Selection of Tuning Parameter and Comparison of Lasso and Adaptive Lasso on ZYZ Condition: A Monte Carlo Study

Amir Shahzad[1], Umar Waqas[2], Nazakat Ali[3] and Hafiz Bilal Ahmad[4]

https://doi.org/10.62345/jads.2024.13.3.21

Abstract

One of the fundamental objectives of statistics is to achieve accurate predictions. In high-dimensional settings (where the number of variables, p, exceeds the number of observations, n), the performance of ordinary least squares (OLS) is often suboptimal due to its high variance, which leads to lower prediction accuracy. Shrinking the variables is a promising approach, and methods such as ridge regression, elastic net, lasso, and adaptive lasso are well-known techniques for this purpose. While variable shrinkage introduces a small bias, it significantly reduces the variance compared to OLS. The effectiveness of shrinkage methods largely depends on the selection of the tuning parameter. Cross-validation and the Bayesian Information Criterion (BIC) are commonly used for this purpose, and an improved version of BIC has shown impressive results. $\beta_o = (5.6,5.6,5.6,0), \beta_o = (3,1.5,0,0,2,0,0,0), \beta_o = (0.85,0.85,0.85,0)$ *are the multiple regression models which are compared.*

Keywords: Lasso Regression, Adaptive Lasso Regression, Ridge Regression, Cross Validation.

Introduction

The penalty estimator, part of the least squares penalized family, optimizes a quadratic function with a penalty term. Notable estimators like lasso belong to this family. In penalized least squares, the penalty term influences regression coefficient calculations, shrinking coefficients of variables with minimal impact toward zero. This process removes unimportant variables, yielding a simplified sub-model that is easier to interpret. Such methods are instrumental in high-dimensional models where the number of predictors exceeds the number of observations (p > n) (Ahmed, 2014).

In shrinking methods, if prior information about the sub-model is given, then the estimates are calculated from it. If initial information is not available, then different information criteria are available for the variable selection. Such information criteria take some variables from the model, drop some variables from the model, and help in the variable selection. Bayesian information criteria (BIC), Akaike information criteria (AIC) and many other information criteria are available for variable selection (Ahmed, 2014).

Penalty estimators and shrinking methods both reduce variables by shrinking coefficients toward zero, aiding in variable selection. The extent of shrinking depends on the tuning parameter. Lasso, SCAD, and adaptive-lasso are popular methods, with lasso excelling in

[1]*Master of Philosophy in Statistics, Department of Mathematics & Statistics, University of Agriculture, Faisalabad, Email: amir.statistics07@gmail.com*

[2]*Lecturer, Shiblee Group of Colleges and M. Phil Statistics from GC University Faisalabad, Email: umarwaqas86@yahoo.com.*

[3]*Master of Philosophy in Statistics, Department of Mathematics & Statistics, University of Agriculture, Faisalabad, Email: nazakatalibandesha@gmail.com*

[4]*Master of Philosophy in Statistics, Department of Mathematics & Statistics, University of Agriculture, Faisalabad, Email: bilal.uafstat@gmail.com*

Compliance with ethical standards: There are no conflicts of interest (financial or non-financial). This study did not receive any funding.

creating sparse models but struggling to distinguish covariates with weak or no effects (Ahmed, 2014).

The correlation between the predictor variables (multicollinearity) in a multiple regression model is a big problem because it causes high variance and gives misleading results. One of the main objectives of statistics and also regression is prediction when variance is so high that prediction will not be acceptable. Ridge regression is a method from the list of penalty estimation methods that deal with multicollinearity and give low variance, which is helpful in the prediction. The main idea of ridge regression is to shrink the coefficient to zero, which provides a sub-model and also helps in model selection (Mahajan et al., 1977).

Ridge regression shrinks the variable toward zero, which gives us low variance and helps solve a very dangerous problem of multicollinearity. Multicollinearity is basically a correlation between the independent variables (X's), and the result of multicollinearity is misleading coefficients and high variance. Due to high variance and misleading coefficient, prediction is not possible because if the result is dishonest, then how can prediction be reliable (Hoerl et al., 1975)?

In multiple regression data, ordinary least squares (OLS) give high variance, and its performance is not so good in the presence of multicollinearity, but OLS is an unbiased estimator. Ridge regression in multiple regression data gives low variance and also handles the problem of multicollinearity, but ridge regression provides a biased estimate.

Lasso, a penalty estimator, shrinks non-significant or weak effect variables to zero, selecting only significant predictors. It is effective in high-dimensional settings ($n < p$) but also performs well elsewhere. Proposed by Tibshirani (1996), lasso outperforms ordinary least squares (OLS) by offering lower prediction variance, despite being slightly biased (Tibshirani, 1996).

Lasso estimators are used for different purposes, but the most commonly used features of lasso are two. (A) Accuracy in the prediction because lasso gives prediction with low variance in the comparison of ordinary least square estimators (OLS). (B) The other reason to use the lasso technique is interpretation because it is a difficult task to do interpretation with a large number of predictors when the lasso selects a subset of the predictors so it becomes easy to interpret. One of the primary needs of statistics is to do prediction with the minimum error or low variance, so lasso provides such an estimator for these properties. Lasso is famous among researchers (Hastie et al., 2015).

Lasso is the same as the ridge regression. Still, the main difference between the ridge regression and lasso is the penalty because the ridge regression includes the coefficient in the model. Still, the lasso consists of the in the model, and this is the main reason why ridge regression and lasso regression are different. Lasso shrinks some coefficients precisely equal to zero. Ridge regression cannot shrink the coefficient equal to zero. It only shrinks near to zero (Chand, 2012).

$$\beta^{*}=\text{argmin}\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \text{abs}(\beta_j)\}$$

Λ is the tuning parameter that is responsible for the shrinking of the variables. If the value of λ is considerable, then variable shrinking will be over, and if λ is small, then the shrinking of the variable will be low. The selection of the λ is also a difficult stage because it is the backbone of the shrinking, and shrinking is the backbone of the actual model estimation.

As every method has some properties, similar to most minor, the absolute shrinking selection operator has two properties. These properties are called oracle properties.

- The zeros (only and only zero components) will calculate equal to zero, and the probability will increase and go to the one when n will increase. Probability → 1 when n → infinity.
- If the correct sub-model is known, then parameters that and not zero will be calculated correctly and very efficiently.

Both properties have great value in correct variable selection (Chand, 2012).

Least angle regression (lars) is an algorithm that performs the lasso procedure. The purpose of the least angle regression is to provide all possible predictors with forward and backward stepwise regression, choose the best subset of the predictor, and offer it as a variable, which is selected through this produce Lars perform model selection (Efron et al., 2004).

