Prior Elicitation and Variable Selection for Bayesian Quantile Regression

A thesis submitted for the degree of

Doctor of Philosophy

by

Rahim Jabbar Thaher Al-Hamzawi

Supervised by

Dr. Keming Yu

Department of Mathematical Sciences
School of Information Systems, Computing and Mathematics
Brunel University, West London

May 2013

Abstract

Bayesian subset selection suffers from three important difficulties: assigning priors over model space, assigning priors to all components of the regression coefficients vector given a specific model and Bayesian computational efficiency (Chen et al., 1999). These difficulties become more challenging in Bayesian quantile regression framework when one is interested in assigning priors that depend on different quantile levels. The objective of Bayesian quantile regression (BQR), which is a newly proposed tool, is to deal with unknown parameters and model uncertainty in quantile regression (QR). However, Bayesian subset selection in quantile regression models is usually a difficult issue due to the computational challenges and non-availability of conjugate prior distributions that are dependent on the quantile level. These challenges are rarely addressed via either penalised likelihood function or stochastic search variable selection (SSVS). These methods typically use symmetric prior distributions for regression coefficients, such as the Gaussian and Laplace, which may be suitable for median regression. However, an extreme quantile regression should have different regression coefficients from the median regression, and thus the priors for quantile regression coefficients should depend on quantiles.

This thesis focuses on three challenges: assigning standard quantile dependent prior distributions for the regression coefficients, assigning suitable quantile dependent priors over model space and achieving computational efficiency. The first of these challenges is studied in Chapter 2 in which a quantile dependent prior elicitation scheme is developed. In particular, an extension of the Zellners prior which allows for a conditional conjugate prior and quantile dependent prior on Bayesian quantile regression is proposed. The prior is generalised in Chapter 3 by introducing a ridge parameter to address important challenges that may arise in some applications, such as multicollinearity and overfitting problems. The proposed prior is also used in Chapter 4 for subset selection of the fixed and random coefficients in a linear mixed-effects QR model. In Chapter 5 we specify normal-exponential prior distributions for the regression coefficients which can provide adaptive shrinkage and represent an

alternative model to the Bayesian Lasso quantile regression model.

For the second challenge, we assign a quantile dependent prior over model space in Chapter 2. The prior is based on the percentage bend correlation which depends on the quantile level. This prior is novel and is used in Bayesian regression for the first time. For the third challenge of computational efficiency, Gibbs samplers are derived and setup to facilitate the computation of the proposed methods.

In addition to the three major aforementioned challenges this thesis also addresses other important issues such as the regularisation in quantile regression and selecting both random and fixed effects in mixed quantile regression models.

Certificate of Originality

I declare that the work considered in this thesis entitled "Prior Elicitation and Variable Selection for Bayesian Quantile Regression" is my own work and was composed by myself during my study at Brunel University except where otherwise acknowledged.

.....

Rahim Alhamzawi

Acknowledgements

I would like to express my appreciation to many people who have been contributed their time, expertise and talent to complete my study and making my time at Brunel University successful. I am very proud and grateful to my supervisor Dr. KemingYu for his support, help, advice and his supervision of this thesis. I also want to express my thanks to Dr. Veronica Vinciotti for her support during my study, Professor Jianxin Pan for giving me the Age-Related Macular Degeneration data which is used in Chapter 4 and Dr Wei Sun for providing me with some information about prior selection, used in Chapter 5. In addition, I am heavily indebted to Dr. Abdallah Ally for being supportive and always offering great advices. I also would like to thank all fellow students and staff at Brunel University. Specifically, I would like to thank Ali Alkenani, Fatmir Qirezi, Hakim Mezali and Hussein Hashem. My thanks also go to Dr. Hamied Alhashimi and Dr. Majed Abdulla Altemimi.

Finally, I extremely appreciate the support, help and love of my family.

Author's Publication

1. Yu, K., Dang, W., Zhu, H. and Alhamzawi, R. (2013). Invited discussion by the Editor-in-Chief, Journal of Statistical Planning and Inference on Article "local quantile regression" by Spokoiny, Wang and Hardle. Accepted.

2. Alhamzawi, R. and Yu, K. (2012). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics and Data Analysis*. In press.

3. Alhamzawi, R. and Yu, K. (2012). Bayesian Lasso mixed quantile regression. *Journal of Statistical Computation and Simulation*. In press.

4. Alhamzawi, R. and Yu, K. and Benoit, D. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, 12: 279-297.

5. Alhamzawi, R. and Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, 39: 799-813.

6. Alhamzawi, R., Yu, K., Vinciotti, V. and Tucker, A. (2011), Prior elicitation for mixed quantile regression with an allometric model. *Environmetrics*, 22: 911-920.

7. Alhamzawi R, Yu, K., Pan, J. (2011). Prior Elicitation in Bayesian Quantile Regression for Longitudinal Data. *J Biomet Biostat* 2:115.

8. Alhamzawi, R. and Yu, K. (2011). Power Prior Elicitation in Bayesian Quantile Regression. *Journal of Probability and Statistics*, vol. 2011, Article ID 874907, 16 pages, 2011.

Revised Manuscripts

1. Alhamzawi, R., Benoit, D. and Yu, K. (2012). Bayesian Lasso Binary Quantile regression. *Computational Statistics*.

2. Alhamzawi, R. and Yu, K. (2012). Bayesian Tobit quantile regression using g-prior distribution with ridge parameter.

R Packages

1. Benoit, D., Alhamzawi, R., Yu, K., Poel, D. V. (2013). bayesQR: Bayesian quantile regression. R package version 2.1.

2. Alhamzawi, R. (2012). *Brq*: Bayesian analysis of quantile regression models. R package version 1.0.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Decades after its introduction by Koenker and Bassett (1978), the technique of quantile regression (QR) has been the subject of great theoretical interest as well as numerous practical applications in a number of fields such as, ecology, econometrics, biology, finance, social sciences and survival analysis; see Koenker (2005), Yu et al. (2003) and Cade and Barry (2003) for an overview. Like standard mean regression models, dealing with parameter and model uncertainty as well as updating information is of great importance for QR and its applications. One of the attractions of QR over its standard mean regression counterpart lies, in its ability to give a more strong investigation of the entire distribution of the relationship between an outcome of interest and its independent variables. To this end, QR is a very important technique and has steadily spread as a comprehensive extension to standard mean regression (Koenker, 2005). To highlight the importance of QR and demonstrate its application, by way of illustration we consider the US girls weight data (Cole, 1988), studied by Yu and Jones (1998). This data describes the relationship between the weight and age of 4011 individuals. In this section, we model the weight as a function of age using the following cubic model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i, \tag{1.1}$$

