure and multiple integrated considerations is necessary. Secondly, the Support Vector Machine (SVM) is a data mining method based on statistical learning theory. Compared with other algorithms, it has unique advantages in solving small samples, non-linear problems, and identifying high-dimensional patterns. The Grid Search Method (GSM) possesses a solid self-adaptive optimizing search ability, which is able to maximize the search for the optimal parameter combination of SVM. To summarize, this paper uses a support vector machine optimized by the grid search method (GSM-SVM). It will achieve a more accurate prediction of the construction industry's gross output.

The contributions of this paper are summarized as follows:

(1) Screening scientific and practical influencing factor indicators are the basis for construction output to conduct predictive analysis research. Based on the literature and related research, the evaluation index system for influencing factors is proposed by combining it with the current Chinese construction industry development situation to screen out the more scientific influencing factors of construction output.
(2) A grid search method is used to optimize the SVM algorithm to find the optimal combination of values of penalty factor C and kernel function parameters g to improve the recognition accuracy and prediction performance of the SVM prediction model.
(3) The SVM algorithm attempts to apply to Chinese construction output forecasting, and related experiments verify that the GSM-SVM forecasting model has higher accuracy and is a critical extension of the economic forecasting method of the construction industry.

The remainder of this paper is organized as follows. Section 2 introduces the related works on the methodology and influencing factor indicators used to forecast construction output gross output. Section 3 presents the index system's support vector machine methods and construction. In Section 4, construction of the GSM-SVM forecasting model and verification analysis of forecasts for gross construction output. To verify the model's accuracy, we compared its forecast errors with those of the BP neural network and grey GM (1,1) models. Finally, it clarifies the GSM-SVM approach's feasibility in the construction output forecasting field in Section 5.

2. Related Works

2.1. Current Status of Research on Factors Influencing Construction Output

Research on the evaluation index system of the construction industry is a prevalent issue. Recently, domestic scholars proposed different models for the gross output value of the construction industry in China and its provinces or cities. Yang [12] proved through the Granger causality test that GDP, the added value of the real estate industry, and investment in construction and installation works to influence the construction industry. Zhang et al. [13,14] selected from the comprehensive ratio of various inputs and outputs of the industry: construction workers, number of construction enterprises, investment in fixed assets in construction, and labor productivity in construction as input indicators

that reflect the efficiency of the industry. Hua et al. [15,16] added the gross construction industry assets in the constructed influencing factors of construction industry development as a comprehensive indicator containing both produce results and operation of the construction industry, to measure the capital stock of the industry. Jiang [17] proposed a more comprehensive composition and evaluation indicators of high-quality development of the construction industry, including scale growth, development efficiency, and innovative development et al. Peng [18] utilized the number of enterprises, housing construction area and completion rate to indicate the existing situation of construction industry and reflect the development process of the industry.

Moreover, scholars proposed different index systems to study and analyze the development of China's construction industry. However, there are few analyses and evaluations on the concept of influencing factors of construction industry economic development in China, and there still needs to be a perfect index system of influencing factors of the gross output value of the construction industry. Therefore, based on literature induction and analysis, combined with the concept of high-quality development of China's construction industry [19]. It is planned to select the added value of the construction industry, the gross income of construction enterprises, and other indicators to evaluate the economic benefits and industrial scale of the construction industry. In addition, the impact of industrial innovation production capacity is also considered, increasing labor productivity, the number of construction enterprises, and other production efficiency indicators. Combining the economic development of the construction industry with 'scale' and 'benefit' contributes to construct the index system of influencing factors of the gross output value of the construction industry from a more comprehensive analysis perspective.

2.2. Current Status of Research on Construction Output Forecasting

Scientific forecast of the construction industry's gross output and development trend plays a significant role in formulating supportive policies and legal regulation of the construction industry in China. In recent years, Chinese scholars have proposed various models to forecast and analyze the gross output of the national, provincial, or municipal construction industry. For example, Lu [4] adopted the Granger Causality Test method to verify the variables influencing economic growth and applied a multiple regression model to forecast and analyze the growth potential. Zhang et al. [7] established an ARIMA (1,1,2) model, whose performance is better than the grey model and triple exponential smoothing. With the time series analysis, Tang et al. [5,6,9] constructed a grey forecast model and the Logistic growth model to perform industrial forecasting research. Li [10] adopted the data envelopment analysis method to establish an evaluation index system for the development efficiency of the construction industry and constructed a BP neural network model to reasonably forecast the output economic index value of the industry in China. Li [8] established linear regression and grey (1,1) forecast models to forecast, thereby expounding the application scope of the two forecast methods by combining the forecast results. Based on the genetic algorithm and grey system, Guang [11] utilized BP neural network theory to establish a GA-GM (1,1)-BP composite model to forecast the added value of China's construction industry.

In all the research mentioned above, the forecast methods for construction output mainly include regression analysis method, time series analysis method, grey GM (1,1) model, and BP neural network. They all have characteristics but require many training samples and do not consider the relevant influencing factors. There are certain limitations. The comparative analysis of commonly used models is shown in Table 1.