There are six steps to perform the lasso with the help of Lars algorithm, which are the following,

- Mean should be zero of predictors, variance should be constant of predictors and residual should equal to zero.
- Predictor should be in standardized form (X_j : j= 1,…,p), mean should zero and variance should be constant, start with the residual r= y-$\bar{y}$, $\beta_1, \dots, \beta_p = 0$
- Find predictors which are most correlated.
- Move β_j to least square coefficients (X_j, r) from zero, until any other X_k competitor match the correlation with the X_j current residual.
- Move (β_j, β_k) by joint least square coefficient in the direction, until any other X_j competitor have such correlation with current residuals.
- If some non-zero becomes zero then drop it and recompute the least square joint direction.
- Repeat all the process until the all p predictors entered in the model and we arrive at the full least squares solution.

The smoothly clipped absolute deviation (SCAD) algorithm shrinks variables and selects them based on specific criteria, incorporating penalized likelihood methods. It is effective in both real-life and simulation-based data (Matsui et al., 2011).

Lasso is good in variable selection and variable shrinking, but sometimes the lasso results in several coefficients becoming identically zero, and it fails in the oracle properties (Fan & Li, 2001). For this problem, a new algorithm was proposed name SCAD for variable selection and estimation of coefficients simultaneously and automatically. This method performs both work variable selection and shrinking of the variables and also makes spare solutions, ensuring the continuity of the model that is selected (Fan et al., 2001).

SCAD is not biased. It is an unbiased technique because the results of the sub-model, which is selected from the SCAD method, are equal to the original models for significant coefficients (Fan & Li, 2001).

This technique has lasso variables selection property along with penalized least square properties; the p>>n situation is not ideal for the lasso because it calculates most n predictors. Elastic net performance is better than lasso in variable selection sometimes (Zou et al., 2005).

The first purity is to estimation of the parameter (θ), and it this an assumption that the population from where the sample is collected is customarily distributed or close to the normal distribution. When the reliability of $(\theta\ )$ is unknown, which is the estimation of the population parameter so, there is a method to check the reliability. The combination of the (θ) and (θ) which define a mean square error (MSE) optimal linear shrinking estimator:

$\hat{\theta}^{ls} = c\theta_0 + (1 - c)\,\hat{\theta}$

"C" is responsible of the shrinking and the selection of the "c" in such a way that it minimizes the mean square error (MSE) and also "c" can be called a confidence degree.

The "c" is mostly between the zero and one, c $\in [0,1]$. When c=0 it's mean that we are only using the sample data because prior information is subtract when c=0. The concept of the linear shrinking estimator (LSE) is that it shrink the sample estimator $(\hat{\theta})$ towards the θ_0 (Ahmed, 2014).

Zou (2006) proposed a new shrinking method which was similar as lasso but it was better in selection and shrinking of variables because lasso was successful in selection of the true model as oracle properties but this new method of shrinking was far better.

When lasso becomes very famous everywhere then (Zou, 2006) proves that lasso gives constant variable selection when a condition fulfills. He was the first who noticed this certain condition and he named this condition as irrepresentable condition. He proves that when this condition is not satisfied then lasso gave inconstant variable selection and lasso don't enjoy the oracle properties named this condition as ZYZ condition (Chand, 2012).
Adaptive lasso is an advanced stage of the lasso. The weight is added in the old lasso and it gives more constant results than the previous. It enjoys the oracle properties in all conditions (Zou, 2006).

β^*=$argmin\{ \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} w_j abs(\beta_j)\}$

W are the weights.

Adaptive lasso does not have any conditions as lasso has it perform similarly in any condition. The condition we are talking about is the irrepresentable condition, which is responsible for the performance of the lasso. If this condition is fulfilled, then the lasso also performs well. Of course, adaptive lasso, but if this condition is violated, then lasso selection is not constant, and the adaptive lasso again gives a continuous selection of the variables.
Chand (2012) named as ZYZ condition as the name of the originals of this condition are Zhao (2006). ZYZ basically is the condition on variance and covariance matrix of the variables when defined following:

$y_i = Xi^T\beta + \epsilon_{ij}$; $y_i \in R$

Xi is the response with p numbers of dimensions. ϵ is the error term which is identical and independent with mean zero and constant variance (σ^2).

A = { j : $\beta_j \neq 0$}; A is non-zero beta's β_j.

$p_0 <$ p ; $p_0$is the non-zero dimensions and p are the total dimensions.

$C = \begin{bmatrix} C11 & C21 \\ C21 & C22 \end{bmatrix}$

C is the variance and covariance matrix.
C11 is $p_0 \times p_0$ matrix; the matrix of non-zero order.
C12 is the remaining matrix of C.
C21 is the transpose of C12.
C22 is the remaining part of C.

$|[c_{21}c_{11}^{-1}\beta(A)]_r| \leq 1$, r=1,…, p -$p_0$

This is the ZYZ condition if its result is less than 1 so mean condition hold and lasso will perform good and if this condition is not hold then lasso variable selection will not be constant and we have to more for the adaptive lasso which is good in both situation is this condition hold or not adaptive lasso variable selection is constant.
The linear regression model

$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$, I = 1,2,…,n.

In the form of the vectors, it can be written as

$y_i = \beta_0 + x_i^T \beta + \varepsilon_i$

$y_i \epsilon$ R and the variables of the response $x_i = (x_{i1}, \ldots, x_{ip})^T \epsilon R^p$, where p are the dimensions of the set of predictors, $\varepsilon_i \sim$ N (0,σ^2) and β= ($\beta_{1,\ldots,\beta_p}$) is the set of the parameters and β_0 *is constant.*

Objectives

In this study, four different techniques for shrinking the variables are compared.
(i) Lasso (ii) Adaptive lasso (iii) Elastic net (iv) Ridge regression
For selection of the tuning parameter three techniques are compared when ZYZ condition hold or not.
(i) Classical BIC (ii) Improved BIC (iii) Cross validation

Review of Literature

Lasso is renowned for its variable selection and shrinking capabilities, improving model dominance by shrinking variables and minimizing the sum of squares. An advanced form of ridge regression, lasso reduces variance, aiding prediction. Although biased compared to ordinary least squares (OLS), its low variance is beneficial (Tibshirani, 1996).

Lasso and its extensions perform variable selection in regression models effectively, including in logistic regression. They maintain consistent performance under orthogonal reparameterizations and excel in high-dimensional settings ($n < p$), offering low variance and improved prediction accuracy (Meier, 2008).

Model selection is crucial, as poor choices lead to misleading results. Lasso, proposed by Tibshirani (1996), helps with this by shrinking variables and minimizing mean square error (MSE). It relies on the irrepresentable condition and tuning parameters; without this condition, lasso may yield inconsistent results (Zhao et al., 2006).

The new regularization technique is useful for minimizing the experimental risk and leads to sparse estimators. The vector sparse support is mostly added of potentially overlapping groups of covariates defined a prior, covariate set which attached to each other when graph of covariate is available. In this study estimation properties are under consideration and illustrate the behavior on simulated and cancer of breast gene expression data (Jacob et al., 2009).

Penalization of least square by the sum of absolute coefficients is called the lasso and this technique allow to shrink some coefficient to exactly zero for the subject model selection. In such data when sample size (n) is less then variable (p) the interpretation is a difficult stage lasso is a good option it shrinks some coefficient to exactly zero and exclude from the model and minimize the mean square error (Tibshirani et al., 2005).

