UDC 519.233.5 Вестник СПбГУ. Прикладная математика. Информатика... 2022. Т. 18. Вып. 1
MSC 62P12

Statistical analysis of investment attractiveness of China's regions*

D. Qi, V. M. Bure

St Petersburg State University, 7–9, Universitetskaya nab., St Petersburg, 199034, Russian Federation

For citation: Qi D., Bure V. M. Statistical analysis of investment attractiveness of China's regions. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2022, vol. 18, iss. 1, pp. 188–194. https://doi.org/10.21638/11701/spbu10.2022.116

This study is devoted to the application of methods of applied statistics to the study of investment attractiveness of regions of China. Methods of applied statistics are widely used in various engineering research, medicine, economics, sociology, astronomy, ecology, physics, information technology, etc. statistical tools is an integral part of the study. Data taken from the World Bank website and the China Statistical Yearbook. The main topic of this research is: What factors are most strongly associated with investment attractiveness in the economy of modern China? Such a study can only be carried out based on the analysis of economic data, especially in the conditions of the modern digital economy. Consequently, the problem under consideration belongs to the field of applied mathematics. First of all, need to choose a mathematical model and research method. This paper uses the method of regression analysis - one of the most important methods for analyzing economic data. Regression models are mathematical models built from empirical data. Observations (empirical data) are the numerical data on the level of investment in the regions of China, as well as the numerical values of various factors. For each year, a multiple regression model is built using the least squares method, its statistical significance is checked, statistically significant factors are selected (statistically significant coefficients for the factors correspond to them). Calculations are carried out in Excel and SPSS, using subroutines and. Mathematical functions, but at the same time, an algorithm for analyzing the initial data based on a stepwise regression algorithm has been developed, in which only one factor with the least significant coefficient (maximum p-value (t)) is discarded at each step and then an algorithm for choosing the most important factors for investment attractiveness was developed, the essence of this algorithm is that the regression models are compared for each year and how many times each factor was included in the model is calculated algorithms can be attributed to the application of methods informatics in this work. The choice of the most important factors that determine the level of investment attractiveness is made to solve the problem of managing the economy. Attracting investments is important for the effective development of every region, every city, every district. With the receipt of significant investments, it is possible to solve the problems of economic and social development. Determination of the most important factors will make it possible to most effectively solve the problem of managing the economic development of each region, each city, each district.

Keywords: investment attractiveness, stepwise regression algorithm, multiple linear regression models, significant coefficients, least squares method.

1. Introduction. Over the past decade, there has been a significant change in the combination of political and economic forces in the world, which has affected not only the single national economy of the country but also the economies of all countries and

* This work was supported by the Ministry of Science and Education of the Russian Federation (agreement N 075-15-2020-805).

regions of the world. The international business environment includes various factors, such as social, political, regulatory, cultural, legal and technological ones that surround a business entity in different sovereign countries. There are exogenous factors associated with the internal economic environment of each country. These factors influence not only the decision-making process about the use of resources and opportunities, but also the country's attractiveness for international business. Our purpose is to identify the most important factors related to investment attractiveness [1].

Firms have no control over the external business environment. Therefore, the success of an international company depends on its ability to adapt to the general environment. Its success also depends on the ability to regulate and manage the internal characteristics of the company to take advantage of the opportunities presented by the external environment. In addition, a company's ability to control various threats posed by the same environment also determines its success. The term "country's attractiveness" is often discussed in the international business community. It is crucial to consider attractiveness before we move on to discussing the environmental factors [2].

A country's "attractiveness" is an indicator of a country's attractiveness to international investors. In international business, investment in foreign countries is the essential aspect. Therefore, firms want to determine the appropriate level for a country in terms of its external business environment. International business firms assess the risks and profitability of doing business in a given country before investing in and starting a business there. This judgment involves examining environmental factors to arrive at a decision. Business prefers a less costly, more profitable, and less risky country. Cost considerations are related to investment and profitability depends on resources. To summarise, a more stable country in terms of political, social, legal, and economic conditions is an attractive place to start a business [3].