Therefore, we introduced SVM theory into the construction industry gross output forecast to establish the GSM-SVM forecast model to provide a more effective forecast method to predict construction output. It has many unique advantages in solving problems in limited samples, such as non-linear and high-dimensional pattern recognition [20–23]. Existing linear prediction methods cannot achieve this. In addition, its generalization ability could be better and reflect the correspondence between the growth of gross construction

output and the various influencing factors. As a result, it can not only be used to solve the problem of limited samples and non-linear gross output value prediction but also to realize the identification of the influence of various external factors [24] on the gross output value of the construction industry.

Table 1. Comparative analysis of existing method.

	Existing Method	Method Limitations	SVM Advantages
Forecast of gross output of construction industry	linear regression, ARIMA Model	For linear relationships only, it cannot predict unstable data;	Kernel function can be used to solve non-linear, high-dimensional problems;
	Grey GM (1,1)	Not suitable for non-linear prediction, does not consider the influence of related factors;	Input influencing factors as feature vectors;
	BP Neural Network	Converge to a local minimum rather than a global minimum, usually results in overfitting and unstable output data;	Obtained the optimal global solution, low generalization error rate, and more suitable for small sample learning;

3. Related Methods

3.1. Definition

It is challenging to solve the problems in the forecast of the gross output value of the construction industry, such as the small sample size, the affected of multiple factors, and the need to satisfy the linear separability of sample characteristic data. SVM [25–27] is a machine learning method based on the VC (Vapnik-Chervonenkis) dimension theory of statistical learning theory and the principle of structural risk minimization. The obtained extreme-value solution is globally optimal, so it has higher accuracy than other machine-learning methods. It can overcome many obstacles in time series forecast, pattern recognition, etc. Currently, this model is barely used in the economic development forecasting of the construction industry.

The mechanism of SVM is to search for an optimal classification hyperplane that satisfies the classification requirements. The system randomly generates a moving hyperplane for a given sample set to classify the samples. When the points of different categories in the training sample are set to fall on both sides of the hyperplane, it can maximize the sum of the distances from the hyperplane to the nearest heterogeneous sample points. The model is described in detail as follows:

A given training sample set is $(x_i, y_i), i = 1, 2, \ldots, n$, in which $x_i \in R^n$, $y_i \in R$.

Let the regression function be:

$$F = \left\{ f \middle| f(x) = w^{\mathrm{T}} x + b, w \in R^n \right\} \tag{1}$$

In Equation (1), w is the n-dimensional weight vector; b is the threshold used to determine the optimal classification hyperplane; the optimal classification hyperplane is $w^{\mathrm{T}} x + b = 0$.

The gross output and relevant influencing factors are nonlinear time series, so when the sample set is linearly inseparable, it is necessary to map the input vector from the low-dimensional feature space to the high-dimensional feature space and construct the optimal classification hyperplane in this space. Replace the input data x with the feature $\phi(x)$, and derive a linear classification: $x \to \phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_n(x))^{\mathrm{T}}$.

Transform from the input space χ to the feature space H. In the space H, the discriminant function is:

$$f(x) = w^{\mathrm{T}} \phi(x) + b \tag{2}$$

The weight vector is expressed as a linear combination of training examples: $w = \sum_{i=1}^{n} \alpha_i x_i$, in which α_i is the Lagrange multiplier introduced into the model, $0 \leq \alpha_i \leq C$; C is an error penalty factor to achieve a compromise between empirical risks and structural risks.

Therefore, in the feature space H, the optimal classification function can be obtained as follows:

$$f(x) = \sum_{i=1}^{n} \alpha_i \phi(x_i)^{\mathrm{T}} \phi(x_j) + b \tag{3}$$

The inner product operation in the high-dimensional feature space is replaced by a kernel function $K(x_i, x_j) = \phi(x_i)^{\mathrm{T}} \phi(x_j)$ that satisfies the Mercer condition.

The kernel function of the support vector machine and its corresponding parameters affect the number of support vectors and the prediction model's internal algorithm and significantly impact the model's performance. Generally, the RBF (radial basis function) kernel function is selected by default, which has the characteristics of few parameters, strong adaptability, and strong anti-interference ability to noise data. That is $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$ [28–31], and the kernel parameter is $g = -\frac{1}{2\sigma^2}$.

The SVR (support vector regression) function can be expressed as a kernel function:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b \tag{4}$$

3.2. Construction of the Index System

In the study, construction output in the Hubei province of China is taken as the research subject, and the influencing factors of the gross output are selected by the following scientific, systematic and operable principles. Eight major influencing factors at the current stage are selected comprehensively based on the current research findings in the related fields of the construction industry [32–37] and the status quo of the industry development in Hubei. The index system is shown in Figure 1.