The important behavior which is asymptotic of regression estimator for minimizing the difference between estimated and original values (residuals) adding a penalty proportional to $\varepsilon|\beta_j|^r$ for γ is greater than zero. Lasso is the special case when is γ equal to one (Knight et al., 2000).

For high dimensional data in machine learning lasso can lead great accuracy, introduction of l_1 and l_2 penalty for linear regression model have great effect to achieve the desire results in group wise and within group. This simulation study showed that this technique helps in variable selection with consistency (Simon et al., 2013).

The linear regression of the lasso communicates to mode posterior when independent double exponential distribution prior applies on the regression coefficient. This study introduced a new Bayesian lasso treatment. Coefficients of regression posterior distribution calculated and hypothesis on the behalf of this characterization. Point estimation is emphasis with the help of posterior mean, which help for the prediction along with the posterior predictive distribution. This study shows that the lasso which is standard don't need to match with model-based (Hans, 2009).

Least absolute shrinking selection operator (lasso) which is used for calculation of coefficient vector regression also minimizing the sum or square of the residuals suggested by Tibshirani. Lasso shrink some coefficient with low effect towards zero and exclude it from the model by this it preforms the model selection. Library name S-plus is developed for the lasso by (Osborne et al., 2000).

Bridge regression is a special extension of the penalty regression, lasso y=1 a new algorithm is developed during the studying of the bridge regression. The selection of the tuning parameter which is a very difficult and important step is done by the generalized cross validation and the result of the simulation shows that bridge regression performance is great when we compare it with the lasso and ridge regression with some computational advantages and limitations (Fu, 1998).

Testing of the lasso is also having great importance which will confirm that the variables which are added in the model is significant or not, a significance test which was purposed for the test of the significance name is covariance test statistic. This test is used for the linear model of lasso testing of the significance and for testing the two variables chi-square test is used. In this analysis shrinkage of the variable paly very important role and shrinking is done by the penalty of the lasso (Lockhart et al., 2014).
Genome-wide forecast become the technique for the choice of animals and plants breeding. Such kind of prediction required millions of observations for the prediction, lasso have great performance in high dimensional data so lasso is great option for the prediction and choice of animals and plants breeding. Different algorithms are suggested like AUTALASSO which is based on adaptive lasso (Waldmann et al., 2019).
This research proposes the self-adaptive lasso for variable selection in linear regression, using a specific tuning parameter for each coefficient. It employs a Bayesian framework and Gibbs sampling. Results are tested on simulation and real-life data, with extensions to elastic net and fused lasso (Kang et al., 2009).
In this paper adaptive least absolute shrinking selection operator (lasso) is studied for the autoregressive model. Adaptive lassos exclude some variable from the model by shrinking them exactly equal to zero. This study proposes a modified Bayesian information criteria (MBIC) for the selection of the tuning parameter of adaptive lasso. A small number study is conducted for verify the results (Kwon et al., 2017).
This paper explores penalized quantile regression for variable selection, where shrinking depends on a tuning parameter. Results show the estimators are unbiased and normal if individual effects have zero-median distribution. Monte Carlo studies demonstrate low variance without bias (Lamarche, 2010).

Materials and Methods

It is a Monte Carlo study; data is generated with R-software (3.6.1). The technique which is used by the researcher is simulation which is very famous for generating the random variables. Simulation technique is well known for producing artificial data, random numbers for analysis and lab study. One of the most common reasons for using the simulation is eliminated the approximations, means situation can vary in real life so simulation help in such situation and eliminate such approximations (Heermann, 1990).
The main idea of the Monte Carlo study is to repeat the random sampling with the help of computer software to solve the deterministic problems. It is becoming very famous technique in different field of study like nuclear physics, statistics and many other fields (Hammersley, 2013).
Three models are chosen for the comparison, the models are:
Model 1: $\beta_0 = (5.6, 5.6, 5.6, 0)$
Model 2: $\beta_0 = (3, 1.5, 0, 0, 2, 0, 0, 0)$
Model 3: $\beta_0 = (0.85, 0.85, 0.85, 0)$
These models are used by (Chand, 2012).
Cross-validation divides data into K equal subsets. One subset is used for testing, while the remaining (K-1) subsets are used for training. The model is trained on the (K-1) subsets and tested on the remaining one, repeating the process until each subset has been used for testing (Ahmed, 2014).
Selecting the tuning parameter is crucial for lasso methods, as it controls shrinking and variable selection. Cross-validation is a popular approach for tuning parameter selection and minimizing mean squared error (Chand, 2012).
Cross validation has several steps which are explained below:

- Data which is consisting of observation (n) classified into random and mutually exclusive subparts and the subparts known as k-folds.
- The whole solution path is obtained as a function of tuning parameter which is standardized, mean zero and constant variance s $\epsilon[0,1]$ and exclude the I^{th} fold.
- The K-1 model is use for the prediction of the excluded I^{th} submodel and error of the prediction is calculated on every picked tuning parameter s $\epsilon[0,1]$.
- The s-value is responsible for the smallest error of prediction and selection of the optimal tuning parameter.

Bayesian information criteria (BIC) is a well know information criteria for model selection and it also play role for selection of the tuning parameter in the shrinking method. There are different methods for the tuning parameter selection, but the problem is the consistency because consistency is the symbol of low variance and low variance is the necessary for the prediction which is the main reason to use the regression and shrinking methods.

Wang (2009) suggested that BIC is a good option for variable selection and the main reason why it is good is consistency. It provides a tuning parameter which is consistent is variable selection and also have low variance.

BIC = $\ln(n)$ K – 2ln(L)

K is the number of regressor in the model.

Number of the data is n.

L is the likelihood function.

The main idea of the BIC is that ln(n) is penalizing and it's a trade of between ln(n) and likelihood the lowest value of BIC is selected as optimal and best.

BIC = $\ln(n)$ K – 2ln(L)($\sqrt{n}$/p)

Addition of $\sqrt{n}$/p by chand (2012) make it more consistent.

The solution path is the entire set of estimates corresponding to different tuning parameter choice, this measure is used for getting the probability of true model in solution path. The measure is used by zou (2006), n $\rightarrow\infty$ as PTSP $\rightarrow$ 1.

After selection of the tuning parameter this measure is used for comparison of the correct model selection, this measure tells the percentage of correct model selection. PCM is used by (Wang et al., 2009).

Generalized linear model (Glmnet) is R-software package which provide different things like generalized linear model with penalized maximum likelihood, linear regression, logistic regression, multiple-response regression etc.

Jerome Friedman, Trevor Hastie, Rob Tibshirani, Noah simon are the maker of the Glmnet and the person who is responsible to maintain this package in R-software is Trevor Hastie. Glmnet is extremely fast then other packages like lars and this package also have a upper edge of prediction, plotting and cross validation technique (Hastie, 2014).