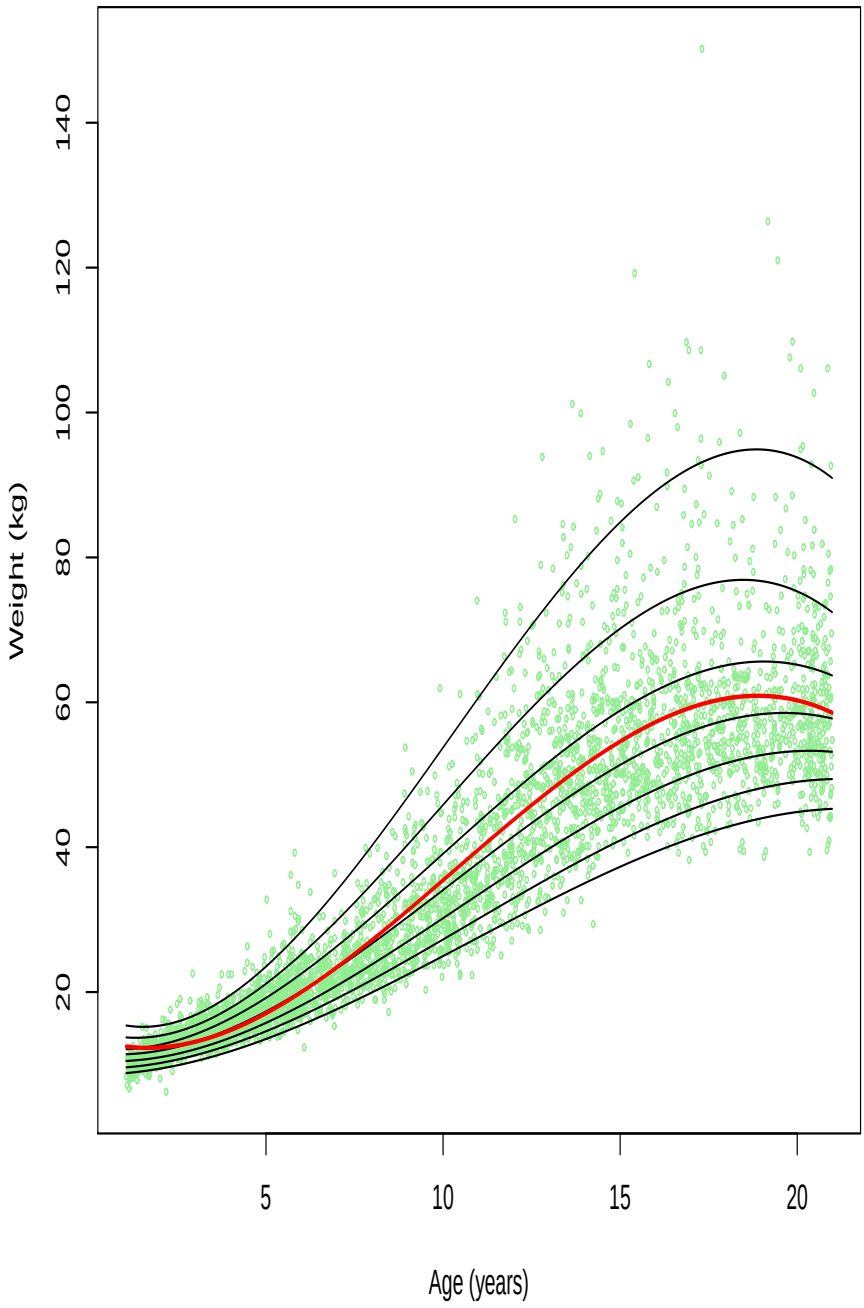

Figure 1.1: The panel depicts the relationship between weight and age of girls. QR curve estimates from the highest to the lowest quantiles are plotted for $p \in \{97\%, 90\%, 75\%, 50\%, 25\%, 10\%, 3\%\}$. The fitted standard mean regression curve is illustrated by the solid red curve.

where y_i is the ith outcome (weight) in kg, x_i is the ith age in years, ε_i is the ith residual term. Here, $\beta_0, \beta_1, \beta_2$ and β_3 are the QR coefficients. The relationship between weight and age is plotted in Figure 1.1 with seven fitted QR curves together with standard mean regression curve. The seven fitted QR curve estimates are fitted using the `rq()` function (Koenker, 2011).

From Figure 1.1 one can clearly indicate that the weight of girls tends to increase as age increases. Furthermore, we can see that the spacing of QR curves decreases from the highest ($p = 0.97$) to the lowest ($p = 0.03$) quantiles indicating that the distribution of $y|x$ is askew to the right, where p denotes the quantile level. Owing to potential outliers present near age 17 and the inherent right skewed conditional distribution, it can be observed that the conditional median (the 0.50 quantile) and mean curves are different and the standard mean regression estimate is insufficient to estimate the relationship between weight and age. Additionally, Figure 1.1 demonstrates that different quantiles, such as the 25th and 75th percentile of the weight, perhaps depend on the age of girls in a different form from the mean or the median. In summary, Figure 1.1 demonstrates that a group of quantiles provides a comprehensive tool to describe the relationship between weight and age compared with the standard mean regression. Moreover, as noted by Yu et al. (2003), the highest and lowest QR curves in Figure 1.1 can respectively be used as a proxy to identify the obesity and weight loss of girls, respectively. Furthermore, QR models are flexible models and insensitive to heteroscedastic errors and outliers in the outcome variable, which are popular in many real world applications (Koenker and Bassett, 1978; Koenker, 2005).