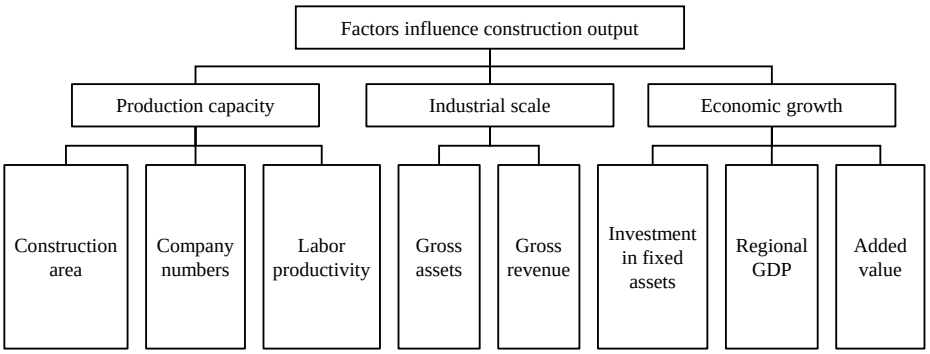


Figure 1. Factors influence construction output.

The hierarchical structure of the influencing factor index system can be divided into three layers. The first is the target layer, namely the factors influencing construction output; the second is the criterion layer, including three dimensions of production capacity, industrial scale and economic growth; the third is the index layer, including: (1) the construction area of buildings, and the number and labor productivity of companies in the construction industry, which are selected to reflect the influence of labor input on the output of the construction industry in Hubei; (2) the gross assets and gross revenue of companies in the construction industry, which are selected to reflect the development level and competition potential of the industry from the perspective of gross volume; (3) the investment in fixed assets of construction and installation projects, the regional GDP and the added value of the construction industry, which are selected to reflect the influence of economic scale on the gross output growth of the industry.

A grey correlation degree algorithm [38,39] was adopted in the research for analysis in order to inspect the degree of correlation between the eight influencing factors and the gross output. As shown in Table 2, the correlation between each index and the gross output is higher than 0.65, representing a high correlation state, indicating the remarkable influence of the selected influencing factors on the construction industry output.

Table 2. Correlation of influencing factors.

Evaluation Items	Construction Area	Company Number	Labor Productivity	Gross Assets	Gross Revenue	Investment in Fixed Assets	Regional GDP	Added Value
Correlation	0.863	0.718	0.725	0.915	0.930	0.857	0.782	0.822

3.3. SVM Prediction Model Based on GSM Optimization

As the RBF kernel function parameter g and penalty factor C are both randomly valued in the positive range, there are many types of two-parameters random combinations. Furthermore, the different combinations significantly influence the prediction model's performance and prediction accuracy. However, the traditional trial and error and experimental methods are time-consuming and laborious and prone to under-fitting and over-fitting [40,41]. Therefore, this paper uses the grid search method (GSM) [42–44] to debug and optimize penalty parameters C and kernel parameters g in the SVM model.

The basic principle of the grid search algorithm is to divide parameters into each grid with equal step length (a grid represents a pair of parameters) and traverse all points in the grid to find the optimal global solution. For the set parameters C and g, the parameter variations are first roughly selected over a wide range, and accuracy was found for different combinations of parameters using K-fold cross-validation. Then, the parameter selection range is refined to obtain more accurate parameter combinations. Finally, the best set of parameters to achieve the minor parameter C in the higher validation classification rate is obtained [45].

Based on the grid search method optimization support vector machine combination forecasting model (GSM-SVM) implementation process, as shown in Figure 2.

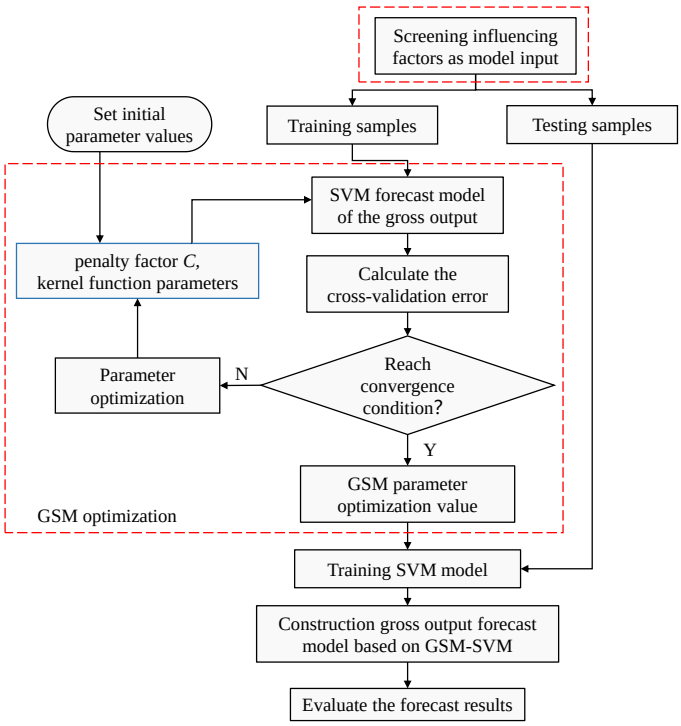


Figure 2. Workflow of GSM-SVM gross output value forecast.

4. Results and Discussion

4.1. Data Sources and Data Pre-Processing

This paper selected construction output in Hubei during the 20 years from 2001 to 2020 as the research object. Furthermore, the eight indexes mentioned above were selected as the characteristic values of the prediction model. The data comes from the "China Statistical Yearbook", "Hubei Statistical Yearbook" and "China Construction Industry Statistical Yearbook" from 2001 to 2020. The original gross output data and relevant indexes are shown in Table 3.