Results and Discussion

Table 1 shows the ratio of true selection estimates (PTSP) for lasso for different samples and sample sizes. Model 2 shows a PTSP of 1.0 consistently across all samples, indicating excellent selection of the true variables. In comparison, the PTSP for model 1 decreased from 0.33 at n = 100 to 0.10 at n = 2000, indicating a decrease in accuracy for larger samples. Model 3 presents a different PTSP, ranging from 0.44 to 0.42 as the sample size increases, indicating a moderate and stable performance. This table shows the superiority and stability of model 2 over the other models in terms of selection accuracy.

Table 1: PTSP of lasso

Sample size (n)	Model 1	Model 2	Model 3
100	0.33	0.98	0.44
311	0.32	1.0	0.44
522	0.21	1.0	0.49
733	0.26	1.0	0.46
944	0.21	1.0	0.49
1156	0.17	1.0	0.47
1367	0.12	1.0	0.47
1578	0.22	1.0	0.45
1789	0.11	1.0	0.42
2000	0.10	1.0	0.47

Table 2: PTSP of adaptive lasso

Sample size (n)	Model 1	Model 2	Model 3
100	1.0	1.0	0.84
311	1.0	1.0	0.96
522	1.0	1.0	0.99
733	1.0	1.0	1.0
944	1.0	1.0	1.0
1156	1.0	1.0	1.0
1367	1.0	1.0	1.0
1578	1.0	1.0	1.0
1789	1.0	1.0	1.0
2000	1.0	1.0	1.0

Table 2 shows the selection probability (PTSP) for the modified lasso across different samples and models. For both model 1 and model 2, the PTSP consistently reached 1.0 for all sample sizes from 100 to 2000, indicating excellent performance in identifying true variables. Model 3 shows a higher PTSP of 0.84 at 100 samples and increasing to 1.0 over 733 samples. This suggests that although model 3 initially performed worse than models 1 and 2, it actually improved to keep pace with the larger models.

Table 3: PTSP of model 1

Sample size (n)	Model 1	Model 2	Model 3
100	1.0	0.97	0.42
311	1.0	1.0	0.47
522	1.0	1.0	0.44
733	1.0	1.0	0.46
944	1.0	1.0	0.46
1156	1.0	1.0	0.40
1367	1.0	1.0	0.54
1578	1.0	1.0	0.45
1789	1.0	1.0	0.38
2000	1.0	1.0	0.44

Table 3 shows the true selection rates (PTSP) for different models for model 1. Model 1 consistently achieved a PTSP of 1.0 across all sample sizes from 100 to 2000, indicating that it is good at detecting the difference accurately. In contrast, model 2 and model 3 do the opposite. Model 2 maintains a high PTSP of 0.97 or 1.0, while model 3 offers lower and more pronounced PTSP ranging from 0.38 to 0.54. This change in performance of model 3 suggests that it is less reliable than model 1 in detecting the true variables, especially in small samples.

Table 4: PTSP of ridge regression

Sample size (n)	Model 1	Model 2	Model 3
100	0	0	0
311	0	0	0
522	0	0	0
733	0	0	0
944	0	0	0
1156	0	0	0
1367	0	0	0
1578	0	0	0
1789	0	0	0
2000	0	0	0

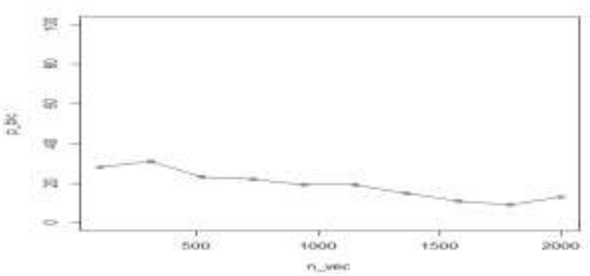
Ridge regression PTSP is equals to zero on every sample size and also in every model because ridge regression doesn't shrink's the variables exactly equal to zero. Table 4 displays the PTSP of ridge regression across different sample sizes. The sample sizes range from 100 to 2000. For all models (model 1, model 2, and model 3), the PTSP values are consistently zero across all sample sizes. This indicates that there is no variation in PTSP values regardless of the sample size or the model used.

Table 5: (Model 1) PCM of Lasso with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0.28	0.31	0.23	0.22	0.19	0.19	0.15	0.11	0.09	0.13

In model 1, lasso's comparative analysis (PCM) against Bayesian Information Criterion (BIC) was analyzed in multiple samples. As the sample size increases, the PCM value decreases from 0.28 at a sample size of 100 to a low of 0.09 at a sample size of 1789, then slightly increases to 0.13 in 2000.

Figure 1: (Model 1) PCM of Lasso with BIC



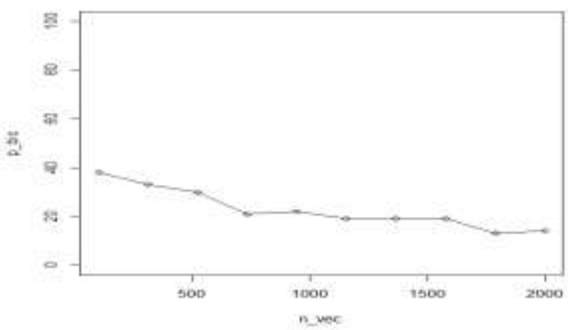
The ZYZ condition does not hold in model 1, so the lasso performance is not consistent with the classical BIC.

Table 6: (Model 1) PCM of Lasso with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0.38	0.33	0.30	0.21	0.22	0.19	0.19	0.19	0.13	0.14

Table 6 presents the performance comparison metric (PCM) of the lasso model with improved Bayesian Information Criterion (BIC) across various sample sizes. The results indicate that as the sample size increases from 100 to 2000, the PCM initially decreases from 0.38 to a low of 0.13 at 1789, before slightly rising to 0.14 at 2000. This suggests that the lasso model's performance improves with larger sample sizes, showing a general trend of enhanced model accuracy and stability with increased data.

Figure 2: (Model 1) PCM of Lasso with improved BIC



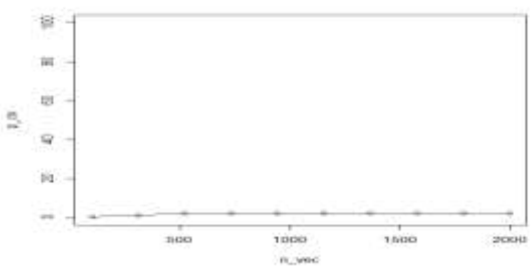
The ZYZ condition does not hold in model 1, so the lasso performance is not consistent with the improved BIC.

Table 7: (Model 1) PCM of lasso with cross validation

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Model 1 in table 7 shows the performance of the lasso model in competition. In the size range (n) from 100 to 2000, the PCM (Predictive Consistency Measure) always shows the minimum value of 0 to 0.02. This shows that the predictive performance of lasso remains constant regardless of sample size, with a slight increase in PCM as the sample size increases. However, the overall effect of sample size on the consistency of sample estimates appears to be small, indicating that the performance of lasso is similar across sample sizes.

