The Republic of Iraq
Ministry of Higher Education &
Scientific Research
University of Al-Qadisiyah
College of Administration & Economics
Department of Statistics- Higher studies

Bayesian group bridge composite quantile regression

A Thesis by

Mayada Jawad Kadhim Al-Jasimee

Submitted to the Council of the College of Administration & Economics / University of Al-Qadisiyah in partial fulfilment of the requirements for a Master Degree (MSc) in Statistical sciences

Supervised by

Prof. Dr. Rahim J. Al-Hamzawi

2022 AD **1444 AH**

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

قَالُوا سُبْحَانَكَ
لَا عِلْمَ لَنَا إِلَّا مَا عَلَّمْتَنَا
إِنَّكَ أَنْتَ الْعَلِيمُ الْحَكِيمُ

صَدَقَ اللَّهُ الْعَلِيُّ الْعَظِيمُ

سورة البقرة الآية (32)

Acknowledgments

In the Name of Allah, the Merciful, the Most Merciful

Praise be to Allah, Lord of the Worlds,

we seek His help and repent to Him. Praise be to Allah, first for having guided us, lighted our insight, illuminated our minds, and facilitated all ways for us to complete this humble thesis. I am thankful to Allah, who has always been with me whenever I face difficulty, wherever I find it. Here I am walking according to what our Lord commanded us when He said (Read).

I would like to extend my thanks who honored me by supervising my thesis… Prof. Dr. Rahim J. Al-Hamzawi, and showed me the right ways to achieve my aim. Through his great patience and useful scientific guidance, which contributed significantly to the completion of this thesis, I hope Allah Almighty will guide him to all that he wishes and requests and to what is good and always progress for the better. Also, many thanks are to my respected teachers who enlightened me through science and knowledge until I became what I am now: Prof. Dr. Mohammed Al-Sharoot, Prof. Dr. Ahmad Naeem, Prof. Dr. Tahir Dikheel, Prof. Dr. Ali Al-kenani, Dr. Muhannad Al-Saadon, Dr. Hassan Sami, Dr. Taha Alshaybawe, Dr. Fadil Alhusseini, Dr. Bahar Kadhim, Dr. Haider Hadi, Dr. Asaad Nasser, Saif Hossam, Afra Abbas, and Hamida Naim. I would also like to thank Mr. Chairman and the members of the discussion committee who will have the greatest impact in evaluating this thesis to show its final form.

I thank my dear parents, who stood by my side in all stages of my life and made my dreams come true. My father, may Allah have mercy on you, and I hope that you are satisfied with me, and I hope it will be good and charity for them. My mother, who enriched me with her prayer hymns, I hope that I have fulfilled even a small amount of fulfillment for them. I would like to thank my brothers. I thank Allah in my prostration for this blessing. My Lord, preserve them for me. I would like to thank my sister, who provided me with advice in this world all my life. I thank Allah Almighty for giving me an ideal sister like her. I would like to thank my soul, the look in my eyes, and my beating heart, my dear daughters.

I would like to extend my thanks and respect to all my loyal friends. Finally, I thank everyone who supported me with a prayer for me and helped me to complete this thesis. I pray to Allah for all the students of the Statistics Department's success and to make this thesis a science that benefits everyone. He is the All-Hearing.

Mayada

DEDICATION

To Allah, my creator, so that my deeds may benefit me when I meet him.

To the beloved, the chosen one, the best of God's creation, Muhammad (may Allah bless him and his family and grant them peace) and to the family of Muhammad (peace be upon them). To the Qa'im of the family of Muhammad (may Allah hasten his reappearance). To the righteous servants of Allah.

To those who instilled in us the love of science and ambition, my respected teachers.

To my generous teacher and my supervisor, Prof. Dr. Rahim J. Al-Hamzawi, who whenever I asked him about the knowledge, he gave it to me. He gave me his precious time, despite his many preoccupations.

To the first who saw my eyes, with their satisfaction, the Most Merciful commanded us, and the chosen one recommended us to them. To my first school in life... To the soul that will never leave me, but flutters in the sky of my life… To the person who makes me proud to be his daughter… My Father. To my heaven in this world… To the one who still raises her hands to pray for me... To the one who I was relieved whenever I remembered her smile in my face… To the wellspring of mercy and kindness…To the angel of my heart and the crown of my head… My mother, may Allah keep her safe and alive. Without them, my dreams would not have come true, so they deserve all the credit and gratitude.

To my dear brothers… To my eyes and my soul mate, my sister. To whom I look forward to seeing their bright future, Allah willing… To whom I continue to come forward for them, my beloved daughters, for their continuous support and always being with me.

To those who stayed up nights and challenged the difficulties in order to advance our beloved Iraq... To my brothers and sisters, students of the Statistics Department. To all my loyal friends and everyone who prayed for my success. I dedicate the fruit of my humble effort and I pray to

Allah to obtain your satisfaction. **Mayada**

Abstract

The high dimensional problem has appeared in many regression applications, i.e., the number of covariates is greater than the sample size. In this case, the Ordinary Least Squares (OLS) method estimates are not stable, as well as having high variance and high bias, which leads to overfitting, multicollinearity of the estimation of the model parameters, very poor prediction, and difficulty interpreting an appropriate model. The traditional statistical methods with this problem become not possible to use for statistical analysis. Thus, the difficulty of estimating coefficients and selecting the important covariates (the covariates that affect on the dependent variable).

We compare the performance of two regularization approaches In this thesis: the least absolute shrinkage and selection operator (lasso) and the reciprocal lasso (rlasso). Also, we propose a new method for removing unimportant covariates in high dimensional data to improve the prediction accuracy and obtain better interpretation. This method is called Bayesian group bridge composite quantile regression (BgBCQR). Specifically, we improve the hierarchical model for the suggested method. We introduce a new Markov Chain Monte Carlo (MCMC) algorithm for posterior inference employing a scale mixture of normals of the asymmetric Laplace distribution (ALD) to carry out the hierarchical Bayesian for the suggested method. We compare our proposed method with other regularization methods to verify the effectiveness of the proposed method through conducting a study of simulation examples as well as in a real data application to compare the performance of these regularization methods.

Simulation results and analyses of real data show that the performance of the proposed method is more efficient and outperforms the current

approaches in terms of prediction accuracy, variable selection (VS) and the estimation of coefficients. Also, it provides a clear interpretation.

List of Abbreviations

Abbreviations	Meaning
OLS	Ordinary least Squares
lasso	least absolute shrinkage and selection operator
rlasso	reciprocal lasso
BgBCQR	Bayesian group bridge composite quantile regression
MCMC	Markov Chain Monte Carlo
ALD	asymmetric Laplace distribution
VS	Variable Selection
FS	Forward Selection
BE	Backward Elimination
C_p	Mallows' C_p criterion
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
DIC	Deviance Information Criteria
SSVS	Stochastic Search Variable Selection
SCAD	smoothly clipped absolute deviation
LARS	the least angle regression
alasso	adaptive least absolute shrinkage and selection operator
MCP	minimax concave penalty
Blasso	Bayesian lasso
Balasso	Bayesian adaptive lasso

SMU	scale mixture of uniform
SMTN	scale mixture of truncated normal
QR	Quantile Regression
CQR	composite QR
MLE	Maximum Likelihood Estimation
Brlasso	Bayesian reciprocal lasso
ARS	Adaptive Rejection Sampling
MSE	Mean squared errors
SD	standard deviations of MAD
BgBR	Bayesian group bridge regression
gBR	group bridge regression
gLR	group lasso regression
MMAD	Median of mean absolute deviations
MMSE	mean squared prediction errors

Contents

Contents

Contents

Figures

Contents

Table

Chapter One

Introduction, Thesis problem, The objective of the thesis and Literature Review

1.1 Introduction

Linear models are statistical models with wide application in economics and engineering, as well as in various sciences such as agricultural, physical, medical, and social sciences.

Best subset selection, which aims to exhaustively search through all 2^p possible models and select the best one (p is a number of covariates), is a natural idea for producing a good model with a reduced number of covariates. In practice, just a small subset of possible covariates has an effect on the response variable, whereas some covariates are not important (not effective or with very little effect), so not important covariates should be removed from the model (without simultaneously losing a lot of information). Regression analysis in some applications becomes very difficult when the number of covariates is large.

Choosing the influencing covariates correctly is a difficult and important problem in any scientific research because the misspecification of a model can have a significant influence on a scientific outcome. So the main goal in many real data analysis studies is to find the best model for the data with the lowest cost and time.

1.2 Thesis problem

The high dimensional problem has appeared in many regression applications, i.e., the number of covariates is bigger than the sample size $(p > n)$, where n the sample size. The covariates selection is a critical issue in the building of a regression model. Selecting a suitable subset of

covariates may typically increase prediction accuracy. The traditional statistical methods with this problem become not possible to use for statistical analysis. Thus, the difficulty of estimating coefficients and selecting the important covariates. So, the high dimensional problem may result in extremely complex models.

1.3 The objective of the thesis

The main objective of this thesis is to propose a new method for removing unimportant covariates in high dimensional data to improve prediction accuracy and obtain better interpretation.

1.4 Literature review

In some multiple regression applications, the number of covariates has become large, where the matrix X does not have the full rank and $(X'X)^{-1}$ cannot be calculated, which leads to overfitting and multicollinearity problem, making data analysis difficult. It is necessary to perform dimensionality shrinkage of data to address this problem. One of the methods to reduce high dimensional data is the Variable Selection method.

One of the most essential goals of regression analysis is the selection of relevant variables. Various methods for dealing with VS in high dimensional linear models have been developed through the years to obtain a model with the fewest important covariates, high prediction accuracy and ease of interpretation of the model as well as providing the model with low cost (Guyon and Elisseeff, 2003).

There are two types of methods for the VS process:

1.4.1. Classical model selection methods for linear models

There are many widely used approaches for model selection in statistical tradition are backward, forward and stepwise selection.

Efroymson (1960) introduced the stepwise method as a VS procedure. It is essentially a modification method for Forward Selection (FS) and Backward Elimination (BE) methods that combined the mechanisms of both (FS) and (BE) procedures. The calculation of the stepwise method depends on the inclusion and deletion of covariates. (see, James et al., 2013) for more details.

Traditional approaches, such as stepwise selection, fail short of one or more of the above. Additionally, it neglects stochastic errors in the VS procedure and can be computationally costly (Fan and Li, 2001). Therefore, several approaches have been proposed to address these problems.

Mallows (1973) proposed the Mallow's C_p criterion, which was used to choose a better model that contains a subset of important covariates. It is noted that C_p is inconsistent in large samples and showed that C_p is a careful model selector, which tends to overfit (Woodroofe, 1982). Also, Nishii (1984) showed that C_p is inconsistent in selecting the correct model, and often selects a larger model when $n \to \infty$.

Akaike (1974) proposed the Akaike information criterion (AIC) is one of the most common criteria used for model selection that gives the most accurate description of the data. Nishii (1984) showed that the AIC provides an inconsistent model. As a result, the model selection by the AIC is

inconsistent when the sample size (n) is big (Dziak et al., 2005; Javed and Mantalos, 2013). Also, the AIC is weak in selecting the best model when the sample size (n) is small (Dziak et al., 2005). See (Burnham and Anderson, 2004) for more information about AIC.

Schwarz (1978) proposed the Bayesian Information Criteria (BIC) to solve the problem in AIC. The BIC differs from the AIC in that it takes into account the sample size (n), making it more efficient than the AIC. The BIC is a consistent model selection method when the sample size (n) is large enough. On the other hand, it is known that both the AIC and the BIC do not work better all the time.

Spiegelhalter et al. (2002) suggested the generalization of AIC and BIC for model selection in Bayesian hierarchical normal linear models, is Deviance Information Criteria (DIC). Ando (2007) has addressed DIC's bias for selecting over-fitted models, despite very little is known about its performance in high dimensional models. It is based on the posterior log-likelihood distribution or the deviance; on the other hand, one important drawback is that they are not well-defined when using incorrect priors (Berg and Meyer, 2004). Also, other problems have been noted with DIC, according to Gelman et al. (2007), but no consensus on a replacement has appeared.

