Study on prediction and influencing factors of energy consumption in Jiangxi province

Hui Xia

School of Information Engineering, Jingdezhen University, Jingdezhen, 333000, China

Abstract: This study aims to determine the main factors affecting energy consumption in Jiangxi Province, and to forecast energy consumption effectively. An indicator system of energy consumption influencing factors containing 17 variables is constructed from 4 aspects of economy, society, technology and energy. Based on relevant data from 1996 to 2021, conduct comprehensive analysis and select 10 main factors. Establish three energy consumption prediction models: multiple regression model (MLR), time series model (ARIMA), and support vector regression (SVR). Evaluate the prediction performance through mean square error (MSE), mean absolute percentage error (MAPE), and goodness of fit (R2). The main influencing factors of energy consumption are determined as GDP, output value of the tertiary industry, industrial output value, residents' consumption level, fixed assets investment of the whole society, total population, urbanization rate, internal expenditure of R&D funds, energy intensity, and energy consumption structure. The best prediction model is SVR model. It provides a basis for the effective formulation of energy policies and the implementation of economic development programs.

1 Introduction

Energy is a material resource in nature that can provide some form of energy for human beings, and it is an indispensable resource for industrial production, economic development, and people's life in every country, and it is vital to the development of all countries in the world [1]. With the rapid development of the economy, the demand for energy is expanding, but energy resources are limited. At the same time, energy consumption also has a serious impact on the environment, such as air pollution, greenhouse gas emissions, etc. [2], with the development of "carbon peak" and "carbon neutral" goals, energy consumption has received more and more attention.

Jiang [3] used the gray correlation model to find out the main influencing factors affecting energy consumption in Chongqing, and established the gray dynamic model, BP neural network model and the composed model to forecast energy consumption. Liu [4] established a VAR model to analyze the relationship between energy consumption and GDP, energy intensity and urbanization level based on the relevant data of Beijing from 1980 to 2010.The authors in [5] proposed two linear models: a hybrid time-series regression model and a Box-Jenkins transfer function model to determine the relationship between energy consumption and a set of control and environmental variables.

Zhao [6] extracted the main factors affecting energy consumption by PCA and then establishing a regression model to predict energy consumption. Zhang [7] established a VAR model based on the relevant data of Shanxi Province from 1995 to 2012, to analyze the relationship between energy consumption and energy consumption structure, industrial structure and energy price, and five suggestions are put forward to improve the energy consumption situation in Shanxi Province. Based on the data related to energy consumption in Jiangxi Province from 2000 to 2013, factor analysis was used to derive the main factors affecting energy consumption in [8], and the GM(1,1) model was used to predict the total energy consumption. The main factors affecting energy consumption in Beijing are the development of tertiary industry, the diversity of energy consumption and the development level of the economy by using PCA in [9], and the gray model, stochastic time series model and multivariate linear model are established to predict the total energy consumption, and by comparing the prediction accuracy, the ARIMA (1,2,1) model is the best. Zhang et al [10] assessed the influence relationship of many factors affecting energy consumption through relevant analysis in Shandong Province based on the statistical data from 1996 to 2017.

At the provincial level, the energy self-sufficiency of Jiangxi Province is seriously insufficient. With the sustained growth of the national economy and the increasing improvement of people's living standards, the energy consumption continues to increase. The external dependence of energy supply is extremely high [11]. The energy problem is particularly significant in Jiangxi Province. Therefore, this paper takes Jiangxi Province as an example, based on the statistical data from 1996 to 2021, analyzes the influencing factors of energy consumption to find the main influencing factors, and

E-mail: 375473215@qq.com

constructs multiple regression model, time series model and support vector regression model to forecast energy consumption, and uses the mean squared error (MSE), the mean absolute percentage error (MAPE) and the goodness of fit (R^2) to evaluate the prediction effect and select the best prediction model. It provides a basis for formulating energy policies and planning economic development programs.

2 Data sources and methodology

The statistical data of 17 indicators and energy consumption data in the impact factor index system of this paper are all from Jiangxi Statistical Yearbook, which are used to screen the impact factors of energy consumption and construct the prediction model.