Figure 3: (Model 1) PCM of lasso with cross validation



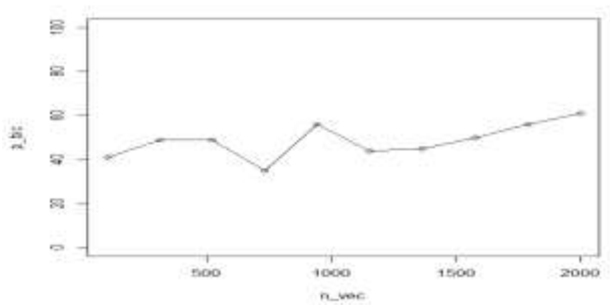
The ZYZ condition does not hold in model 1, and cross-validation is not an ideal method for selecting tuning parameters, leading to inconsistent performance of lasso when using cross-validation.

Table 8: (Model 2) PCM of Lasso with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 2	0.41	0.49	0.49	0.35	0.56	0.44	0.45	0.50	0.56	0.61

In table 8, the Performance Comparison Metric (PCM) for model 2 using lasso with BIC across various sample sizes (n) is presented. The PCM values range from 0.35 to 0.61. Notably, PCM improves with increasing sample size, peaking at 0.61 for a sample size of 2000. This indicates that lasso's performance, as measured by PCM, generally enhances with larger sample sizes when selecting the BIC tuning parameter. However, PCM exhibits some variability across smaller sample sizes, suggesting that while larger samples yield more stable performance, smaller samples may result in less consistent outcomes.

Figure 4: (Model 2) PCM of Lasso with BIC

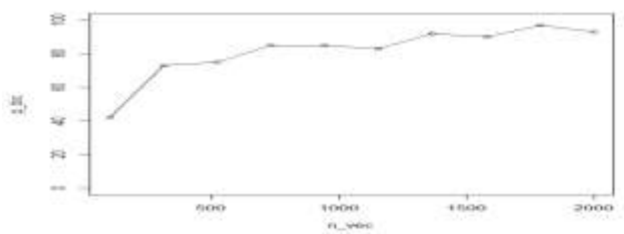


ZYZ condition hold in model 2 so lasso performance is consistent with classical BIC but slow to approch to probability 1. The ZYZ condition holds in model 2, making the lasso performance consistent with the classical BIC, but it is slow to approach a probability of 1.

Table 9: (Model 2) PCM of Lasso with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 2	0.42	0.73	0.75	0.85	0.85	0.83	0.92	0.90	0.97	0.93

Figure 5: (Model 2) PCM of Lasso with improved BIC

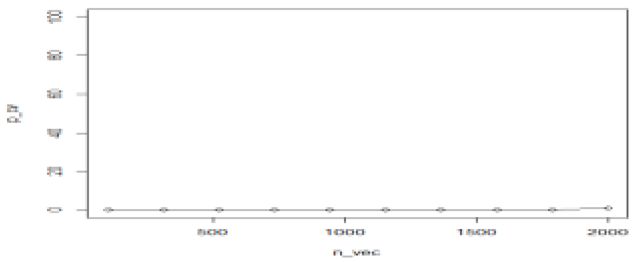


The ZYZ condition holds in model 2, making the lasso performance consistent with an improved BIC and leading to a rapid convergence towards a probability of 1.

Table 10:(Model 2) PCM of lasso with cross validation

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 2	0	0	0	0	0	0	0	0	0	0.01

Figure 6: (Model 2) PCM of lasso with cross validation

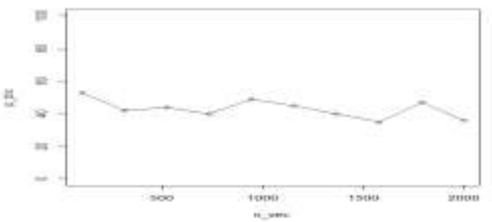


ZYZ condition hold in model 2 but lasso performance is not consistent with cross validation.

Table 11:(Model 3) PCM of Lasso with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 3	0.53	0.42	0.44	0.40	0.49	0.45	0.40	0.35	0.47	0.36

Figure 7: (Model 3) PCM of Lasso with BIC

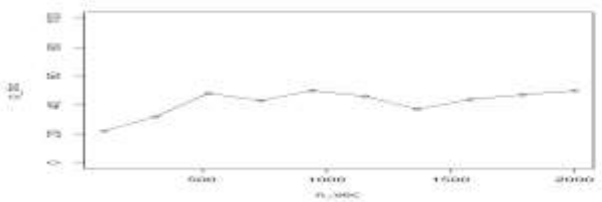


ZYZ condition not hold in model 3 so lasso performance is not consistent with Classical BIC.

Table 12:(Model 3) PCM of Lasso with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 3	0.22	0.32	0.48	0.43	0.50	0.46	0.37	0.44	0.47	0.50

Figure 8: (Model 3) PCM of Lasso with improved BIC

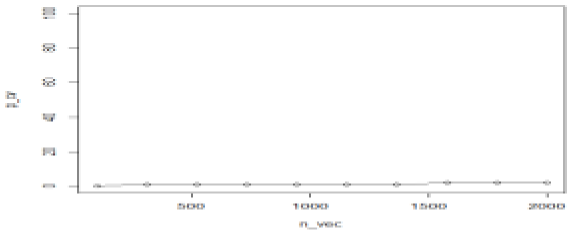


ZYZ condition not hold in model 3 so lasso performance is not consistent with improved BIC but little better than classical.

Table 13: (Model 3) PCM of lasso with cross validation

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 3	0	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02

Figure 9: (Model 3) PCM of lasso with cross validation

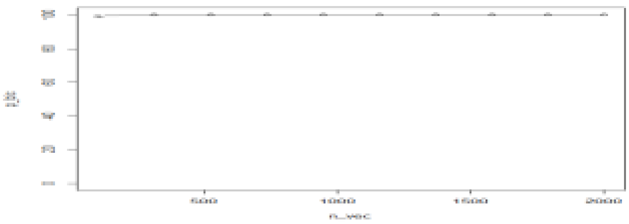


ZYZ condition not hold in model 3 so lasso performance is not consistent with cross validation.

Table 14: (Model 1) PCM of adaptive lasso with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0.99	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Figure 0-10: (Model 1) PCM of adaptive lasso with BIC

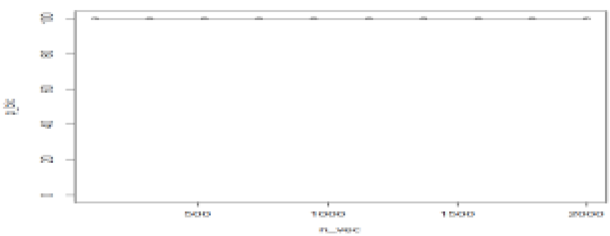


ZYZ condition don't affect the adaptive lasso performance, adaptive lasso performance is consistent with classical BIC.

