https://doi.org/10.31272/jae.i149.1389
https://admics.uomustansiriyah.edu.iq/index.php/admeco
P-ISSN: 1813–6729 E-ISSN: 2707-1359

The Employment of Lasso and Rlasso's Method to Check the Stability of the Multiple Linear Regression Model

Haitham Hasson Majed
Dept of Mathematics, Open Educational College, Al-Najaf Study Centre, Iraq
Email: haithamnkhozaee@gmail.com, ORCID ID:\ https://orcid.org/0009-0000-8286-1960
Jasim Hassan Lazem
Dept of Tourism, Technical College of Management, The Middle Technical University, Baghdad, Iraq.
Email: jhlazem@mtu.edu.iq, ORCID ID:\ https://orcid.org/0000-0001-9667-1009

Article Information

Article History:
Received: 17 / 04 / 2025
Accepted: 26 / 05 / 2025
Available Online: 01 / 09 / 2025
Page no: 68 – 77

Keywords:
Lasso , Rlasso , Chow Test, CUSUM Test, Multicollinearity problem

Correspondence:
Researcher name:
Haitham Hassoon Majed
Email:
haithamnkhozaee@gmail.com

Abstract

The instability of regression models is one of the most prominent issues facing researchers when estimating the model. On the other hand, regularisation techniques (Lasso and Rlasso) are some of the most notable techniques applied in the estimation of multiple regression models because of their ability to reach an explanatory regression model through which accurate predictions can be made. In this research, the simulation method was applied to conduct a comparison between a set of regression models estimated using the two regularisation techniques (Lasso & Ro), where the Chow Test and CUSUM Test were used to test the stability of the estimated models and the presence of multicollinearity. The results showed that both techniques performed excellently in building a stable regression model. The results showed the superiority of the Rlasso technique, especially in the presence of multicollinearity.

1. Introduction

The multiple linear regression model is one of the essential statistical models used to study and analyse the linear relationship between the response variables and a set of explanatory variables. The conditions for the fulfilment of the analysis assumptions, the most important of which is that the random errors are independently distributed in a normal distribution with a zero mean and a constant variance and that there is no strong linear relationship between the explanatory variables. The regression model is one of the most important and widely applied statistical tools in the field of scientific research.

The multiple regression model can be expressed as [9]

$$Y = X\beta + U \quad (1)$$

Whereas:

Y: represents a vector (n×1) of observations of the response variable.

X: is a matrix (n×(k+1)) of observations of the explanatory variables (the first column represents the constant term).

$\boldsymbol{\beta}$: is a vector (k+1) ×1) represents the parameters of the regression model, noting that the constant term of the model is the first element in this vector.

U: Represents a vector (n×1) of random errors.

As is known, the process of building a multiple regression model aims to use explanatory variables (independent) to predict the value of the response variable (dependent). Therefore, the process of selecting explanatory variables that have a tangible impact on the response variable is considered one of the essential steps in building a regression model due to its importance in improving the accuracy of prediction and reaching an explanatory model. Choosing essential and modern methods or techniques in estimating model parameters and selecting explanatory variables should be one of the main steps in building a regression model. On the other hand, testing the stability of the regression model is considered one of the most critical steps that should be carried out before applying the model in analysing the data for the study. Despite the importance of this step, we find that in many studies, the researcher's interest is focused on choosing the appropriate methods for estimating the model without paying attention to the importance of ensuring the stability of the regression model.

Many methods have been applied in selecting the essential variables for the multiple regression model, and perhaps the first of these methods was in 1974, when the researcher Akaike [1], and in 1978, when the researcher Schwarz [16] selected the variables by comparing a group of candidate models.