Table 3. Raw data of gross output and relevant indexes of the construction industry in Hubei from 2001 to 2020.

Year	Construction Area (Thousand m^2)	Company Number (Unit)	Labor Productivity (Yuan/Person)	Gross Assets (Billion Yuan)	Gross Revenue (Billion Yuan)	Investment in Fixed Assets (Billion Yuan)	Regional GDP (Billion Yuan)	Added Value (Billion Yuan)	Gross Output (Billion Yuan)
2020	852,682.0	4633	752,079	1581.790	1480.225	2341.263	4344.35	282.80	1613.611
2019	920,422.3	4566	675,791	1488.138	1608.693	2886.884	4542.90	307.31	1697.967
2018	882,381.0	4196	595,205	1303.266	1435.736	2603.141	4202.20	278.14	1513.387
2017	792,576.9	3692	525,908	1145.165	1248.490	2377.298	3723.50	234.25	1339.073
2016	728,350.5	3368	458,776	985.331	1165.986	2287.658	3335.30	210.92	1186.240
2015	622,047.2	3218	454,961	889.491	1113.957	2182.603	3034.40	195.78	1059.286
2014	622,278.8	3217	487,456	755.753	971.267	1818.542	2824.21	187.54	1005.959
2013	489,379.6	3197	486,265	635.188	834.340	1431.225	2537.80	165.70	846.527
2012	391,139.0	2774	410,437	585.253	708.284	1105.416	2259.09	141.79	704.342
2011	310,237.5	2640	261,332	487.432	568.842	805.652	1994.25	124.68	558.645
2010	250,467.0	2846	268,719	375.635	465.363	670.191	1622.69	102.27	434.520
2009	204,993.3	2860	234,352	311.270	364.885	502.562	1319.21	84.23	342.189
2008	182,728.6	2972	193,189	225.493	264.597	356.897	1149.75	67.57	260.508
2007	166,796.8	2490	167,141	196.809	215.524	285.731	945.14	54.49	211.080
2006	144,782.4	2271	145,998	134.349	170.648	225.272	753.18	41.80	166.700
2005	120,911.6	2072	122,576	114.010	136.572	169.68	646.97	35.74	134.932
2004	107,201.3	2417	96,520	106.860	113.538	140.925	554.68	32.52	111.213
2003	89,334.4	1746	81,761	90.862	94.460	105.998	475.75	27.39	87.300
2002	72,275.0	1579	69,478	74.870	65.506	97.473	421.28	23.69	63.655
2001	66,626.0	1637	63,664	60.080	52.588	89.060	388.05	21.43	52.902

In the SVM algorithm, unifying the magnitude of original data will significantly improve the SVM recognition rate and training efficiency and eliminate the forecast errors caused by the significant value differences of indexes [46]. The mapminmax function was used to normalize the data to [0, 1]. Let the training sample set to be $\{x_i\}$, the test sample to be $\{x'_i\}$, $x_{\max}$ and $x_{\min}$ to be the maximum and minimum values of the indexes in the training sample set, respectively. A sample x_i in the training sample set was selected for calculation with the normalization equation $x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$. The normalization of the test sample set x'_i is the same as that of x_i. The results after simulation need to be de-normalized to obtain the final forecast value.

4.2. Model Fitting Based on GSM-SVM

After data sorting and preprocessing, the eight influencing factors and the gross output of the construction industry were selected as the input and output of the model. The time series data of construction output in Hubei from 2001 to 2016 was selected as the training samples, and the data from 2017 to 2020 was selected as the testing samples.

In this study, Radial Basis Function (RBF) kernel function was selected to establish a predictive SVR model, which transformed the optimization problem into solving equations. The grid search method was used to optimize the parameters *C* and g of SVM since the accuracy and performance depend on *C* and *g*. Firstly, the large step length is used for

rough search, and then the small step length is used for accurate search. The accuracy of different parameter combinations is compared by 5-fold cross-validation, and the optimal SVM parameter set is finally determined.

The process of parameter optimization is shown in Figure 3. The *x*-axis represents the value of parameter *C* taken as the logarithm of base 2, the *y*-axis represents the value of the kernel parameter *g* taken as the logarithm of base 2, and the contours represent the accuracy rates corresponding to taking the corresponding *C* and *g*. The optimal parameters for Support Vector Regression (SVR) were obtained, bestc = 11.3173, bestg = 0.0078, and CVmse = 0.0073.