There are two techniques to estimate the QR coefficients of independent variables of the linear QR model. Both techniques of QR are considered in the next section.

1.1 Estimation

1.1.1 Classical quantile regression

Let $(\boldsymbol{x}_1', y_1), \cdots, (\boldsymbol{x}_n', y_n)$ represent a sample of observations. Then, the pth QR equation can be denoted as

$$Q_{y_i|\boldsymbol{x}_i}(p) = \boldsymbol{x}_i'\boldsymbol{\beta}_p, \quad p \in (0,1), \tag{1.2}$$

where y_i is the outcome of interest, $\boldsymbol{x}_i'$ is a $1 \times k$ vector denoting the ith row of the $n \times k$ design matrix $\boldsymbol{X}$, the unknown quantity $\boldsymbol{\beta}_p$ is a vector of k QR coefficients and $Q_{y_i|\boldsymbol{x}_i}(\cdot) = F^{-1}_{y_i|\boldsymbol{x}_i}(\cdot)$ is the inverse distribution function. Koenker and Bassett (1978) indicated that the unknown QR coefficients vector $\boldsymbol{\beta}_p$ can be evaluated as the solution to

$$\min_{\boldsymbol{\beta}_p} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_p), \tag{1.3}$$

where

$$\rho_p(\varepsilon) = \begin{cases} p\varepsilon & \text{if } \varepsilon \geq 0, \\ -(1-p)\varepsilon, & \text{if } \varepsilon < 0. \end{cases} \tag{1.4}$$

Equivalently, (1.4) is sometimes expressed as

$$\rho_p(\varepsilon) = \frac{|\varepsilon| + (2p-1)\varepsilon}{2}. \tag{1.5}$$

Figure 1.2 shows the check function at three different quantiles, namely 0.30, 0.20 and 0.10. Since the empirical check function, which is defined in (1.3), is not differentiable at 0, a closed-form solution is not available for the QR parameters vector $\boldsymbol{\beta}_p$ (Koenker, 2005). However, the minimisation of (1.3) can be achieved through an algorithm proposed by Koenker and D'Orey (1987). From a computational perspective, many well known statistical packages such as STATA and SAS can accommodate the estimation of the QR parameters and confidence intervals. In this

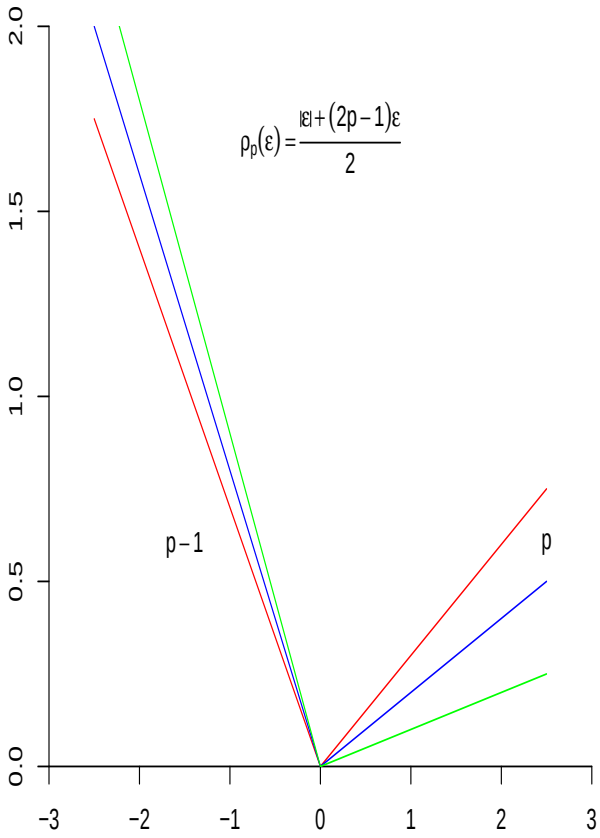

Figure 1.2: The panel shows the check function at $p = 0.30$ (red line), $p = 0.20$ (blue line) and $p = 0.10$ (green line).

thesis, the estimators are implemented using the `rq()` function in the R package `quantreg` (Koenker, 2011), and further information and explanations are given in Koenker (2005).

Alternatively, Koenker and Machado (1999) observed that minimising the empirical loss function of Koenker and Bassett (1978) is closely related to maximising the likelihood of the Asymmetric Laplace Distribution (ALD) and consequently the unknown quantile coefficients vector $\boldsymbol{\beta}_p$ can be estimated through exploiting this link. This observation, discussed in the proceeding subsection, opens new avenues when dealing with QR and its application.

1.1.2 Bayesian quantile regression (BQR) based on ALD

Following Yu and Zhang (2005), a random variable ε has an ALD(μ, τ, p) with $\mu = 0$, $\tau > 0$ and $0 < p < 1$ if its pdf is given by

$$f(\varepsilon; \mu = 0, \tau, p) = \frac{p(1-p)}{\tau} \exp\{-\rho_p(\varepsilon)\}, \tag{1.6}$$

where $\rho_p(\varepsilon) = (|\varepsilon| + (2p-1)\varepsilon)/(2\tau)$. Here, μ and τ represent the location and scale parameters, respectively. It is known that when $p = 0.5$ the probability density function in (1.6) is reduced to the standard symmetric form of the Laplace density, that is

$$f(\varepsilon; \mu = 0, \tau, p = 0.5) = \frac{1}{4\tau} \exp\{-\frac{|\varepsilon|}{2\tau}\}. \tag{1.7}$$