George and McCulloch (1993) introduced the Stochastic Search Variable Selection (SSVS) approach, which is a traditional Bayesian variable selection method. It's a procedure for selecting a subset of covariates based on a mixture prior distribution that allows several coefficients equal to zero.

SSVS is using MCMC sampling, to sample indirectly from this posterior distribution on the set of all possible subset selections. Subsets that have a

larger posterior probability are determined by their more frequent occur in the MCMC sample.

1.4.2. Regularization methods

Regularization approaches may be defined as an approach to addressing the problem of model complexity by penalizing models of higher complexity. Shrinkage regression methods (also known as regularization methods or penalized likelihood) have been proposed recently to address the overfitting issue in high dimensional linear models. Thus, VS and the coefficient estimate may be done at the same time.

The model with high complexity has low bias and high variance, but the low complexity model has high bias and low variance. As a result, regularization approaches are frequently used to control the model complexity.

Hoerl and Kennard (1970) proposed the method of Ridge regression. It is an approach for estimating the linear regression coefficients vector (β) based on adding small positive values to the diagonal of X´X (ℓ_2-norm) in order to get biased estimations with less mean square error. However, ridge estimators perform poorly when real coefficient sizes differ greatly (Jolliffe, 1982).

Frank and Friedman (1993) proposed that Bridge regression is a large class of penalized regression. It has attractive properties such as impartiality and Oracle, as well as the VS and coefficients estimation of the model. However, the convergent covariance matrix and bootstrap studied standard errors are unstable.

Tibshirani (1996) developed the method of VS by proposing a new method of estimation in linear models. It is the least absolute shrinkage and selection operator (lasso) method by adding a penalty function (ℓ_1 norm) to the least squares loss function that puts the coefficients of unimportant covariates equal to zero. Thus, VS can be achieved automatically, as can obtaining interpretable models.

The lasso penalty function has also been widely used in many statistical applications (see for example; Zheng, 2008; Nardi and Rinaldo, 2011; Bien et al., 2013; Wu et al., 2014; Kaul, 2014).

Fan and Li (2001) introduced a new approach to regularization known as smoothly clipped absolute deviation (SCAD). It is a particularly important method due to its computational features. SCAD estimated has the Oracle property if the penalization parameter is chosen correctly.

Over the years, several most computationally efficient algorithms have been suggested, for example, the least angle regression (LARS, Efron et al., 2004) and the coordinate descent algorithm (Friedman et al., 2010) to select a linear model based on the same set of data that will be used to apply the model. These algorithms are suggested since the lasso methods' estimations of regression coefficients are not analytically derivable due to the ℓ_1-penalty term is not differentiable.

Zou and Hastie (2005) Suggested the elastic net regression model is another regularization regression and VS method. It is a type of linear regression regularized improvement. It may be thought of as a VS approach that works as VS and shrinkage method at the same time to obtain better results in situations when the number of covariates (p) is larger than the

sample size (n), but the lasso method is not a good variable selection method in the $p > n$ case. Also, The elastic net is very useful when there is grouping among covariates and multicollinearity problem.

In an elastic net, the penalty term is an addition the ridge penalty and the lasso penalty to the least squares loss function. In addition, unlike the lasso, the elastic net deals the grouping effect of correlated covariates well, where these covariates are either in or out of the model at the same time.

The elastic net estimator may be thought of as a more stable version of the lasso. The researchers demonstrated that the elastic net method outperforms the lasso, especially when there are groups of covariates that are highly correlated. However, it complicated that requires to the high calculation cost.

Tibshirani et al.(2005) proposed the fused lasso as a method of regularization.

Zou (2006) proved that the lasso estimator is inconsistent in VS since the lasso penalizes all coefficients equally. The adaptive least absolute shrinkage and selection operator (alasso) method controls the lasso estimate's bias by adding adaptive weights, which are employed to penalize different coefficients in the lasso method. Thus, the coefficients of unimportant covariates are reduced to 0 more efficiently. This method minimizes bias and improves VS accuracy, which creates estimates that are consistent and unbiased, as well as performs a better job of estimating significant coefficients than lasso (Wang et al. 2007; Zou 2006).

The alasso estimator has been widely used in many statistical applications due to its good theoretical properties. See for example (Zhang and Lu, 2007; Zeng et al., 2014; Yang and Wu, 2016). However, it needs consistent initial

estimations of the regression coefficients, which are commonly unavailable in the high dimension, small sample size setting.

Yuan and Lin (2006) proposed the group lasso as a new regularization method, which Kim et al. (2006) extended to include general loss functions. The group lasso method considers the issue of choosing grouped covariates for accurate prediction in regression, which means that the basic idea is that important covariates in this group should be selected at the group level, but at the same time, unimportant covariates cannot be entirely eliminated since they determine covariates in the same group. The grouped lasso is useful when there are meaningful groups of covariates, such as polynomial regression.

This method has the Oracle property. Yaun and Lin (2006) also demonstrated that lasso cannot identify the effects of grouped covariates.

Shimamura et al. (2006) discussed the issue of selecting the best penalty parameter from a group of possible values in the group lasso presented by Yuan and Lin (2006) since it affects the prediction accuracy of the fitted model.

Meinshausen (2007) proposed a new regularization approach for controlling the bias of the lasso parameter. This new approach is known as the relaxed lasso method. All regular lasso solutions are included in the relaxed lasso solutions. Thus, calculating all relaxed lasso solutions is often as costly as computing all regular lasso solutions.

Wang and Leng (2008) proposed the penalty function for the adaptive group lasso method to address the problems of the group lasso method, where the group lasso method selects covariates in a grouped manner.

However, it suffers from inefficient estimation and inconsistent selection. The two researchers theoretically demonstrated that the new method can consistently determine the real model, and the resulting estimator can be just as efficient as Oracle.

Zou and Zhang (2009) proposed a new regularization method is adaptive elastic net regularization method that focused on the limitations of the lasso method in dealing with the presence of grouped covariates and the inconsistency of estimators.

Kang and Guo (2009) suggested a self-adaptive lasso approach for VS and parameter estimation at the same time. They also developed an effective Gibbs sampling approach to choose tuning parameters and estimate regression coefficients automatically.

Zhang (2010) proposed a minimax concave penalty (MCP) as a quick, continuous, almost unbiased, and accurate approach of penalized VS in high dimensional linear regression. The lasso is quick and continuous, but it is biased. The lasso's bias might preclude consistent VS. Subset selection is unbiased, but it is computationally expensive.

Simon and Tibshirani (2012) proposed the penalty function for the standardized group lasso method, and they showed the effectiveness of the proposed method and its preference over the usual group lasso method by analyzing real data and simulating a set of examples.

Song (2014) was the first to study rlasso estimators that have the oracle property.

The rlasso approach suggested by Song and Liang (2015), as well as Song (2018) for VS and the coefficients estimate together, which is based on a

novel class of penalty functions that are discontinuous at 0, decreasing in $(0,\infty)$ and also give near 0 coefficients of infinity penalties, while the traditional penalty functions are close to 0 penalties when the coefficients are close to 0 (for example, lasso and SCAD) or constant penalties (for example, ℓ_0 - penalty). Because of this characteristic property, rlasso is highly desirable for selecting models. It is able to successfully avoiding model selection that are too dense.

Among the new approaches that are based on regularization and selection operator are dantzig selector (Candes and Tao, 2007), and matrix completion (Cand`es and Tao, 2010; Mazumder et al., 2011), among others.

Similarly, from a Bayesian viewpoint, several researchers have used the Bayesian process in their approaches.

When identical and independent Laplace prior is assigned to every coefficient of regression, the estimations of lasso may be interpreted as a Bayesian posterior mode estimation (Tibshirani, 1996). In consequence, Park and Casella (2008) proposed the Bayesian lasso (Blasso) for models of linear regression, which uses the Laplace prior like a mixture of exponential and normal priors.

After Park and Casella (2008), more Blasso techniques have been suggested by researchers through the years (see, Yi and Xu, 2008; Huang et al., 2008; Brown and Griffin 2010; Li et al., 2011; Legarra et al., 2011; Cai et al., 2011). Also, Blasso regression was introduced by Hans (2009).

The Bayesian alasso (Balasso) and the iterative alasso were developed by

Sun et al. (2010) using different adaptive weights and updating these adaptive weights iteratively. The iterative alasso approach is also much more computationally efficient than the widely utilized stepwise regression and marginal regression approaches.

Li and Lin (2010) proposed a Bayesian elastic net to solve the elastic net model utilizing a Gibbs sampler, whereas the marginal posterior mode of the regression coefficients is equivalent to non-Bayesian elastic net estimates. The two penalty parameters are selected together in this proposed method, solving the "double shrinkage problem" of the elastic net method.

Over the years, different methods to the Bayesian elastic net have been suggested (see for example; Bornn et al., 2010; Alhamzawi, 2014; Huang et al., 2015).

Chen et al. (2011) proposed a new full hierarchical Bayesian version of the lasso model by improving a reversible - jump the MCMC algorithm for joint posterior inference to get the Blasso estimation.

Hans (2011) proposed a new Gibbs sampler algorithm for calculating Bayesian estimates utilizing the elastic net approach, where the estimate that results from the elastic net method may be considered as a Bayesian posterior mode under a prior distribution estimated by the elastic net penalty form.

Malik and Yi (2014) followed Park and Casella (2008) and developed a new Bayesian lasso, where Malik and Yi (2014) introduced a new hierarchical formulation of the Blasso by using the scale mixture of uniform (SMU) representation of the Laplace density. As well as, They suggested a new Gibbs sampler for the Blasso. The suggested approach outperformed

compared with the Blasso approach.

Leng et al. (2014) presented the Balasso method for VS as well as a coefficient estimate in linear regression. Furthermore, Prompted through the lasso's hierarchical Bayesian interpretation, they gave a model selection mechanism for the Balasso through evaluating the posterior conditional mode estimations.

in 2018, Alhamzawi and Ali introduced several Bayesian methods for obtaining the alasso as well as relevant estimators to address the drawbacks of the traditional alasso method. They took into account a fully Bayesian approach treating to the alasso, which leads to a novel Gibbs sampler with tractable full conditional posteriors. They used a scale mixture of truncated normal (SMTN) representation of the Laplace density to propose a novel expanded hierarchy representation of Balasso.

Flaih et al. (2020) introduced a new Bayesian lasso, where a new hierarchical model is developed by employing a new scale mixture of the Laplace distribution, which is a mixture of normal mixing and Rayleigh distribution. Thus, a novel Gibbs sample algorithm was improved to compute the mode of the posterior density of the lasso regression model parameter.

Mallick et al. (2021) considered a fully Bayesian approach to of the rlasso issue, based on the observation that whenever the parameters of regression are allocated independent inverse Laplace priors, the rlasso estimation for linear regression parameters may be interpreted as a Bayesian posterior mode estimation.

Koenker and Basset (1978) introduced the Quantile Regression (QR) model to assess effects of covariates on outcome variable at various quantile levels, where Quantile Regression (QR) has acquired growing popularity since it makes few assumptions about the error distribution. For the θth quantile ($0 < \theta < 1$). Compared to standard mean regression, QR is more strong and efficient to data outliers than standard mean regression.

Koenker and Dorey (1987) developed and improved an algorithm for the least absolute error estimation of linear regression to calculate the quantile regression statistics of Koenker and Bassett (1978).

Li and Zhu (2008) proposed the lasso quantile regression as a regularization approach for coefficient estimate and VS. They take into account the ℓ_1-penalty (lasso) shrank quantile regression, which employs the sum of the absolute values of the coefficients as the penalty.

Zou and Yuan (2008) indicated that QR may result tiny relative efficiency when compared with the standard mean regression. Since, QR at one quantile can provide more efficient estimators than QR at another quantile, Zou and Yuan (2008) suggested a composite QR (CQR) approach to simultaneously study multiple QR models. They proved that, irrespective of the error distribution, the relative efficiency of the CQR estimator is higher than 70% when compared to the mean regression estimator.

Huang et al. (2009) suggest a group bridge method capable of selecting at both the group and within-group individual covariate levels. The suggested method is a penalized regularization approach that employs a group bridge penalty. It has the property of Oracle group selection.

