

# Prediction and Influencing Factors Analysis of Natural Gas Consumption in China Based on SPSS

Guang-jing He<sup>1</sup>, Rong-ge Xiao<sup>1</sup> and Shuai Liang<sup>2</sup>

<sup>1</sup>School of Earth Sciences and Resources, China University of Geosciences, Beijing, 100000, China

<sup>2</sup>School of Mining, Liaoning Technical University, Liaoning, 123000, China

**Keywords:** natural gas consumption, factor analysis, linear regression, prediction, SPSS

**Abstract.** Natural gas is the third largest energy pillar in the world, the best energy that all countries are scrambling to develop. Five main influencing factors of natural gas consumption are analyzed by collecting relevant information, including GDP, the gross industrial output value, the increased value of the third industrial production, the urban population, and the proportion of natural gas in primary energy. Then based on data from 2001 to 2011, factor analysis is taken by using the SPSS software. Then a linear regression model is obtained to predict the natural gas consumption. At last, the natural gas consumption in 2011-2013 is predicted by the proposed model, and the result is analyzed which shows that the model based on SPSS is reasonable and efficient.

## 1. Introduction

In 2013, the apparent consumption of natural gas in China has reached 167.6 billion cubic meters, increasing 13.9% on year-on-year basis. Nowadays, China has become the third largest natural gas consumption country after America and Russia. The proportion of natural gas in primary energy consumption structure of China has been increased from 2.4% in 2000 to 5.9% in 2013.

At present, scholars have done a lot of researches on influencing factors of natural gas consumption[1,2]. Zhao Xiaoqin et al[3] held the opinion that the influencing factors of natural gas consumption contain national economy, energy policy, population and so on, on the basis of which they extracted GDP, per capita GDP, total population, industrial production and the output of the second industry as main factors of gray correlation analysis to extract long-term influence factors. Zhou Yuezhong et al[4] pointed that the stage of economic development, price of natural gas and alternative energy are the main factors of natural gas consumption. Yuan Shuming et al[5] took combination prediction on the natural gas consumption of Guangzhou by extracting GDP, the resident population, the output of the second and third industry. Yeqian[6] found four kinds of factors including economic factors, population, restricted factors and development factors. However, it is necessary to analyze the influencing factors more comprehensive and to realize the accurate forecasting of natural gas consumption, providing a basis for future energy planning.

## 2. Analysis of Factors Influencing the Demand for Natural Gas

Main influencing factors of natural gas consumption are analyzed base on relevant documents and studies.

(1) Gross domestic production (GDP). Studies show that there is a casual relationship between the consumption of natural gas and the economic development level of the whole country or a region[7], which means that the faster the national economic development is, the greater the natural gas consumption will be. For example, the annual GDP growth rate of China in 2004-2006 is 14.4%, and over the same period the one of natural gas consumption is 18.1%. Meantime, Sun Yanping[8] concluded that there is a long-term equilibrium between natural gas demand and GDP. Therefore, GDP is chosen as an important influencing factor of the natural gas consumption.

(2) Gross industrial output value. Natural gas is important industrial raw material and fuel, whose consumption has a positive correlation with the gross industrial output value, that is to say,

the greater output value is, the higher the consumption will be. Therefore, the gross industrial output value should be regarded as a main factor of natural gas consumption.

(3) The increased value of the third industrial production. Restaurants, hotels and other commercial gas consumption as well as gas consumption from transport are embodied by the increased value of the third industrial production. To some extent, the increased value of the third industrial production stands for the increased natural gas consumption.

(4) Urban population. At present, natural gas can be applied as city gas to the urban residents cooking or hot water heating in daily life. The population number directly affects the total consumption, per capita and use pattern of natural gas. In general, the more the urban population is, the more the natural gas demand will be. Therefore, urban population is selected as a main factor.

(5) The proportion of natural gas in primary energy. With fast growing energy consumption, natural gas accounts for a greater proportion of primary energy, indicating that the greater amount of natural gas consumption. There is a clear positive correlation between them, as a result of which the proportion of natural gas in primary energy is regarded as a main factor.

### 3. Factor Analyses of Influencing Factors about Natural Gas Consumption Based on SPSS

#### 3.1 Data Sources

In order to analyze influencing factors of natural gas consumption, factor analysis is applied to extract the principle component, and the relationship between it and the natural gas consumption is analyzed to predict future gas consumption based on the statistical package for social sciences (SPSS), combining with data from "China Statistical Yearbook" in 2001 to 2013[9].