It is known that the expected value, variance, skewness (Sk) and kurtosis (Ku) of ε are respectively given by (Yu and Zhang, 2005)

$$\begin{aligned}
\mathbb{E}(\varepsilon) &= \frac{\tau(1-2p)}{p(1-p)}, \\
\mathbb{V}\text{ar}(\varepsilon) &= \frac{\tau^2(1-2p+2p^2)}{p^2(1-p)^2}, \\
\text{Sk} &= \frac{2(p^3-(1-p)^3)}{((1-p)^2+p^2)^{3/2}}, \\
\text{Ku} &= \frac{9p^4+6p^2(1-p)^2+9(1-p)^4}{(1-2p+2p^2)^2}.
\end{aligned}$$

Yu and Moyeed (2001) suggested a BQR approach where the errors are independently ALD distributed. This framework is developed by the parity of the maximum a posteriori estimator under the ALD and the check function estimator of Koenker and Bassett (1978); (see, Koenker and Machado, 1999). Later, this Bayesian framework has been extended by a number of researchers, and the evidence indicates that the ALD is simply a working model with artificial assumptions (Yuan and Yin, 2010). For example, Yu and Stander (2007) developed a Bayesian estimation procedure for a left censored QR, Geraci and Bottai (2007) proposed a Bayesian QR

framework for a random-intercept model using the ALD for the errors, Yuan and Yin (2010) implemented Bayesian mixed-effects QR for correlated data, Benoit and Poel (2011) considered binary QR from a Bayesian perspective, Lee and Neocleous (2010) proposed a Bayesian framework for QR for count data and Härdle et al. (2011) who developed an adaptation method for local QR based on the ALD, among others.

One of the attractive properties of the ALD is that it can be written as a member of the location-scale mixture of normals family; see Kozumi and Kobayashi (2011) and Reed and Yu (2009) who have independently shown that any variable that is ALD distributed can be expressed as

$$\mathrm{w} =^{d} \frac{1-2p}{p(1-p)}\tau \mathrm{w}_1 + \sqrt{\frac{2\mathrm{w}_1\tau}{p(1-p)}}\mathrm{w}_2, \quad \mathrm{w}_2 \sim N(0,1), \tag{1.8}$$

where w_1 is called the mixing variable with the standard exponential distribution, Exp(1). Here, $N(0,1)$ denotes the density function of a standard Gaussian distribution and the variables w_1 and w_2 are supposed to be independent. This approach connects the linear QR model for the outcome variable to the classical normal linear regression model. In addition, under this representation, the regression coefficients of independent variables, the scale parameter (τ) and the mixing variable (w_1) have desirable conjugacy features for constructing a simple Markov chain Monte Carlo (MCMC) technique for fitting the model to the data. This MCMC algorithm is implemented in an R function called `MCMCquantreg()` (Martin et al., 2011). The mixture representation appeared in papers by Li et al. (2010), Yue and Rue (2011), Burgette and Reiter (2012), Ji et al. (2012), Lum and Gelfand (2012), among others, to conduct Bayesian QR methods via Gibbs sampler. Recently, Khare and Hobert (2012) showed that the Gibbs sampler algorithm defined by Kozumi and Kobayashi (2011) and Reed and Yu (2009) converges at a geometric rate.

Another perspective of Bayesian QR appeared in papers by Reich et al. (2010) and Kottas and Krnjajić (2009), among others. For example, Kottas and Krnjajić (2009) proposed a semiparametric formulation based on the Dirichlet process and

Reich et al. (2010) considered a mixture model by assuming the errors are defined by a mixture of two Gaussian distributions.

1.2 Subset selection

Finding the significant independent variables plays the most crucial role in building a multiple regression model in many real world applications. Owing to removing irrelevant independent variables, the selection process provides a very good prediction performance as well as highlighting those independent variables, which are most important in fitting the model to the data (Griffin and Brown, 2010). However, classical subset selection methods, such as AIC (Akaike, 1973), Mallow's C_p (Mallows, 1973) and BIC (Schwarz, 1978) are often highly time consuming and maybe suffer from instability (Breiman, 1996). Recently, MCMC-based computation techniques for subset selection using stochastic search variable selection (SSVS) algorithms have become widely used in linear regression, generalised linear models, QR models and other modeling frameworks (George and McCulloch, 1993; Lee et al., 2003; Kinney and Dunson, 2007; Meligkotsidou et al., 2009; Reed et al., 2009; Fahrmeir et al., 2010; Ji et al., 2012, among others). However, SSVS consumes a lot of time in some applications, such as chemometrics or bioinformatics, and consequently the method suffers from computational difficulties (Griffin and Brown, 2010). Despite these undesirable properties, in practice SSVS produces good promising models compared to other approaches.

Subset selection by shrinkage and selection of the coefficients of independent variables has attracted much interest in recent years; see for instance, Lasso (Tibshirani, 1996), fused Lasso technique (Tibshirani et al., 2005), SCAD (Fan and Li, 2001), the elastic net method (Zou and Hastie, 2005), group Lasso method (Yuan and Lin, 2005b) and the graphical Lasso (Yuan and Lin, 2007), among others. From a Bayesian point of view, Park and Casella (2008) proposed Lasso-based model, Sun et al. (2010) suggested Bayesian regression with the adaptive version of Lasso penalty,

and Polsen and Scott (2011) considered and developed the global-local regularised technique.

With regard to QR, Koenker (2004) introduced an l_1-regularisation QR approach for clustered data to shrink the random coefficients towards the origin and Geraci and Bottai (2007) developed a Bayesian QR approach for clustered dataset using an automatic technique to shrink the random coefficients to the origin. Additionally, Wang et al. (2007) proposed the least absolute deviation technique, Zou and Yuan (2008) introduced the idea of the composite QR and the authors have shown the composite QR with adaptive version of the Lasso penalty enjoys the oracle properties, Li and Zhu (2008) studied the QR with l_1 penalty, Wu and Liu (2009) developed regularised QR using the SCAD penalty and the adaptive version of the Lasso penalty. Recently, Yuan and Yin (2010) introduced an l_2 norm check function to shrink the random effects towards the origin and Li et al. (2010) developed Bayesian shrinkage techniques for QR.