Similarly, from a Bayesian viewpoint, several researchers have used the Bayesian process in their approaches to quantile regression.

Yu and Moyeed (2001) introduced Bayesian quantile regression, which makes use of a likelihood function based on the asymmetric Laplace distribution. They have also demonstrated that using the asymmetric Laplace distribution is a highly natural and effective method for modeling Bayesian quantile regression.

Li et al. (2010) studied regularization in quantile regression from a Bayesian point of view. By introducing a hierarchical model framework, they gave general treatment to lasso, elastic net and group lasso penalties.

Kozumi and Kobayashi (2011) considered quantile regression models that used an asymmetric Laplace distribution (ALD) from a Bayesian standpoint. They improved an efficient Gibbs sampling algorithm for Bayesian QR by assuming that the random variable follows the ALD.

Huang and Chen (2015) studied composite quantile regression from a Bayesian standpoint through using the ALD for the errors. In the literature, composite quantile regression approaches that are robust to heavy-tailed errors or outliers in response have been presented.

Alhamzawi (2016) presented a Bayesian method for composite quantile regression using the skewed Laplace distribution for the error distribution. An effective Gibbs sampling algorithm is improved to modify the unknown quantities from the posteriors.

Mallick and Yi (2017) developed the Bayesian group bridge to choose bi-

level variables for regularized regression. As an alternate to frequentist group VS approaches, the Bayesian group bridge combines structural information among covariates by a group-wise shrinkage prior.

Mallick and Yi (2018) suggested bridge regression from a Bayesian point of view because the bridge regularization lacks a systematic method of inference, although the bridge regularization has several desired statistical properties such as unbiasedness, sparsity, and Oracle. The suggested Bayesian approach gave regression parameters uncertainty estimates, allowing for consistent inference using the posterior distribution.

The remainder of this thesis is organized as follows: In chapter two, we introduce variable selection methods for linear models that include classical model selection methods and regularization methods. We introduce the CQR with the group bridge penalty in chapter three. We also outline the Bayesian sampler algorithm for CQR. In Chapter four, we run examples of simulation to demonstrate the performance of lasso and rlasso approaches as well as to investigate the performance of the suggested approach, and we investigate the performance of the lasso method and the rlasso method using wheat crop production rate data as well as explain our approach employing the prostate cancer data in Chapter five. Finally, in Chapter six, conclusions and future research were reported.

Chapter Two

Variable selection methods for linear models

2. Variable selection methods for linear models

2.1. Classical model selection methods

One important problem in building a linear regression model is selection of covariates. Thus, a large amount of work has been done on the topic of model selection. See, for example, backward, forward, stepwise selection, Mallow's C_P, Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), Deviance Information Criteria (DIC) and Stochastic Search Variable Selection (SSVS). Actually, the prediction accuracy can be improved by removing unimportant covariates. In this chapter, we briefly discuss Mallow's C_P, AIC, BIC, DIC and SSVS.

2.1.1 Mallow's C_p

Mallow's C_p criterion proposed by Mallows (1973), was used to select a better model that contains a subset of important covariates. The formula of Mallow's C_p procedure is:

$$C_p = \frac{RSS(K)}{S^2} - n + 2p \quad , \tag{1}$$

where $RSS(k)$ is a residuals sum of squares for the subset model involving k covariates, n is the number of observations. S^2 is the mean squared error of the model. p is a number of covariates. In practice, the parsimonious mode is the mode with C_p close to p, in the sense of minimizing the total bias of the predicted values. It is noted that C_p is inconsistent in large samples and showed that C_p is a careful model selector, which tends to overfit

(Woodroofe, 1982). Also, Nishii (1984) showed that C_p is inconsistent in selecting the correct model, and often selects a larger model when $n \to \infty$.

2.1.2 Akaike's Information Criteria (AIC)

One of the most common criteria used for model selection that gives the most accurate description of the data is the Akaike information criterion (AIC), that proposed by Akaike (1974). AIC can be written as

$$AIC=-2 \log L+2p \,, \tag{2}$$

where L be the maximum likelihood estimation function (MLE). The best model among a set of available models is the one with the lowest AIC value. The same Mallow's C_p. Nishii (1984) showed that the AIC provides a model that is inconsistent. As a result, the model selection by the AIC is inconsistent when the sample size (n) is big (Dziak et al., 2005; Javed and Mantalos, 2013). Also, the AIC is weak in selecting the best model when the sample size (n) is small (Dziak et al., 2005).

2.1.3 Bayesian Information Criteria (BIC)

To solve the problem in (AIC), Schwarz (1978) proposed the Bayesian Information Criteria (BIC). It is characterized by its computational simplicity in a variety of modeling frameworks. The BIC is defined as

$$BIC = -2 \log L + p \log n \,, \tag{3}$$

in the above criteria, we can see that the BIC difference from the AIC by taking into consideration the sample size, making it more efficient than the

AIC. However, the same the AIC, the model with the lowest BIC value is the best among a set of candidate models. The BIC is a consistent model selection method when the sample size (n) is large enough. On the other hand, it is known that both the AIC and the BIC do not work better all the time.

2.1.4 Deviance Information Criteria (DIC)

Spiegelhalter et al. (2002) suggested the generalization of AIC and BIC for model selection in Bayesian hierarchical normal linear models defined as

$$DIC = -2 \log L - 4p \log L \ , \tag{4}$$

the DIC is especially useful when MCMC samples are readily available, and it is only effective when the parameters' joint distribution is approximately multivariate normal. Similar to AIC and BIC, the best model with the lowest DIC value. Also, Ando (2007) has addressed DIC's bias for selecting over-fitted models, despite very little is known about its performance in high dimensional models. DIC is simple to calculate and can be used in a variety of statistical models. It is based on the posterior log-likelihood distribution or the deviance; on the other hand, one important drawback is that they are not well-defined when using incorrect priors (Berg and Meyer, 2004). Also, other problems have been noted with DIC, according to Gelman et al. (2007), but no consensus on a replacement has appeared.

2.1.5 Stochastic Search Variable Selection (SSVS)

Stochastic Search Variable Selection (SSVS) method is a conventional Bayesian variable selection methods which is proposed by George and McCulloch (1993). It's a procedure for selecting a subset of covariates based on a mixture prior distribution that allows several coefficients equal to zero.

SSVS is using MCMC sampling, to sample indirectly from this posterior distribution on the set of all possible subset selections. Subsets that have a larger posterior probability are determined by their more frequent occur in the MCMC sample.

2.2. Regularization methods

2.2.1 Regularization regression using lasso

Suppose that model of the multiple linear regression is defined as

$$\mathbf{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{5}$$

where $\mathbf{y} = (y_1,\ldots, y_n)'$ is the vector of response, $\mathrm{X} = (\mathbf{x}_1,\ldots, \mathbf{x}_p)$ is the matrix of covariates, $\boldsymbol{\beta} = (\beta_1,\ldots ,\beta_p)'$ a regression coefficient vector, and $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots, \varepsilon_n)'$ a random errors vector where the error distribution $\varepsilon_i \sim N(0, \sigma^2)$. The regression coefficients $\boldsymbol{\beta}$ can be estimated by minimizing

$$\min_{\boldsymbol{\beta}} \left\| y - \sum_{k=1}^{p} x_k \ \beta_k \right\|_2^2 = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}), \tag{6}$$

problem (6) leads to poor prediction performance, overfitting, and difficulty interpreting an appropriate model of least squares estimation when the number of covariates is large. Therefore, using the Ordinary Least Squares

(OLS) method with the above problems produced unstable and high variance estimates. These problems are the main reasons behind the use of shrinkage and subset selection approaches. To ease model interpretation and improve prediction accuracy, it is usually necessary to conduct a VS, so that a parsimonious regression model is built by just using a few important covariates.

Regularization is a penalized regression method which is used for the regularization process for VS and the estimate of coefficients together in issues of regression. It is the least absolute shrinkage and selection operator (lasso). The lasso method was suggested by Tibshirani (1996) to solve the overfitting problem when there are many studies have many more covariates than the sample size $(p > n)$ through the addition of a penalty function (ℓ_1 norm) to the least squares loss function, which puts the coefficients of unimportant covariates equal to zero. Thus, VS can be achieved automatically. The lasso estimator is obtained as follows:

$$\hat{\beta}_{lasso} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} |\beta_k|, \qquad (7)$$

where $\lambda \geq 0$, λ is the regularization (tuning or shrinkage) parameter controlling the quantity of penalty, the highest value of λ gives a highest level of shrinkage (Alkenani and Yu, 2013). Small penalties result in the selection of big models having possibly high variance but less bias; big penalties result in the selecting models having less covariates with lower variance. The lasso regression has some notable properties:

1. It can solve the problem of multicollinearity.
2. It reduces the prediction error of the model by putting the coefficients of unimportant covariates equal to zero (Ranstam and Cook, 2018).

3. It deals with regression models that have a large number of covariates (p > n).

The lasso penalty function has also been widely used in many statistical applications (Zheng, 2008; Nardi and Rinaldo, 2011; Bien et al., 2013; Wu et al., 2014; Kaul, 2014).

Although these are the attractive advantages of lasso and it has proven successful in a variety of situations, lasso has some problems that occur in the lasso method's work and it doesn't work well in certain of the following cases:

1. In practice, there are many studies that $p > n$, but more covariates than the sample size cannot be chosen by the lasso method (Zou and Hastie, 2005).

2. When there is a set of covariates that are highly correlated, the lasso will select just one of the group while ignoring the others (Zou and Hastie, 2005).

3. In the case of $(n > p)$ and when the covariates are highly correlated, as shown by experiments, the lasso regression does not choose covariates correctly (Zou and Hastie, 2005).

Although lasso performs well in putting the coefficients of unimportant covariates equal to zero, it has some drawbacks. Empirically, lasso chooses more covariates than required. The lasso approach has a bias for the estimate of large coefficients, indicating that the lasso approach is inconsistent since this approach penalizes all coefficients equally.

Consequently, lasso doesn't have oracle properties (the definition of

oracle properties: according to Fan and Li (2001), the method with this property can select the true model with a probability of 1 quantity when $n \to \infty$).

The lasso can be calculated using the fast algorithm which is available in the statistical program (R) called the least angle regression (LARS, Efron et al., 2004). It is the most efficient algorithm in terms of computation and is extremely fast. As a result, making penalized regression models highly popular in high dimensional data analysis.

2.2.2 Bayesian interpretation of lasso

Bayesian regularization methods were developed for difficulty obtaining statistical inference on the regression coefficients. On the other hand, a Bayesian method provides an exact inference even with a small sample size, as well as an exact estimate when p is bigger than n ($p > n$, Alhamzawi and Ali, 2018b; Li et al., 2010).

There are two steps in Bayes regularization models: First, determining the prior distribution of regression coefficients, which is the most important step in the Bayes method for VS and estimation of coefficients together, is the main idea in Bayesian analysis minimize estimator variance while increasing bias. Therefore, the choice of the prior distribution must be exact because choosing an inaccurate or incorrect prior distribution without caution will lead to many problems, including Gibbs sample convergence issues and posterior estimation instabilities (Alhamzawi and Yu, 2012). Second, computing the posterior distribution (Agresti, 2010).

