

# Prediction of Rural Residents' Consumption Expenditure Based on Lasso and Adaptive Lasso Methods

Xiaoting Tao, Haomin Zhang

College of Science, Guilin University of Technology, Guilin, China

Email: 394644616@qq.com

**How to cite this paper:** Tao, X.T. and Zhang, H.M. (2016) Prediction of Rural Residents' Consumption Expenditure Based on Lasso and Adaptive Lasso Methods. *Open Journal of Statistics*, 6, 1166-1173. <http://dx.doi.org/10.4236/ojs.2016.66094>

**Received:** October 20, 2016

**Accepted:** December 23, 2016

**Published:** December 27, 2016

Copyright © 2016 by authors 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

When the variable of model is large, the Lasso method and the Adaptive Lasso method can effectively select variables. This paper prediction the rural residents' consumption expenditure in China, based on respectively using the Lasso method and the Adaptive Lasso method. The results showed that both can effectively and accurately choose the appropriate variable, but the Adaptive Lasso method is better than the Lasso method in prediction accuracy and prediction error. It shows that in variable selection and parameter estimation, Adaptive Lasso method is better than the Lasso method.

## Keywords

Lasso Method, Adaptive Lasso Method, Consumption, Prediction

---

## 1. Introduction

Consumption, investment and export have always been referred to as the "troika" of economic growth. In China, investment and export are the main power of economic growth for a long time. But compared with the investment and the export, the level of residents' consumption in our country is very low, especially the rural residents' consumption which depressed for a long time. So, the government also stressed several times that we need to focus on expanding domestic demand, especially consumer demand. Therefore, discovering the main factors influencing the rural residents' consumption in China has a very important practical significance.

There are many empirical studies of literature about consumer spending in China. For example, by using panel data of 31 Chinese provinces from 2000 to 2011 through regression analysis, Zhao [1] found that the main factors influenced the rural residents'

consumption level is the net income of rural residents. From the perspective of total and category expenditure, Liu and Wu [2] analyzed the local government expenditure on livelihood and rural consumption from the macro perspective, and made an empirical test by using the panel data of 31 Chinese provinces from 1998 to 2011. Chu and Yan [3] established a variable intercept model of the fixed effect of local government expenditure on rural residents' consumption demand. They found that the effect of local government expenditure of supporting agriculture on rural residents' consumption demand is positive, but transfer payment is not significant. Chen and Liu [4] set up different types of rural credit model which affects the rural residents' consumption experience, found that whether short-term or long-term, consumer credit can boost rural consumption better than production. Li [5] empirically analyzed the relations between the social security expenditure and the rural consumption in China. Wen and Meng [6] used ELES model from the urban and rural residents' consumption structure comparison, rural residents' marginal propensity to consume, marginal budget share, basic consumer spending, the actual consumer spending, consumption income elasticity of demand and consumption structure changes, analyzed China's rural residents' consumption structure and its evolution in across time respectively. Hu, Tian and Xia [7] empirically analyzed of China's fiscal expenditure for supporting agriculture's impact on rural residents' consumption by using the annual time series data from 1978 to 2010. Yu and Zhang [8] created a multivariate prediction model based on the combination of Lasso method and BP neural network, and prediction of China's urban and rural resident's consumption expenditure.

In this paper, we are using R statistical software's *Lars* packages and *Msgps* packages, modeling and projections for rural residents' consumption expenditure about China by using the Lasso method and the Adaptive Lasso respectively.

## 2. Lasso Method and Adaptive Lasso Method

Supposing this paper has model data  $(X_{i1}, X_{i2}, \dots, X_{ip}; Y_i)$ , considering usually linear model:

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i, \quad (1)$$

Which  $i = 1, 2, \dots, n, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  is independent identically distributed, and  $E[\varepsilon_i] = 0$ .

Usually write model (1) as the following form:

$$Y = X\beta + \varepsilon, \quad (2)$$

Which  $Y_{n \times 1}$  as the response vector,  $X_{n \times p}$  as the independent variable matrix,  $\beta_0$  as constant,  $\beta_{p \times 1}$  as the coefficient vector, assume that the data has been standardized, remember  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ , so