1.3 Thesis outline

In Chapter 2, we suggest a modification of Zellners g-prior in QR as well as presenting the Bayesian MCMC estimation procedure. For subset selection, we propose a novel prior based on percentage bend correlation over model space. Most of Chapter 2 has been published in Alhamzawi and Yu (2012b).

The focus of Chapter 3 is on Bayesian subset selection and coefficient estimation in Tobit QR model. In this chapter, the modified g-prior has been used and generalised by introducing a ridge parameter inside the variance covariance matrix to deal with some problems such as multicollinearity and overfitting problems that may arise with left-censored data. Some possible extensions of the proposed technique are also considered and outlined, including the continuous and binary responses in QR. The performance of the proposed techniques are examined via simulation scenarios and using leukemia dataset described in subsection 1.4.2. This Chapter is a revised

manuscript in Alhamzawi and Yu (2013).

In Chapter 4, the selection of both fixed and random coefficients in quantile mixed effects models is presented from Bayesian framework. Some possible extensions of the proposed technique are also considered. Several advantages of the proposed approach over existing approaches are discussed. Simulation studies and an age-related macular degeneration data are given to demonstrate the methodology. Most of Chapter 4 has been published in Alhamzawi and Yu (2012a).

Chapter 5 addresses the QR with the adaptive version of the Lasso penalty from Bayesian framework. In particular, we propose Bayesian adaptive Lasso QR (BALQR) using an ALD-based model. The performance of the BALQR is considered via simulation scenarios and using prostate cancer data described in subsection 1.4.4. This chapter has been published in Alhamzawi et al. (2012).

Finally, Chapter 6 summaries the thesis and providing recommendations for future researches in the QR area.

Each chapter of this thesis is presented in the form of an article, thus enabling the reader to clearly understanding the aims, techniques, main findings and conclusions of each chapter.

1.4 Real Data

This section provides brief descriptions of the real data sets that will be used in this thesis to illustrate the applications of the proposed methods throughout the thesis.

1.4.1 Air pollution data

Data measured by the *Public Roads Administration in Norway* is used to test the behavior of the proposed methods in Chapter 2. Specifically, a subsample consisting of 500 observations on 7 independent variables plus an outcome variable, collected between October 2001 and August 2003 are used. The outcome variable is hourly

values of the log(concentration of NO_2) and the dataset is available in the R package `truncSP` reported by Lindmark and Karlsson (2012).

1.4.2 Leukemia data

The popular leukemia data (Golub et al., 1999), which can be retrieved from `http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi`, is used to test the behavior of the proposed method in Chapter 3. This dataset contains 7129 gene expression values taken over 72 leukemia patients and was previously analysed using various Bayesian and non Bayesian approaches (Golub et al., 1999; Bae and Mallick, 2004; Yang and Song, 2010, among others).

1.4.3 Age-related macular degeneration (ARMD) data

To test the performance of the proposed method in Chapter 4, we use ARMD which has been previously analysed by Chaili (2008). The data has a total of 203 patients who were randomly selected from three cities in the United Kingdom (70 patients from London, 84 from Belfast and the remaining from Southampton) to measure the treatment effects of teletherapy on the loss of vision. The outcome of interest in this study is the change in Distance Visual Acuity (DVA) which was measured four times for each patient over a period of 24 months.

1.4.4 Prostate cancer data

In Chapter 5, we analyse prostate cancer data available in the R-package "`bayesQR`" (Benoit et al., 2011) to test the performance of the proposed method. This data reported by Stamey et al. (1989) and analysed by many authors (see for example, Tibshirani, 1996; Yuan and Lin, 2005a). A number of clinical measures were recorded based on 97 male patients who were suffering from prostate cancer. The outcome

variable in this dataset is the level of prostate-specific antigen [PSA] (also called gamma-seminoprotein) which is commonly used as an indicator of prostate cancer.

Chapter 2

Conjugate priors and variable selection for Bayesian QR

Bayesian subset selection in quantile regression (QR) models is usually a difficult issue due to the computational challenges and non-availability of conjugate prior distributions that depend on the quantile level. These challenges are rarely addressed via either penalised likelihood functions or stochastic search variable selection (SSVS). These methods typically use symmetric prior distributions such as normal or Laplace distributions for regression coefficients, which may be suitable for median regression. However, an extreme QR should have different regression coefficients from the median regression, and thus the priors for QR should depend on the quantile. In this chapter an extension of the Zellner's prior that allows for a conditional conjugate prior and quantile dependent prior on Bayesian QR is proposed. Secondly, a novel prior based on the percentage bend correlation for model selection is also used in Bayesian regression for the first time. Thirdly, a MCMC-based computation algorithm is developed to facilitate the calculations. The proposed approaches are illustrated with both simulation scenarios and air pollution data.

2.1 Introduction

Bayesian subset selection in standard mean regression suffers from the following three difficulties: assigning a prior $p(\mathcal{S})$ for each subset $\mathcal{S}$ in the model space, assigning a prior $p(\boldsymbol{\beta}_{\mathcal{S}}|\mathcal{S})$ for $\boldsymbol{\beta}_{\mathcal{S}}$ given a specific model $\mathcal{S}$ and Bayesian computational efficiency (Chen et al., 1999). These difficulties become more challenging in QR framework when one is interested in assigning prior distributions characterised by a p-dependent parameter. As mentioned in Chapter 1, from a Bayesian point of view, Yu and Moyeed (2001) suggested a BQR aproach assigning the ALD-based working model and sampling the unknown quantile coefficients vector $\boldsymbol{\beta}_p$ using a MCMC algorithm. The authors assigned flat priors for all components of the unknown quantile coefficients vector $\boldsymbol{\beta}_p$. A serious challenge in Bayesian QR lies in specifying a quantile dependent prior for $\boldsymbol{\beta}_p$. It is well known that a conjugate prior distribution that depends on the quantile level is not available for regression coefficients in QR models (Yu and Moyeed, 2001; Yu and Stander, 2007). Thus, Bayesian quantile inference models, including Bayesian parametric, semiparametric and nonparametric models, either set priors independently of the values of the quantiles, or assume the prior to be the same for modelling different order of quantiles. In doing so, this approach may result in inflexibility in quantile modelling. For example, a 95% QR model should have different parameter values from the median quantile, and thus the priors used for modelling the quantiles should be different. It is therefore more reasonable to set different priors for different quantiles.