Park and Casella (2008) noted the penalty term's form (7). In 1996, Tibshirani proposed that when the coefficients of regression possess identical and independent Laplace (i.e., double-exponential) , estimations of lasso may be interpreted as posterior mode estimations, In consequence, many Bayesian lasso (Blasso) techniques have been suggested through the years by some other researchers subsequently employing Laplace-like priors (see for example, Figueiredo, 2003; Bae and Mallick, 2004; Yuan and Lin, 2005). Park and Casella (2008) considered a fully Bayesian analysis based on a conditional Laplace prior description of the form

$$\pi(\boldsymbol{\beta}|\sigma^2) = \prod_{k=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_k|/\sqrt{\sigma^2}} \quad , \tag{8}$$

for the regression coefficients vector $\boldsymbol{\beta}$ as well as the scale-invariant marginal prior for σ^2

$\pi(\sigma^2) = 1/\sigma^2$ on σ^2. The prior for $\boldsymbol{\beta}$ can be written as a scale mixture of normals (Andrews and Mallows, 1974; Park and Casella, 2008).

$$\frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta|/\sqrt{\sigma^2}} = \int_0^{\infty} \frac{1}{\sqrt{2\pi s}} e^{-\beta^2/(2s)} \frac{\lambda^2}{2\sigma^2} e^{-\lambda^2 s/2\sigma^2} ds \, . \tag{9}$$

Under these assumptions, the Bayesian hierarchical modeling is given by (Andrews and Mallows, 1974; Park and Casella, 2008):

$$\mathbf{y}|\mu, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \mathrm{N}_n(\mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n),$$

$$\boldsymbol{\beta}|\sigma^2, s_1^2, \ldots, s_p^2 \sim \mathrm{N}_p\left(\mathbf{0}_p, \sigma^2 \boldsymbol{w}_s\right),$$

$$\boldsymbol{w}_s = diag\left(s_1^2, \ldots, s_p^2\right), \tag{10}$$

$$\sigma^2, s_{1,\ldots,}^2 s_p^2 \sim \pi(\sigma^2)\, w\sigma^2 \prod_{k=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 s_k^2/2} w s_k^2 ,$$

a scale mixture of normal (Andrews and Mallows, 1974; Park and Casella, 2008). Following this representation, Park and Casella (2008) proposed an efficient algorithm for Blasso.

2.2.3 Regularization regression using reciprocal lasso

To avoid overfitted models, Song and Liang (2015) suggested the rlasso approach for VS and the estimate of coefficients together, which is based on a novel class of penalty functions that are discontinuous at 0, decreasing in (0,∞) and also give near 0 coefficients of infinity penalties. Therefore, rlasso is high desirable for model selection due to this characteristic property (Song and Liang, 2015; Song, 2018). The rlasso estimator is obtained as follows:

$$\hat{\beta}_{\text{rlasso}} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} \frac{1}{|\beta_k|} I\{\beta_k \neq 0\}, \qquad (11)$$

where $\lambda > 0$ is the tuning parameter controlling the penalization degree as well as $I(.)$ is an indicator function, the lowest value for λ gives a highest level of shrinkage and gives coefficients that are near to zero. Compared to the lasso penalty, which is nondecreasing in (0,∞) and continuous, the penalty of rlasso decreases in (0,∞) and discontinuous at zero. Additionally, the lasso gives near 0 coefficients of 0 penalties, but the rlasso gives near 0 coefficients of infinity penalties. Also, the VS method of rlasso is very different from that of lasso. The lasso selects smaller coefficients, whereas the rlasso selects bigger ones. Theoretically, rlasso has the same oracle property (Mallick et al., 2021). The rlasso can perform much better at VS

compared to the lasso. However, rlasso requires a lot of computational intensive (Song and Liang, 2015).

2.2.4 Bayesian interpretation of rlasso

Mallick et al. (2021) considered a fully Bayesian approach to the rlasso issuse, based the observation that whenever the parameters of regression are determined by independent inverse Laplace priors, the rlasso estimation for linear regression parameters may be interpreted as a Bayesian posterior mode estimation by using the following:

$$\pi(\boldsymbol{\beta}) = \prod_{k=1}^{p} \frac{\lambda}{2\beta_k^2} \exp\{-\frac{\lambda}{|\beta_k|}\}\ I\{\beta_k \neq 0\}\,, \tag{12}$$

where $\lambda > 0$ denotes a scale parameter determining the prior's dispersion around 0. As a result, λ should be tiny in order to ease sparse recovery. This is rather counterintuitive considering that the majority of lasso-type shrinkage approaches penalize coefficients by a big value of λ. In specifically, the Bayesian rlasso (Brlasso) requires a limited value of λ while the Blasso prefers a great value of λ to the best performance, which can be written as (Mallick et al., 2021)

$$\begin{aligned}
&\mathbf{y}^{n\times 1} | \ \mathrm{X}, \boldsymbol{\beta}, \sigma^2 \sim \mathrm{N_n}(\mathrm{X}\boldsymbol{\beta}, \sigma^2 \mathrm{I_n}),\\
&\boldsymbol{\beta}^{p\times 1} | \mathbf{u} \sim \prod_{\mathbf{k=1}}^{\mathbf{p}} \frac{\mathbf{1}}{\mathrm{Uniform}(-\mathrm{u_k}, \mathrm{u_k})}\,,\\
&\mathbf{u}^{p\times 1} | \lambda \sim \prod_{k=1}^{p} \mathrm{Gamma}(2, \lambda)\,,\\
&\sigma^2 \sim \pi(\sigma^2),
\end{aligned} \tag{13}$$

following this representation, Mallick et al. (2021) proposed an efficient algorithm for Bayesian rlasso. Thus, they showed that the Bayesian approach performs better than its conventional counterpart in VS, prediction, and estimate. On the other hand, there are two reasons for using a Bayesian method to the rlasso issue. First, one of the greatest reasons for using a fully Bayesian method is the optimization issue's multimodal nature (7). Second, the Bayesian rlasso method is computationally efficient, resulting in scalable MCMC algorithms having perfect convergence as well as mixing characteristic (11).

2.2.5 The Adaptive lasso Regression

In 2006, Zou proved that the lasso estimator is inconsistent in VS, although the lasso performs well in putting the coefficients of unimportant covariates equal to zero, and the lasso approach is a common approach for VS as well as the estimate of coefficients at the same time. To address this problem, Zou (2006) proposed a new regularization method by assuming different regularization weights for different coefficients. It is called the adaptive least absolute shrinkage and selection operator (alasso) method. Lasso estimations are known to be biased to large coefficients since lasso penalizes all coefficients equally. The alasso method controls the lasso estimate's bias by adding adaptive weights, which are employed to penalize various coefficients in the lasso method. Thus, the coefficients of the unimportant covariates are reduced to 0 more efficiently. This method minimizes bias and improves VS accuracy, which creates estimates that are consistent and unbiased as well as performs a better job of estimating the coefficients of the important covariates than lasso (Zou, 2006; Wang et al.

2007). The Alasso estimator is obtained as follows:

$$\hat{\beta}_{alasso} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} \widehat{w}_k |\beta_k| , \quad (14)$$

where $\lambda \sum_{k=1}^{p} \widehat{w}_k |\beta_k|$ is known as the alasso penalty, $\widehat{\boldsymbol{w}}_k = (\widehat{w}_1, \widehat{w}_2, \ldots, \widehat{w}_p)$ denotes the adaptive weight vector, which is defined as follows:

$$\widehat{\boldsymbol{w}}_k = \frac{1}{|\widehat{\boldsymbol{\beta}}_k|^{\gamma}} , \quad (15)$$

for k=1,…, p and $\gamma > 0$. The parameter γ is a tuned parameter (Zou, 2006), that may be calculated by using the cross-validation method.

The alasso regression has the following advantages:

1. It is computationally more appealing (Zou, 2006).
2. It can be solved using the same efficient algorithm that was utilized to solve the lasso, i.e., the LARS algorithm (Efron et al., 2004).
3. It does have oracle properties (i.e., it selects the right subset of variables from a larger set on a consistent basis and includes asymptotic guarantees of unbiasedness and normality) by employing the ℓ_1 penalty, which is adaptively weighted (Zou, 2006).

The alasso estimator has been widely used in many statistical applications due to its good theoretical properties. (See, Zhang and Lu, 2007; Zeng et al., 2014; Yang and Wu, 2016). On the other hand, it needs consistent initial estimations of the regression coefficients, which are commonly unavailable in the high dimension, small sample size setting. Additionally, none of the algorithms employed to compute the

alasso estimators gives a correct estimate of standard error.

2.2.6 The Bayesian adaptive lasso regression

A result of the alasso's suffering from collinearity induced by highly corrected covariates due to the illness of the OLS initial estimates in $\widehat{\boldsymbol{w}}_k$'s. When the correlation between the covariates is high, the illness (ill-condition) happens, implying which $(X'X)^{-1}$ does not have full rank, resulting in coefficients estimations to be unstable.

Several authors used a Bayesian formulation in their studies, such as Sun et al. (2010) as well as Leng et al. (2014), who recently suggested Bayesian adaptive lasso for VS as well as the estimate of coefficients together in linear regression to avert the OLS initial estimations for the coefficients of regression because they estimation the adaptive weight $\widehat{\boldsymbol{w}}_k$'s automatically (Alhamzawi and Ali, 2018). Additionally, on the basis of a geometrically ergodic Markov Chain, the Bayesian approach to the alasso gives a valid standard error measure. (Casella et al., 2010).

The Balasso is similar to the alasso (Zou, 2006), that develops the lasso (Tibshirani, 1996) through adding covariate-specific penalties (Sun et al., 2010). The Balasso may be obtained by using the following conditional Laplace prior with coefficient-specific tuning parameters (Feng et al., 2017):

$$\pi(\boldsymbol{\beta}|\sigma^2) = \prod_{k=1}^{p} \frac{\lambda_k}{2\sqrt{\sigma^2}} e^{-\lambda_k|\beta_k|/\sqrt{\sigma^2}} \quad . \qquad (16)$$

The Balasso, like the alasso, applies different penalties to different coefficients to improve its ability to produce better estimate and model

selection outcomes.

The Balasso is a Bayesian hierarchical model. The priors are defined as follows (Sun et al., 2010):

$$p(B_0) \propto 1, \tag{17}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{18}$$

$$p(B_k|\lambda_k) = \frac{1}{2\lambda_k}\exp\left(-\frac{|B_k|}{\lambda_k}\right), \tag{19}$$

$$p(\lambda_k|\delta,\tau) = inv - Gamma(\lambda_k;\delta,\tau) = \frac{\tau^{\delta}}{\Gamma(\delta)}{\lambda_k}^{-1-\delta}\exp\left(-\frac{\tau}{\lambda_k}\right), \tag{20}$$

where $\delta > 0$ as well as $\tau > 0$ represent two hyperparameters. The Balasso was found to be very efficient, conceptually simple, simple to apply, and there is no need for any initial estimations of the regression coefficients that are useful. The Balasso, on the other hand, does not specify a point mass at zero. The regression coefficient samples would not be precisely zero. Thus, the Balasso method does not select variables, but if we observe the posterior distribution mode, it may be precisely 0.

Chapter Three

Bayesian group bridge composite quantile regression

3. Bayesian group bridge composite quantile regression

3.1 Introduction

The normal linear regression model supposes that an outcomes vector $\boldsymbol{y} = (y_1, \cdots, y_n)'$ can be written as

$$\boldsymbol{y} = b_0\mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{21}$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$ is a $n \times p$ covariates matrix, b_0 is the intercept, $\mathbf{1}$ is an $n \times 1$ unit vector, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$ are independent, as well as ε_i has a Gaussian distribution having mean 0 and variance σ^2. According to model (21), it's supposed that only an unfamiliar subset from covariates are effective in the regression; therefore, the issue of covariate selecting is to find this unfamiliar subset of covariates.

Traditional approaches to model selection based on the observed data log likelihood, comparing a set of candidate models include Mallows's C_p (Mallows, 1973), Akaike information criterion (AIC; Akaike, 1973), and Bayesian information criterion (BIC; Schwarz, 1978). Among the new approaches that are based on regularization and selection operator involve the bridge regression (Frank and Friedman, 1993), lasso (Tibshirani, 1996), smoothly clipped absolute deviation (Fan and Li, 2001), fused lasso (Tibshirani et al., 2005), adaptive lasso (Zou, 2006), graphical lasso (Yuan and Lin, 2006), dantzig selector (Candes and Tao, 2007), and matrix completion (Cand`es and Tao, 2010; Mazumder et al., 2011), among others. These approaches are setup for selecting individual covariates. However, covariates are naturally grouped in many real studies. An important example appears in association studies, genes may form overlapping sets where each gene can be involved in multiple tracks (Jacob et al., 2009). For this and

other situations, Yuan and Lin (2006) suggested the group lasso penalty for choosing covariates groups by introducing a suitable expansion of the lasso penalty. Since Yuan and Lin (2006), over the years, various group lasso methods have been improved for dealing with chosen groups of covariates (see for example, Breheny, 2015; Huang et al., 2012, 2009; Meier et al., 2008; Park and Yoon, 2011; Qian et al., 2016; Simon et al., 2013; Simon and Tibshirani, 2012).

