

The LASSO Estimation Method for Linear EIV Model Parameters

Mingqing Zhao¹ and Tiantian Xi²

¹College of Mathematics and Systems Science, Shandong University of Science and Technology, Shandong, Qingdao, China

²College of Mathematics and Systems Science, Shandong University of Science and Technology, Shandong, Qingdao, China

Keywords: EIV model, Parameter estimation, Structural risk minimization principle, LASSO.

Abstract. With regard to the linear EIV model, for the problem the weighted total least squares (WTLS) method only considers the goodness of fit, but ignores the complexity, which reducing its generalization ability, the LASSO estimation method for linear EIV model parameters (LE) was proposed that adding an L1 norm penalty to random error matrix of observation vector and coefficient matrix. Through the empirical study of the factors affecting the percentage of China's personal health expenditure in 2001-2017, compared with WTLS and least squares (LS) methods, the LE method could significantly improve the prediction accuracy, achieve stronger generalization ability and realize variable selection to achieve dimensionality reduction.

Introduction

In practical applications, there are random errors in both interpreted and explanatory variables in many cases. Models that considering the both random errors are called Errors-in-Variables(EIV) models.

As for the problem of optimal estimation of linear EIV model parameters, many scholars at home and abroad have carried out extensive and in-depth research on it, proposed the total least squares (TLS) method and WTLS method.

Schaffrin et al. (2005) [1] proposed a TLS estimation method with constraints; Zhang et al. (2013) [2] proposed a TLS method with inequality constraints based on the exhaustive method. Schaffrin et al. (2008) [3] first proposed WTLS and its solving algorithm; Shen (2011) et al [4] proposed the Newtown-Guass iteration based WTLS algorithm; Mahboub (2012) [5] studied the WTLS algorithm with the rules in the structure matrix; Xu et al. (2012) [6] proposed a partial-EIV model derived from the EIV model, and calculate the parameters of partial-EIV model using the WTLS algorithm.

However, the above study only considers the goodness of fit, but ignores its complexity, which leads to over-fitting, thus reduces the ability of prediction. In this paper, the goodness and complexity of the model are considered at the same time. Based on the principle of structural risk minimization, the LASSO estimation method for linear EIV model parameters (LE) is proposed. Experiment shows that the LE method can significantly improve the prediction accuracy, can achieve stronger generalization ability and can realize variable selection to achieve dimensionality reduction.

Linear EIV Model and Its WTLS Estimation

Let the EIV model be the following definition:

$$y - e_y = (A - E_A)\beta \quad (1)$$

where y is the $n \times 1$ observation vector affected by the random error vector e_y , A is the $n \times m$ coefficient matrix affected by the random error matrix E_A , and β is the (unknown) $m \times 1$ parameter vector.

Let the stochastic properties of the errors be characterized by

$$\begin{bmatrix} e_y \\ e_A \end{bmatrix} = \begin{bmatrix} e_y \\ \text{vec}(E_A) \end{bmatrix} \sim N_{n+n(m+1)} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} \right) \quad (2)$$

where "vec" represents an operator that stacks a column of the matrix below the previous matrix. Q_y and Q_A are the symmetric and non-negative-definite cofactor matrices of e_y and e_A , respectively, and σ_0^2 is an (unknown) variance component.

We assume that:

$$Q_y = P_y^{-1}, Q_A = P_A^{-1} = Q_0 \otimes Q_1, Q_0 = P_0^{-1}, Q_1 = P_1^{-1} \quad (3)$$

Where Q_0 has size $m \times m$, and Q_1 has size $n \times n$. Therefore, if the cofactor matrix is reversible and the structure of the cofactor matrix of E_A can be expressed in the form of "Kronecker-Zehfuss product" (\otimes), then we allow heteroscedasticity and even related observations.

We compute the WTLS in the sense that

$$\begin{cases} \min_{\beta} e_y^T P_y e_y + e_A^T P_A e_A \\ \text{s.t. } y - e_y - (A - E_A)\beta = 0 \end{cases} \quad (4)$$

The analytical solution of the optimization problem cannot be obtained. The iterative algorithm for its numerical solution is given in [3].

LASSO Estimation of Linear EIV Model Parameters

Based on the principle of structural risk minimization, considering the goodness of fit and complexity of the model, and referring to the LASSO regression idea, we add an L1 norm penalty to the sum of the weighted residuals of all data, called the LE method, the model is as follows:

$$\begin{cases} \min_{\beta} e_y^T P_y e_y + e_A^T P_A e_A + \mu \|\beta\|_1 \\ \text{s.t. } y - e_y - (A - E_A)\beta = 0 \end{cases} \quad (5)$$

When $\mu = 0$, Eq. 5 can be regarded as Eq. 4. The optimal solution for this optimization problem can be analyzed by the Lagrange multiplier method.