2. The foundation for building the model. This study consists of building several empirical models based on statistical data to identify the degree of influence of certain socio-economic factors on investment attractiveness. Research objects are regions of China.

The purpose of the work includes the following points: the selection of statistical factors influencing the "investment attractiveness" indicator; building a Multiple Linear Regression Model; verification of the estimate of the quality of the model using the coefficient of determination and the corrected coefficient of determination; check the significance of the equation and coefficients for factors; analysis of the received data.

Investment attractiveness models are built under the following assumptions. Primarily, the volume of investments per region characterizes the level of investment activity, therefore, the level of investment attractiveness. Another is that the level of investment attractiveness depends on several indicators (factors) reflecting the social, financial, economic, legal and geopolitical characteristics of the region.

3. Logarithm investment attractiveness model. We try to establish the following econometric model:

$$\begin{gathered}\hat{y}_t = b_0 + b_1 x_{1,t} + b_2 x_{2,t} + {}_3\, x_{3,t} + b_4 x_{4,t} + b_5 x_{5,t} + \\ + b_6 x_{6,t} + b_7 x_{7,t} + b_8 x_{8,t} + b_9 x_{9,t} + b_{10} x_{10,t},\end{gathered} \tag{1}$$

where t is year number, $t = 1, \ldots, 10$; $\hat{y}_t$ is logarithm of the estimate of the volume of investments in the current year; $x_{1,t}$ is electricity consumption; $x_{2,t}$ is average per capita money income in the current year; $x_{3,t}$ is debt on loans provided by credit institutions to legal entities in the current year; $x_{4,t}$ is the cost of fixed assets in the current year; $x_{5,t}$ is research costs in the current year; $x_{6,t}$ is the amount of work performed by the type of activity "Construction" in the current year; $x_{7,t}$ is the number of enterprises and

organizations in the current year; $x_{8,t}$ is retail trade turnover in the current year; $x_{9,t}$ is GNP per capita in the current year; $x_{10,t}$ is unemployment rate in the current year (in percent) [4].

For each year with numbers t ($t = 1, \ldots, 10$), there is information on individual regions (a total of 26 regions that have a high level of investment): $y_{1,t}, \ldots, y_{26,t}$ are the level of investments in the regions; $i = 1, \ldots, 26$, i is the region number.

According to the information for the year with numbers t, a regression model is constructed using the least squares method (1):

$$\sum_{i=1}^{26}(y_{i,t} - b_0 - b_1 x_{i,1,t} - b_2 x_{i,2,t} - \cdots - b_{10} x_{i,10,t})^2 \to \min_{b_0,\ldots,b_{10}}. \tag{2}$$

The method of multiple linear regression was used to carry out the calculations. Using the "regression" tool of the "data analysis" add-in in MS Excel, we will perform a regression analysis of the available values of the y and x.

The first step is to build the logarithm investment attractiveness of China's regions model by checking data assumptions: normality, linearity, and heteroscedasticity. From the first assumption all of the variables in our research must have a normal distribution. The normal distribution can be seen by the histogram graph, plot Q — Q [*)], kurtosis, and skewness. If the distribution of data is not normal, we need transformation. In our research all data has a normal distribution. Then multiple linear regression should have a linear relationship between the response variable and the controlled variables. The third assumption in multiple linear regression is that any data should be free from heteroscedasticity, wich will happen whenever there is an interruption in the model that is not fulfilled. If the important variable in the model is missing, hence heteroscedasticity will happen.

3.1. Multiple regression model for 2017. The first row of the Table corresponds to a model that includes all factors for which $R^2 = 0.935$, therefore, according to the initial data, approximately 94 % of the variation in the studied indicator y can be "explained" by the influence of a combination of factors contained in the model (see Table). Taking into account the penalty imposed for a large number of explanatory variables, $R^2_{adj} = 0.892$, that is the regression equation determines 89 % of the variance of the dependent variable (within the observed values of y). The multiple correlation coefficient $R = 0.967$, which is extremely close to 1, which means that there really is a close relationship between y and a number of factorial features $x_1, x_2, \ldots, x_{10}$.