Although covariate selection methods in standard mean regression models have been well developed, we frequently require to assess effects of covariates on outcome variable at various quantile levels. Koenker and Bassett (1978) suggested quantile regression (QR) to overcome this issue. Compared to standard mean regression, QR is more strong to data outliers than standard mean regression, and can provide a more clear picture of the relation between covariates and outcome of interest. However, for linear regression models, Zou and Yuan (2008) indicated that QR may result in an arbitrarily tiny relative efficiency when compared with the standard mean regression. Since, QR at one quantile can provide more efficient estimators than QR at another quantile, Zou and Yuan (2008) suggested a composite QR (CQR) approach to simultaneously study multiple QR models. They proved that, irrespective of the error distribution, the relative efficiency of the CQR estimator is higher than 70% when compared to the mean regression estimator. Recently, when p is finite, CQR has been employed in covariate selection methods; for example see, Zou and Yuan (2008), Bradic et al. (2011) and Jiang et al. (2012). In this thesis, we suggest a Bayesian framework to combine CQR and group bridge penalty together to perform model selection and estimation of coefficients simultaneously.

3.2 Methods

3.2.1 QR

QR (Koenker and Bassett, 1978) has acquired growing popularity since it makes few assumptions about the error distribution. For the θth quantile ($0 < \theta < 1$), the linear QR model is $\boldsymbol{y} = b_0 + X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ are independent, and their θth quantiles equal to zero. The θth QR model takes the form of

$$Q_{y_i}(x_i) = b_\theta + x_i'\beta \tag{22}$$

where b_θ is the quantile intercept. The regression parameters b_θ and $\boldsymbol{\beta}$ are estimated by minimizing (Koenker and Bassett, 1978):

$$(\hat{b}_\theta, \hat{\beta}) = \min_{b_\theta, \beta} \sum_{i=1}^{n} \rho_\theta(y_i - b_\theta - x_i'\beta), \tag{23}$$

where $\rho_\theta(\varepsilon) = \frac{|\varepsilon| + (2\theta - 1)\varepsilon}{2}$ denotes the quantile check (Loss) function.

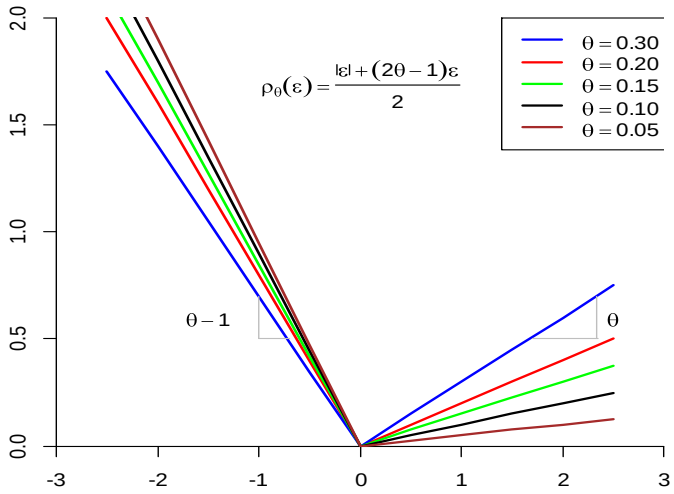


Figure 1: The panel shows the check function at theta = 0.30 (blue line), theta = 0.20 (red line), theta = 0.15 (green line), theta = 0.10 (black line), and theta = 0.05 (brown line).

This function can also written $as\ \rho_\theta(\varepsilon) = \varepsilon\theta - \varepsilon I(\varepsilon \leq 0)$ where $I(.)$ denotes the indicator function, and

$$\rho_\theta(\varepsilon) = \begin{cases} \theta\varepsilon & \text{if } \varepsilon \geq 0, \\ -(1-\theta)\varepsilon, & \text{if } \varepsilon < 0 \end{cases}$$

Figure (1) shows the check function at five different quantiles, namely 0.30, 0.20, 0.15, 0.10 and 0.05.

The asymmetric Laplace distribution (ALD) provides a possible parametric correlation between the minimization issue in (23) and the maximum likelihood theorem (Koenker and Machado, 1999; Yu and Moyeed, 2001). The ALD density function for the response y is

$$f(\mu,\sigma) = \frac{\theta(1-\theta)}{\sigma} exp\left\{-\frac{\rho_\theta(y-\mu)}{\sigma}\right\}, \qquad (24)$$

where σ is the scale parameter and μ is the location parameter. Yu and Moyeed (2001) introduced a Bayesian framework for QR employing the ALD for the errors, and the MCMC Metropolis-Hastings sampling algorithm is utilized to (approximately) draw $\boldsymbol{\beta}$ from it's conditional distribution. Kozumi and Kobayashi (2011) improved an efficient Gibbs sampling algorithm for Bayesian QR by assuming that the random variable $\epsilon_i = (1-2\theta)w_i + \sqrt{2\sigma w_i}z_i$ follows the ALD, where w_i and z_i have an exponential distribution having scale parameter $(\theta(1-\theta)/\sigma)$ and a standard normal distribution, respectively (see, Alhamzawi and Yu, 2012; Alshaybawee et al., 2017; Alhamzawi and Ali, 2018; Alhamzawi et al., 2019; Alhamzawi, Taha Mohammad Ali, 2020). As the conditional distribution of y_i given w_i is normal having mean $b_\theta + x_i'\beta + (1-2\theta)w_i$ and variance $2\sigma w_i$, the density of y_i is given by

$$p(y_i|x_i,\beta,b_\theta,w_i,\sigma)=\frac{1}{\sqrt{4\pi\sigma w_i}}exp\left\{-\frac{\left(y_i-b_\theta-x_i^{'}\beta-(1-2\theta)w_i\right)^2}{4\sigma w_i}\right\}. \quad (25)$$

3.2.2 CQR

CQR (Zou and Yuan, 2008) has acquired growing popularity as it can combine information of numerous quantiles simultaneously to get a group of good estimations. Denote $0<\theta_1<\theta_2<\cdots<\theta_K<1$, where $\theta_k=k/(K+1)$. The CQR estimators of $\boldsymbol{b}_\theta=(b_{\theta1},\cdots,b_{\theta K})$ and $\boldsymbol{\beta}$ can be estimated by minimizing

$$\left(\hat{b}_\theta,\hat{\beta}\right)=\underset{b_\theta,\beta}{min}\ \sum_{i=1}^{n}\left\{\sum_{k=1}^{K}\rho_{\theta_k}\left(y_i-b_{\theta_k}-x_i^{'}\beta\right)\right\}, \quad (26)$$

Huang and Chen (2015) and Alhamzawi (2016) proposed Bayesian formulations for CQR using the ALD for the errors. Under these formulations, the joint distribution of $\boldsymbol{y}$ is

$$p(X,\beta,b_\theta,w,\sigma)=\prod_{k=1}^{K}\prod_{i=1}^{n}\left(\frac{1}{\sqrt{4\pi\sigma w_{ik}}}\right)exp\left\{-\frac{\left(y_i-b_{\theta_k}-x_i^{'}\beta-\xi_k w_{ik}\right)^2}{4\sigma w_{ik}}\right\}, \quad (27)$$

where $w=(w_1,\cdots,w_K)$, $w_k=(w_{1k},\cdots,w_{nk})$ and $\xi_k=1-2\theta_k$.

Alsaadi and Alhamzawi (2022) used the above formula and proposed a Bayesian formulation for bridge and reciprocal bridge CQR.

3.2.3 CQR with the group bridge penalty

Assume that the covariates are collected into G groups so that $x_i=\left(x_{i1}^{'},\ldots,x_{iG}^{'}\right)^{'}$, $\beta=\left(\beta_1^{'},\ldots,\beta_G^{'}\right)^{'}$, β_g is the m_g-dimensional coefficient vector of the gth group covariates x_{ig}, $\sum_{g=1}^{G}m_g=p$ and $G<p$. In this thesis, we define the following group bridge regularized CQR:

$$(\hat{b}_\theta, \hat{\beta}) = \min_{b_\theta, \beta} \sum_{i=1}^{n}\{\sum_{k=1}^{K} \rho_{\theta_k}(y_i - b_{\theta_k} - x_i'\beta)\} + \sum_{g=1}^{G} \lambda_g \|\beta_g\|_1^\alpha, \quad (28)$$

where $\|\beta g\|_1$ is the L_1 norm of β_g , $\lambda_g > 0$, $g = 1, \cdots, G$ are the group-specific shrinkage parameters and $\alpha > 0$ denotes the concavity parameter. The bridge parameter α does covariate selection when $\alpha \in (0, 1]$, and shrinks the coefficients of regression when $\alpha > 1$. From a Bayesian point of view, one may define the following group bridge prior on the coefficients (G´omez-S´anchez-Manzano et al., 2008; G´omez-Villegas et al., 2011; Mallick and Yi, 2018):

$$p(\alpha, \lambda_1, \cdots, \lambda_G) \propto \prod_{g=1}^{G} exp(-\lambda_g \|\beta_g\|_1^\alpha). \quad (29)$$

If we remove the dependence on the group index g, the prior for a group bridge may be written as follows

$$p(\beta) = \frac{\lambda^{\frac{p}{\alpha}}\Gamma(p+1)}{2^p\Gamma\left(\frac{p}{\alpha}+1\right)} \exp(-\lambda\|\beta\|_1^\alpha). \quad (30)$$

If we put the group bridge prior (29) on $\boldsymbol{\beta}$ and assume the errors ϵ_i is from the ALD (24), the conditional distribution of $\boldsymbol{\beta}$ is

$$p(X, \beta, b_\theta, w, \sigma) \propto exp\left\{-\sum_{i=1}^{n}\sum_{k=1}^{K} \frac{(y_i - b_{\theta_k} - x_i'\beta - \xi_k w_{ik})^2}{4\sigma w_{ik}} - \sum_{g=1}^{G} \lambda_g \|\beta_g\|_1^\alpha\right\}. \quad (31)$$

So minimizing the group bridge regularized CQR (28) is equivalent to maximizing the composite likelihood (31). Mallick and Yi (2018) show that the group bridge prior may be expressed as a scale mixture of multivariate uniform (SMU) distribution, the mixing density is a specific Gamma distribution, in other words, $\boldsymbol{\beta}|u \sim$ Multivariate Uniform (A), where $A = \{\boldsymbol{\beta} \in R^q : \|\beta_g\|_1^\alpha < u\}$, $u > 0$ and $u \sim$ Gamma $(\frac{q}{\alpha}+ 1, \lambda)$.