#### 3.2 Factor Analysis Steps of Natural Gas Consumption Influencing Factors Based on SPSS

Firstly, correlation analysis is taken to get the correlation matrix, seeing in table 1, from which we know that there is great correlation among the five influencing factors, which means that if taking direct analysis, there is a serious problem of colinearity. Therefore, it is necessary to take the principle component analysis.

Table1: Correlation matrix

		x1	x2	x3	x4	x5	
correlation	x1	1.00	.999	.999	.982	.968	
			0				
	x2	.999	1.00	.998	.980	.966	
				0			
	x3	.999	.998	1.00	.977	.967	
					0		
	x4	.982	.980	.977	1.00	.939	
						0	
	x5	.968	.966	.967	.939	1.00	
							0

Secondly, SPSS is used to get the list of principle component, seeing in table 2, in which all of influencing factors are arranged according to the characteristic root. The characteristic root of the first principle component is 4.911, accounting for 98.2% of the total variability.

Table 2: List of principle component

component	initial eigenvalues			extract and load quadratic sum		
	grand total	variance%	accumulation%	grand total	variance%	accumulation%
1	4.911	98.214	98.214	4.911	98.214	98.214
2	.064	1.270	99.485			
3	.024	.473	99.958			
4	.002	.039	99.997			
5	.000	.003	100.000			

Extracting method: principle component analysis.

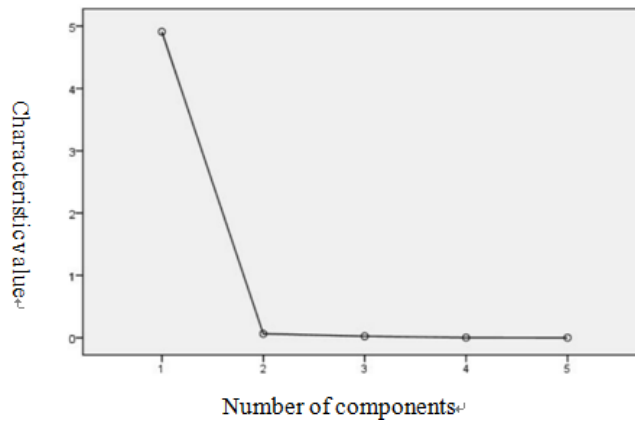


Fig 1. The scree plot

Then combined with the scree plot (seeing in figure 1), we can know that the characteristic value of other main principles are low except the first main principle. Therefore, it is only necessary to get the first principle component.

Finally, we get the component matrix seeing in table 3, and then the first component can be represented by the linear combination of all the variables, that is:

$$z = 0.203 * zx_1 + 0.203 * zx_2 + 0.201 * zx_4 + 0.199zx_5 \quad (1)$$

Where  $zx_i$  stands for the five influencing factors of natural gas consumption,  $i=1,2,3,4,5$ .

Table 3: Component score matrix

	component 1
GDP	.203
the gross industrial output value	.203
the increased value of the third industrial production	.203
the urban population	.201
the proportion of natural gas in primary energy	.199

Extracting method: principle component analysis.

Score of components.

### 3.3 To establish a linear regression model by using SPSS

Because there exists significant linear correlation between natural gas consumption and the value of the first component, so the linear regression analysis can be taken next. Based on the first principle component in the above, we analyze the relationship between it and natural gas consumption by using the linear regression model.

(1) Standardize all the raw data by range standardization method. The first main component is a linear combination of normalized values of the five factors, and we should firstly standardize all the data according to certain principles to get the value of the first principle component. In order to simplify the calculation, we chose range standardization method to process data, the processing result sees in table 4.

Table 4: Data of main influencing factors after range standardization and natural gas consumption

Year s	$zx_1$	$zx_2$	$zx_3$	$zx_4$	$zx_5$	z	Natural gas Consumption
2001	-0.39	-	-	-	-	-0.41	274.3
		0.39	0.39	0.49	0.38		
2002	-0.36	-	-	-	-	-0.36	291.84
		0.36	0.34	0.38	0.31		
2003	-0.31	-	-	-	-	-0.33	339.08
		0.31	0.30	0.27	0.45		
2004	-0.22	-	-	-	-	-0.24	396.72
		0.21	0.23	0.16	0.38		
2005	-0.14	-	-	-	-	-0.15	467.63
		0.13	0.15	0.06	0.25		
2006	-0.03	-	-	0.05	0.02	0.00	561.41
		0.02	0.04				
2007	0.14	0.13	0.13	0.17	0.22	0.23	705.23
2008	0.31	0.29	0.29	0.26	0.48	0.35	812.94
2009	0.40	0.36	0.42	0.38	0.49	0.45	895.2
2010	0.61	0.61	0.61	0.51	0.55	0.58	1069.41