A second serious challenge in QR lies in Bayesian variable selection, due to the challenge in specifying a quantile dependent prior over model space. At the present time, all Bayesian variable selection approaches in QR set priors independent of the value of quantiles over model space (see, Meligkotsidou et al., 2009; Reed et al., 2009; Ji et al., 2012, among others). Finally, another serious challenge encountered in modelling with Bayesian QR lies in computational efficiency.

These three challenges are addressed in the rest of this chapter. For the first,

it is crucial to elicit a prior distribution for QR coefficients that is as informative as possible, and more crucially, that depends on the quantile level. To address this challenge a quantile dependent conjugate prior distribution is proposed. For the second, the percentage bend correlation is used to find suitable prior distributions over subset space and to address the third difficulty a MCMC-based computation algorithm is derived to facilitate the computations.

The rest of this chapter is organised as follows. Section 2.2 introduces a modification of Zellner's g-prior in QR as well as presenting the Bayesian MCMC-based computation estimation. An outline of the prior assumptions and a simple MCMC algorithm for model selection are addressed in Section 2.3, and in Section 2.4 simulation scenarios are implemented to test the behavior of the proposed techniques for subset selection and estimation. Section 2.5 provides an illustration of the proposed methods using the air pollution data described in subsection 1.4.1. A chapter summary follows in Section 2.6.

2.2 Methods

2.2.1 Zellner's informative g-prior

It is well known that conjugate priors play very crucial roles in Bayesian probability theory as it is attractive to have conditional distributions that have a closed form under sampling (Chen and Ibrahim, 2003). In standard mean regression, various approaches for assigning prior distributions for regression coefficients of independent variables and variance in the nature of closed form under sampling have been proposed over the years. However, it is not easy to assess the prior covariance matrix for regression coefficients (Zellner, 1983; Agliari and Parisetti, 1988).

Zellner (1983, 1986) proposed a procedure for evaluating a conjugate prior distribution referred to as Zellner's informative g-prior, or simply g-prior. The g-prior has been vastly used in the situation of Bayesian analysis for the mean regression

models, due to the verity that analytical results are more readily available, better computational efficiency and its simple interpretation (Krishna et al., 2008).

Let $\boldsymbol{y} = (y_1, \cdots, y_n)'$ be a vector of outcomes, $\boldsymbol{x}_i'$ a $1 \times k$ vector denoting the ith row of the $n \times k$ matrix of predictors $\boldsymbol{X}$. A standard linear regression model can be denoted as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \qquad \boldsymbol{e} \sim \mathrm{N}(0, \tau^2 \boldsymbol{I}_n),$$

with a vector of regression coefficients $\boldsymbol{\beta}$, Zellner's informative g-prior based on a sample of n observations and k regression coefficients of independent variables can be written as (Zellner, 1983)

$$\begin{aligned} p(\boldsymbol{\beta}, \tau | \boldsymbol{\beta}_a, \tau_a, \boldsymbol{y}, \boldsymbol{X}) \propto & \\ & \tau^{-(n-k+1)} \exp\{-(n-k-2)\tau_a^2/2\tau^2\} \\ & \times \tau^{-k} \exp\left\{-(\boldsymbol{\beta} - \boldsymbol{\beta}_a)'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_a)/2g\tau^2\right\}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\beta}_a$ and τ_a are anticipated values, and $g > 0$ is a known scaling factor. The choices of the scaling factor g are discussed later in subsection 2.2.3. Agliari and Parisetti (1988) proposed an extension of g-prior by allowing different possible weights for different independent variables. This extension can be written as (Agliari and Parisetti, 1988)

$$\begin{aligned} p(\boldsymbol{\beta}, \tau | \boldsymbol{\beta}_a, \tau_a, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{C}) \propto & \\ & \tau^{-(n-k+1)} \exp\{-(n-k-2)\tau_a^2/2\tau^2\} \\ & \times \tau^{-k} \exp\left\{-(\boldsymbol{\beta} - \boldsymbol{\beta}_a)'\boldsymbol{C}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{C}(\boldsymbol{\beta} - \boldsymbol{\beta}_a)/2g\tau^2\right\}, \end{aligned} \tag{2.2}$$

where $\boldsymbol{C} = \mathrm{diag}[c_j \geq 0], \quad j = 1, \cdots, k.$

In the next subsection, we suggest a modification of Zellner's informative g-prior in QR to take into account different priors for different quantile levels.

2.2.2 Conditional conjugate prior distribution

In this thesis, we assume an ALD-based working model such that $\varepsilon_i \sim \text{ALD}(0, \tau, p)$, with a likelihood given by

$$\ell(\boldsymbol{\varepsilon}|\tau) \propto \tau^{-n} \exp\{-\sum_{i=1}^{n} \frac{|\varepsilon_i| + (2p-1)\varepsilon_i}{2\tau}\}, \tag{2.3}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$. Following Reed and Yu (2009) and Kozumi and Kobayashi (2011), the likelihood function (2.3) can be written as a member of the scale mixture of normals family as follows. For any $a_1, a_2 > 0$, we have (Andrews and Mallows, 1974)

$$\exp\{-|a_1 a_2|\} = \int_0^\infty \frac{a_1}{\sqrt{2\pi v}} \exp\{-\frac{1}{2}(a_1^2 v + a_2^2 v^{-1})\} dv. \tag{2.4}$$