Putting Beta prior on α and Gamma priors on λ_g and σ_k, the Bayesian hierarchical model for CQR with group bridge penalty (28) is as follows

$$y_i = \prod_{k=1}^{K}\left(b_{\theta_k} + x_i'\beta + \xi_k w_{ik} + \sqrt{2\sigma w_{ik}}z_i\right), i = 1,\cdots,n,$$

$$w|\sigma \sim \prod_{k=1}^{K}\prod_{i=1}^{n}\frac{\theta_k(1-\theta_k)}{\sigma}exp\left(-\frac{\theta_k(1-\theta_k)}{\sigma}w_{ik}\right),$$

$$z \sim \prod_{i=1}^{n}N(0,1)$$

$$\beta_g|u_g, \alpha \sim Multivariate\ Uniform\left(\Omega_g\right) independently\ for\ g = 1,\cdots,G,$$

where $\Omega_g = \{\beta_g \in R^{m_g} : \|\beta_g\|_1^{\alpha} < u_g\}$, (32)

$$u_1,\ldots,u_G|\lambda_1,\ldots,\lambda_G,\alpha \sim \prod_{g=1}^{G}Gamma(\frac{m_g}{\alpha}+1,\lambda_g),$$

$$\lambda_1,\ldots,\lambda_G \sim \prod_{g=1}^{G}Gamma(a,b),$$

$\alpha \sim$ Beta(c, d),

$\sigma \sim$ Gamma(r, δ),

where $\boldsymbol{u} = (u_1, \cdots, u_G)$, and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_G)$. It's clear that the full conditional posteriors may be obtained by employing easy algebra for the prior description and the parameters of interest ($\boldsymbol{b}_\theta$, $\boldsymbol{\beta}$, σ, w, $\boldsymbol{u}$, $\boldsymbol{\lambda}$, α) can be sampled as listed in Figure 2

Input:(y, X**)**

Initialize: $(\boldsymbol{b}_\theta, \boldsymbol{\beta}, \sigma, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{\lambda}, \alpha)$

for $t = 1, \ldots, (t_{\max} + t_{\text{burn-in}})$ **do**

1. Sample $\boldsymbol{\beta}|. \sim N_p(\boldsymbol{\beta}, B) \prod_{g=1}^{G} I\left\{\|\boldsymbol{\beta}_g\|_2^{\alpha} < u_g\right\}$, where

$$B^{-1} = \left(\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{x_i x_i^{'}}{2\sigma w_{ik}}\right) \text{ and}$$

$$\boldsymbol{\beta} = B\left(\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{x_i\left(y_i - b_{\theta_k} - \boldsymbol{x_i^{'}\beta} - \xi_k w_{ik}\right)}{2\sigma w_{ik}}\right)$$

2. Sample $b_{\theta_k}|. \sim N\left(\frac{\sum_{i=1}^{n}(y_i - \boldsymbol{x_i^{'}\beta} - \xi_k w_{ik})/2\sigma w_{ik}}{\sum_{i=1}^{n} 1/2\sigma w_{ik}}, \frac{1}{\sum_{i=1}^{n} 1/2\sigma w_{ik}}\right)$

3. Sample $w_{ik}|. \sim$ inverse Gaussian $\left(\frac{1}{2\sigma}, \sqrt{\frac{1}{\left(y_i - b_{\theta_k} - \boldsymbol{x_i^{'}\beta}\right)^2}}\right)$

4. Sample $\sigma|. \sim$ inverse Gamma $\left(\frac{3nK}{2} + r, \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\left(y_i - b_{\theta_k} - \boldsymbol{x_i^{'}\beta} - \xi_k w_{ik}\right)^2}{2w_{ik}} + \sum_{i=1}^{n}\sum_{k=1}^{K}\theta_k(1 - \theta_k)w_{ik} + \delta\right)$

5. Sample $u|. \sim \prod_{g=1}^{G}$ Exponential$(\lambda_g) I\left\{u_g > \|\boldsymbol{\beta}_g\|_1^{\alpha}\right\}$

6. Sample $\boldsymbol{\lambda}|. \sim \prod_{g=1}^{G}$ Gamma $\left(a + m_g/\alpha, b + \sum_{g=1}^{G}\|\boldsymbol{\beta}_g\|_1^{\alpha}\right)$

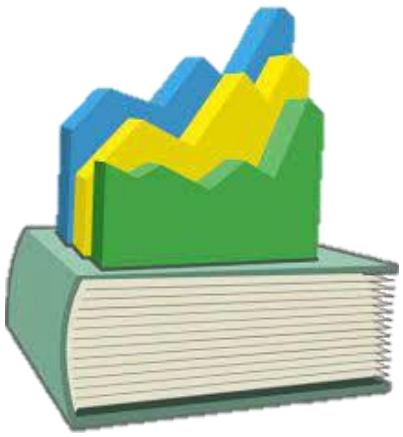
7. Sample $\alpha|. \sim \alpha^{c-1}(1-\alpha)^{d-1}\prod_{g=1}^{G}\frac{{\lambda_g}^{m_g/\alpha}}{\Gamma\left(\frac{m_g}{\alpha}+1\right)} exp\left(-\lambda_g\|\boldsymbol{\beta}_g\|_1^{\alpha}\right)$, which has no closed form. Since $p(.)$ is a log-concave, we update α using Adaptive Rejection Sampling (ARS; Gilks, 1992)

end for

Figure 2: MCMC sampling for the Bayesian group bridge CQR.

Chapter four

Simulation Studies



4. Simulation Studies

4.1. Example 1

We carry out simulation studies in this section using the statistical program (R) to demonstrate the performance of lasso and rlasso approaches.

Where:

- Lasso: the least absolute shrinkage and selection operator.
- Rlasso: reciprocal lasso.

The data in the simulation examples were generated by

$$\boldsymbol{y}_i = x_i'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, i = 1, 2, \ldots, n. \quad (1)$$

We setup the error distribution $\varepsilon_i \sim N(0, \sigma^2)$. The design matrix rows $\boldsymbol{X}$ were generated from $N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ has an autoregressive correlation matrix, where $\boldsymbol{\Sigma_{ij}} = 0.50^{|\boldsymbol{i}-\boldsymbol{j}|}$ for all $1 \leq \boldsymbol{i} \leq \boldsymbol{j} \leq p$. We consider three cases for $\boldsymbol{\beta}$:

Simulation 1: $\boldsymbol{\beta}$ = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)

Simulation 2: $\boldsymbol{\beta}$ = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1)

Simulation 3: $\boldsymbol{\beta}$ = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

Table 1: Mean squared errors (MSE) and standard deviations (SD) for Simulation 1

	lasso	rlasso
MSE σ^2 =1	1.57	1.66
SD σ^2 = 1	0.30	0.32
MSE σ^2 = 4	5.64	5.95
SD σ^2 = 4	0.81	0.92
MSE σ^2 = 9	12.50	12.89
SD σ^2 = 9	2.18	2.06

Table 2: Mean squared errors (MSE) and standard deviations (SD) for Simulation 2

	lasso	rlasso
MSE σ^2 = 1	1.26	1.45
SD σ^2 = 1	0.22	0.23
MSE σ^2 = 4	5.50	6.27
SD σ^2 = 4	1.22	1.21
MSE σ^2 = 9	11.78	13.43
SD σ^2 = 9	1.95	2.25

Table 3: Mean squared errors (MSE) and standard deviations (SD) for Simulation 3

	lasso	rlasso
MSE $\sigma^2 = 1$	1.27	1.51
SD $\sigma^2 = 1$	0.26	0.30
MSE $\sigma^2 = 4$	5.10	6.16
SD $\sigma^2 = 4$	1.06	1.81
MSE $\sigma^2 = 9$	11.15	13.68
SD $\sigma^2 = 9$	1.83	2.30

The results are listed in Tables 1, 2, and 3. The results of both approaches (lasso and rlasso) are very similar. Our outcomes demonstrate that lasso and rlasso perform comparably in choosing a high dimensional model in various simulation studies.

Overall, the simulations show that the both approaches have the same accuracy of the prediction in most of the cases, so often outperform their frequentist counterparts in terms of prediction accuracy all over a wide range of scenarios.

4.2. simulations for BgBCQR

Here, we use simulations of Monte Carlo to illustrate the performance of Bayesian group Bridge CQR (BgBCQR) with comparison to the Bayesian group bridge regression (BgBR, Mallick and Yi, 2018), group bridge regression (gBR, Huang et al., 2009) and group lasso regression (gLR, Yuan and Lin, 2006). The Bayesian estimations are posterior means employing 20,000 draws of the MCMC algorithm following burn-in the first 10,000 draws. For our approach, we set $a = 1$, $b = 0.1$, $r = 10$, $\delta = 10$, $c = 0.1$, and $d = 0.1$.

We generate data using the following real model

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

In each generated data, we consider three different choices for the error distribution: $N(0,9)$, $t(3)$ distribution having (3) freedom degrees, and $\chi^2_{(3)}$ distribution having (3) freedom degrees. Additionally, we run 100 replications. In each replication, we simulate a training set of 20 observations and a testing set of 200 observations.

Example 2 (Li et al., 2010). In this example, the rows of the design matrix X are provided by $(I(S_1 = 0), I(S_1 = 1), I(S_1 = 2), \cdots, I(S_5 = 0), I(S_5 = 1), I(S_5 = 2))$, where the latent variables $S = (S_1, \cdots, S_5)'$ are simulated independently from $N(0, \boldsymbol{\Sigma})$ with the (i, j)th element of $\boldsymbol{\Sigma}$ is $\rho^{|i-j|}$ and $\rho = 0.5$. Each latent variable S_j for $j = 1, \cdots, 5$ is trichotomized as zero, one or two, depending on whether it's less than $F^{-1}(1/3)$, between $F^{-1}(1/3)$and $F^{-1}(2/3)$, or greater than $F^{-1}(2/3)$, where F^{-1} is the quantile function to standard normal distribution. We set the regression coefficients vector as $\boldsymbol{\beta} = ((-1.2, 1.8, 0), (0, 0, 0), (0.5, 1, 0), (0, 0, 0), (1, 1, 0))$. Thus, the regression parameters in a group may be

either all zero, all nonzero or partly. We use $\{n_T, n_P\} = \{20, 400\}$, $\{50, 400\}$ and $\{100, 400\}$ respectively, to simulate datasets, where n_T stands for the number of the observations in the training set, while n_P stands for the number of the observations in the testing set. The experimental outcomes are presented in Table 4. Here, in terms of prediction accuracy, our suggested approach outperforms current Bayesian and non-Bayesian approaches.

Table 4: Median of mean absolute deviations (MMAD) with the standard deviations of MAD (SD) for Example 2. The bold numbers of MMAD stands for the least MMAD in each category.

		Error					
		$N(0, 9)$		$t(3)$		$\chi^2_{(3)}$	
Method	n_T	MMAD	SD	MMAD	SD	MMAD	SD
gLR	20	1.4278	1.4021	**1.4325**	1.4690	1.7025	1.6536
gBR	20	1.4166	1.4722	1.6533	1.8253	1.9837	2.3613
BgBR	20	1.3728	1.2083	1.5241	1.3422	1.6572	1.5344
BgBCQR	20	**1.3213**	1.4082	1.5221	1.3314	**1.5267**	1.4362
gLR	50	1.4099	1.5504	1.3722	1.4797	1.4359	1.4685
gBR	50	1.5313	2.1991	1.4850	2.0498	1.5083	2.0028
BgBR	50	1.3121	1.4467	1.2901	1.3788	**1.3283**	1.3936
BgBCQR	50	**1.2614**	1.1231	**1.1751**	1.1238	1.3781	1.5865
gLR	100	1.2543	1.4199	1.2347	1.3747	1.2459	1.3555
gBR	100	1.3013	1.9240	1.2365	1.8437	1.2331	1.7848
BgBR	100	1.1841	1.3446	1.1328	1.3026	1.1281	1.2801
BgBCQR	100	**1.0021**	1.5278	**1.1206**	1.4711	**1.1061**	1.4311
gLR	200	1.1197	1.3299	1.0750	1.2976	1.0859	1.2699
gBR	200	1.0892	1.7387	1.0194	1.6856	1.0292	1.6394
BgBR	200	1.0148	1.2540	0.9735	1.2209	0.9844	1.1984
BgBCQR	200	**0.9893**	1.3966	**0.9059**	1.3601	**0.9137**	1.3264

Example 3 (High Correlation Example). The setup for this example is identical to the first, excepting we set ρ = 0.95. The experimental outcomes are presented in Table 5. Here also, in terms of prediction accuracy, our suggested approach outperforms the other methods.

Table 5: MMAD with the standard deviations of MAD (SD) for Example 3. The bold numbers of MMAD stands for the least MMAD in each category.