Pearson correlation coefficient and Spearman correlation coefficient were applied for correlation analysis to screen the factors affecting energy consumption. In view of the time-series characteristics of energy consumption data and the multiplicity of influencing factors, multiple regression model (MLR), time-series model (ARIMA) and support vector regression (SVR) were selected to fit the data and forecast energy consumption.

3 Analysis of impact factors

3.1 Constructing the indicator system

There are many factors affecting energy consumption in Jiangxi Province. By combing and drawing on the influencing factors in the relevant literature, construct a system containing 17 indicators from four aspects: economic, social, technological and energy, as shown in Table 1.

Table 1. Indicator system of factors influencing energy consumption

Catalog	Factor	Sign	Unit (of measure)
Economic aspects	Gross Domestic Product GDP	x_1	billion CNY
	Primary industry output	x_2	billion CNY
	Secondary sector output	x_3	billion CNY
	Tertiary output	x_4	billion CNY
	industrial output	x_5	billion CNY
	Construction output	x_6	billion CNY
	Secondary sector as a share of GDP	x_7	%
	Tertiary sector as a share of GDP	x_8	%
	Consumption level of the population	x_9	CNY
	Total investment in fixed assets	x_{10}	billion CNY
Social aspects	Total population	x_{11}	all the people
	urbanization rate	x_{12}	%
Technology	Internal expenditure on R&D funds	x_{13}	billion CNY
Energy	Energy production	x_{14}	million tons of standard coal
	Energy prices (price index for fuel and power categories)	x_{15}	%
	Energy intensity (energy consumption per unit of GDP)	x_{16}	Million tons of standard coal/ billion CNY
	Energy consumption structure (share of coal consumption)	x_{17}	%

3.2 Screening main factors

Among the 17 indicators, the output value of the primary industry(x_2) and the output value of the construction industry(x_6) account for a relatively small proportion of GDP compared with the secondary industry and the tertiary industry, and both of them consume less energy and have a smaller impact on the total energy consumption, so the factors x_2 and x_6 are removed. The energy consumption is represented by the symbol y in this paper. Table 2 shows the correlations between y and x_7, x_8, x_{14}, x_{15} are low. So the factors x_7, x_8, x_{14} and x_{15} are not impact factors.

Table 2. Pearson and Spearman correlations

	y&x_7	y&x_8	y&x_{14}	y&x_{15}	y&x_3	y&x_5	x_3&x_5
Pearson	0.598	0.540	0.224	-0.212	0.985	0.986	0.999
Spearman	0.541	0.370	0.226	-0.194	0.994	0.994	1

The mainstay of the secondary sector is industry, and the correlation between industrial output and the output value of the secondary sector is close to 1. The correlation between industrial output value and total energy consumption is slightly higher than that of the output value of the secondary sector, as shown in Table 2, so the variable x3 is removed.

After comprehensive analysis, it is determined that 10 indicators are the main influencing factors of total energy

consumption. They are the regional GDP (x_1), the output value of tertiary industry (x_4), the industrial output value (x_5), the consumption level of residents (x_9), the amount of investment in the whole society's fixed assets (x_{10}), the total number of population (x_{11}), the rate of urbanization (x_{12}), the internal expenditure of the R&D funds (x_{13}), the intensity of energy (x_{16}) and the structure of energy consumption (x_{17}).

① In the economic aspect, economic growth requires the consumption of energy, and the expanding scale of the tertiary sector and the industry-led economic model have a significant effect on the growth of energy consumption. The rising consumption level of the population and the increase in investment lead to the growth of energy consumption while stimulating economic growth. ② In the social aspect, the growth of population size and the increase of urbanization rate lead to the increase of energy consumption. ③ On the technological side, the internal expenditure of R&D funds and energy consumption show two aspects, on the one hand, technological progress promotes social development and improves people's living standards, and good living conditions come at the cost of consuming energy. On the other hand, technological progress improves the efficiency of energy utilization and reduces energy consumption. It is manifested in the increase of internal expenditure on R&D, the growth of energy consumption and the decrease of energy consumption per unit of GDP. ④ In terms of energy. Energy intensity and energy consumption structure have a significant impact on energy consumption, the negative correlation between energy intensity and energy consumption indicates that the energy consumption per unit of GDP is decreasing, and the reduction of the proportion of coal consumption has a positive effect on the total energy consumption.