Lasso method to estimate is defined as:

$$\begin{cases} (\hat{\beta}_0^{(lasso)}, \hat{\beta}^{(lasso)}) = \underset{(\beta_0, \beta)}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j)^2 \\ \sum_{j=1}^p |\beta_j| \leq s \end{cases} \quad (3)$$

here  $s \geq 0$  is the penalty parameter. The optimal solution of (3) is called the Lasso solution; the entire Lasso solution can be obtained by changing the  $s$  values, at this time, this paper uses  $k$ -fold CV and Mallows  $C_p$  criteria to choose the best model.  $k$ -fold CV is a common method of evaluation model, it roughly put all of the observation data divided into  $k$  equal parts, and then take turns to use one of the  $k - 1$  parts for the training set, used to fitting data, the remaining part is a test set, totally calculating  $k$  times, get the  $k$  index of the mean square error of fitting test set, do an average, then repeat all of the steps of the above, then select the model of the minimum average mean square error.  $C_p$  criteria is also a standard which used to assess a regression model, if select a  $p$  independent variables from the  $k$  independent variables to involved in regression, then  $C_p$  criteria is defined as:

$$\begin{cases} C_p = \frac{SSE_p}{S^2} - n + 2p \\ SSE_p = \sum_{i=1}^n (Y_i - Y_{pi})^2 \end{cases} \tag{4}$$

Therefore, this paper can choose a model with minimum  $C_p$ .

Lasso method selects variables, at the same time, it is good in estimates the unknown parameters, can solve the multicollinearity problems that exist in the model better, especially the high-dimensional data processing.

Adaptive Lasso method to estimate is defined as:

$$\begin{cases} (\hat{\beta}_0^{(alasso)}, \hat{\beta}^{(alasso)}) = argmin_{(\beta_0, \beta)} \sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j)^2 \\ \sum_{j=1}^p \omega_j |\beta_j| \leq s \end{cases} \tag{5}$$

In Type (5),  $\omega_i = \frac{1}{(\hat{\beta}_i)^\gamma}$  as the weight coefficient, and  $\gamma > 0$  as a adjustment parameters,

$\hat{\beta}_i$  is the initial estimate about the parameters,  $\hat{\beta}_i$  can use the least squares estimate, ridge estimate and Lasso estimates. The optimal solution of (5) is called Adaptive Lasso solution, all Adaptive Lasso solution can be obtained by changing the  $s$  values, at this moment, Mallows  $C_p$  criteria, AIC criteria, GCV criteria and BIC criteria can be used to choose the best model [9]. The Adaptive Lasso method using different weight coefficient, with a smaller weight punish the variable which regression coefficient is larger, with a larger weight punish the variable which regression coefficient is smaller, makes the selected variables more accurately.

Due to Lasso method use the same weight of all coefficient, and the Adaptive Lasso methods based on different variables given different weights, with a smaller weight punish the variable which regression coefficient is larger, with a larger weight punish the variable which regression coefficient is smaller, improved the Lasso method in variable selection, which cannot meet the model selection of consistency and parameter estimation lack of convergence speed to  $\sqrt{n}$ , makes selected variables more accurate. For this purpose, this article respectively building the forecasting model based on the method of Lasso and Adaptive Lasso, to forecast the rural residents' consumption expenditure in China.

### 3. The Empirical Analysis

#### 3.1. Variable Selection and Data Sources

In this paper, on the basis of the theory of economics and the research of Yu and Zhang [8], 16 variables which influenced rural residents' consumption expenditure( $y$ ) are selected. Name of the 16 variables specific in **Table 1**.

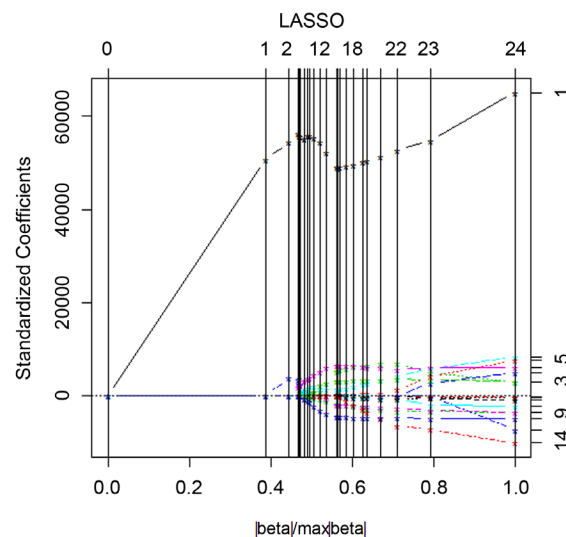
In this article, the dependency ratio data was from *Statistical Yearbook of China Population*. The interest rate data was from the website of the people's bank of China, interest rates will be subject to the one-year rate stipulated by the central bank, if there are multiple interest rate in a year, then use weighted average, the weight of the interest rate used in accounted for the proportion of 12 months. Other variable data are from 1981-2015 periods, *China Statistical Yearbook*.