		Error					
		$N(0, 9)$		$t(3)$		$\chi^2_{(3)}$	
Method	n_T	MMAD	SD	MMAD	SD	MMAD	SD
gLR	20	1.2157	1.0958	1.4045	1.5105	1.6392	3.4381
gBR	20	1.1169	1.6627	1.3276	2.2563	1.5980	3.8811
BgBR	20	1.2516	1.0141	1.3050	1.2258	1.4053	1.4456
BgBCQR	20	**1.1087**	1.3581	**1.2732**	1.7005	**1.4008**	2.1010
gLR	50	1.3368	3.0464	1.3128	2.8754	1.3578	2.6771
gBR	50	1.2352	3.4627	1.1757	3.2631	1.2064	3.0724
BgBR	50	1.1389	1.3623	1.1205	1.4771	1.1517	1.4497
BgBCQR	50	**1.1123**	1.9312	**1.0463**	1.9399	**1.0749**	1.8490
gLR	100	1.2226	2.5129	1.1945	2.3731	1.2195	2.2664
gBR	100	1.0464	2.8894	1.0008	2.7298	1.0234	2.6102
BgBR	100	1.0228	1.3897	0.9612	1.3315	0.9882	1.3033
BgBCQR	100	**0.9636**	1.7549	**0.8931**	1.6710	**0.9162**	1.6120
gLR	200	1.1040	2.1808	1.0616	2.0965	1.0670	2.0238
gBR	200	0.8993	2.5071	0.8571	2.4100	0.8449	2.3241
BgBR	200	0.8784	1.2708	0.8372	1.2350	0.8332	1.2030
BgBCQR	200	**0.8304**	1.5592	**0.7751**	1.5073	**0.7689**	1.4594

Example 4. The setup for this example is identical to the first, excepting we set the coefficients of regression vector as $\boldsymbol{\beta}$ = ((0.5, 1, 1.5, 2, 2.5), (2, 2, 2, 2, 2), (0, 0, 0, 0, 0)). Thus, in each group, the regression parameters are either all nonzero or all zero. The experimental outcomes are shown in Table 6. Again, we may observe that in terms of prediction accuracy, our proposed approach outperforms the other approaches.

Table 6: MMAD with the standard deviations of MAD (SD) for Example 4. The bold numbers of MMAD stands for the least MMAD in each category.

		Error					
		$N(0, 9)$		$t(3)$		$\chi^2_{(3)}$	
Method	n_T	MMAD	SD	MMAD	SD	MMAD	SD
gLR	20	0.9023	1.3315	**0.9852**	1.1570	**0.9991**	1.9758
gBR	20	1.2054	2.0352	1.3890	1.9173	1.5677	3.5526
BgBR	20	**0.8993**	0.9335	1.0620	0.9569	1.1418	1.4970
BgBCQR	20	1.0882	1.4828	1.1909	1.4525	1.3595	1.8149
gLR	50	**0.8698**	1.7703	**0.8549**	1.6444	**0.8512**	1.5493
gBR	50	1.2195	3.1602	1.1014	2.8997	1.1144	2.7275
BgBR	50	0.9995	1.3676	0.9356	1.2845	0.9457	1.2530
BgBCQR	50	1.1121	1.6576	1.0524	1.5576	1.0498	1.4973
gLR	100	**0.8051**	1.4692	0.7990	1.4078	**0.8051**	1.3559
gBR	100	0.9750	2.5649	0.9377	2.4384	0.9430	2.3289
BgBR	100	0.8438	1.1981	0.7990	1.1597	0.8272	1.1301
BgBCQR	100	0.8306	1.4276	**0.7750**	1.3702	0.8164	1.3255
gLR	200	0.7708	1.3108	0.7582	1.2707	0.7685	1.2389
gBR	200	0.8369	2.2343	0.7950	2.1482	0.7827	2.0760
BgBR	200	0.7502	1.0969	0.7218	1.0677	0.7216	1.0481
BgBCQR	200	**0.7493**	1.2841	**0.7159**	1.2453	**0.7111**	1.2131

Overall, the simulations show that all of the Bayesian approaches have the same accuracy of the prediction in most of the cases, so often outperform their frequentist counterparts in terms of prediction accuracy all over a wide range of scenarios.

Chapter five

The real data

5. Real Data Analyses

5.1. The rate of production of wheat yield example

We applied two regularization approaches (lasso and rlasso) with real data (data approved and registered in Al-Diwaniyah Agriculture Directorate for the year 2021). These data are related to the rate of production of wheat yield. The reasons for the increase in the rate of wheat production are considered one of the most important criteria required for researchers in agricultural sciences.

Wherefore, the application of both approaches, which attempts to identify the factors that positively affect the increase in wheat production. Determining the covariates affecting the increase in production may contribute significantly to predicting an increase in the yield of the crop in the future. These covariates are types of fertilizers, new standards in the cultivation of agricultural crops, in addition to modern agricultural techniques. Wheat production data includes 9 covariates, with a dependent variable represented in the rate of wheat production per dunum, where the dunum is the unit of the cultivated land area and is equal to 2500 square meters.

Table 7: The mean squared prediction errors (MMSE) for wheat crop production rate data analyses.

Method	MMSE
lasso	0.85
rlasso	0.84

We compare MMSE for wheat crop production rate data analyses in Table 7, which shows that both approaches are very close in terms of MMSE.

The covariates

symbol	covariates description	covariates definition
U	Urea fertilizer	The urea fertilizer is a simple fertilizer that provides the main element nitrogen.
DS	Date of sowing	The date of planting wheat seeds in the field
SQ	Sowing seed quantity	The amount of wheat seeds in the field
TL	laser field leveling	The technique of Laser field leveling is a smoothing and leveling operation for farm land
NPK	Compound fertilizer	Compound fertilizer contains nitrogen, phosphorus, and potassium
SM	Seed sowing machine	Technicality of seed sowing machine is a machine that plants seeds in the ground
SP	Successive crop planting	Successive crop planting is a method of extending the harvest of the corps through staggered crop planting
K	High Potassium	Fertilizer with a high potassium content: Potassium is necessary for crop health
ME	Micro-Elements	Micro-Elements Fertilizer are mineral elements that crops need in extremely small quantities

5.2. The prostate cancer example

In this section, we implement the suggested approach for the standard datasets, namely the data on prostate cancer (Stamey et al., 1989). This dataset has been utilized for illustration in previous regularization papers. In this dataset, the logarithm of prostate-specified antigen is the outcome of interest. Here is a list describing briefly the response variable and 8 covariates.

lcavol	Log(volume of cancer)
lweight	Log(weight of the prostate)
age	Age
lbph	Log(The quantity of benign prostatic hyperplasia)
svi	Invasion of seminal vesicles
lcp	Log(capsular breakthrough)
gleason	The Gleason result
pgg45	The rate of Gleason results is four or five
lpsa	Log(prostatic specified antigan)

Table 8: MMSE for Prostate data analyses.

Method	MMSE
gLR	0.48
gBR	0.48
BgBR	0.47
BgBCQR	**0.45**

We compare the mean squared prediction errors (MMSE) for Prostate data analyses in Table 8, which shows that our suggested approach outperforms both the existing Bayesian and non-Bayesian approaches in terms of prediction accuracy.

Chapter six

Conclusions and Future Research

6. Conclusions and Future Research

6.1. Conclusions

This thesis has reviewed the literature on some Bayesian and non-Bayesian regularization methods. We have compared between lasso and rlasso using a real data example. The results show that both approaches have similar results. We have also proposed a Bayesian analysis of group bridge composite quantile regression, which includes the group lasso composite quantile regression as a special case.

We have proposed Gibbs sampler algorithm for posterior inference using a scale mixture of normals of the asymmetric Laplace distribution. The proposed algorithm uses prior distributions for the regression coefficients that are scale mixtures of multivariate uniform distributions with a particular Gamma distribution as a mixing distribution. Simulation examples show that the proposed algorithm is effective in regularization under a variety of scenarios. We have also illustrated the advantages of the new method on prostate data example. Hence, both the simulation and the prostate cancer data show strong support for the use of Bayesian group bridge composite quantile regression.

6.2. Main Contributions

We have made the following contributions:

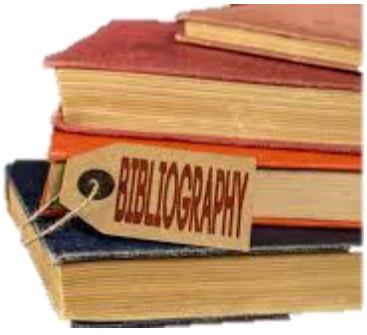
- We have summarized the literature review of some Bayesian and non-Bayesian regularization methods.
- We have proposed a Bayesian group bridge composite quantile regression.
- We have proposed a Bayesian group Lasso composite quantile regression.

- We have proposed an efficient Gibbs sampler algorithm for posterior inference.

6.3. Recommendations for Future Research

The work considered in this thesis can be extended in many directions, for example: one can extend the idea of Bayesian group bridge composite quantile regression to Bayesian composite Tobit quantile regression with group bridge penalty; Bayesian composite left censored quantile regression with group bridge penalty; Bayesian composite right censored quantile regression with group bridge penalty; and Bayesian composite interval censored quantile regression with group bridge penalty.

BIBLIOGRAPHY



Bibliography

1. Agresti, A. (2010). Analysis of ordinal categorical data (Vol. 656). John Wiley and Sons.
2. Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In Second International Symposium on Information Theory, B. N. Petrov and F. Csaki (eds), Budapest: Akademiai Kiado.
3. Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.
4. Alhamzawi, R. (2014). Bayesian elastic net tobit quantile regression. Communications in Statistics-Simulation and Computation (just-accepted), 00–00.
5. Alhamzawi, R. (2016). Bayesian analysis of composite quantile regression. Statistics in Biosciences 8 (2), 358–373.
6. Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian elastic net regression. Communications in Statistics-Simulation and Computation, 47(4), 1168-1178 .
7. Alhamzawi, R., & Taha Mohammad Ali, H. (2020). A new Gibbs sampler for Bayesian lasso. Communications in Statistics-Simulation and Computation, 49(7), 1855-1871 .
8. Alhamzawi, R., & Yu, K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. Computational Statistics & Data Analysis, 64, 209-219.Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian adaptive lasso regression. Mathematical biosciences, 303, 75-82.
9. Alhamzawi, R., Alhamzawi, A., & Ali, H. T. M. (2019). New Gibbs sampling methods for Bayesian regularized quantile regression. Computers in biology and medicine, 110, 52-65.

10. Alhamzawi, R., and Ali, H. T. M. (2018b). Bayesian tobit quantile regression with penalty. Communications in Statistics-Simulation and Computation, 47(6), 1739– 1750.
11. Alhamzawi, R., and Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, *39*(4), 799–813.
12. Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, *12*(3), 279-297.
13. Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34,pp. 85–105.
14. Alsaadi, Z., & Alhamzawi, R. (2022). Bayesian bridge and reciprocal bridge composite quantile regression. *Communications in Statistics-Simulation and Computation*, 1-18.
15. Alshaybawee, T., Midi, H., & Alhamzawi, R. (2017). Bayesian elastic net single index quantile regression. Journal of Applied Statistics, 44(5), 853-871.
16. Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika, 94(2), 443-458.
17. Andrews, D. F., and Mallows, C. L. (1974), "Scale Mixtures of Normal Distri- butions," Journal of the Royal Statistical Society, Ser. B, 36, 99–102.
18. Bae, K., and Mallick, B. K. (2004), "Gene Selection Using a Two-Level Hierarchical Bayesian Model," Bioinformatics, 20, 3423–3430.
19. Berg, A., Meyer, R., & Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. Journal of Business & Economic Statistics, 22(1), 107-120.

20. Bien, J., Taylor, J., & Tibshirani, R. (2013). A lasso for hierarchical interactions. *Annals of statistics*, *41*(3), 1111.
21. Bornn, L., Gottardo, R., & Doucet, A. (2010). Grouping priors and the Bayesian elastic net. arXiv preprint arXiv:1001.4083.
22. Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52(3), 345-370.
23. Bradic, J., Fan, J., & Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(3), 325-349.
24. Breheny, P. (2015). The group exponential lasso for bi-level variable selection. Bio- metrics 71 (3), 731–740.
25. Brown, P. J., & Griffin, J. E. (2010). Inference with normal-gamma prior distributions in regression problems. Bayesian analysis, 5(1), 171-188.
26. Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.
27. Cai, X., Huang, A., & Xu, S. (2011). Fast empirical Bayesian LASSO for multiple quantitative trait locus mapping. BMC bioinformatics, 12(1), 1-13.
28. Candès, E. J., & Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, *56*(5), 2053-2080.
29. Candes, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The annals of Statistics*, 35(6), 2313-2351.

