

The Analysis of Impact Factors of Foreign Investment Based on Relaxed Lasso

Yanqi He

College of Mathematics and Information, China West Normal University, Nanchong, China

Email: heyanyq11@gmail.com

How to cite this paper: He, Y.Q. (2017) The Analysis of Impact Factors of Foreign Investment Based on Relaxed Lasso. *Journal of Applied Mathematics and Physics*, 5, 693-699.
<https://doi.org/10.4236/jamp.2017.53058>

Received: January 21, 2017

Accepted: March 28, 2017

Published: March 31, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Relaxed Lasso method is used for variable selection to 15 main economic factors affecting foreign investment. While Relaxed Lasso method, method of least squares and regression are compared, the result further reveals the main problem facing foreign investment at the present stage.

Keywords

Relaxed Lasso, Variable Selection, Foreign Investment

1. Introduction

Variables selection will often be put forward when building statistical models. It is not conducive to study the problems that the variables in the model are more or less than the actual variables. In the process of optimizing the models, most explanatory and influential subset of variables need to be found, in order to make the model more reasonable and high forecast precision. In the traditional method, the variable selection and parameter estimation are separated, such as AIC criterion proposed by Akaike [1], BIC criterion proposed by Schwarz G based on Bayes method [2]. But Lasso method regards absolute coefficient function as a penalty term to compress the coefficients of the model, and coefficients which absolute value is relative smaller than others are compressed to 0, so as to achieve the purpose of variable selection and parameter estimation [3]. Zou noted that different coefficients were compressed in same degree in Lasso method, which leading to new problems because of excessive compression coefficient [4]. In view of this situation, a number of scholars have proposed a series of improvements to the Lasso method, Meinshansen Relaxed Lasso proposed by Meinshansen was one of them [5]. Variable selection and coefficient compression were achieved by 2 independent parameters in this method, while preventing excessive compression to coefficient. Meinshansen presented corresponding al-

gorithms, reduced the computational complexity and it can effectively prevent noise variables into the model while improving the accuracy of forecasts.

At present our country is in a new stage of development, the role of foreign direct investment can not be underestimated in China's economic development in a fairly long period of time. Therefore, it still has very important practical significance that the main influencing factors of foreign direct investment in China are deeply discussed. The research on the causes of foreign direct investment and its decision making has been paid much attention by scholars. Xu Jinliang used ordinary least squares method to study the influencing factors of attracting foreign direct investment in Jiangxi Province [6]. Xia fan used the generalized least squares method to study the spatial difference of FDI in China [7]. Zhou Guofu used principal component analysis to study the impact of foreign direct investment in Bohai [8]. Xu Helian used partial least squares method to study the influencing factors of foreign direct investment in China [9]. In this paper, we will use Relaxed Lasso method to study the influencing factors and the degree of influence of foreign investment in China, and find the main problems of foreign investment.

Given a set of observed data a , $i = 1, 2, \dots, n$, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ is a vectors consist of variables, y_i is dependent variable. Linear regression model can be expressed as: $y_i = x_i\beta + \varepsilon_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i$ ($Y = X\beta + \varepsilon$), $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ is the vector of unknown regression coefficients, ε_i is a random error, $Y = (y_1, y_2, \dots, y_n)^T$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$, X is $n \times p$ -order matrix, line i is $x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$, $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2 I$,

$E(Y|X) = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$. Assuming that the observations are independent, or dependent variable y_i is independent in the case of the given observations, While x_{ij} is standardized, that is to say, $\frac{1}{N} \sum_i x_{ij} = 0$, $\frac{1}{N} \sum_i x_{ij}^2 = 1$.

Many regression coefficients in the model are 0, Relaxed Lasso method is used to identify those variables with a coefficient of 0 in the model based on the data obtained, and estimate non-zero coefficient, so as to find out the sparse model.