#### 3.2. The Prediction of Rural Residents' Consumption Expenditure Model Based on Lasso Method

From **Figure 1**, this article just need 24 steps to get all the Lasso solution, when parameters  $s = 1$ , all variables into the model. Because of the value of  $k$ -fold CV is bigger

**Table 1.** The set of independent variables.

variable	meaning	variable	meaning
$x_1$	Residents' disposable income	$x_9$	Young dependency ratio
$x_2$	GDP growth rate	$x_{10}$	Education situation
$x_3$	Inflation	$x_{11}$	Spending on social security
$x_4$	The first industrial output value	$x_{12}$	Employment figure
$x_5$	The tertiary industry output value	$x_{13}$	Income distribution gap
$x_6$	The annual fixed asset investment	$x_{14}$	Spending habits
$x_7$	The interest rate	$x_{15}$	Highway mileage
$x_8$	population	$x_{16}$	Post and telecommunications business



**Figure 1.** The variable selection step with *Lasso* method.

than the value of  $C_p$  criteria, so this paper selects the most optimal Lasso solution according to Mallows  $C_p$  criteria. This article gets minimum  $C_p$  value when step is 20, then the model is optimal, in the end, this paper chooses 11 variables such as

$x_1, x_2, x_3, x_5, x_6, x_9, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}$ . From the result, inflation is the most important factors influencing the rural residents' consumption expenditure; And highway mileage, Residents' disposable income, the annual fixed assets investment and the tertiary industry output value have a positive impact on rural residents' consumer spending; But the GDP growth rate, young dependency ratio, employment figure, income distribution gap, consumption habits and post and telecommunications business has a negative effect on the rural residents' consumption expenditure; Other factors are not significant in affect the rural residents' consumption expenditure, are elected to the model.

LASSO model:

$$y = 0.6557x_1 - 0.131x_2 + 9.195x_3 + 0.0047x_5 + 0.4407x_6 - 0.0042x_9 - 0.0695x_{12} - 0.0785x_{13} - 0.0208x_{14} + 5.2057x_{15} - 0.0924x_{16} \tag{6}$$

On this basis, this paper predicts China rural residents' consumer spending from 2008 to 2014, forecasting results are shown in **Table 2**.

From **Table 2**, the forecasts of rural residents' consumption expenditure is widespread undervalued, but from the perspective of the relative error, the estimate result is more stable, the prediction error decreases year by year, proves the estimate effect of the model of Lasso method is good.

### 3.3. The Prediction of Rural Residents' Consumption Expenditure Model Based on Adaptive Lasso Method

Through the use of Adaptive Lasso method (see **Figure 2**), this article can predict the rural residents' consumption expenditure. Because of the value of Mallows  $C_p$  criteria, AIC criteria, GCV criteria and BIC criteria is 9.767, 10.11, 11.14, 8.263, the BIC value is minimum which can fitting the model best, so this paper chooses BIC criteria to modify the model and finally selects 13 variables such as

$x_1, x_2, x_3, x_6, x_7, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}$ . From the result, inflation is the most important factors influencing the rural residents' consumption expenditure; And highway mileage, the interest rates, residents' disposable income, expending on social security and the annual fixed assets investment are have a positive impact on rural residents'

**Table 2.** Rural residents' consumer spending predictions based on the lasso method.

year	Y (billion)	Predicted value (billion)	relative error
2008	2767.726	2531.852	0.08522
2009	2900.533	2702.854	0.06815
2010	3197.46	2993.954	0.06365
2011	3896.959	3667.178	0.05896
2012	4231.038	4023.456	0.04906
2013	4712.732	4601.321	0.02364
2014	5185.961	5089.899	0.01852

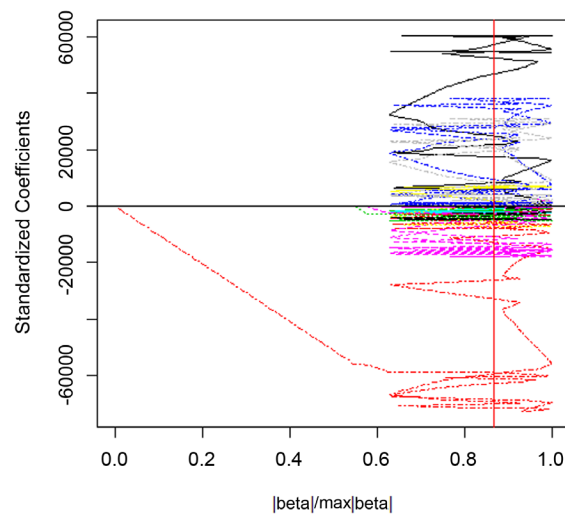
consumption expenditure; But the GDP growth rate, young dependency ratio, education, employment figure, income distribution gap, consumption habits and post and telecommunications business are have a negative effect on the rural residents' consumption expenditure; Other factors are not significant in affect the rural residents' consumption expenditure, are not elected to the model.