Table. **The dynamics of select final model by backward elimination stepwise**

Variable x	R	R^2	R^2_{adj}	F-value	$P(F)$*
1, 2, 3, 4, 5, 6, 7, **8**, 9, 10	0.967	0.935	0.892	21.81	$3.7254 \cdot 10^{-7}$
1, 2, 3, 4, 5, 6, 7, 9, **10**	0.967	0.935	0.899	25.85	$6.9766 \cdot 10^{-8}$
1, 2, 3, 4, 5, 6, **7**, 9	0.967	0.935	0.905	30.89	$1.1947 \cdot 10^{-8}$
1, 2, 3, 4, 5, 6, 9	0.967	0.935	0.910	37.13	$1.9585 \cdot 10^{-9}$
2, 3, **4**, 5, 6, 9	0.964	0.930	0.908	42.50	$5.1728 \cdot 10^{-10}$
2, 3, 5, 6, 9	0.958	0.919	0.899	45.69	$2.8909 \cdot 10^{-10}$

* If $P(F) < \alpha$, then the regression model is considered statistically significant at a given significance level. For the analyzed model, $P(F) = 3.72 \cdot 10^{-7}$ (less 0.05).

*)] In statistics, a Q — Q (quantile — quantile) plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.

We conclude that the assessment of the quality of the regression model, which includes all factors, is quite high, it has high statistical significance according to Fisher's criterion ($P(F)$) and well approximates y according to the initial data.

Let's set the level of significance at α. Within the framework of statistical research, traditionally $\alpha = 0.05$ [5, 6].

We will sequentially exclude factors that correspond to statistically insignificant estimates of the coefficients until all of them become significant for a given α [7].

The last row of the table corresponds to the final model obtained by the stepwise regression process. For 2017 we have

$$y = 10712.19 - 0.45x_2 - 6.704x_3 - 0.00263x_5 + 0.7599x_6 + 1.2368x_9.$$

3.2. Multiple regression models. By analogy with the process of building the final regression model for 2017 discussed above, we will build models for each year in the period from 2008 to 2016:

for 2008 ($P(F) = 2.38 \cdot 10^{-17}$):

$$y = 1687.516 - 0.10266x_2 + 0.966363x_4;$$

for 2009 ($P(F) = 4.36 \cdot 10^{-13}$):

$$y = 4642.47 - 0.268x_2 + 0.924488x_4 + (3.96 \cdot 10^{-5})x_6;$$

for 2010 ($P(F) = 3.01 \cdot 10^{-12}$):

$$y = 4682.416 - 0.22744x_2 + 1.067083x_4;$$

for 2011 ($P(F) = 4.51 \cdot 10^{-13}$):

$$y = 6043.807 - 0.28657x_2 + 0.888326x_4 + (4 \cdot 10^{-5})x_6;$$

for 2012 ($P(F) = 2.14 \cdot 10^{-12}$):

$$y = 8342.07 - 0.34x_2 + 0.878616x_4 + (5.32 \cdot 10^{-5})x_6;$$

for 2013 ($P(F) = 2.61 \cdot 10^{-12}$):

$$y = 10791.72 - 0.3948x_2 + 0.913893x_4 + (5.94 \cdot 10^{-5})x_6;$$

for 2014 ($P(F) = 5.79 \cdot 10^{-12}$):

$$y = 7537.944 - 0.33494x_2 + 0.960512x_4 + (5.52 \cdot 10^{-5})x_6;$$

for 2015 ($P(F) = 2.87 \cdot 10^{-12}$):

$$y = 8714.362 - 0.3407x_2 + 1.034456x_4 + (6.82 \cdot 10^{-5})x_6;$$

for 2016 ($P(F) = 3.10 \cdot 10^{-11}$):

$$y = 9857.724 - 0.38742x_2 + 1.123665x_4 + (5.95 \cdot 10^{-5})x_6.$$

4. Result and discussion. Data obtained from the World Bank [8] and the China Statistical Yearbook [9] takes inflation into account (all quoted at comparable prices).