4 Prediction of total energy consumption

To explore the future development trend of energy consumption in Jiangxi Province, based on the influencing factors confirmed in section 3.2, and combined with relevant data from 1996 to 2021, MLR, ARIMA and SVR models were constructed for prediction. The prediction effects of the three models were compared using MSE, MAPE and R^2.

4.1 MLR model

The total energy consumption is taken as the dependent variable Y, and the 10 influencing factors are taken as the explanatory variables, and the correlation coefficient matrix shows that the correlation coefficients between the variables are large, and there is a strong correlation between the variables. In order to overcome the existence of multi-collinearity between the variables, stepwise regression is used to select the variables with significant effects for modeling[12].

At the significance level of $\alpha_{entry} = 0.1$, $\alpha_{removal} = 0.15$, the optimal regression subset is obtained, which contains five variables: GDP (X1), urbanization rate (X12), internal expenditure on R&D funding (X13), and energy consumption structure (X17). The corresponding equation of the model is:

$$y = -9.03e-12 + 0.648x_1 + 0.681x_{12} - 0.271x_{13} + 0.084x_{17} \quad (1)$$

The t-statistics of the coefficients of X1, X12, X13 and X17 in the equation are all greater than 2, and the p-values are all less than 0.05（see Figure1）, which passes the test of significance, indicating that the effect of each of the independent variables on Y is significant.

The F-statistic of equation is 3372.64 and p value is less than 0.005, the regression equation passes the test of significance indicating that it is significant. The goodness of fit of the equation $R^2 = 0.9988$, adjusted $R^2 = 0.9985$, indicates that the regression equation as a whole is well fitted.

DW = 1.96, indicating that there is no autocorrelation in the equation. The residual plot shows that there is no heteroskedasticity（see figure 1）.

After testing and analyzing, the multiple regression model is valid and can be used for the prediction of total energy consumption.

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
C	-9.03E-12	0.007749	-1.17E-09	1.0000
X12	0.681015	0.034491	19.74479	0.0000
X1	0.648137	0.059754	10.84680	0.0000
X17	0.084256	0.018625	4.523788	0.0002
X13	-0.271055	0.034008	-7.970423	0.0000

R-squared	0.998689	Mean dependent var	-3.85E-11
Adjusted R-squared	0.998439	S.D. dependent var	1.000000
S.E. of regression	0.039512	Akaike info criterion	-3.453362
Sum squared resid	0.032786	Schwarz criterion	-3.211420
Log likelihood	49.89371	Hannan-Quinn criter.	-3.383692
F-statistic	3998.000	Durbin-Watson stat	2.091345
Prob(F-statistic)	0.000000		

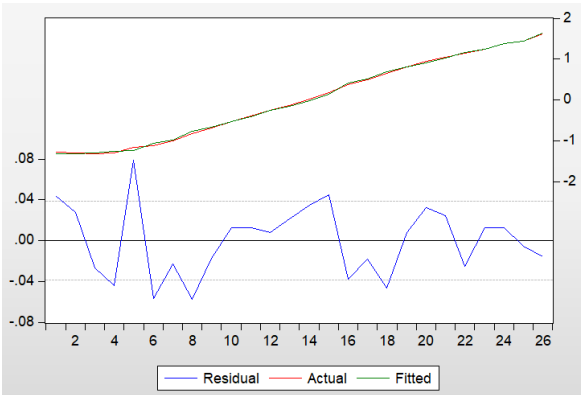
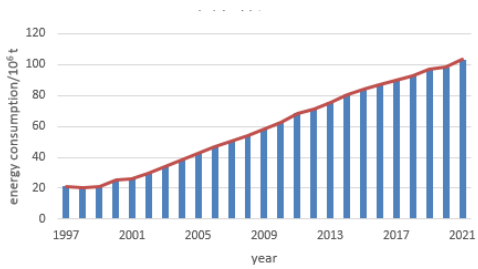