2. Methodology

2.1. Models and Algorithm

Actually the Relaxed Lasso method of variable selection for linear model is equivalent to take into account the following questions:

$$\hat{\beta}^{\lambda, \phi} = \arg \min \left\{ \sum_{i=1}^n \left(y_i - x_i \{ \beta \cdot 1_{M_\lambda} \} \right)^2 + \phi \lambda \sum_{j=1}^p |\beta_j| \right\} \tag{1}$$

Which parameter is $\lambda \in [0, \infty)$, $\phi \in (0, 1]$, and 1_{M_λ} is the characteristic function regarding the set of variables subscript, that is for all $k \in \{1, 2, \dots, p\}$,

$$\{ \beta \cdot 1_{M_\lambda} \}_k = \begin{cases} 0, & k \in M_\lambda \\ \beta_k, & k \in M_\lambda \end{cases} \tag{2}$$

It is not difficult to find that we consider only the variable which subscripts in

the collection in estimating the Relaxed Lasso. The same as Lasso estimation, parameter λ controls variable selection section, and the second parameter ϕ controls the parts of the coefficient compression. So we have the following conclusions:

- 1) $\phi = 1$, Relaxed Lasso and Lasso are completely equivalent.
- 2) $\phi < 1$, compared with Lasso estimates, the coefficient of compression ratio in Relaxed Lasso are weakened. It can be prevented that some of the significant variables in the model coefficients become 0 because of excessive compression.
- 3) The case of $\phi = 0$ needs special consideration, as the definition above would produce a degenerate solution. In general we define the relaxed Lasso estimator for $\phi = 0$ as the limit of the above definition for $\phi = 0$. In this case, all coefficients are estimated by the OLS-solution in the model.

In conditions of orthogonal design, Relaxed Lasso is:

$$\hat{\beta}_k^{\lambda, \phi} = \begin{cases} \hat{\beta}_k^0 - \phi\lambda, & \hat{\beta}_k^0 > \lambda \\ 0, & |\hat{\beta}_k^0| \leq \lambda \\ \hat{\beta}_k^0 + \phi\lambda, & \hat{\beta}_k^0 < -\lambda \end{cases} \quad (3)$$

where $\hat{\beta}^0$ is the OLS solution.

For the General linear model, Relaxed Lasso algorithm is based on the LARS algorithm, it is actually a two-stage approach, The theoretical description of the algorithm is as follows:

- 1) Compute all ordinary Lasso solutions with the Lars-algorithm. Let M_1, M_2, \dots, M_m be the resulting set of final models. Let $\lambda_1 > \lambda_2 > \dots > \lambda_m = 0$ be a sequence of penalty values so that $M_\lambda = M_k$ if and only if $\lambda \in (\lambda_k, \lambda_{k-1}]$, where $k = 1, 2, \dots, m$, $\lambda_0 := \infty$. (the models are not necessarily distinct, so it is always possible to obtain such a sequence of penalty parameters.
- 2) For each $k = 1, 2, \dots, m$, compute all Lasso solutions on the set M_k of variables, varying the penalty parameter between $[0, \lambda_k]$. The obtained set of solutions is identical to the set of relaxed Lasso solutions $\hat{\beta}^{\lambda, \phi}$ for $\lambda \in (\lambda_k, \lambda_{k-1}]$. The Relaxed Lasso solutions for all penalty parameters are given by the union of these sets.

We find that this method gives all the Relaxed Lasso solutions when the parameter $\lambda > 0$, $\phi \in [0, 1]$ in theory. This simple algorithm is not optimal, however, can be further improved [5].

2.2. Data

According to economic theory and research findings, 14 variables were selected from the infrastructure, human resources, labor cost, market size, exchange rates, labor productivity, concentration factor, trade openness and trade barriers.