We form the target function using the traditional Lagrange approach:

$$\begin{aligned} \Phi(e_y, e_A, \lambda, \beta) &= e_y^T P_y e_y + e_A^T P_A e_A + \mu \|\beta\|_1 + 2\lambda^T [y - A\beta - e_y + E_A\beta] \\ &= e_y^T P_y e_y + e_A^T P_A e_A + \mu \|\beta\|_1 + 2\lambda^T [y - A\beta - e_y + (\beta^T \otimes I_n)e_A] \end{aligned} \quad (6)$$

where the $n \times 1$ vector λ is "Lagrange multipliers". Deriving each parameter and making it equal to zero, namely

$$\frac{\partial \Phi}{\partial e_y} = 2(P_y e_y - \lambda) = 0 \quad (7)$$

$$\frac{\partial \Phi}{\partial e_A} = 2(P_A e_A + (\beta \otimes I_n)\lambda) = 0 \quad (8)$$

$$\frac{\partial \Phi}{\partial \lambda} = 2(y - A\beta - e_y + (\beta^T \otimes I_n)e_A) = 0 \quad (9)$$

$$\frac{\partial \Phi}{\partial \beta} = -2(A^T \lambda - E_A^T \lambda) + \mu r = 0 \quad (10)$$

Where

$$r = (r_0, r_1, \dots, r_j, \dots, r_m)^T$$

$$r_j = \begin{cases} 1, & \beta_j > 0 \\ [-1, 1], & \beta_j = 0 \\ -1, & \beta_j < 0 \end{cases} \quad j = 0, 1, \dots, m \quad (11)$$

From Eq. 7 and Eq. 8, the residual vectors can be computed as follows:

$$e_y = Q_y \lambda \quad (12)$$

$$e_A = -(Q_0 \otimes Q_1)(\beta \otimes I_n)\lambda = -(Q_0 \beta \otimes Q_1)\lambda \quad (13)$$

According to [3], we can obtain the fact that

$$\text{vec}(HFG^T) = (G \otimes H)\text{vec}(F) \quad (14)$$

the error matrix can be retrieved from Eq. 13:

$$E_A = -Q_1 \lambda (Q_0 \beta)^T = -Q_1 \lambda \beta^T Q_0 \quad (15)$$

Inserting Eq. 12, Eq. 13 into Eq. 9 leads to

$$y - A\beta = e_y - (\beta^T \otimes I_n)e_A = Q_y \lambda + (\beta^T Q_0 \beta \otimes Q_1)\lambda \quad (16)$$

and thus to

$$\lambda = (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1}(y - A\beta) \quad (17)$$

inserting Eq. 17 into Eq. 12 and Eq. 15, we obtain

$$e_y = Q_y (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1}(y - A\beta) \quad (18)$$

$$E_A = -Q_1 (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1}(y - A\beta)\beta^T Q_0 \quad (19)$$

exploiting Eq. 10 in conjunction with Eq. 17 and Eq. 19:

$$\left. \begin{aligned} -2E_A^T \lambda &= -2A^T \lambda + \mu r \\ &= 2A^T (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1}(A\beta - y) + \mu r \\ &= 2Q_0 \beta v \end{aligned} \right\} \quad (20)$$

with the scalar

$$v = (y - A\beta)^T (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1} Q_1 (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1} (y - A\beta) \quad (21)$$

β can be derived by using Eq. 20 and Eq. 21:

$$\beta = (2A^T (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1} A - 2vQ_0)^{-1} (2A^T (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1} y - \mu r) \quad (22)$$

Eq. 22 is the conditional equation that is satisfied to the optimal solution of the Eq. 5.

Let

$$U = 2A^T (Q_y + (\beta^T Q_0 \beta)Q_1)^{-1} A - 2vQ_0 \quad (23)$$

$$V = 2A^T (Q_y + (\beta^T Q_0 \beta)Q_1) \quad (24)$$

Then Eq. 22 can be written as

$$U\beta = V^{-1}y - \mu r \quad (25)$$

that is

$$VU\beta = y - \mu Vr \quad (26)$$

Considering the j -th component on both sides of Eq. 26

$$(VU\beta)_j = y_j - \mu(Vr)_j \quad j = 0, 1, \dots, m \quad (27)$$

Actually, we can not get the analytical solution of the Eq. 5 by Eq. 27. Therefore, Inspired by the SHOOTING algorithm in [7], we give an iterative algorithm to calculate the numerical solution of the optimization problem. In Eq. 27, fix all parameters but β_j , then the function image on the left is the curve① in Fig. 1, the function image on the right is the polyline② in Fig. 1. (Note: both the left and right sides of the ordinate axis are parabola, the middle is the straight line segment on the ordinate axis).