The key point of applied statistical research is the interpretation of the results obtained. After processing the obtained data, the mathematical model of the relationship between annual investments in the region and several factors characterizing various aspects of its socio-economic potential should be interpreted in terms of the economy.

In the statistical analysis of China's investment attractiveness, a multivariate regression model was built based on the method of step-by-step removal of unimportant factors based on data for 2008–2017. For 26 regions of China with the largest investments in 2018, 10 regression equations were built. Compare how several socio-economic indicators have changed each year over the past decade, and their cumulative effect has significantly influenced the average cost of investment. The results are shown in Figure.

Year	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	
2008											2
2009											3
2010											2
2011											3
2012											3
2013											3
2014											3
2015											3
2016											3
2017											5
	0	9	1	8	1	7	0	0	1	0	

Figure. Regression result

Factors correspond to the columns of the Figure (for clarity, each factor is associated with a certain cell color), rows correspond to sets of exogenous variables of the final regression model for the corresponding year. At the bottom of the Figure for the k-th factor ($k = 1, \ldots, 10$), the number of inclusions in the regressors of the final models built for 2007–2016 is indicated. The number in the far right column of the Figure is the number of explanatory variables in the final model for the corresponding year.

As for the factor x_2, it is included in all models, that is, this factor (socio-economic indicator) has a continuous impact on the formation of regional investments from 2008 to 2017. The fact is that, the factors x_1, x_7, x_8, x_{10}, excluded from the regression equation means these indicators will hardly affect the volume of investments in this period.

Factor x_2 is average per capita money income is included in all models. It is shown that monetary income per capita is an important factor affecting investment attractiveness and has a positive impact on investment attractiveness.

The next most frequent variable is x_4, which is the cost of fixed assets. It is quite logical that the cost of fixed assets is one of the important factors affecting the investment attractiveness of the country and occupying an important place in the socio-economic plan of the country as a whole. However, investment growth slowed in 2017, increasing by 7.5 % in the first three quarters and declining by 0.7 % year on year. Investment operations are showing new signs of "stability positive changes" in the number of indicators, such as sources financing, investments in the civil industry, investments in equipment manufacturing. Currently, economic benefits from investments have significantly decreased, fiscal consolidation is holding back public and infrastructure investments, financial cuts in financing costs, the real estate market has slowed down, some investments in the sector continue to decline, and other problems more noticeable [10].

The factor x_6 is also important in the regression variables, which is the amount of work performed by the type of activity "Construction". It is likely that the construction

market grew in the studied regions during the period under review. Apparently, many consider real estate investment to be one of the most reliable. The absence of the factor x_6 in the model for 2008 and 2010 may be associated with the peak point of the 2008 crisis and high economic risks.

5. Conclusion. In general, the results obtained allow us to conclude that the unstable economic situation in the world at the micro and median levels of the country significantly affects the formation of investment prerequisites. Equally important is the current stage of the industrial cycle [11]. Indicators of economic risk are especially important in the pre and post-crisis periods, but in the context of stable exports, a significant increase in industrial profits, strong investment demand for scientific and technological reforms and the accelerated development of new individual investments, the volume of investment attraction should continue to grow.

References

1. *International Business Ecosphere*. Available at: https://www.tutorialspoint.com/international/business/management country/attractiveness.htm (accessed: October 06, 2021).

2. Prakash A., Suvojit C. B. *Managing social and economic change with information technology*. Harrysburg, Idea group Publ., 1994, 191 p.

3. Barbora G. The host country attractiveness from perspective of selected FDI indices. *Intern. J. of Interdisciplinarity in Theory and Practice*, 2016, vol. 9, pp. 58–64.