Selected variables are as follows: Highway mileage (x_1), Freight turnover (x_2), throughput of post and telecommunications (x_3) reflect infrastructure; The number of students in Colleges and Universities(x_4) reflects the human resource

situation; Average wage of workers (x_5) reflects labor costs; GDP(x_6), GDP growth rate (x_7) the total investment in fixed assets (x_8) and the total retail sales of consumer goods (x_9) reflect the size of the market; The dollar-Yuan exchange rate (x_{10}) reflects exchange rate; The ratio of GDP to employment (x_{11}) reflects the level of labor productivity; Third industry accounted for the proportion of GDP (x_{12}) reflects the agglomeration effect ; Proportion of total imports and exports to GDP (x_{13}) reflects the degree of trade openness Tariff (x_{14}) reflects the degree of trade barriers; We expect wages and tariffs have a negative impact on foreign investment, and other variables have a positive effect.

Data were selected in this article range from 1995 to 2014. Exchange rate data were obtained from the State administration of foreign exchange, other variable data were obtained from *China Statistical Yearbook 1995-2015*. The article utilized the exchange rate of dollar against RMB on the last day from 1995-2014. In order to eliminate dimensional effect among variables, and relatively easy to get a smooth sequence, we take the natural logarithm of the time series data, and then take Centralized criterion, so does not affect the relationship between variables.

2.3. Variable Selection

Then Relaxed Lasso was used to select the 14 variables. All solutions were found just need 20 steps by r-language Relaxed Lasso algorithm. Solution path as shown in **Figure 1**.

Can be seen from the results corresponding to each step in the path of the estimation of model parameters, it only took 15 steps before the results of optimal solution. Results showed that Relaxed Lasso method can realize the estimates of the model parameters and variables selection.

Relaxed Lasso variable selection results showed that the number of students in Colleges and Universities, GDP, the growth rate of GDP have a significant positive effect on foreign direct investment, tariff has a significant negative effect on

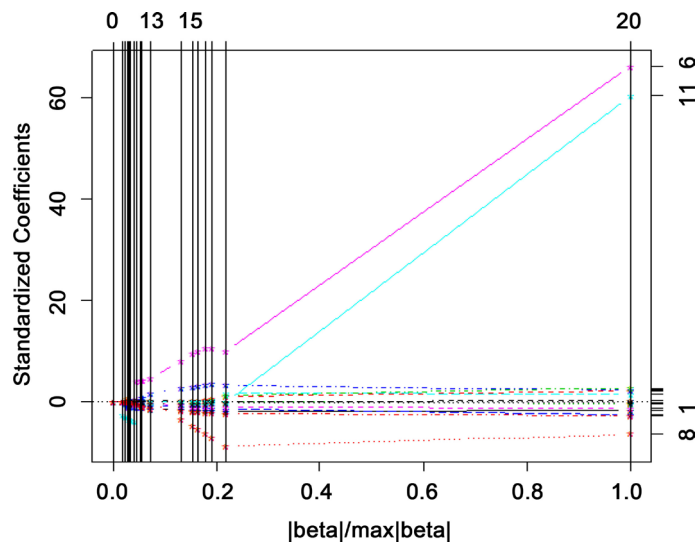


Figure 1. Relaxed Lasso.

foreign direct investment. GDP has most significant positive effect on foreign investment. This shows that foreign investment in China is mainly due to the huge domestic market. The impact of freight turnover, throughput of post and telecommunications, average wage of workers, the total retail sales of consumer goods, the ratio of GDP to employment are not significant, have not been selected into the model. In order to see the Relaxed Lasso advantage of variable selection, we compare the result with least square method and stepwise regression method, the parameter estimates are shown in **Table 1**.

Can be seen from **Table 1**: the parameters can only be estimated and the variable selection can not be realized by Least-squares regression and some notation is obviously inconsistent with the actual situation so that Model cannot be explained. 7 variables were selected by Stepwise regression and 9 variables were selected by the Relaxed Lasso. 9 variables selected by Relaxed Lasso method contain all variables selected by stepwise regression method. The Symbols for parameter estimation are same and numerical values are also relatively close, to some extent it shows the rationality of the results obtained by Lasso method. The factors which have significant influence on foreign investment are only selected by Stepwise regression method and it cut too many variables.