Adaptive LASSO model:

$$y = 0.728x_1 - 0.0145x_2 + 7.007x_3 + 0.3634x_6 + 1.064x_7 - 0.0047x_9 - 0.3195x_{10} + 0.4005x_{11} - 0.0617x_{12} - 0.0126x_{13} - 0.064x_{14} + 6.052x_{15} - 0.0839x_{16} \quad (7)$$

On this basis, this paper predicts China rural residents' consumer spending from 2008 to 2014, forecasting results are shown in **Table 3**.

From **Table 3**, the prediction of rural residents' consumption expenditure is also has the phenomenon of widespread undervalued, but from the perspective of the relative error, the estimate result is more stable, the prediction error decreases year by year, and the most important point is that the prediction result of the Adaptive Lasso method is more close to the real value than the Lasso method, proves the estimate effect of the model of Adaptive Lasso method is better than the Lasso method.



**Figure 2.** The variable selection with Adaptive Lasso method.

**Table 3.** Rural residents' consumer spending predictions based on the Adaptive Lasso method.

year	Y (billion)	Predicted value (billion)	relative error
2008	2767.726	2590.408	0.064066
2009	2900.533	2776.138	0.042887
2010	3197.46	3068.92	0.040201
2011	3896.959	3716.852	0.046217
2012	4231.038	4099.892	0.030996
2013	4712.732	4701.812	0.002317
2014	5185.961	5178.577	0.001424

## 4. Conclusions and Recommendations

Based on the Lasso method and the Adaptive Lasso method to construct predictive model of rural residents' consumption expenditure in China respectively, inflation is the most important factors influencing the rural residents' consumption expenditure; and residents' disposable income, inflation, the annual fixed asset investment and highway mileage factors are all have positive effects on the two models; and the GDP growth rate, population, employment figure, income distribution gap, consumption habits and post and telecommunications business all have a reverse effect on the two models. To this end, this article puts forward the following suggestions:

1) Control the level of inflation in China. Inflation has a great influence on rural residents' life. On the one hand, inflation makes the price of agricultural and sideline products increase, and increase the income of farmers. On the other hand, inflation makes the rural consumer goods and services prices increase and the farmers' real income down. Thus, the inflation can be controlled by using these ways, such as the tightening of monetary policy, fiscal policy and income policy, positive supply policy, currency reform and other measures to restrain inflation.

2) Raising the income level of rural residents. Raise the income level of rural residents, so that they will have more consumption. The government can implement active employment policy, provide more employment platform, give more job training for rural surplus labor about mount guard, expand the source of raising income, and provide modest fiscal policy of subsidies for agricultural production to increase the income of rural residents.

## References

- [1] Zhao, Y.D. (2013) The Influence of Local Fiscal Expenditure on Rural Residents' Consumption. *Public Finance Research*, **5**, 66-69.
- [2] Liu, Z.Z. and Wu, F. (2014) The Livelihood Process of Local Government Fiscal Expenditure and Rural Consumption—Theoretical Analysis and Empirical Test Based on Total and Category Expenditure. *The Theory and Practice of Finance and Economics*, **1**, 75-80.
- [3] Chu, D.Y. and Yan, W. (2009) Local Government Expenditure and Rural Residents' Consumption Demand: An Empirical Analysis Using Panel Data of Provinces from 1998 to 2007. *Statistical Research*, **8**, 38-44.
- [4] Chen, D. and Liu, J.D. (2013) Rural Credit Impact on Rural Residents' Consumption—Based on State Space Model and the Mediation Effect Test of the Dynamic Analysis for a Long Time. *Journal of Financial Research*, **6**, 160-172.
- [5] Li, X.J. (2013) Dynamic Effect Analysis on Rural Residential Consumption and Expenditure for Social Security. *Collected Essays on Finance and Economics*, **4**, 22-28.
- [6] Wen, T. and Meng, Z.L. (2012) Rural Residents' Consumption Structure Evolution Research in China. *Journal of Agrotechnical Economics*, **7**, 4-14.
- [7] Hu, D.L., Tian, K. and Xia, J.C. (2013) The Influence on Rural Residents' Consumption of China's Fiscal Expenditure for Supporting Agriculture—Empirical Analysis and Policy Recommendations. *Public Finance Research*, **1**, 50-53.
- [8] Yu, S.H. and Zhang, J. (2016) The Study on Prediction of Residents Consumption Expenditure based on Lasso and BP Neural Network. *The Theory and Practice of Finance and Economics*, **3**, 11-15.

*Economics*, **37**, 123-128.

- [9] Wu, X.Z. (2012) *Complex Data Statistical Method—Based on the Application of R*. China Renmin University Press, Beijing.



**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 [ojs@scirp.org](mailto:ojs@scirp.org)