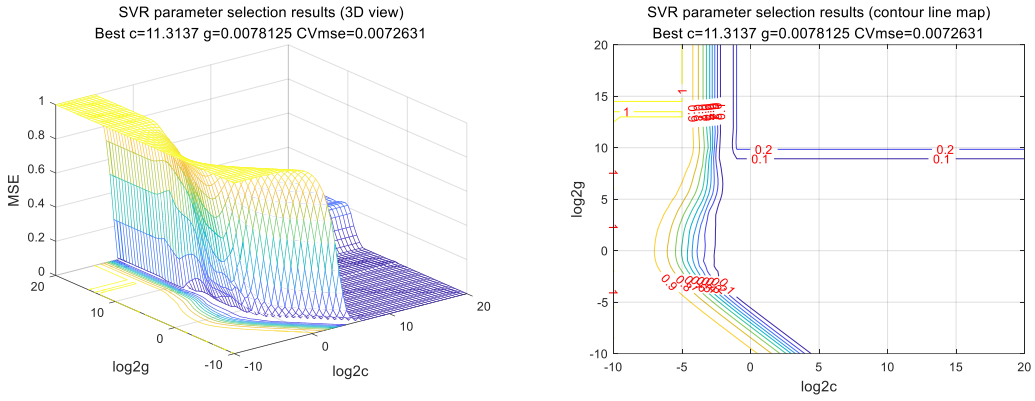


Figure 3. Diagram of parameter optimization process. (**a**) 3D view, (**b**) contour map.

The SVM network was trained with the optimal parameters *C* and *g* of regression analysis, and the network regression forecast was carried out on construction output in Hubei from 2017 to 2020. The output data were processed by de-normalization, and the final forecast value of the construction output in Hubei Province was obtained. The comparison between the value forecasted by the model and the actual value is shown in Figure 4.

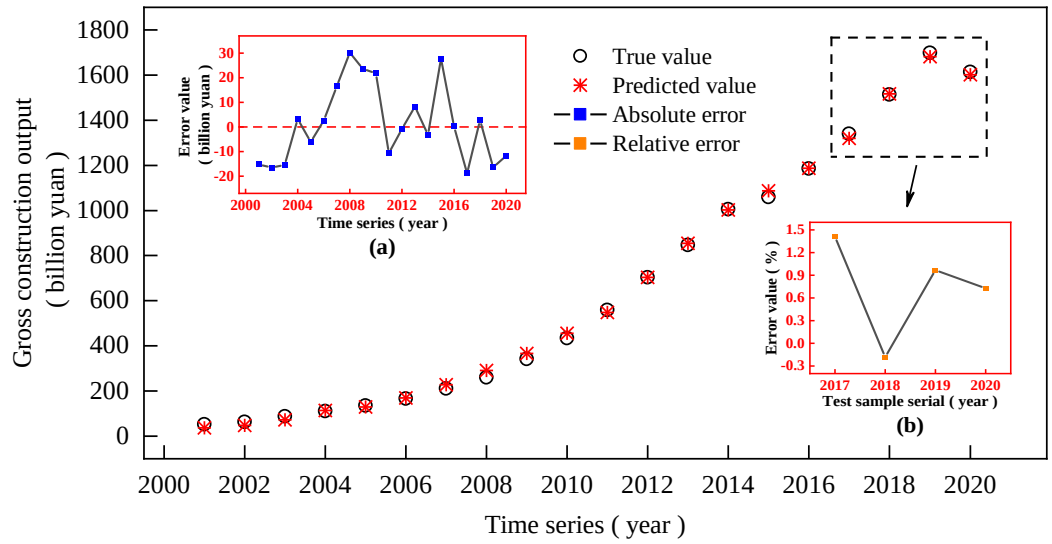


Figure 4. Fitting results of GSM-SVM forecast model for the construction industry output in Hubei. (**a**) total samples' absolute error diagram, (**b**) testing samples' relative error diagram.

Figure 4 shows the forecasted output value of Hubei's construction industry given by the GSM-SVM model, which better fits the actual value. The overall trend of their curves is consistent, and the prediction accuracy of the overall sample is 93.90%. Of the accuracy test of all samples from 2001 to 2020, the prediction results in 2008 and 2015 are slight

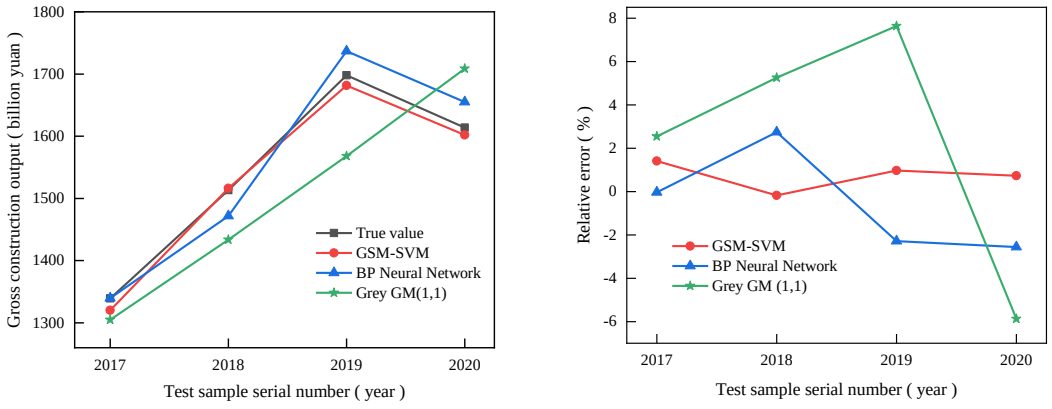
deviations, while the rest predicted values that were consistent with the actual results. It can be seen more intuitively from Figure 4a absolute error diagram that the maximum absolute error of the overall sample is no more than 30 billion Yuan. Among the training samples, the prediction effect of 2004, 2012, and 2016 is excellent, and the absolute error is no more than 4 billion Yuan. Figure 4b testing samples' relative error diagram shows that the relative error of the industry's gross output forecasted by the GSM-SVM model in the past four years (2017–2020) is relatively stable. The maximum relative error of the testing sample is no higher than 1.5%, and this forecast result has high accuracy in macroeconomic forecasts. To sum up, it shows that the GSM-SVM model can be used to forecast the gross output of the construction industry, and our model forecasts that the construction output has high credibility.