Relaxed Lasso method does not select too many variables, nor excessive delete variables. The final model can be explained better.

3. Conclusions

Based on the above analysis, the major conclusions are as follows:

First, it can be proved theoretically that least squares estimation of parameter is too long on average when data has serious multicollinearity. Parameter

Table 1. Comparison of three methods for parameter estimation.

| Parameters | Least-squares regression | Stepwise regression | Relaxed Lasso |
|--------------|--------------------------|---------------------|---------------|
| β_1 | -0.4077 | -0.4457 | -0.3756 |
| β_2 | 0.4636 | 0 | 0 |
| β_3 | -0.09055 | 0 | 0 |
| β_4 | 0.5249 | 0.7081 | 0.6728 |
| β_5 | 0.3295 | 0 | 0 |
| β_6 | 15.2 | 2.677 | 2.18 |
| β_7 | 0.06574 | 0 | 0.0406 |
| β_8 | -1.459 | -1.616 | -1.078 |
| β_9 | 0.6041 | 0 | 0 |
| β_{10} | -0.5728 | -0.3663 | -0.2346 |
| β_{11} | 13.83 | 0 | 0 |
| β_{12} | -0.3028 | -0.2101 | -0.185 |
| β_{13} | -0.04689 | 0 | -0.0864 |
| β_{14} | -0.6188 | -0.4868 | -0.4175 |

estimates in value obtained by Relaxed Lasso is significantly less than the least squares estimate. It is the compression of the least squares estimation, which can largely eliminate the adverse effects of multicollinearity in the model. At the same time, Relaxed Lasso also has obvious advantages in the selection of high dimensional variables, neither like the least squares method that chooses too many variables, nor like the stepwise regression method to eliminate the excessive variables. The deleted variables are not significant variables to the model, and thus improve the accuracy of the model.

Second, foreign investment is greatly affected by the size of the domestic market. Foreign investment will increase by about 1 percentage points with an increase of two percentage points of GDP. At the same time, GDP growth rate has a certain role in promoting foreign investment.

Third, human resources also have a certain impact on foreign investment. The state of human resources represented by the number of students in Colleges and universities reflects the level of education to a certain extent. In fact, the area with higher education has obtained more foreign investment than other places. At the same time, the technology content of foreign capital is also high. It is most evident in the eastern region. With the increase of the number of students in Colleges and universities, it provides more technical talents for foreign investment, and further enhances our competitiveness in attracting foreign investment with other countries.

Fourth, research shows that: lower tariffs will help our country to attract more foreign investment; therefore, China should further increase the intensity of reform and opening up, accelerate the pace of negotiations on the free trade area, to create conditions to fully participate in the international competition.

References

- [1] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., *Proceedings of the 2nd International Symposium on Information Theory*, Akadémiai Kiado, BudaPest, 267-281.
- [2] Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464. <https://doi.org/10.1214/aos/1176344136>
- [3] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B*, **58**, 267-288.
- [4] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. <https://doi.org/10.1198/016214506000000735>
- [5] Meinshausen, N. (2007) Relaxed Lasso. *Computational Statistics and Data Analysis*, **52**, 374-393. <https://doi.org/10.1016/j.csda.2006.12.019>
- [6] Xu, J. and Bu, W. (2007) An Empirical Study on the Economical Determinants of Inward-FDI in Jiang-xi Province. *International Trade Journal*, **2**, 57-61.
- [7] Xia, F. (2007) The Spatial Difference of FDI Influencing Factors in China. *Integrated Management*, **12**, 98-99.
- [8] Zhou, G. (2008) An Empirical Study on the influencing factors of FDI-Taking Bohai

as an Example. *Journal of Hebei University of Economics and Business*, **6**, 59-63.

- [9] Xu, H. and Lai, M. (2002) Partial Least Squares Regression Analysis of Influencing Factors of Foreign Direct Investment. *Chinese Journal of Management Science*, **5**, 20-25.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jamp@scirp.org