Table 15:(Model 1) PCM of adaptive lasso with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Figure 11: (Model 1) PCM of adaptive lasso with improved BIC



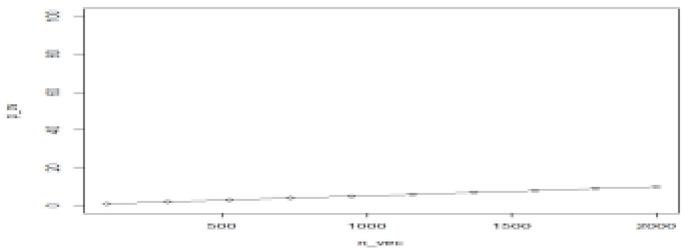
ZYZ condition don't affect the adaptive lasso performance, adaptive lasso performance is more consistent with improved BIC.

Table 16: (Model 1) PCM of adaptive lasso with cross validation

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10

The results presented in table 16 show the performance comparison metric (PCM) of model 1 using lasso adjustment and cross-validation across different models. The PCM value gradually increases with sample size, for example from 0.01 for a sample size of 100 to 0.10 for a sample size of 2000. The sample performance improves. This finding highlights the importance of sufficiently large sample sizes to optimize the performance of lasso transformations in predictive modeling.

Figure 12: (Model 1) PCM of adaptive lasso with cross validation

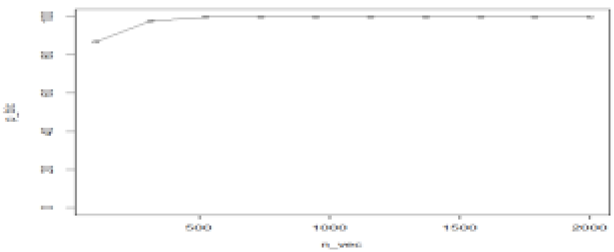


ZYZ conditions don't affect the adaptive lasso performance, but adaptive lasso performance is not consistent with cross validation.

Table 17: (Model 2) PCM of adaptive lasso with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 2	0.87	0.98	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Figure 13: (Model 2) PCM of adaptive lasso with BIC

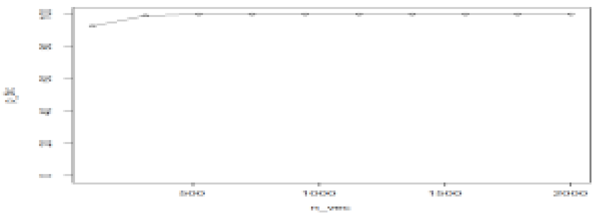


ZYZ condition don't affect the adaptive lasso performance, adaptive lasso performance is consistent with classical BIC.

Table 18: (Model 2) PCM of adaptive lasso with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 2	0.93	0.99	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Figure 14: (Model 2) PCM of adaptive lasso with improved BIC

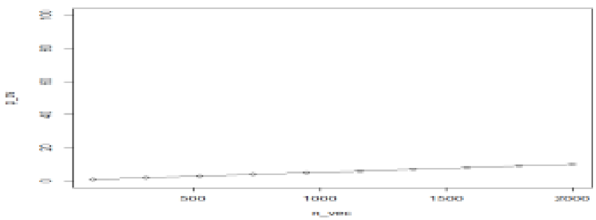


ZYZ condition don't affect the Adaptive lasso performance, Adaptive lasso performance is more consistent with improved BIC.

Table 19: (Model 2) PCM of adaptive lasso with cross validation

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 2	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10

Figure 15: (Model 2) PCM of adaptive lasso with cross validation



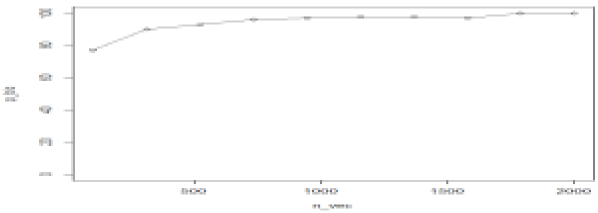
ZYZ conditions don't affect the adaptive lasso performance, but adaptive lasso performance is not consistent with cross validation.

Table 20: (Model 3) PCM of adaptive lasso with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 3	0.77	0.90	0.93	0.96	0.97	0.98	0.98	0.97	1.0	1.0

Table 20 presents the performance of model 3, adaptive lasso with BIC, across various sample sizes. The results show an increasing trend in model performance as sample size increases. For sample sizes ranging from 100 to 2000, the performance measure (PCM) starts at 0.77 and improves steadily to 1.0. Specifically, PCM values are 0.77 for n = 100, reaching 1.0 at n = 1789 and 2000. This indicates that as the sample size grows, the adaptive lasso model's performance with BIC significantly enhances, achieving perfect performance for larger sample sizes.

Figure 0-16: (Model 3) PCM of adaptive lasso with BIC

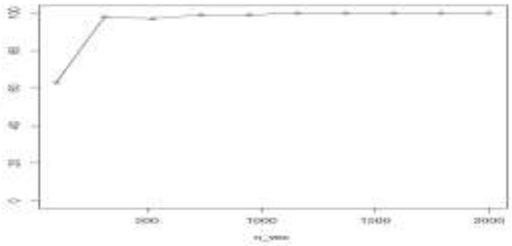


ZYZ condition don't affect the adaptive lasso performance, adaptive lasso performance is consistent with classical BIC but probability little slowly approach to 1 due to magnitude of the variables.

Table 21: (Model 3) PCM of adaptive lasso with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 3	0.63	0.98	0.97	0.99	0.99	1.0	1.0	1.0	1.0	1.0

Figure 0-17: (Model 3) PCM of adaptive lasso with improved BIC



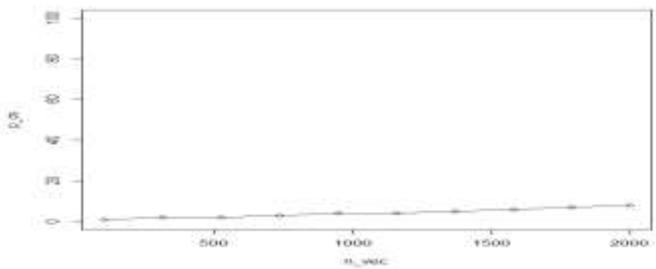
The ZYZ conditions don't impact the performance of the adaptive lasso. Its performance remains consistent, with the BIC improving steadily and the probability approaching 1 rapidly.

Table 22: (Model 3) PCM of adaptive lasso with cross validation

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 3	0.01	0.02	0.02	0.03	0.04	0.04	0.05	0.06	0.07	0.08

Table 22 shows the proportion of correct model (PCM) for model 3 with lasso adjustment and cross-validation among different models. PCM values range from 0.01 when sample size is 100 to 0.08 when sample size is 2000, indicating that the performance of the model gradually increases with the dataset. Despite the increase, PCM is still low for all sample sizes, indicating that lasso adjustment with competition has difficulty in identifying the correct model. This shows that although larger samples can increase the power, the effectiveness of this approach is still limited.