Let $\beta_j = 0$, we can get the intercept of ①, called *intercept*; On the right side of Eq. 27, let $\beta_j = 0$, $r_j = -1$, and get the ordinate of A, called y_A , let $\beta_j = 0$, $r_j = 1$, and get the ordinate of B, called y_B . The intersection of curves ① and ② is β_j^* , if $\text{intercept} > y_A$, then $\beta_j^* < 0$; if $y_B \leq \text{intercept} \leq y_A$, then $\beta_j^* = 0$; if $\text{intercept} < y_B$, then $\beta_j^* > 0$.

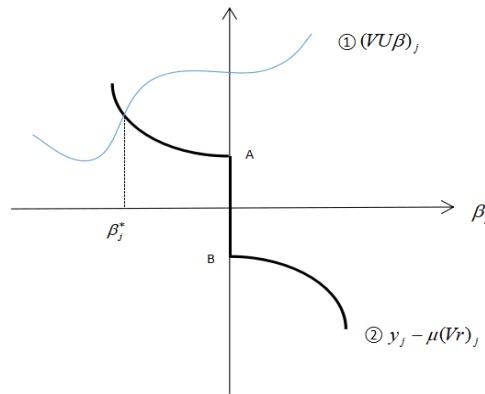


Figure 1. Schematic diagram of algorithm

From the analysis above, we can give the solution algorithm for the numerical solution of the LE method as follows:

-
- (1) $v^{(0)} = 0$, $\beta^{(0)} = N^{-1}c$, $[N, c] = A^T P_y [A, y]$, $i = 0$.
 - (2) calculate $\beta^{(i+1)}$.
 - ① $j = 0$;
 - ② Let $\beta = (\beta_0^{(i)}, \dots, \beta_{j-1}^{(i)}, \beta_j^{(i+1)}, \beta_{j+1}^{(i)}, \dots, \beta_m^{(i)})$, calculate *intercept*, y_A , y_B ;
 - ③ if $\text{intercept} > y_A$, then $r_j = -1$, go to ⑥;
 - ④ if $\text{intercept} < y_B$, then $r_j = 1$, go to ⑥;
 - ⑤ $\beta_j^{(i+1)} = 0$, go to ⑦;
 - ⑥ insert r_j into Eq. 27, calculate $\beta_j^{(i+1)}$ by Newton's method;
 - ⑦ $j = j + 1$;
 - ⑧ if $j \leq m$, go to ②;
 - (3) if $\|\beta^{(i+1)} - \beta^{(i)}\| < \delta$, go to (5).
 - (4) $i = i + 1$, go to (2).
 - (5) end.
-

In practical applications, the weights in the algorithm are usually given as a special form [3].

Empirical Analysis

This paper uses the LE method to analyze the factors affecting the percentage of China's personal health expenditure in 2001-2017, and compared with WTLS and LS methods, for which verifying

the effectiveness of the method. The Python language is used for all data processing in this paper.

Construction of Indicator System

There are many factors affecting the percentage of China's personal health expenditure. The following 12 indicators are selected from several different observation points: total health expenditure, population, policy, economy, medical technology level and medical service supply income, which are: the percentage of government health expenditure in total health expenditure (X_1), the percentage of social health expenditure in total health expenditure (X_2), 65 years and older accounted for the total population (X_3), urban population ratio (X_4), unemployment rate (X_5), the percentage of total health expenditure in GDP (X_6), GDP per capita (X_7), infant mortality rate (X_8), number of beds per thousand population (X_9), number of doctors per thousand population (X_{10}), the per capita net income of urban residents (X_{11}) and the per capita net income of rural residents (X_{12}).

The Dataset

We use the dataset about the percentage of China's personal health expenditure from 2001-2017. The dataset are derived from the China Statistical Yearbook and the China Health Statistics Yearbook. we use the data from 2001-2014 as traing data, and the data from 2015-2017 as test data.

LE Model Establishment

For the convenience of calculation, this paper has simplified the weight matrix as follows:

$$P_0 = \text{diag}(0,1,1,1,1,1,1,1,1,1,1,1), P_1 = I_{14}, P_y = I_{14}$$

The penalty parameters are selected by the cross-validation method [8]. The regression equation is

$$\hat{y} = 105.2622 - 0.9736X_1 - 1.0029X_2 - 0.1230X_4 + 0.00002X_7 - 0.0362X_8 \quad (28)$$

It can be seen that the coefficients of 7 variables have been compressed to 0, the impact is completely ignored, leaving only 5 variables, respectively, the percentage of government health expenditure to total health expenditure, the percentage of social health expenditure in total health expenditure, urban population ratio, per capita GDP, and infant mortality.

(1) The percentage of government health expenditure to total health expenditure has a negative impact on the percentage of personal health expenditure.