Recently, new methods have been introduced for selecting variables in the regression model, called regularisation techniques. These techniques can estimate the parameters of the regression model and choose the significant explanatory variables in the regression model, even in the case of high-dimensional or multicollinearity problems. Perhaps the most famous of these methods is the Lasso technique. (Least absolute shrinkage and selection operator) This technique was produced in 1996 by researcher Tibshirani [19]. This method is essential in estimating the model parameters and choosing critical explanatory variables, in addition to improving the accuracy of prediction in the regression model; it can also deal with the problem of multicollinearity in the model [19]. After the Lasso technique, developments in the field of selecting variables for the multiple regression model included the elastic net technique, which was introduced by Zou and Hastie in 2005. This technique combines the Ridge Regression technique and the Lasso technique mentioned above [22]. The group Lasso technique was introduced by Yuan and Lin in 2006. This has great importance in many applications in machine learning and cancer prediction by removing the unimportant data set and choosing the important set of variables in the model [21].

The Rlasso technique introduced in 2015 by Song, Q., & Liang, F., is an improvement on the Lasso technique, as it uses a decreasing penalty for the model parameters instead of the increasing penalty applied in the Lasso technique [18].

On the other hand, the accuracy of the regression model and its ability to analyse the phenomenon and make accurate predictions are based primarily on assumptions, the most important of which is the stability of the model parameters from one sample to another. When the values of the estimated parameters of the regression model differ significantly from one sample to another in the phenomenon data, this is an indication of the instability of the model and that the results of the analysis of this model will be unreliable and cannot be generalised. The predictions resulting from this model will also be unreliable. Therefore, testing the stability of the model is considered one of the important pillars applied before using the model in analysing the phenomenon.

Testing the stability of the regression model is one of the important statistical topics that has been recently applied in many studies and research. Reaching a stable regression model means that the model can predict consistently in different samples, whether for the same period or different periods. This means that the parameters of the regression model are consistent for different data [8]. Many tests have been presented in the field of stability testing, for example: The CUSUM test, developed by E.S.Page at Cambridge University in the early fifties of the last century, is an important statistical tool for testing statistical models, especially in the field of economic studies whose data are in the form of a time series, in which cumulative values are calculated over time based on the difference between the actual and estimated values of observations [14].

The Chow test in 1960, presented by the scientist Gregory Chow, is based on the analysis of the use of two different samples of the phenomenon's data to demonstrate the stability of the model's parameter estimates [6]. If the parameter estimates differ in these two samples relatively significantly, then the regression model is considered unstable. This test is characterised by the possibility of applying it to cross-sectional data and time series data. Also, the Hausman test, presented by Jerry A. Hausman In 1978, based on the comparison of two types of models (constrained and unconstrained), A. Wald also presented the Wald test in 1983, which uses the comparison of multiple constraints on the parameters of the regression model[20], and in 1986, Ronald F. Engle presented the Lagrange Multiplier test, which compares a constrained and unconstrained model with the application of non-linear constraints in the test. With the development of computer technology and software, many tests have been developed in recent years for the stability of the regression model. In 2014, researcher Mills, T. C., presented a study on testing the stability of a linear regression model estimated using the least squares method [14]. In 2022, Liao, J., Wan, A. T. K., & He, S. presented a study on choosing the optimal estimation for a multiple linear regression model based on the optimal model mean [13].

In section two (Lasso and Rlasso), the importance of estimating the parameters of the regression model was discussed. The Chow test and the CUSUM test were also discussed, as well as their importance in conducting a stability test for the regression model.

In the third section, the simulation study was used to test the stability of two groups of regression models that were estimated using (Lasso and Rlasso), where the test was repeated (1000Times).

2. The aim of the research:

The research aims to reveal the importance of the Lasso & Rlasso techniques in estimating the regression model and to show the extent of their impact in maintaining the stability of the model in the presence of the multicollinearity problem.

3. Theoretical aspect

3.1 Lasso (Least Absolute Shrinkage and Selection Operator) technique: in estimating a multiple regression model

It is known that the basic method applied in estimating the parameters of the multiple linear regression model is the ordinary least squares (OLS) method, which is based on finding the values that make the sum of the squares of the deviations of the actual values of the response variable from its expected values as small as possible, as follows [4]:

$$\min_{\beta}(y-\hat{y})'(y-\hat{y}) = \min_{\beta}(y-X\beta)'(y-X\beta) \quad (2)$$