If we assume $a_1 = 1/\sqrt{2\tau}$, $a_2 = \varepsilon/\sqrt{2\tau}$ and multiplying by $\exp\{-(2p-1)\varepsilon/2\tau\}$ yields

$$\begin{aligned} &\tau^{-n} \exp\{-\sum_{i=1}^{n} \frac{|\varepsilon_i| + (2p-1)\varepsilon_i}{2\tau}\} \\ &\quad = \prod_{i=1}^{n} \int_0^\infty \frac{1}{\tau\sqrt{4\pi\tau v_i}} \exp\{-\frac{(\varepsilon_i - \xi v_i)^2}{4\tau v_i} - \zeta v_i\} dv_i, \\ &\quad \propto \prod_{i=1}^{n} \int_0^\infty \text{N}(\varepsilon_i|\xi v_i, 2\tau v_i)\text{Exp}(v_i|\zeta) dv_i \end{aligned} \tag{2.5}$$

where $\xi = (1-2p)$ and $\zeta = p(1-p)/\tau$. Here, $\text{N}(x_1|\mu_1, \sigma_1^2)$ and $\text{Exp}(x_2|\theta)$ denote the densities of a Gaussian distribution with mean μ_1 and variance σ_1^2 and an exponential distribution with rate parameter θ, respectively. Following Zellner (1983), we consider an imaginary sample, $\boldsymbol{y}_0 = (y_{01}, y_{02}, ..., y_{0n})'$, generated by

$$y_{0i} = \boldsymbol{x}_i' \boldsymbol{\beta}_p + \varepsilon_{0i}, \quad i = 1, 2, ..., n, \tag{2.6}$$

and we assume that errors ε_{0i} are asymmetric Laplace distributed such that $\text{ALD}(0, g\tau, p)$. For simplicity of notation, henceforth we will omit the quantile level

p in the unknown quantile coefficients vector, $\boldsymbol{\beta}_p$. Then, the conditional posterior distribution of $p(\boldsymbol{\beta}, \tau | v_1, \cdots, v_n, \boldsymbol{y}_0, \boldsymbol{X})$ based on (2.6) and a prior distribution $p(\boldsymbol{\beta}, \tau) \propto \tau^{-1}$ is given by

$$\begin{aligned}
&p(\boldsymbol{\beta}, \tau | v_1, \cdots, v_n, \boldsymbol{y}_0, \boldsymbol{X}) \\
&\quad \propto (\frac{1}{\tau})^{\frac{3n}{2}+1} \exp\left\{ -\sum_{i=1}^{n} [\frac{(y_{0i} - \boldsymbol{x}_i'\boldsymbol{\beta} - \xi v_i)^2}{4g\tau v_i} + \zeta v_i] \right\}, \\
&\quad \propto (\frac{1}{\tau})^{\frac{3n}{2}+1} \exp\left\{ -\sum_{i=1}^{n} [\frac{(y_{0i} - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - \xi v_i)^2 + (\boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{x}_i'\boldsymbol{\beta}_0)^2}{4g\tau v_i} + \zeta v_i] \right\}. \\
&\quad = (\frac{1}{\tau})^{\frac{3n-k}{2}+1} \exp\left\{ -\sum_{i=1}^{n} \frac{(y_{0i} - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - \xi v_i)^2 + 4gp(1-p)v_i^2}{4g\tau v_i} \right\} \\
&\qquad \times \quad (\frac{1}{\tau})^{\frac{k}{2}} \exp\left\{ -\frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{4g\tau} \right\},
\end{aligned} \tag{2.7}$$

where

$$\hat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{y}_0 - \xi\boldsymbol{v}), \tag{2.8}$$

and $\hat{\boldsymbol{\beta}}_0$ is the estimated value of $\boldsymbol{\beta}_0$. Here, $\boldsymbol{V} = \text{diag}(v_1^{-1}, \cdots, v_n^{-1})$ and $\boldsymbol{v} = (v_1, \cdots, v_n)'$. Similar to Zellner (1983) and Agliari and Parisetti (1988), conditional conjugate prior distribution in the normal-inverse gamma family form for the unknown quantile coefficients vector $\boldsymbol{\beta}$ and the scale parameter τ can be obtained by substituting $\boldsymbol{\beta}_{ap} = \hat{\boldsymbol{\beta}}_0$ and $(3n - k - 2)\tau_{ap} = \sum_{i=1}^{n}[(y_{0i} - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - \xi v_i)^2 + 4gp(1 - p)v_i^2]/(2gv_i)$, where $\boldsymbol{\beta}_{ap}$ and τ_{ap} are anticipated values at a given quantile level for $\boldsymbol{\beta}$ and τ, respectively. That is,

$$\begin{aligned}
p(\boldsymbol{\beta}, \tau | \boldsymbol{v}, \boldsymbol{y}_0, \boldsymbol{X}) \quad &\propto (\frac{1}{\tau})^{\frac{3n-k}{2}+1} \exp\left\{ -\frac{(3n - k - 2)\tau_{ap}}{2\tau} \right\} \\
&\times (\frac{1}{\tau})^{\frac{k}{2}} \exp\left\{ -\frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_{ap})'\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ap})}{4g\tau} \right\}.
\end{aligned} \tag{2.9}$$

Given p and $\boldsymbol{v}$, the prior mean vector of $\boldsymbol{\beta}$ in (2.9) is $\mathbb{E}(\boldsymbol{\beta}|\boldsymbol{v}) = \boldsymbol{\beta}_{ap}$ and the covariance matrix of QR coefficients $\boldsymbol{\beta}$ is $\text{Cov}(\boldsymbol{\beta}|\boldsymbol{v}) = 2g\tau_{ap}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}$. Thus, given p, $\boldsymbol{v}$, τ_{ap} and $\boldsymbol{\beta}_{ap}$, the standard conditional prior distribution for $\boldsymbol{\beta}$ and τ in (2.7) is

readily available. As in Zellner (1986), we suggest a conjugate normal-inverse gamma distribution at a given quantile level p for $\boldsymbol{\beta}$ and τ given by

$$\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N(0, 2g\tau(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}), \qquad p(\tau) \propto \tau^{-1}. \tag{2.10}$$

This choice for the regression coefficients of independent variables $\boldsymbol{\beta}$ has several attractive features. First, under this setting this prior has the very attractive property that is dependent on the quantile level. Thus, we have different priors for different quantiles. Second, the prior (2.10) is conditionally conjugate prior, a feature that is employed in constructing an efficient MCMC-based computation technique. Third, in the case of $0.5\tau\boldsymbol{V} = \tau^2 I_n$, the proposed prior is reduced to the original g-prior, i.e. $p(\boldsymbol{\beta}|\tau^2, \boldsymbol{X}) = N(0, g\tau^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$. Finally, as $g \longrightarrow \infty$, the modified g-prior distribution for $\boldsymbol{\beta}$ converges to Jeffrey's prior of the form $p(\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X}) \propto |\boldsymbol{X}'\boldsymbol{V}_0\boldsymbol{X}|^{1/2}$, where $\boldsymbol{V}_0 = \text{diag}((2\tau v_1)^{-1}, \cdots, (2\tau v_n)^{-1})$.