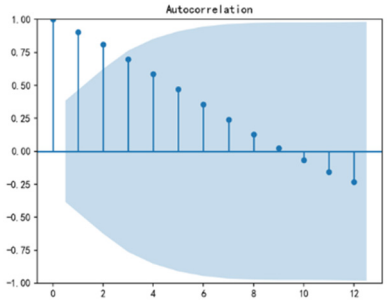
Figure 1. Regression model and residual plots with Eviews

4.2 ARIMA model

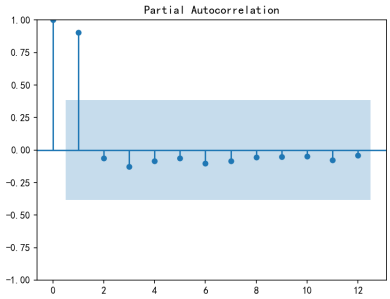
The total energy consumption of Jiangxi Province from 1996 to 2021 was used as sample data for a random time series model, and Eviews was used for modeling. The time series, autocorrelation, and partial autocorrelation were used to determine that the sequence was non-stationary (see figure 2). The p-values of the 6th and 12th order tests in the white noise test were far less than 0.05, indicating that the sequence was a non pure random sequence.



(a) Total energy consumption



(b) Autocorrelation plot



(b) Partial autocorrelation plot

Figure 2. Energy consumption series, autocorrelation and partial autocorrelation plot

Combining the first-order difference autocorrelation and partial autocorrelation plots and the BIC criterion, the model was determined to be ARIMA(1,1,1). The equation of the model is:

$$Y_t = 0.97Y_{t-1} + e_t - 0.5063e_{t-1} \tag{2}$$

According to the test result (as shown in Figure 3),The p-values of the coefficients AR(1) , ma(1) in the equation and the variance of the model estimation Sigma2 converge to 0, which is less than the significance level (0.05) and passes the test.

The Q-stat= 0.32, prob=0.57, indicating that there is no autocorrelation in the residuals. The JB-stat= 1.06, prob=0.59, indicating that the residuals are consistent with the assumption of normality. The H-stat=0.58, prob=0.46, indicating that the residuals are not heteroskedasticity.