This method, as proven by many studies, is inappropriate in many cases, including in the presence of multicollinearity, dimensional issues, or overfitting. Also, the model estimated by this method is often poor in the prediction process, especially when the number of explanatory variables is larger than the sample size (p>n), and the OLS estimators for multiple regression model parameters in the presence of the aforementioned issues often have a significant variance and are also unstable [2] In order to develop the process of estimating and selecting the parameters of the multiple regression model and overcoming the issues facing the estimation process, many techniques have been presented, perhaps the most famous of which is the Lasso technique or(Least Absolute Shrinkage and Selection Operator), which is one of the important regularization techniques introduced by the scientist (Tibshirani) in (1996) As a development of the ridge regression technique in the estimation of the multiple regression model and as an important tool in dealing with the issue of multicollinearity and dimensionality in the regression model, this technique works to make the insignificant regression coefficients equal to exactly zero, so it has the ability to improve the accuracy of prediction and access to an explanatory model[19] .

The Lasso technique or L1-norm, reduces the parameters of the regression model and makes the unimportant parameters equal to zero completely through the restriction imposed in this method, which distinguishes this method by the possibility of reaching an explanatory model better than the ridge regression method (L2-norm), as the latter method, despite reducing the values of the unimportant regression parameters, does not make them equal to zero completely.

The parameters of a multiple linear regression model can be estimated by the Lasso technique as follows (Tibshirani 1996):

$$\hat{\beta} = (y - X\beta)^{\prime}(y - X\beta) + \tau \sum_{k=1}^{p} |\beta| \tag{3}$$

Where: τ is the regularisation parameter

Determining the value of the regularisation parameter τ is of great importance in estimating the model parameters. When this value is equal to zero, the estimation process becomes free of restrictions and is equivalent to the OLS estimation process. Also, its considerable value makes the model parameters approach zero more, which leads to the loss of independent variables that are important in the model, according to the researchers Alkenani and Yu. In 2013, it was mentioned that large values of this parameter lead to a greater reduction in the selection of explanatory variables in the model [3].

Therefore, the cross-validation technique is used to choose the best value in the estimation process.

Although the Lasso technique is of great importance in selecting variables, improving the accuracy of prediction, overcoming the problem of dimensions, and other important characteristics previously mentioned, the experimental aspect showed some disadvantages of this method, including that this technique neglects some important explanatory variables in the model, especially when the number of explanatory variables is greater than the sample size. This technique selects only one variable from a group of explanatory variables that are highly correlated with each other, some of which may be of great importance in the prediction process for this model (Zou & Hastie, 2005).

3.2. Rlasso (reciprocal Lasso) technique in estimating the multiple regression model:

The Rlasso technique is considered one of the modern and effective techniques in estimating the multiple regression model, which was presented by the researchers (Song, Q. & Liang, F.) in 2015 [18]. This technique is characterised by its ability to select critical explanatory variables in the model, maintain the stability of the model parameters, and improve the accuracy of prediction, especially when there is a problem with high dimensions [17]. The concept of this technique is based on modifying the loss function in a way that allows important parameters with small values to be preserved without eliminating them from the model.

This technique uses the inverse of the Lasso technique, where the loss function is expressed as the sum of the squares of the residuals plus the inverse of the Lasso technique, as follows [10]:

$$\hat{\beta} = (y - X\beta)^{\prime}(y - X\beta) + \tau \sum_{k=1}^{p} \frac{1}{|\beta|} \quad \beta \neq 0 \tag{4}$$

3.3 Stability of Multiple Regression Model

The process of building a regression model involves three important steps [6].

The first is to determine the explanatory variables and dependent variables that enter into the model.

The second is estimating the model parameters using appropriate statistical methods.

The third is testing the stability of the model or testing the stability of the model's parameter estimates across different sets of data.

The topic of regression model stability is considered one of the important topics of statistics, as it is closely related to the accuracy of the results of statistical analysis of different phenomena [8]. For the estimated model to be appropriate for the data of the problem under study, the estimation of the model parameters must be constant for the different groups of data specific to the study community [11]. Sometimes the researcher may find that the value of the estimated parameters of the model differs from one sample to another, which is an indicator of the instability of the model, which is reflected in the reliability of the results obtained from the model. Therefore, a model stability test must be conducted to ensure the development of an explanatory regression model that can be used to find accurate predictions about the phenomenon under study [12].