4. Zhou C. The index system of national economy and socialty. *Contemporary China*, 2010, pp. 89–109.

5. Bure V. M., Parilina E. M., Sedakov A. A. *Applied statistics methods in R and Excel*. 3rd ed. St Petersburg, Lan Publ., 2019, 196 p.

6. Bure V. M., Parilina E. M. *Probability theory and mathematical statistics*. 1st ed. St Petersburg, Lan Publ., 2013, 416 p.

7. Kovaleva M. A., Voloshin S. B. *Data analysis*. Moscow, World of Science Publ., 2019, 48 p.

8. *The World Bank*. Available at: http://data.worldbank.org/indicator (accessed: October 01, 2021).

9. *National Bureau of Statistics*. Available at: http://www.stats.gov.cn/tjsj/ndsj/ (accessed: September 30, 2021). (In Chinese)

10. Iakushev V. P., Bure V. M., Mitrofanova O. A., Mitrofanov E. P. Theoretical foundations of probabilistic and statistical forecasting of agrometeorological risks. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2021, vol. 17, iss. 2, pp. 174–182. https://doi.org/10.21638/11701/spbu10.2021.207

11. Iakushev V. P., Bure V. M., Mitrofanova O. A., Mitrofanov E. P.. K voprosu avtomatizatsii postroeniia variogramm v zadachakh tochnogo zemledeliia [On the issue of semivariograms constructing automation for precision agriculture problems]. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2020, vol. 16, iss. 2, pp. 177–185. https://doi.org/10.21638/11701/spbu10.2020.209

Received: October 19, 2021.
Accepted: February 01, 2022.

Authors' information:

Dongfang Qi — Postgraduate Student; st073409@student.spbu.ru

Vladimir M. Bure — Dr. Sci. in Technics, Professor; vlb310154@gmail.com

Статистический анализ инвестиционной привлекательности регионов Китая*

Д. Ци, В. М. Буре

Санкт-Петербургский государственный университет, Российская Федерация,
199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Qi D., Bure V. M.* Statistical analysis of investment attractiveness of China's regions // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2022. Т. 18. Вып. 1. С. 188–194.
https://doi.org/10.21638/11701/spbu10.2022.116

Настоящее исследование посвящено применению методов прикладной статистики к изучению инвестиционной привлекательности регионов Китая. Данные, используемые в работе, взяты с веб-сайта Мирового банка (The World Bank) и Китайского статистического ежегодника (China Statistical Yearbook). Основной темой такого исследования является выявление наиболее значимых факторов, связанных с инвестиционной привлекательностью экономики современного Китая. В статье используется регрессионный анализ, который относится к наиболее важным методам анализа экономических данных. Наблюдения (эмпирические данные) представляют собой числовые данные об уровне инвестиций в регионах Китая, а также числовые значения различных факторов. На основе данных каждого рассматриваемого года построена модель множественной регрессии методом наименьших квадратов, проверена ее статистическая значимость и выбраны статистически значимые факторы (им соответствуют статистически значимые коэффициенты при факторах). Разработан алгоритм анализа исходных данных на основе алгоритма пошаговой регрессии, в котором на каждом шаге отбрасывается только один фактор с наименьшим значимым коэффициентом (максимальным $P(F)$). Суть этого алгоритма заключается в том, что окончательные модели регрессии сравниваются для каждого рассматриваемого года и рассчитывается время включения каждого фактора в окончательные модели. Выбор наиболее важных факторов, определяющих уровень инвестиционной привлекательности, производится для решения задачи управления экономикой.

Ключевые слова: инвестиционная привлекательность, алгоритм пошаговой регрессии, модели множественной линейной регрессии, значимые коэффициенты, метод наименьших квадратов.

Контактная информация:

Ци Дунфан — аспирант; st073409@student.spbu.ru

Буре Владимир Мансурович — д-р техн. наук, проф.; vlb310154@gmail.com

* Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (соглашение № 075-15-2020-805).