Figure 18: (Model 3) PCM of adaptive lasso with cross validation



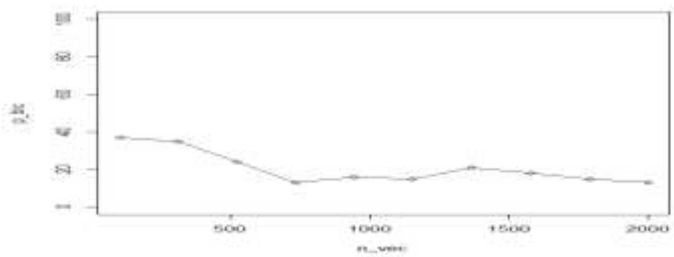
The ZYZ conditions don't impact the performance of the adaptive lasso, but its performance is not consistent with that of cross-validation.

Table 23: (Model 1) PCM of elastic net with BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0.37	0.35	0.24	0.13	0.16	0.15	0.21	0.18	0.15	0.13

Table 23 shows the proportion of correct models (PCMs) using elastic nets with BIC for model 1 across different models. The PCM value decreases as the sample size increases, starting from 0.37 at n = 100 and decreasing to 0.13 at n = 2000. The capacity gradually decreases. This model suggests that larger sample sizes may lead to decreased model accuracy, possibly due to more samples or noisy data, which may affect the performance of the model on larger datasets.

Figure 0-19: (Model 1) PCM of elastic net with BIC



The ZYZ conditions don't impact the performance of the adaptive lasso, but its performance is not consistent with that of cross-validation.

Table 24: (Model 1) PCM of elastic net with improved BIC

Sample size (n)	100	311	522	733	944	1156	1367	1578	1789	2000
Model 1	0.41	0.33	0.21	0.26	0.12	0.20	0.23	0.19	0.19	0.17

The results presented in table 24 show the proportion of correct model (PCM) for model 1 using an elastic net with improved BIC covering different models. The general decrease of PCM as the sample size (n) increases from 100 to 2000 indicates that the accuracy of the model in selecting the correct model decreases. The high correlation (0.41 for n = 100 and 0.33 for n = 311) indicates that the model is better at identifying the correct model when the sample size is small. However, as the sample size increases to 522 and above, the PCM decreases, reaching 0.12 at n = 944. This indicates that problems such as overfitting or noise continue to increase as the sample size increases. Preserving sample accuracy in large data sets may be important for researchers using this method for sample selection in large-scale studies.

Conclusion

In conclusion, the ZYZ condition plays a pivotal role in variable selection and shrinkage, particularly for lasso. When the ZYZ condition is not met, lasso's ability to select variables becomes inconsistent, leading to the failure of achieving oracle properties, regardless of the method used for tuning parameter selection—whether it's classical BIC, improved BIC, or cross-validation. Classical BIC is generally effective for tuning parameter selection in lasso, with improved BIC offering slight improvements, but cross-validation underperforms when the ZYZ condition holds.

For adaptive lasso, the ZYZ condition is less significant, as the method performs well regardless of its presence. Classical BIC continues to be a reliable method for tuning parameters, with improved BIC showing marginal advantages. However, cross-validation remains an ineffective choice.

In the case of elastic Net, the ZYZ condition does enhance the likelihood of selecting the true model, particularly when using classical BIC, although this improvement is gradual. Improved BIC performs similarly to classical BIC, while cross-validation is again found to be a poor option for tuning. Without the ZYZ condition, elastic net becomes inconsistent due to the lasso penalty, rendering classical BIC, improved BIC, and cross-validation ineffective.

Ridge regression, on the other hand, fails to achieve effective variable selection regardless of the ZYZ condition. Neither classical BIC, improved BIC, nor cross-validation succeeds in selecting the true model, resulting in a zero probability of success. This highlights the limitations of ridge regression in scenarios where variable selection is critical, emphasizing the importance of considering the ZYZ condition and appropriate tuning methods when applying these statistical techniques.

References

- Acquah, H.D.-G. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, *2*, 001-006.
- Ahmed, S.E. (2014). *Penalty, shrinkage and pretest strategies: variable selection and estimation.* Springer.
- Algamal, Z.Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications, 42,* 9326-9332.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, *4,* 40-79.
- Arslan, O. (2012). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis, 56*, 1952-1965.
- Bach, F.R. (2008). Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research*, *9*, 1179-1225.
- Benkeser, D., & Van Der Laan, D. (2016). *The highly adaptive lasso estimator. In: 2016 IEEE international conference on data science and advanced analytics (DSAA).* IEEE: pp: 689-696.
- Boulesteix, A.-L., De Bin, R., Jiang, X., & Fuchs, M. (2017). IPF-LASSO: Integrative-penalized regression with penalty factors for prediction based on multi-omics data. *Computational and mathematical methods in medicine*, 2017.
- Breiman, L. (1992). The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. *Journal of the American Statistical Association, 87*, 738-754.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The annals of statistics*, *24*, 2350-2383.
- Breiman, L. (1996). Stacked regressions. *Machine learning*, *24*, 49-64.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123-140.
- Browne, M.W. (2000). Cross-validation methods. *Journal of mathematical psychology, 44*, 108-132.
- Bunea, F., Tsybakov, A., & Wegkamp, M. (2007). Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics*, *1,* 169-194.
- Burnham, K.P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*, 261-304.
- Chand, S. (2012). *On tuning parameter selection of LASSO-TYPE methoss-A Monte Carlo Study.* International bhurban conferance on applied sciences & technology (IBCAST) Islambad, Pakistan: 120-129.
- Chand, S. (2011). D*iagnsotic checking and lasso variable selection in time series analysis.* PhD Thesis, School of Mathematical Sciences, The University of Nottingham, United Kingdom.