To accurately test the stability of the model, it must first be taken into consideration that the sample is representative of the study community and sufficient size to estimate the model parameters and conduct the hypothesis test correctly. Sometimes, it may be necessary to draw successive random samples to ensure that the sample represents the community correctly. It is worth noting that there are many statistical techniques used in testing the stability of the model that differ from each other according to the methods of analysis and hypothesis testing, including cross-validation [15]. (Bootstrap,

3.4. Chow Test for the stability of the multiple regression model:

The Chow Test, or what is known as the Chow Breakpoint Test, is one of the important and widely applied statistical tests in demonstrating the stability or instability of the regression model parameters across different periods or for different sets of data. The Chow Test is a statistical tool to determine whether the relationship between the dependent variable and the explanatory variables changes with time or across different samples of the phenomenon's community data.

This test is based on the fact that a stable model corresponds to a small value of the test statistic [6]. Chow. The considerable value of this statistic is an indication of the instability of the model or that the estimates of the model parameters differ from one sample to another or from one time period to another. The Chow Test is a widely used tool in many fields such as economics, sociology, and others.

Chow Test application mechanism: Chow, GC 1960,[6] showed that the test steps can be carried out by estimating three models:

The first model: is the joint model, and the parameters are estimated using all the data of the phenomenon.

The second model: In it, the parameters are estimated from the first group or sample of the phenomenon data.

The third model: In it, the parameters are estimated from the second group or sample of the phenomenon data.

Note that the sample in the second and third models is divided according to a breakpoint in the study data that is believed to have an impact on the relationship between the dependent variable and the independent variables. This breakpoint may represent a specific breakpoint after which there was a change in security stability, legal amendments, a change in economic policy, or other influencing factors. Furthermore, the size of each sample should preferably be large enough to ensure the accuracy of the estimation results. It is also preferable for the sample size to be approximately equal in both groups to reduce bias in estimating the parameters of each model [6].

The second step is to calculate the sum of squared errors (SSE) for each model.

The test hypothesis can be written as

H 0: There is no significant difference between the model parameters.

H1: There is a significant difference between the model parameters.

The third step is to calculate the test statistic and my agencies. (Chow, GC 1960) [6]:

$$F = \frac{(SSE_C - SSE_1 - SSE_2)/k}{(SSE_1 + SSE_2)/(N - 2k)} \tag{5}$$

Whereas: F: represents the Chow statistic (F distribution)، SS C: represents the sum of squared errors of the joint model., SS 1: represents the sum of squared errors of the model for the first sample.، SS 2: represents the sum of squares of the error of the second sample.، K: represents the total number of parameters in the model (or number of explanatory variables + 1)، N: represents the total number of observations (total number of observations of the first sample + total number of observations of the second sample)

Decision: We reject the null hypothesis when the value of the F-statistic is greater than the table value $F_{(k, N-k, \alpha)}$, or when the p-value is less than the value of α, and we conclude that the model is unstable across different sets (different samples) of data or different periods.

3.5. CUSUM Test:

The CUSUM test, also known as the cumulative or increasing summation test, is one of the important statistical tests applied in testing the stability of the coefficient model across different periods. This test is characterised by its ease of use and its excellent performance in detecting the stability of the model coefficients.

This test was first applied in the 1950s. It has many applications in the fields of economics, engineering, and others.

The researchers Box & GEP in 2015] explained that this test is based on its analysis of the accumulation of differences between the expected values and the observed values of the data [5]. If these differences are significant and have a specific direction, then this is evidence of the instability of the parameters. However, if these differences are minor and change randomly (have no specific direction), then this is evidence of the stability of the parameters.

Statistical hypothesis and CUSUM test statistic:

H0: There is no significant difference between model parameters

H1: There is a significant difference between the model parameters.