(2) Linear fitting is carried out on consumption of natural gas and the first principle component, with confidence interval being 95%. Then we can get the correlation coefficient R is equal to 0.995, which is the determination coefficient of regression analysis, representing the proximity of between scatters and regression curve. Meantime, the goodness of fit  $R^2$  is equal to 0.990, the adjusted one is equal to 0.989, and the error of standard estimation is 28.92043. After that, through analyzing the scatter diagram, we know the strong correlation between them, and then we get the result of variance analysis seeing in table 5.

Table 5: Variance analysis

model		quadratic sum	df	mean square	F	Sig
1	regression	683202.008	1	683202.008	816.845	.000
	residual error	6691.129	8	836.391		
	grand total	689893.136	9			

- a. Predictive variable: (constant), the first principle component z.
- b. Dependent variable: the natural gas consumption.

Where Sig. in table 5 stands for the significance coefficient of regression relationship, which is smaller than 0.5, showing that the establishment of regression model is statistically significant.

(3) Then, according to the coefficient table 6, a regression equation is obtained between natural gas consumption and the first principle component to predict future of natural gas consumption, that is:

$$Y=754.681X+572.320. \quad (2)$$

Table 6: The coefficient table

model		unstandardized		standard	t	Sig
		B	standard error	trial version		
1	constant	572.320	9.151		62.542	.000
	the first principle component z	754.681	26.405	.995	28.581	.000

a. Dependent variable: the natural gas consumption.

(4) At last, we predict the consumption of natural gas in 2011-2013 on the basis of the raw data by using the proposed model. The predicted consumption in 2011 is 115.23 billion cubic meter, 132.5 billion in 2012 and 178.9 billion in 2013.

(5) Compared with the real value of natural gas consumption (130.5, 147.1, 167.6 billion cubic meter) in 2011-2013, we find that the predicted value has a little variance with the real one, but the variance is reasonable, because it may result from government policies and other factors that we didn't consider.

#### 4. Conclusion

Five influencing factors of natural gas consumption are concluded, including GDP, the gross industrial output value, the increased value of the third industrial production, the urban population, and the proportion of natural gas in primary energy. And based on data in 2001-2011, factor analysis is taken by using SPSS to get a first principle component. Then a linear regression model is obtained to predict future natural gas consumption. At last, the natural gas consumption in 2011-2013 is predicted by the model. However, as a result of time limitation, factors such as price factor, infrastructure construction and government policies have not been taken into consideration, which affects the accuracy of the evaluation results.

#### References

- [1] Hong Jianjun. Supply and demand situation of natural gas in world. *Power and Energy*, 2012, 33(5): 467-470.
- [2] LI De-sheng, LI Bo-hua. Create a new and multielement era for energy sources development by integration of petroleum geology with environmental geology. *Xinjiang Petroleum Geology*, 2009, 30(4): 403-406.
- [3] Zhao Xiaoqin, Kang Zhengkun, Wu Fengrong. Influencing factors on piped natural gas market and gray correlation analysis. *Oil and Gas Storage and Transportation*, 2008, 27(8): 5-8.
- [4] Zhou Yuezhong, Liting. Exploration on the influencing factors of natural gas consumption demand. *Market Modernization*, 2008, 15(10): 187-188.
- [5] Yuan Shuming, Lai Jianbo, Zhang Dekun, et al. Forecast on long term planning of gas consumption of city gas pipeline network. *Gas and Heat*, 2009, 29(5): 24-27.
- [6] Yeqian. City nature gas demand forecast research and application. Chongqing: Chongqing University, 2010: 25-31.
- [7] Wang Lijuan. Research on forecasting methods of natural gas load based on particle swarm optimization-least squares support vector machine. Shanghai: East China University of Science and Technology, 2012: 36-40.
- [8] Sun Yanping. Research on cointegration analysis and error correction model of the relationship between natural gas demand and GDP in China. *Special Zone Economy*, 2007, 12: 264-265.