4.3. Model Comparison and Analysis

As is well known, the GSM-SVM model is compared with the BP neural network and the grey GM (1,1) model, which are commonly used in the construction of the gross output value prediction, and to verify the reliability and practicability of the GSM-SVM model. Table 4 and Figure 5 show the comparison of construction output in Hubei from 2017 to 2020 forecasted by the three models.

Table 4. Comparison between the gross output forecasted by the three models and the actual value of the construction industry.

Year	Actual Value (Billion Yuan)	Value Forecasted by the GSM-SVM Model (Billion Yuan)	Relative Error (%)	Value Forecasted by the BP Neural Network (Billion Yuan)	Relative Error (%)	Value Forecasted by the Grey GM (1,1) (Billion Yuan)	Relative Error (%)
2020	1613.611	1601.869	0.7277	1654.968	−2.5630	1708.434	−5.8764
2019	1697.967	1681.539	0.9675	1736.754	−2.2843	1568.251	7.6395
2018	1513.387	1516.158	−0.1831	1471.862	2.7438	1433.838	5.2563
2017	1339.073	1320.183	1.4107	1339.562	−0.0366	1304.956	2.5478



(**a**) (**b**)

Figure 5. Comparison of the predictive performance of the three models. (**a**) forecast values and targets for the gross output of construction, (**b**) relative error diagram.

From Table 4 and Figure 5, it is clear that the grey GM (1,1) forecast model has the worst performance. Moreover, the value forecasted by the GSM-SVM model and the BP neural network model has a similar variation tendency with the actual construction output, indicating that they can effectively forecast the output of an industry. Among them, the curve of the values forecasted by GSM-SVM was more consistent with the actual values, proving its better forecast performance.

Furthermore, in this paper, the mean absolute percentage error (MAPE) and Theil inequality coefficient (TIC) [47] are used to analyze the errors of the three models. The comparison of the model forecast accuracy is shown in Table 5. The equation of MAPE is $\text{MAPE} = \frac{1}{\text{n}} \sum_{i=1}^{\text{n}} \left| \frac{p_i - \hat{p}_i}{p_i} \right|$, where $\hat{p}_i$ is the forecasted construction industry output, and p_i is the actual value, n is the number of forecasted samples. MAPE reflects the overall closeness of the forecasted value to the actual. The smaller the value is, the higher the forecast accuracy of the model is. The equation of Theil's inequality coefficient is $\text{T} = \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{(p_i - \hat{p}_i)^2}}{\sqrt{(p_i)^2} + \sqrt{(\hat{p}_i)^2}}$, reflecting the difference between the real value and the simulation result. The smaller the T value is, the better the fitting is. Herein, T = 0 means complete fitting.

Table 5. Comparison of the forecast accuracy of the three models.

	Value Forecasted by the GSM-SVM Model	Value Forecasted by the BP Neural Network	Value Forecasted by the Grey GM (1,1)
(MAPE) %	0.823	1.905	5.333
(TIC) %	0.413	0.951	2.704

It can be concluded by comparing the errors in Table 5 that the MAPE and TIC of the GSM-SVM model are 0.823% and 0.413%, respectively. It has remarkable prediction accuracy compared with 1.905% and 0.951% of the BP neural network model and 5.333% and 2.704% of the grey GM (1,1) forecast model. The correlation coefficient between the forecasted value by the GSM-SVM model and the actual value is also as high as 0.99612, overall better than the other two algorithm models.

According to the above analysis and comparison, the GSM-SVM algorithm model proposed has the best forecast performance in terms of forecasting construction output in Hubei. The grey GM (1,1) algorithm model has the worst performance. In conclusion, the GSM-SVM forecast model can forecast construction output more accurately, which is an efficient and reasonable forecast method and has the opportunity to be applied to a larger scale of output value prediction and has significant promotion and application value. Additionally, it can provide a decision-making basis for China to formulate construction industry plans scientifically and effectively, thus boosting its output through multiple channels.

5. Conclusions

This study presents the influencing factors of the gross value of construction output. The principle of the GSM-SVM is further employed to establish a forecast model for the gross output of the construction industry in Hubei. The relevant industry data from 2001 to 2020 is used for simulation and forecast. The maximum relative error of the test samples is no higher than 1.5%. According to a systematic comparison among three forecast models, including the GSM-SVM model, BP neural network, and grey GM (1,1), the MAPE of the GSM-SVM forecast model is 0.823%, and the TIC coefficient is 0.413%. The prediction effect is superior to BP neural network and grey GM (1,1) prediction model, which has better forecast and convergence performance for forecasting the construction industry output.