The test statistic can be expressed as follows [8]:

$$CU_t = \max\{0, CU_{t-1} + (y_t - \hat{y}_t)\} \tag{6}$$

Where:

CU_t: is the test statistic at time t, CU_{t-1}: is the test statistic at time t-1, y_t: Observed value of the dependent variable at time t, y_{t-1}: Observed value of the dependent variable at time t-1

Decision: We do not reject the null hypothesis and conclude that the model parameters are stable when the value of the test statistic is within certain limits (determined by the researcher and according to the accuracy of the test).

We reject the null hypothesis and conclude that the model parameters are unstable when the test statistic is outside these limits.

4. Simulation Studies

In this section, the simulation method was used to compare the stability of multiple regression models estimated by using the Lasso (least absolute shrinkage and selection operator) technique with the stability of regression models estimated by using the Rlasso (reciprocal Lasso) technique [23]. the programs were written using the R language by the researcher [7].

The work steps can be summarised as follows:

1. The simulation method was used to generate data for the multiple regression model referred to in equation (1).

$$U \sim N(0, \sigma_u^2 I_n),\ X \sim N(0, \Sigma) \tag{7}$$

$\mathbf{I_n}$ is the identity matrix, Σ representing the Autocorrelation matrix assuming a strong correlation between the explanatory variables, such that

$$\Sigma_{ij} = 0.6^{|i-j|} \quad for\ all\ 1 \le i \le j \le p \tag{8}$$

2. We considered three cases for **β**

Simulation 1: **β**= (1,1,0.5,0.2,0.3,0,5,6,7,0,1,1,1,2,2)

Simulation 2: **β**= (1,2.4,0,0,0,1,0.3,0.002,11,12,0,0,1,3,0)

Simulation 3: **β= (2**,2,1,1,0,0,1,1,0.04,0.001,1,1,2,3,2**)**

We chose different size of sample (100, 50, 25)

The parameter τ was estimated using (cross-validation), and it was $\tau = 0.02$ for the Lasso technique and $\tau = 0.01$ for Rlasso

The multiple regression models were estimated using the Lasso method and the Rlasso method. Then the Chow Test and the CUSUM Test were used to compare the stability of the estimated model parameters in both methods.

3. The previous steps were repeated 1000 times.

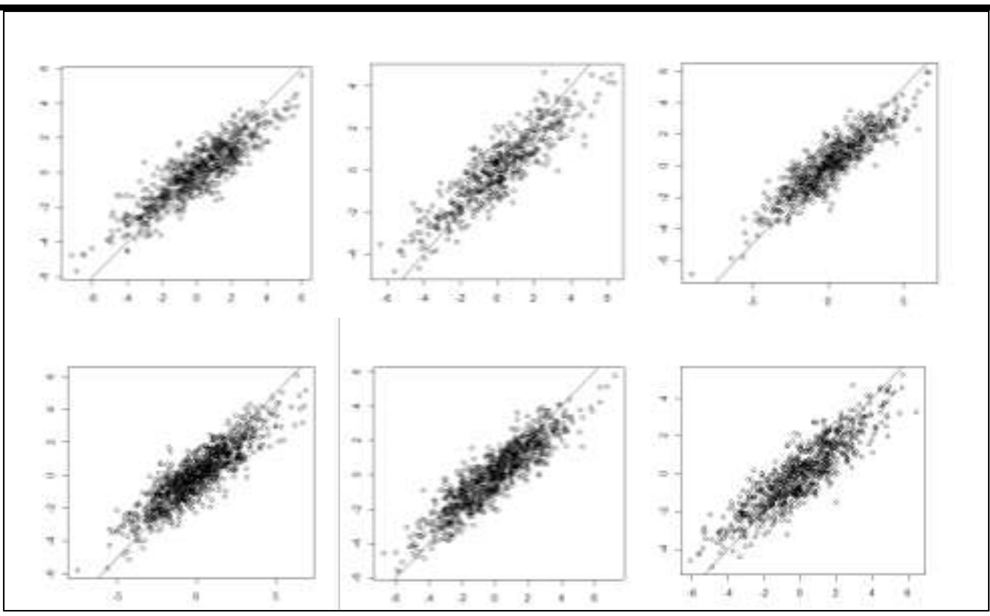

Figure (1): represents the actual values and estimated values for a set of models estimated using the Lasso technique and in the presence of the multicollinearity problem and with different sample sizes