(2) The percentage of social health expenditure to total health expenditure has a negative impact on the percentage of personal health expenditure.

(3) Urban population ratio has a negative impact on the percentage of personal health expenditure.

(4) Per capita GDP has a positive impact on the percentage of personal health expenditure.

(5) Infant mortality has a negative impact on the percentage of personal health expenditure.

Comparative Analysis

The parameters of the regression model estimated by WTLS and LS respectively are shown in the following regression equations:

$$\begin{aligned} \hat{y}_{WTLS} = & 102.7128 - 0.8678X_1 - 1.0431X_2 + 1.5981X_3 - 0.2734X_4 - 0.4924X_5 - 0.0860X_6 - \\ & 0.0003X_7 - 0.0091X_8 - 2.3275X_9 + 0.5043X_{10} + 0.0007X_{11} - 0.0001X_{12} \end{aligned} \quad (29)$$

$$\begin{aligned} \hat{y}_{LS} = & 102.4310 - 0.8713X_1 - 1.0426X_2 + 1.5489X_3 - 0.2614X_4 - 0.4665X_5 - 0.0914X_6 - \\ & 0.0003X_7 - 0.0072X_8 - 2.3275X_9 + 0.5192X_{10} + 0.0007X_{11} + 0.0001X_{12} \end{aligned} \quad (30)$$

Table 1. Comparison of predicted and actual value of the percentage of personal health expenditure

Year	Actual value	LE		WTLS		LS	
		predicted value	residual	predicted value	residual	predicted value	residual
2015	29.27	28.99	0.28	29.29	0.02	30.22	0.95
2016	28.78	28.86	0.08	29.48	0.70	30.47	0.69
2017	28.77	28.81	0.04	29.47	0.70	30.52	0.75

We use the regression equations obtained by the three estimation methods above to predict the data in the test dataset, and compared with the actual data. The results are shown in Table 1.

we compare the LE method proposed in this paper with the WTLS and LS methods as follows:

(1) From an economic point of view, all the signs of the coefficients generated by the LE method are in line with reality. As for WTLS and LS methods, the partial coefficients' sign is not realistic.

(2) From the perspective of prediction accuracy, the LE method has higher prediction accuracy. From the definition of standard error $\delta = \sqrt{\sum_{i=1}^n v_i^2 / n}$, the errors of LE, WTLS, and LS with respect to

the actual value are 0.1697, 0.5717, and 0.8066, respectively. Compared with the WTLS and LS methods, the accuracy of the LE method is higher, which is increased by 70.3%, 79.0%.

(3) From the goodness of fit, the LE method has a better goodness of fit. The determination coefficients R^2 of each method are as follows: 0.9904 for WTLS, 0.9605 for LS, and 0.9991 for LE method, from which the LE method is the highest, the fitting effect is better.

Summary

This paper proposes the LE method, which is based on the principle of structural risk minimization. The empirical analysis of the factors affecting the percentage of health expenditure in China illustrates the superiority of the method. The results showed that the LE method can achieve the purpose of dimensionality reduction, can significantly improve the prediction accuracy, has a stronger generalization ability, and can achieve a higher degree of goodness of fit. However, the selection of penalty parameters, the statistical properties of parameter estimation and the convergence of algorithms need to be further studied.

Acknowledgement

This research was financially supported by the Shandong University of Science and Technology Graduate Tutor Guidance Ability Improvement Program Project(KDYC17018).

References

- [1] Schaffrin B, Felus Y A. On Total Least-Squares Adjustment with Constraints[M]. A Window on the Future of Geodesy.Berlin:Springer Berlin Heidelberg, 2005:417-421.
- [2] Zhang S,Zhang K.A solution to EIV model with inequality constraints and its geodetic applications[J].Journal of Geodesy,2013,87(1):23-28.
- [3] Schaffrin B, Wieser A. On weighted total least-squares adjustment for linear regression[J]. Journal of Geodesy,2008,82(7):415-421.
- [4] Shen Y,Li B,Chen Y.An iterative solution of weighted total least-squares adjustment[J].Journal of Geodesy,2011,85(4):229-238.
- [5] Mahboub V.On weighted total least-squares for geodetic transformations[J].Journal of Geodesy, 2012,86(5):359-367.

- [6] Xu P, Liu J, Shi C. Total least squares adjustment in partial errors-in-variables models: algorithm and statistical analysis[J]. *Journal of Geodesy*, 2012, 86(8):661-675.
- [7] Fu W J. Penalized Regressions: The Bridge versus the Lasso[J]. *Journal of Computational and Graphical Statistics*, 1998, 7(3):397-416.
- [8] Gandek B , Ware J E , Aaronson N K , et al. Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries[J]. *Journal of Clinical Epidemiology*, 1998, 51(11):1171-1178.