This paper is only a tentative study to apply the SVM model to forecasting the gross construction output in China and to verify the method's feasibility. As is well known, the gross construction output is also affected by various other complex factors, such as the level of technology and equipment and industry policies. Therefore, it is necessary to add new explanatory variables further to improve the model in the future to build a more scientific and practical forecasting model.

Author Contributions: Conceptualization, M.L.; methodology, Y.F. and Y.H.; formal analysis, Y.H.; investigation, Y.H.; resources, Z.Q. and D.H.; data curation, Y.F. and Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H. and Y.F.; visualization, Y.H. and L.C.; supervision, M.L.; project administration, D.W.; funding acquisition, D.H. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Teaching and Research Project of Hubei Provincial Department of Education (grant no. 2018284).

Data Availability Statement: All necessary data are provided in the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giang, D.T.; Pheng, L.S. Role of construction in economic development: Review of key concepts in the past 40 years. *Habitat Int.* **2011**, *35*, 118–125. [CrossRef]
2. Anaman, K.A.; Osei-Amponsah, C. Analysis of the causality links between the growth of the construction industry and the growth of the macro-economy in Ghana. *Constr. Manag. Econ.* **2007**, *25*, 951–961. [CrossRef]
3. Yang, X.; Cheng, J. Economic Forecasting: Characteristics and Quantitative Methods. *Syst. Sci. Math.* **2019**, *39*, 1553–1582.
4. Lu, X. The Growth Path and Affecting Factors of Construction Industry in China. Doctoral Dissertation, Xi'an University of Architecture and Technology, Xi'an, China, 2003.
5. Yi, Z. The Research of the Development Prediction and the Space of Construction Industry's Growth-Based on the Analysis of Chongqing's Construction Industry. Master's Thesis, Chongqing University, Chongqing, China, 2014.
6. Tang, J.J.; Liang, W.Z.; Hu, S.H.; Zhao, T.S. Present Situation Analysis and Development Trend Forecast of Construction Industry. *J. Civ. Eng. Manag.* **2012**, *121*, 84–88. [CrossRef]
7. Zhang, L.; Li, H. Application of ARIMA model to forecast the total output value of construction industry in China. *Enterp. Econ.* **2011**, *11*, 93–96. [CrossRef]
8. Li, X. Study on the Comprehensive Evaluation and Prediction of Construction Industry in Anhui Province. Master's Thesis, Anhui University of Construction, Hefei, China, 2017.
9. Liu, L.; Wu, L. Predicting the output value of assembled buildings based on grey mean model. *Math. Pract. Underst.* **2019**, *15*, 104–111.
10. Li, H. Research on the Evaluation of the Development of Construction Industry in China Based on Multi-Technology. Master's Thesis, Northeast Forestry University, Harbin, China, 2012.
11. Guang, H. Research on Forecasting Value Added in Construction Industry Based on GA Optimized Grey Neural Network Model. *China Constr. Met. Struct.* **2021**, *5*, 12–13.
12. Yang, S. The Empirical Research on Factors Affecting the Development of the Construction Industry in Anhui. Master's Thesis, Anhui University of Finance and Economics, Bengbu, China, 2017.
13. Zhang, Y. Research on the Economic Status Evaluation and Comparison of the Technical Efficiency for Construction Industry in Guangdong Province. Doctoral Dissertation, South China University of Technology, Guangzhou, China, 2015.
14. Dai, Y.A.; Chen, C. Technical Efficiency in China's Construction Industry and Its Influencing Factors. *China Soft Sci.* **2010**, *1*, 87–95.
15. Hua, R. Research on Evaluation and Influencing Factors of High-Quality Development of China's Construction Industry. Master's Thesis, Anhui University of Construction, Hefei, China, 2021.
16. Wang, X.; Wang, Z. Study on the Temporal and Spatial Difference of Construction Efficiency in the Yangtze River Economic Belt. *Constr. Econ.* **2021**, *42*, 14–18. [CrossRef]
17. Jiang, J. Research on the Evaluation System of High-Quality Development of Provincial Construction Industry in China. Master's Thesis, Guangzhou University, Guangzhou, China, 2021.
18. Peng, X. Research on the High-Quality Development of the Construction Industry in Anhui Province under the Background of the Integration of the Yangtze River Delta. Master's Thesis, Anhui University of Construction, Hefei, China, 2022.
19. Wang, Y.; Wu, X. Research on High-Quality Development Evaluation, Space–Time Characteristics and Driving Factors of China's Construction Industry under Carbon Emission Constraints. *Sustainability* **2022**, *14*, 10729. [CrossRef]
20. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
21. Behzad, M.; Asghari, K.; Eazi, M.; Palhang, M. Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Syst. Appl.* **2009**, *36*, 7624–7629. [CrossRef]
22. Yuan, Y.; Wang, C.; Zhou, A. Prediction Model for Stability Classification of Roadway Surrounding Rock Based on Grid Search Method and Support Vector Machine. *Saf. Coal Mines* **2017**, *48*, 200–203. [CrossRef]
23. Zhang, C.; Tian, Y.; Deng, N. The new interpretation of support vector machines on statistical learning theory. *Sci. China Math.* **2010**, *53*, 151–164. [CrossRef]
24. Wang, L.; Zhang, S.; Li, J. Time Series Prediction Based on Support Vector Regression. *Inf. Technol. J.* **2006**, *5*, 353–357.
25. Zhang, T. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. *AI Mag. Artif. Intell.* **2001**, *22*, 103.