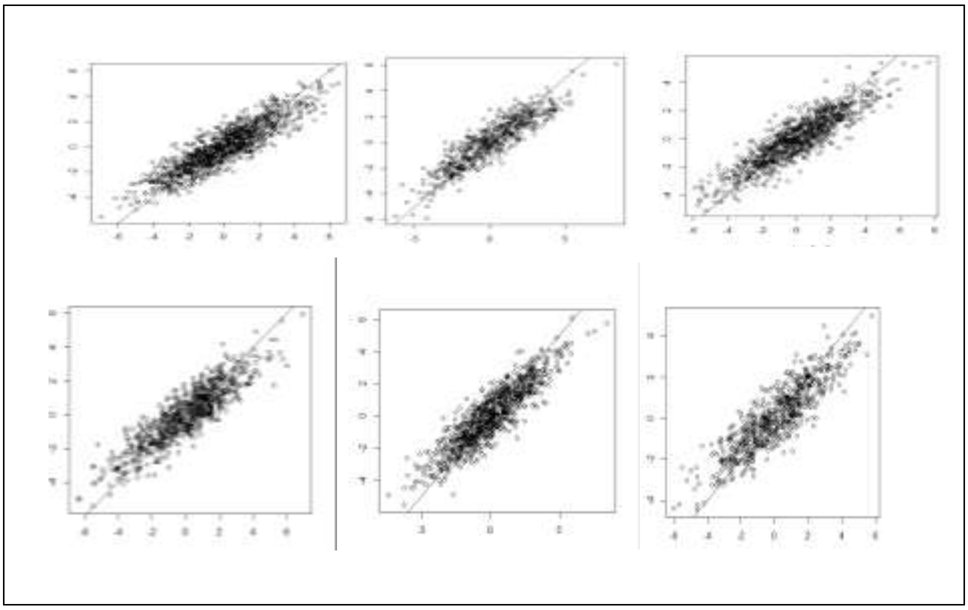

Figure (2) represents the actual values and estimated values for a set of models estimated using the Rlasso technique and in the presence of the multicollinearity problem, and with different sample sizes.

Note that the two figures (1,2) reflect the precision of the models' estimation.

Table (1): The results of the stability test for the regression models are shown below in ascending order

Results of the stability test for the models estimated using the Lasso technique					**Results of the stability test for the models estimated using the Rlasso technique**				
Index	**MSE**	**BIC**	**CUSUM(p) value**	**CHOW(p) value**	Index	**MSE**	**BIC**	**CUSUM(P) value**	**CHOW(P) value**
1	19.146	5.11	0.0042	0.0004	1	25.207	5.33	0.0102	0.0049
2	23.469	5.33	0.0122	0.0008	2	25.396	5.4	0.0299	0.0057
3	24.422	5.4	0.013	0.0058	3	25.991	5.5	0.0345	0.0071
4	24.907	5.36	0.0138	0.0108	4	26.133	5.45	0.0411	0.009
5	25.269	5.5	0.0155	0.0148	5	26.383	5.6	0.0426	0.0273
6	25.698	5.6	0.0346	0.0394	6	26.493	5.7	0.0551	0.0273
7	25.984	5.7	0.0453	0.0408	7	26.637	5.9	0.0578	0.0434
8	26.521	5.77	0.0707	0.0427	8	27.035	5.88	0.0598	0.0435
9	26.545	5.78	0.0724	0.0449	9	27.038	5.99	0.0652	0.0588
10	26.85	5.9	0.0947	0.0504	10	27.124	6.3	0.0718	0.0708