After verification and analysis, the ARIMA model is effective and can be used for predicting the ttal energy consumption.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      Y   No. Observations:                   26
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -160.735
Date:                Thu, 15 Feb 2024   AIC                            327.469
Time:                        14:44:31   BIC                            331.126
Sample:                             0   HQIC                           328.483
                                 - 26
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9700      0.047     20.488      0.000       0.877       1.063
ma.L1         -0.5063      0.222     -2.278      0.023      -0.942      -0.071
sigma2      2.063e+04   5516.839      3.739      0.000    9814.077    3.14e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.32   Jarque-Bera (JB):                 1.06
Prob(Q):                              0.57   Prob(JB):                         0.59
Heteroskedasticity (H):               0.58   Skew:                             0.48
Prob(H) (two-sided):                  0.46   Kurtosis:                         2.69
===================================================================================
```

Figure 3. ARIMA model results

4.3 SVR model

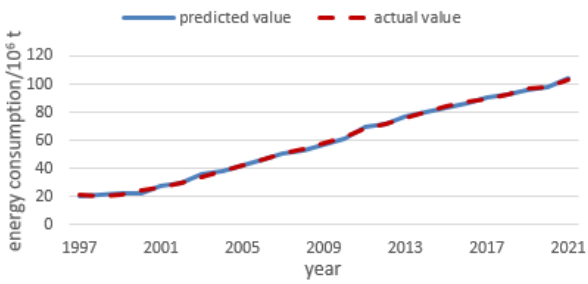
Based on the Python Sklearn package, the SVR model was trained and tested[13]. The standardized data was randomly divided into a training set and a testing set in an 8:2 ratio. Three types of kernel functions (linear kernel function, polynomial kernel function, and radial basis kernel function) were used for training. After parameter tuning, the model performed best when the penalty parameter C was set to 100 and the distance threshold epsilon was set to 0.001. Table 3 shows that the SVR with "Rbf" kernel has the best performance. Therefore, it is determined to establish an SVR prediction model with radial basis function kernel.

Table 3. Evaluation indexes of SVR based on three different kernel functions

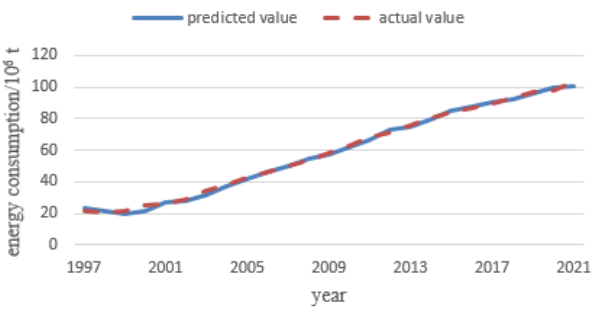
Kernel	MSE	MAPE	R^2
Linear	0.00800	19.76	0.9898
Polynomial	0.04681	40.27	0.9402
Rbf	0.00072	4.64	0.9991

4.4 Comparison of 3 forecasting models

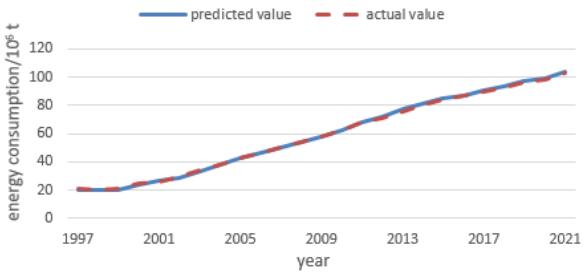
Compare the actual and predicted values of three methods (see Figure 4), SVR has the best effect, goodness of fit R^2 also indicates this point. The MSE and MAPE of the three models shows that: the SVR model is superior to the ARIMA model and the ARIMA model is superior to the MLR model, as shown in Table 4. Through the comprehensive evaluation of fitting accuracy and prediction effect, SVR model has the best performance, and SVR model is selected as the final prediction model.



(a) Comparison between the predicted values and the actual values of MLR



(b) Comparison between the predicted values and the actual values of ARIMA



(c) Comparison between the predicted values and the actual values of SVR

Figure 4. Comparison of results of MLR, ARIMA and SVR

Table 4. Comparison of indicators for evaluating three prediction models

Model	MSE	MAPE	R^2
MLR	0.03375	13.792	0.99882
ARIMA	0.00268	7.329	0.99699
SVR	0.00072	4.64	0.9991

5 Conclusions and suggestions

5.1 Conclusion

This paper constructs an index system of energy consumption influencing factors that includes 17 indicators from four aspects: economic, social, technological, and energy. Based on comprehensive analysis of statistical data from 1996 to 2021, 10 main factors are determined, and three prediction models are constructed: multiple regression model (MLR), time series model (ARIMA), and support vector regression (SVR). The following conclusion is drawn:

(1) Energy consumption is mainly affected by 10 factors, including GDP, output value of tertiary industry, industrial output value, consumption level of residents, fixed assets investment of the whole society, total population, urbanization rate, internal expenditure of R&D funds, energy intensity and energy consumption structure.

(2) MLR, ARIMA, and SVR models can all be used for energy consumption prediction in Jiangxi Province, among which SVR has the best prediction effect.

5.2 Suggestions

(1) Promote intensive economic growth and improve the quality of economic development.