30. Casella, G., Ghosh, M., Gill, J., & Kyung, M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian analysis*, *5*(2), 369-411.
31. Chen, X., Wang, Z. J., & McKeown, M. J. (2011). A Bayesian Lasso via reversible-jump MCMC. *Signal Processing*, *91*(8), 1920-1932.
32. Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. biometrika, 81(3), 425-455.
33. Dziak, J, Li R, Collins L (2005) Critical review and comparison of variable selection procedures for linear regression (Technical report)
34. Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of statistics, 32(2), 407-499.
35. Efroymson, M. A. (1960). Multiple regression analysis. Mathematical methods for digital computers, 191-203.
36. Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348-1360.
37. Feng, X. N., Wu, H. T., & Song, X. Y. (2017). Bayesian adaptive lasso for ordinal regression with latent variables. Sociological Methods & Research, 46(4), 926-953.
38. Figueiredo, M. A. T. (2003), "Adaptive Sparseness for Supervised Learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, 1150– 1159.
39. Flaih, A. N., Alshaybawee, T., & Alhusseini, F. H. H. (2020). Sparsity via new Bayesian Lasso. Periodicals of Engineering and Natural Sciences, 8(1), 345-359.
40. Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. Technometrics, 35(2), 109-135.

41. Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1), 1.
42. G´omez-S´anchez-Manzano, E., G´omez-Villegas, M., and Mar´ın, J. (2008). Multivariate exponential power distributions as mixtures of normal distributions with bayesian applications. Communications in Statistics—Theory and Methods 37 (6), 972–985.
43. Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. Statistics and computing, 24(6), 997-1016.
44. George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88(423), 881-889. Bayesian Analysis (2009) 4, Number 1, pp. 85–118
45. Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for gibbs sampling. Bayesian statistics 4 (2), 641–649.
46. Gómez-Villegas, M. A., Gómez-Sánchez-Manzano, E., Maín, P., & Navarro, H. (2011). The effect of non-normality in the power exponential distributions. In *Modern mathematical tools and techniques in capturing complexity* (pp. 119-129). Springer, Berlin, Heidelberg.
47. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.
48. Hans, C. (2009). Bayesian lasso regression. Biometrika, 96(4), 835-845.
49. Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. Journal of the American Statistical Association, 106(496), 1383-1393.

50. Himel Mallick, N. Y. (2013). Bayesian methods for high dimensional linear models. Journal of biometrics & biostatistics, 1, 005.
51. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
52. Huang, A., Xu, S., & Cai, X. (2015). Empirical Bayesian elastic net for multiple quantitative trait locus mapping. Heredity, 114(1), 107-115.
53. Huang, H., & Chen, Z. (2015). Bayesian composite quantile regression. *Journal of Statistical Computation and Simulation*, *85*(18), 3744-3754.
54. Huang, J., Breheny, P., & Ma, S. (2012). A selective review of group selection in high-dimensional models. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *27*(4).
55. 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, 1-20.
56. Huang, J., Ma, S., Xie, H., & Zhang, C. H. (2009). A group bridge approach for variable selection. *Biometrika*, *96*(2), 339-355.
57. Jacob, L., Obozinski, G., & Vert, J. P. (2009, June). Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning* (pp. 433-440).
58. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
59. Javed, F., & Mantalos, P. (2013). GARCH-type models and performance of information criteria. Communications in Statistics-Simulation and Computation, 42(8), 1917-1933.

60. Jiang, X., Jiang, J., & Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, 1479-1506.
61. Jolliffe, I. T. (1982). A note on the use of principal components in regression. Journal of the Royal Statistical Society: Series C (Applied Statistics), 31(3), 300-303.
62. Kang, J., & Guo, J. (2009). Self-adaptive Lasso and its Bayesian estimation. Working Paper.
63. Kaul, A. (2014). Lasso with long memory regression errors. Journal of Statistical Planning and Inference, 153, 11-26.
64. Kim, Y., Kim, J., & Kim, Y. (2006). Blockwise sparse regression. Statistica Sinica, 375-390.
65. Koenker, R. W., & d'Orey, V. (1987). Algorithm AS 229: Computing regression quantiles. Applied statistics, 383-393.
66. Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.
67. Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448), 1296-1310.
68. Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of statistical computation and simulation*, *81*(11), 1565-1578.
69. Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., & Fritz, S. (2011). Improved Lasso for genomic selection. Genetics research, 93(1), 77-87.
70. Leng, C., Tran, M. N., & Nott, D. (2014). Bayesian adaptive lasso. Annals of the Institute of Statistical Mathematics, 66(2), 221-244.

71. Li, J., Das, K., Fu, G., Li, R., & Wu, R. (2011). The Bayesian lasso for genome-wide association studies. Bioinformatics, 27(4), 516-523.
72. Li, Q., & Lin, N. (2010). The Bayesian elastic net. Bayesian analysis, 5(1), 151-170.
73. Li, Q., Xi, R., and Lin, N. (2010). Bayesian regularized quantile regression. Bayesian Analysis, 5(3), 533–556.
74. Li, Y., & Zhu, J. (2008). L 1-norm quantile regression. Journal of Computational and Graphical Statistics, 17(1), 163-185.
75. Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571-582.
76. Mallick, H., & Yi, N. (2017). Bayesian group bridge for bi-level variable selection. Computational statistics & data analysis, 110, 115-133.
77. Mallick, H., & Yi, N. (2018). Bayesian bridge regression. *Journal of applied statistics*, *45*(6), 988-1008.
78. Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V. (2021). The reciprocal Bayesian lasso. Statistics in medicine, 40(22), 4830-4849.
79. Mallows CL (1973) Some comments on CP. Technometrics 15 : 661–675.
80. Mazumder, R., Friedman, J. H., & Hastie, T. (2011). Sparsenet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, *106*(495), 1125-1138.
81. Meier, L., S. Van De Geer, and P. Bu¨hlmann (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Method- ology) 70 (1), 53–71.
82. Meinshausen, N. (2007). Relaxed lasso. Computational Statistics & Data Analysis, 52(1), 374-393.

83. Nardi, Y., & Rinaldo, A. (2011). Autoregressive process modeling via the lasso procedure. Journal of Multivariate Analysis, 102(3), 528-549.
84. Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. The Annals of Statistics, 758-765.
85. Park, C., & Yoon, Y. J. (2011). Bridge regression: adaptivity and group selection. *Journal of Statistical Planning and Inference*, *141*(11), 3506-3519 .
86. Park, T., & Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686.
87. Qian, W., Yang, Y., & Zou, H. (2016). Tweedie's compound Poisson model with grouped elastic net. *Journal of Computational and Graphical Statistics*, *25*(2), 606-625.
88. Ranstam, J., & Cook, J. A. (2018). LASSO regression. Journal of British Surgery, 105(10), 1348-1348.
89. Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.
90. Shimamura, Teppei, Hiroyuki Minami, and Masahiro Mizuta. "Regularization parameter selection in the group lasso." COMPSTAT, Capri, Italy (2006).
91. Simon, N., & Tibshirani, R. (2012). Standardization and the group lasso penalty. Statistica Sinica, 22(3), 983.
92. Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. *Journal of computational and graphical statistics*, *22*(2), 231-245.
93. Song, Q. (2018). An overview of reciprocal L 1-regularization for high dimensional regression data. Wiley Interdisciplinary Reviews: Computational Statistics, 10(1), e1416.

94. Song, Q., & Liang, F. (2015). High-dimensional variable selection with reciprocal l 1-regularization. Journal of the American Statistical Association, 110(512), 1607-1620.
95. Spiegelhalter DJ, Best NG, Carlin BP, Der Linde AV (2002) Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat ethodol 64: 583-639.
96. Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E., and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate, II: Radical prostatectomy treated patients. Journal of Urology 141, 1076–1083.
97. Sun, W., Ibrahim, J. G., & Zou, F. (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics, 185(1), 349-359.
98. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.
99. 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(1), 91-108.
100. Wang, H., & Leng, C. (2008). A note on adaptive group lasso. Computational statistics & data analysis, 52(12), 5277-5286.
101. Wang, H., Li, G., & Tsai, C. L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(1), 63-78.
102. Woodroofe, M. (1982). On model selection and the arc sine laws. The Annals of Statistics, 1182-1194.

103. Wu, L., Yang, Y., & Liu, H. (2014). Nonnegative-lasso and application in index tracking. Computational Statistics & Data Analysis, 70, 116-126.
104. Xu, X., & Ghosh, M. (2015). Bayesian variable selection and estimation for group lasso. Bayesian Analysis, 10(4), 909-936.
105. Yang, Y., & Wu, L. (2016). Nonnegative adaptive lasso for ultra-high dimensional regression models and a two-stage method applied in financial modeling. Journal of Statistical Planning and Inference, 174, 52-67.
106. Yi, N., & Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. Genetics, 179(2), 1045-1055.
107. Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437-447.
108. 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.
109. Yuan, M., and Lin, Y. (2005), "Efficient Empirical Bayes Variable Selection and Esimation in Linear Models," Journal of the American Statistical Association, 100, 1215–1225.
110. Zeng, P., Wei, Y., Zhao, Y., Liu, J., Liu, L., Zhang, R., ... & Chen, F. (2014). Variable selection approach for zero-inflated count data via adaptive lasso. Journal of Applied Statistics, 41(4), 879-894.
111. Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. The Annals of statistics, 38(2), 894-942.
112. Zhang, H. H., & Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. Biometrika, 94(3), 691-703.
113. Zheng, S. (2008). Selection of components and degrees of smoothing via lasso in high dimensional nonparametric additive models. Computational statistics & data analysis, 53(1), 164-175.

114. Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association 101 (476), 1418–1429 .
115. Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. The Annals of Statistics 36 (3), 1108–1126.
116. 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.
117. Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.

الخلاصة

ظهرت مشكلة عالية الأبعاد في العديد من تطبيقات الانحدار ، أي أن عدد المتغيرات المشتركة أكبر من حجم العينة. في هذه الحالة ، تكون تقديرات طريقة المربعات الصغرى العادية (OLS) غير مستقرة ، فضلاً عن وجود تباين كبير وتحيز مرتفع ، مما يؤدي إلى overfitting ووجود مشكلة التعدد الخطي لتقدير معلمات النموذج ، والتنبؤ السيئ للغاية ، وصعوبة تفسير نموذج المناسب. أصبحت الأساليب الإحصائية التقليدية مع هذه المشكلة غير ممكنة لاستخدامها في التحليل الإحصائي. وبالتالي ، صعوبة تقدير المعاملات واختيار المتغيرات المشتركة المهمة (المتغيرات المشتركة التي لها تأثير على المتغير التابع).

قارنا أداء طريقتين للتنظيم في هذه الرسالة : أقل الانكماش المطلق والاختيار عامل (lasso) ، ومعكوس لاسو (rlasso). أيضًا نقترح طريقة جديدة لإزالة المتغيرات المشتركة غير المهمة في البيانات عالية الأبعاد لتحسين دقة التنبؤ والحصول على تفسير أفضل. تسمى هذه الطريقة الانحدار القسيمي المركب البيزي بمقدرات مجموعة برج (BgBCQR). على وجه التحديد ، نقوم بتحسين النموذج الهرمي للطريقة المقترحة. نقدم خوارزمية MCMC جديدة للاستدلال اللاحق باستخدام مزيج مقياس من توزيعات الطبيعية لتوزيع لابلاس غير المتماثل (ALD) لتنفيذ بيزي الهرمي للطريقة المقترحة. قارنا طريقتنا المقترحة مع طرق التنظيم الأخرى للتحقق من فعالية الطريقة المقترحة من خلال إجراء دراسة لأمثلة المحاكاة وكذلك في تطبيق بيانات حقيقية لمقارنة أداء طرق التنظيم هذه.

تظهر نتائج المحاكاة وتحليلات البيانات الحقيقية أن أداء الطريقة المقترحة أكثر كفاءة وتتفوق على الأساليب الحالية من حيث دقة التنبؤ والاختيار المتغير وتقدير المعاملات. كما أنه يوفر تفسيرًا واضحًا.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية الادارة والاقتصاد
قسم الإحصاء
الدراسات العليا

الانحدار القسيمي المركب البيزي بمقدرات مجموعة برج

رسالة مقدمة

إلى مجلس كلية الادارة والاقتصاد ـ جامعة القادسية

جزءاً من متطلبات نيل درجة الماجستير في علوم الإحصاء

من الطالبة

ميادة جواد كاظم الجاسمي

إشراف

أ.د. رحيم جبار الحمزاوي

2022 م

1444 هـ