11	26.895	5.89	0.1179	0.0552	11	27.147	6.34	0.0761	0.088
12	27.083	5.99	0.1209	0.0581	12	27.211	6.29	0.0765	0.0882
13	27.343	6.1	0.1653	0.0688	13	27.349	6.59	0.0816	0.1121
14	27.456	6.12	0.1716	0.0757	14	27.364	6.7	0.0859	0.1401
15	27.475	6.3	0.1791	0.0828	15	27.383	7.2	0.1176	0.155
16	27.585	6.22	0.1821	0.0942	16	27.49	6.7	0.1287	0.1581
17	27.687	6.4	0.1843	0.1016	17	27.529	6.8	0.1394	0.1614
18	27.7	6.45	0.1933	0.1071	18	27.57	7.1	0.1435	0.1762
19	27.772	6.8	0.2043	0.1167	19	27.636	7.4	0.1513	0.1789
20	27.876	6.7	0.2053	0.1233	20	27.678	7.43	0.1528	0.1837
21	27.9.2014	6.77	0.2331	0.1364	21	27.738	7.8	0.1672	0.1938
22	27.93	6.9	0.2819	0.1371	22	27.79	7.67	0.1757	0.2056
23	28.356	7.1	0.2822	0.1394	23	27.841	7.77	0.1765	0.2124
24	28.842	7.2	0.315	0.147	24	27.866	7.4	0.1836	0.2134
25	28.857	7.5	0.3199	0.1475	25	27.916	7.88	0.1905	0.2331

From observing Table (1) and the test criteria, we find that the models estimated using the Rlasso method are more stable and less affected by structural changes, especially when there are sudden changes in the data. As for the accuracy of estimation and prediction, the Lasso technique is better in that it provides an accurate method for estimating the model parameters and reduces the effect of the independent variables that are not important in the model.

5. Conclusions

The results of the simulation studies demonstrated the significant role of estimation techniques (Lasso & Rlasso) in enhancing model stability and improving prediction accuracy. The Rlasso technique exhibited superior performance when applied to complex data and high-dimensional problems. On the other hand, the Lasso technique has been proven to be an easy and fast implementation for less complex data.

6. Supplementary material

(None).

7. Author’s Contributions

Haitham Hassoun Majid: Designed the study, formulated the theoretical framework, and outlined the practical methodology. Jasim Hassan Lazim: Drafted and edited the initial manuscript and developed the practical implementation program.

8. Funding

(None).

9. Data availability statement

Simulated data were generated using Python to test the LASSO and RLASSO techniques under multicollinearity problem.

10. Acknowledgments

(None).

11. Conflict of interest

The authors declare no conflict of interest.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

[2] Aljasimee, A. K., & Alhamzawi, R. J. (2023). Bayesian group bridge composite quantile regression. Journal of Namibian Studies, 33, 433-444. https://repository.qu.edu.iq/wp-content/uploads/sites/31/2022/10/Bayesian-group-bridge-composite-quantile-regression
[3] Alkenani, A., & Yu, B. (2013). Penalized single-index quantile regression. International Journal of Statistics and Probability, 2(3), 12-25. https://doi.org/10.5539/ijsp.v2n3p12
[4] Aylin Alin (2010)."Multicollinearity" . WIREs ComputationalStatistics.2(3), 370-374 https://doi.org/10.1002/wics.84
[5] Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). Statistics for experimenters: Design, innovation ,and discovery (2nd ed.). Wiley-Interscience. https://doi.org/10.1002/0471718130
[6] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28(4), 591-605. https://doi.org/10.2307/1910133.
[7] Dalgaard, P. (2008). Introductory statistics with R. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-79054-1
[8] Dufour, J. M. (1982). Recursive Stability Analysis of Linear Regression Relationships: An Exploratory Approach. Journal of Econometrics, 19(1), 31-76.
[9] Engle, R. F. (1982). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. Econometrica, 50(6), 1417–1438. https://doi.org/10.2307/1913352
[10] McLachlan, G. J. (2004). Discriminant Analysis and Statistical Pattern Recognition. Wiley-Interscience. https://doi.org/10.1002/0471725293
[11] Mills, T. C. (2014). Testing for Stability in Regression Models. In Analyzing Economic Data (pp. 243-259). Palgrave Macmillan. https://doi.org/10.1057/9781137401908_17
[12] Huberty, C. J. (1994). Applied discriminant analysis. New York: Wiley and Sons. ISBN: 0471311456
[13] Liao, J., Wan, A. T. K., He, S., & Zou, G. (2022). Optimal model averaging for multivariate regression models. Journal of Multivariate Analysis, 189, 1–11. https://doi.org/10.1016/j.jmva.2021.104858
[14] Rencher, A. C. (2002). Methods of Multivariate Analysis (2nd ed.). Wiley-Interscience. DOI: 10.1002/0471271357
[15] Plonsky, L., & Oswald, F. L. (2017). "MULTIPLE REGRESSION AS A FLEXIBLE ALTERNATIVE TO ANOVA IN L2 RESEARCH." Studies in Second Language Acquisition. 2017;39(3):579-592. https://doi.org/10.1017/S0272263116000231
[16] Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), 461-464. https://doi.org/10.1214/aos/1176344136
[17] Song, Q. (2018). An overview of reciprocal L 1-regularization for high dimensional regression data. Wiley Interdisciplinary Reviews: Computational Statistics, 10(1), e1416. https://doi.org/10.1002/wics.1416
[18] Song, Q., & Liang, F. (2015). High-dimensional variable selection with reciprocal l 1-regularization. Journal of the American Statistical Association, 110(512), 1607-1620. https://doi.org/10.1080/01621459.2014.984812
[19] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso . Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
[20] Wald, A. (1983). The classical test of hypothesis and the problem of multiple decisions. In Selected Works of A. Wald (pp. 1-100). Springer, New York, NY. https://doi.org/10.1007/978-1-4613-8196-6_1
[21] Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67. https://doi.org/10.1111/j.1467-9868.2005.00532.x
[22] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
[23] Zucchini, W., & MacDonald, I. L. (2009). An introduction to simulation and Monte Carlo methods using R. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-10462-7