26. Raghavendra, N.S.; Deka, P.C. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* **2014**, *19*, 372–386. [CrossRef]
27. Virmani, J.; Kumar, V.; Kalra, N.; Khandelwal, N. SVM-Based Characterization of Liver Ultrasound Images Using Wavelet Packet Texture Descriptors. *J. Digit. Imaging* **2012**, *26*, 530–543. [CrossRef]
28. Wu, M. Parameter Optimization Method Research and Application of RBF Neural Network and SVM. Master's Thesis, Central South University, Changsha, China, 2007.
29. Scholkopf, B.; Sung, K.-K.; Burges, C.; Girosi, F.; Niyogi, P.; Poggio, T.; Vapnik, V. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Process.* **1997**, *45*, 2758–2765. [CrossRef]
30. Wang, Q.; Zheng, J. Study of network parameter model base on support vector machine. *Mod. Electron. Tech.* **2015**, *12*, 23–24+28. [CrossRef]
31. Boolchandani, D.; Ahmed, A.; Sahula, V. Efficient kernel functions for support vector machine regression model for analog circuits' performance evaluation. *Analog. Integr. Circuits Signal Process.* **2011**, *66*, 117–128. [CrossRef]
32. Gao, Y. Research on Factors Affecting the Economic Growth of the Construction Industry in Henan Province. Master's Thesis, Zhengzhou University, Zhengzhou, China, 2014.
33. Hu, W.; Kong, D.; He, X. Analysis on Influencing Factors of Green Building Development Based on BP-WINGS. *Soft Sci.* **2020**, *34*, 75–81. [CrossRef]
34. Cui, X. Empirical Analysis on Influence Factor of Economic Growth in Henan Construction Industry. *Constr. Econ.* **2012**, *3*, 99–101. [CrossRef]
35. Ding, Z.; Fan, Z.; Tam, V.W.; Bian, Y.; Li, S.; Illankoon, I.C.S.; Moon, S. Green building evaluation system implementation. *Build. Environ.* **2018**, *133*, 32–40. [CrossRef]
36. Chen, C.; Cao, X.; Zhang, S.; Lei, Z.; Zhao, K. Dynamic Characteristic and Decoupling Relationship of Energy Consumption on China's Construction Industry. *Buildings* **2022**, *12*, 1745. [CrossRef]
37. Li, H.; Han, Z.; Zhang, J.; Philbin, S.P.; Liu, D.; Ke, Y. Systematic Identification of the Influencing Factors for the Digital Transformation of the Construction Industry Based on LDA-DEMATEL-ANP. *Buildings* **2022**, *12*, 1409. [CrossRef]
38. Tian, M.; Liu, S.; Bu, Z. Review of research on grey relational degree algorithm model. *Stat. Decis.* **2008**, *1*, 24–27.
39. Yin, K.; Xu, T.; Li, X.; Cao, Y. A study of the grey relational model of interval numbers for panel data. *Grey Syst. Theory Appl.* **2020**, *11*, 200–211. [CrossRef]
40. Cui, J.; Yuan, W. Optimization of Support Vector Machine Parameters Based on Intelligent Algorithm. *J. Hebei Norm. Univ. Sci. Technol.* **2017**, *1*, 34–38.
41. Gaspar, P.; Carbonell, J.; Oliveira, J.L. On the parameter optimization of Support Vector Machines for binary classification. *J. Integr. Bioinform.* **2012**, *9*, 33–43. [CrossRef]
42. Xu, C.; Cao, H.; Zhao, X. Speaker Recognition Parameter Selection Method Based on SVM. *Comput. Eng.* **2012**, *38*, 175–177.
43. Xu, W.; Liu, C. A Regression Model for Forecasting Regional Annual Water-consumed Quantity Based on GSM and SVM. *J. Shenyang Agric. Univ.* **2011**, *42*, 238–240.
44. Tahyudin, I.; Nambo, H.; Goto, Y. An Optimization of the Autoregressive Model Using the Grid Search Method. *Int. J. Eng. Technol.* **2018**, *7*, 84–86. [CrossRef]
45. Wang, X.; Li, Z. Identifying the Parameters of the Kernel Function in Support Vector Machines Based on the Grid-Search Method. *Period. Ocean. Univ. China* **2005**, *35*, 859–862. [CrossRef]
46. Wang, S.; Jin, Z. Intrusion detection classification algorithm based on fuzzy SVM model. *Comput. Appl. Res.* **2020**, *2*, 501–504. [CrossRef]
47. Wang, Y. Research on the Methods of Combining Forecasts Based on Correlativity. *Forecasting* **2002**, *21*, 58–62.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.