- Chatterjee, A., & Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association, 106*, 608-625.
- Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147-167.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics, 32*, 407-499.
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association, 96*, 1348-1360.
- Fu, W.J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics, 7*, 397-416.
- Hall, P., Lee, E.R., & Park, B. U. (2009). B*ootstrap-based penalty choice for the lasso, achieving oracle performance*. Statistica Sinica: 449-471.
- Hammersley, J. (2013). *Monte carlo methods.* Springer Science & Business Media.
- Hans, C. (2009). Bayesian lasso regression. *Biometrika, 96*, 835-845.
- Hastie, T., & Qian, J. (2014). *Glmnet vignette.* Retrieved June, 9: 1-30.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations.* CRC press.
- Heermann, D.W. (1990). Computer-simulation methods. In: Computer Simulation Methods in Theoretical Physics. *Springer*, 8-12.
- Hesterberg, T., Choi, N.H., Meier, L., & Fraley, C. (2008). Least angle and ℓ1 penalized regression: A review. *Statistics Surveys, 2,* 61-93.
- Hoerl, A.E., Kannard, R. W., & Baldwin, K. F. (1975). Ridge regression: some simulations. *Communications in Statistics-Theory and Methods, 4,* 105-123.
- Huang, J., Ma, S., & Zhang, C. H. (2008). *Adaptive Lasso for sparse high-dimensional regression models. Statistica Sinica*: 1603-1618.
- Huang, J., Ma, S., & Zhang, C. H. (2008). *The iterated lasso for high-dimensional logistic regression.* The University of Iowa, Department of Statistics and Actuarial Sciences.
- Hui, F.K., Warton, D. I., & Foster, S. D. (2015). Tuning parameter selection for the adaptive lasso using ERIC. *Journal of the American Statistical Association, 110,* 262-269.
- Ivanoff, S., Picard, F., & Rivoirard, V. (2016). Adaptive Lasso and group-Lasso for functional Poisson regression. *The Journal of Machine Learning Research, 17,* 1903-1948.
- Jacob, L., Obozinski, G., & Vert, J. P. (2009). *Group lasso with overlap and graph lasso. In: Proceedings of the 26th annual international conference on machine learning.* pp: 433-440.
- Jolliffe, I.T., Trendafilov, N.T., & Uddin, M. (2003). A modified principal component technique based on the LASSO. *Journal of computational and Graphical Statistics, 12,* 531-547.
- Johnson, V.E., & Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association, 107,* 649-660.
- Kamarianakis, Y., Shen, W., & Wynter, L. (2012). Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied stochastic models in business and industry, 28,* 297-315.
- Kang, J., & Guo, J. (2009). *Self-adaptive lasso and its Bayesian estimation.* Working Paper.
- Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. A*nnals of statistics*, 1356-1378.
- Kock, A.B. (2012). *On the oracle property of the adaptive lasso in stationary and nonstationary autoregressions.* CREATES Research Papers, 5.
- Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* In: Ijcai. Montreal, Canada: pp: 1137-1145.
- Krogh, A., & Vedelsby, J. (1995). *Neural network ensembles, cross validation, and active learning.* In: Advances in neural information processing systems. pp: 231-238.
- Kwon, S., Lee, S., & Na, O. (2017). Tuning parameter selection for the adaptive LASSO in the autoregressive model. *Journal of the Korean Statistical Society, 46,* 285-297.
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological methods & research, 33,* 188-229.

- Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. Journal of Econometrics, 157, 396-408.
- Leng, C., Lin, Y., & Wahba, G. (2006). *A note on the lasso and related procedures in model selection.* Statistica Sinica: 1273-1284.
- Li, J.M & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting, 30,* 996-1015.
- Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the lasso. *Annals of statistics, 42,* 413.
- Mahajan, V., Jain, A.K., & Bergier, M. (1977). Parameter estimation in marketing models in the presence of multicollinearity: An application of ridge regression. *Journal of Marketing Research, 14,* 586-591.
- Matsui, H., & Konishi, S. (2011). Variable selection for functional regression models via the L1 regularization. *Computational Statistics & Data Analysis, 55,* 3304-3310.
- Meier, L., S. Van De Geer, & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70,* 53-71.
- Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics, 34,* 1436-1462.
- Osborne, M.R., Presnell, B., & Turlach, B. A. (2000). On the lasso and its dual. *Journal of Computational and Graphical statistics, 9,* 319-337.
- Oyeyemi, G.M., Ogunjobi, E.O., & Folorunsho, A.I. (2015). On performance of shrinkage methods–a Monte Carlo Study. *International Journal of Statistics and Applications, 5,* 72-76.
- Pötscher, B.M, & Schneider, U. (2009). On the distribution of the adaptive LASSO estimator. *Journal of Statistical Planning and Inference, 139,* 2775-2790.
- Ren, Y., & Zhang, X. (2010). Subset selection for vector autoregressive processes via adaptive Lasso. *Statistics & probability letters, 80,* 1705-1712.
- Roth, V. (2004). The generalized LASSO. *IEEE transactions on neural networks, 15,* 16-28.
- Roy, S.S., Mittal, D., Basu, A., & Abraham, A. (2015). *Stock market forecasting using LASSO linear regression model.* In: Afro-European Conference for Industrial Advancement. Springer: pp: 371-381.
- Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association, 88,* 486-494.
- Simon, N., Friedman, J., Hastie T., & Tibshirani, R. (2013). A sparse-group lasso. *Journal of computational and graphical statistics, 22,* 231-245.
- Smyth, P. (1996). *Clustering Using Monte Carlo Cross-Validation.* In: Kdd. pp: 26-133.
- So, I. (2009). Comparison of criteria for estimating the order of autoregressive process: a Monte Carlo approach. *European Journal of Scientific Research, 30,* 409-416.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58,* 267-288.
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in medicine, 16,* 385-395.
- Tibshirani, R.J., & Taylor, J. (2011). The solution path of the generalized lasso. *The Annals of Statistics, 39,* 1335-1371.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73,* 273-282.
- Tibshirani, R.J., & Taylor. J. (2012). Degrees of freedom in lasso problems. *The Annals of Statistics, 40,* 1198-1232.
- Tibshirani, R.J. (2013). The lasso problem and uniqueness. *Electronic Journal of statistics, 7,* 1456-1490.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67,* 91-108.

- Taylor, J., Lockhart, R., Tibshirani, R.J., & Tibshirani, R. (2014). *Post-selection adaptive inference for least angle regression and the lasso.* arXiv preprint arXiv:1401.3889, 354.
- Van de Geer, S.A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics, 36*, 614-645.
- Wagener, J., & Dette, H. (2013). The adaptive lasso in high-dimensional sparse heteroscedastic models. *Mathematical Methods of Statistics, 22,* 137-154.
- Waldmann, P., Ferenčaković, M., Mészáros, G., Khayatzadeh, N., Curik, I., & Sölkner, J. (2019). AUTALASSO: an automatic adaptive LASSO for genome-wide prediction. *BMC bioinformatics, 20,*1-10.
- Wang, H., & Leng, C. (2008). A note on adaptive group lasso. *Computational statistics & data analysis, 52,* 5277-5286.
- Wang, H., & Leng, C. (2007). Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association, 102*, 1039-1048.
- Wang, H., Li, R., & Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika, 94*, 553-568.
- Wang, H., Li, B., & Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71,* 671-683
- Wang, S., Nan, B., Rosset, S., & Zhu, J. (2011). Random lasso. *The annals of applied statistics, 5,* 468.
- Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research, 77,* 220-225.
- Wu, T.T., & Lange, K. (2008). Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics, 2,* 224-244.
- Xu, H., Caramanis, C.,& Mannor, S. (2009). *Robust regression and lasso. In: Advances in neural information processing systems.* pp: 1801-1808.
- Xue, J., Luo, Y., & Liang, F. (2017). *Average (E) BIC-like criteria for Bayesian model selection.* Work Pap [Internet].
- 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*, 49-67.
- Yüzbası, B., Ahmed, S.E., & Güngör. S. (2017). Improved penalty strategies in linear regression models. *REVSTAT–Statistical Journal, 15,* 251-276.
- Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine learning research, 7,* 2541-2563.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association, 101*, 1418-1429.