https://doi.org/10.31272/jae.i149.1389
https://admics.uomustansiriyah.edu.iq/index.php/admeco
P-ISSN: 1813–6729 E-ISSN: 2707-1359

JAE

توظيف طريقة Lasso و Rlasso لفحص استقرارية نموذج الانحدار الخطي المتعدد

هيثم حسون ماجد الخزاعي
قسم الرياضيات، الكلية التربوية المفتوحة ، مركز النجف الدراسي ، العراق
Email: haithamnkhozaee@gmail.com, ORCID ID:\ https://orcid.org/0009-0000-8286-1960
جاسم حسن لازم الشويلي
قسم السياحة، الكلية التقنية الادارية، الجامعة التقنية الوسطى ، بغداد ، العراق
Email: jhlazem@mtu.edu.iq, ORCID ID:\ https://orcid.org/0000-0001-9667-1009

المستخلص

ان مشكلة عدم استقرارية نماذج الانحدار تعتبر من ابرز المشاكل التي تواجه الباحثين عند تقدير النموذج. من جانب اخر تقنيتي التنظيم (Lasso & Rlasso) تعد من ابرز التقنيات المطبقة في تقدير نماذج الانحدار المتعدد لما لها من قابلية في الوصول الى نموذج انحدار تفسيري يمكن من خلاله اجراء التنبؤات بشكل دقيق . في هذا البحث تم تطبيق اسلوب المحاكاة لإجراء مقارنة بين مجموعة من نماذج الانحدار تم تقديرها باستخدام تقنيتي التنظيم (Lasso & Rlasso) حيث تم استخدام اختباري Chow Test & CUSUM Test) لاختبار استقرارية النماذج المقدرة و بوجود مشكلة التعدد الخطي و اظهرت النتائج ان كلا التقنيتين كان لها اداء متميز في بناء نموذج انحدار مستقر ، كما اظهرت النتائج تفوق تقنية Rlasso ولاسيما عند وجود مشكلة التعدد الخطي .

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 2025 / 04 / 17
تاريخ قبول البحث: 2025 / 05 / 26
تاريخ نشر الكتروني: 2025 / 09 / 01
عدد صفحات البحث 68 - 77

الكلمات المفتاحية:

Lasso ,Rlasso , Chow test , CUSUM test , Multicollinearity problem

المراسلة:
أسم الباحث: هيثم حسون ماجد الخزاعي
Email: haithamnkhozaee@gmail.com