Economic growth needs energy support, and the increase in energy consumption brings vitality to the economy, both increasing and decreasing. The GDP of Jiangxi Province is increasing year by year and the economic situation is good, but many of them are extensive industries with high input and high production. Therefore, changing high energy consumption production and consumption patterns in the past and promoting intensive production methods with low energy consumption, low input, and high income are beneficial for economic growth and energy consumption reduction.

(2)Increase investment in technological innovation and reduce industrial energy consumption.

Increase investment in technological innovation, take the path of innovative development, improve energy utilization efficiency, carry out industrial reforms, and reduce the energy consumption of GDP. To increase efforts to develop some high-tech, high-efficiency and environmentally friendly, and low energy consuming industries, optimize industrial structure and resource allocation reasonably, gradually reduce energy consumption in the industrial sector, and achieve industrial transformation.

(3)Raise the entry threshold for investment, prohibit investment in energy-intensive production projects that lack energy and environmental support conditions, and allocate more investment to promote green technology progress. Increase investment in high-end manufacturing, energy-saving and environmental protection industries, and infrastructure green.

(4) Optimize energy consumption structure and improve energy utilization efficiency.

Improve the energy supply system, increase the supply of natural gas and raise the proportion of oil and gas energy consumption. Deepen the exploration of solar and wind energy resources, further develop clean energy sources such as hydropower and geothermal, and reduce reliance on traditional energy. Introduce advanced energy production and consumption technologies to improve energy conversion rates.

References

1. Zhao Z C,Liu Q Y. (2019)China's energy strategy planning based on prediction of energy consumption, production and structure [J]. Resources & Industries.21(6):1-8. http://www.resourcesindustries.net.cn.
2. Wu D, Pu M Y. (2024) Study on Relationship Between Energy Consumption and Economic Growth in Anhui Province[J]. Journal of Shijiazhuang University.26(1):76-82. DOI:10.13573/j.cnki.sjzxyxb.2024.01.020.
3. Jiang G H. (2008) Study on the Energy Consumption Factors and the Trends of Chongqing [D]. Chongqing University.
4. Liu J. (2012) Empirical Study on Energy Consumption and Influencing Factors in Beijing [D]. Capital University of Economics and Business.
5. Vaghefi A, Jafari M A, Zhu J M,et al.(2013)A Statistical Approach to Modeling and Forecast of the Building Energy Consumption[J]. Proceedings of the 2013 Industrial and Systems Engineering Research Conference. PP 2828-2845.
6. Zhao J H.(2014) Analysis of influential factors of energy consumption in China based on Principal Component Regression model[J]. China mining magazine. 23(1),44-49.
7. Zhang G Y.(2014) Energy consumption in Shanxi province and its influencing factors [D].North Central University.
8. Li H. (2015) Analysis of Influencing Factors of Energy Consumption and Demand Forecast in Jiangxi Province [J]. Marketing Research. 8(23):45-47. doi:10.13999/j.cnki.scyj.2015.08.023.
9. Luo Z.P. (2019) Analysis of influencing factors of energy consumption in Beijing and short-term forecast [D]. Hebei University of Economics and Trade.
10. Zhang L, Li Y A, Wang C. (2020) Research on the Influencing Factors of Energy Consumption Based on Correlation Analysis and Regression Analysis [J], Research and Approach. 42(6):42-47. doi: 10.3969/j.issn.1003 - 2355.2020.06.008.
11. Yuan J H, Liu Z H, Peng T D. (2023) Potential Difficulties and Its Solution for Jiangxi ' s Energy Transformation Under the Goal of Carbon Neutrality[J].LAOQUJIANSHE.(12):3-11
12. Yusri S A, Saiful M, Sri M S,et al. (2023)Factors influencing urban consumers on selecting electricity saving home appliances for managing energy consumption in Indonesia[J]. Cogent Engineering.2023.10(2):1075-1097. DOI: 10.1080/23311916.2023.2287299.
13. Sachin K, Zairu N, Jagvinder S,et al. (2022)Anuj Kumar. Sensor network driven novel hybrid model based on feature selection and SVR to predict indoor temperature for energy consumption optimization in smart buildings[J]. International Journal of System Assurance Engineering and Management.2022.13(6):3048-3061. DOI: 10.1007/S13198-022-01795-y