2.2.3 Choices of g

Various values of g have been assigned in the context of estimation of the regression coefficients of independent variables and subset selection. For instance, Kass and Wasserman (1995) proposed the idea of the unit-information by assuming the scaling factor g is equal to the sample size, that is $g = n$. Smith and Kohn (1996) considered Bayesian subset selection using splines and suggested that the reasonable value of the scaling factor g is in the range $10 \leq g \leq 1000$. Following this suggestion, a number of authors set $g = 100$ (see for example, Lee et al., 2003; Gupta et al., 2007; Chen et al., 2011, among others). Although for normal linear regression, placing an Inverse Gamma prior on the scaling factor g, $g \sim \text{InvGa}(1/2, n/2)$, produces a Cauchy prior on $\boldsymbol{\beta}$, which is a robust prior for regression coefficients (Clyde and George, 2004; Zellner and Siow, 1980; Kinney and Dunson, 2007), the marginal likelihood of the data $f(\boldsymbol{y}|\boldsymbol{\gamma})$ has no closed form, where $\boldsymbol{\gamma}$ is a latent k-vector with binary entries: $\gamma_j = 1$ if the jth independent variable (x_j) is active in the regression equation and

$\gamma_j = 0$ otherwise. Owing to this undesirable property, Liang et al. (2008) suggested the idea of the hyper-g prior which has attracted much interest in recent years. Celeux et al. (2012) suggested a Jeffrey's prior on the scaling factor g. In this thesis, we follow the suggestion given by Smith and Kohn (1996) and choose $g = 100$.

2.2.4 Posterior inference

The conditional posterior distribution, combining the likelihood function of the data $\ell(\boldsymbol{y}|\boldsymbol{\beta}, \tau, \boldsymbol{v})$, and the proposed prior for $\boldsymbol{\beta}$ and τ (2.10) is given by

$$p(\boldsymbol{\beta}, \tau, \boldsymbol{v}|\boldsymbol{y}) \propto \ell(\boldsymbol{y}|\boldsymbol{\beta}, \tau, \boldsymbol{v}) p(\boldsymbol{\beta}|\tau, \boldsymbol{v}) p(\boldsymbol{v}|\tau) p(\tau). \tag{2.11}$$

A MCMC based computation technique is constructed to update the parameters $\boldsymbol{\beta}, \tau$, and $\boldsymbol{v}$ from their full conditional distributions. Let N_k, InvGa and GIG denote a k-dimensional multivariate normal, Inverse Gamma and generalised inverse Gaussian distributions, respectively.

- **Updating $\boldsymbol{\beta}$**

  The full Conditional Distribution (CD) of $\boldsymbol{\beta}$ is $\mathrm{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

  $$\boldsymbol{\Sigma} = \frac{2\tau g}{(g+1)}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1} \quad \text{and} \quad \boldsymbol{\mu} = \frac{g}{g+1}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{y} - \xi\boldsymbol{v}).$$

- **Updating τ**

  $$\tau|\boldsymbol{\beta}, \boldsymbol{v} \sim \mathrm{InvGa}((3n+k)/2, \frac{1}{4}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \xi\boldsymbol{v})'\boldsymbol{V}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \xi\boldsymbol{v}) + \frac{1}{4g}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})\boldsymbol{\beta} + p(1-p)\sum_{i=1}^{n} v_i).$$

- **Updating $\boldsymbol{v}$**

  Each $v_i, i = 1, \cdots, n$, has a full CD proportional to

  $$v_i^{-1} \exp\left\{-\frac{1}{2}(v_i^{-1}\varrho_1^2 + v_i\varrho_2^2)\right\},$$

where $\varrho_1^2 = ((y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \boldsymbol{\beta}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\beta}/g)/2\tau$ and $\varrho_2^2 = 1/2\tau$, which can be expressed as a GIG $(\nu, \varrho_1, \varrho_2)$. Recall that if $x \sim$ GIG $(\nu, \varrho_1, \varrho_2)$ then the pdf of x is given by (Barndorff-Nielsen and Shephard, 2001)

$$f(x|\nu, \varrho_1, \varrho_2) = \frac{(\varrho_2/\varrho_1)^\nu}{2K_\nu(\varrho_1\varrho_2)} x^{\nu-1} \exp\left\{-\frac{1}{2}(x^{-1}\varrho_1^2 + x\varrho_2^2)\right\},$$

where $x > 0$, $-\infty < \nu < \infty$, $\varrho_1, \varrho_2 \geq 0$ and $K_\nu(.)$ is so called "*modified Bessel function of the third kind*".

2.3 Stochastic search variable selection (SSVS)

The Bayesian SSVS reported in George and McCulloch (1993) opens an avenue to subset selection by using higher posterior probability to identify promising models. Meligkotsidou et al. (2009), Reed et al. (2009) and Ji et al. (2012) extend this approach for subset selection in QR models. However, all these QR methods had the disadvantages of depending on prior distributions that are independent of the value of quantiles. These methods may lead in an inflexibility in QR modelling. Our goal in this chapter is to use the modified g-prior (2.10) in a subset selection problem. In particular, we extend the idea of Bayesian subset selection in QR reported in Reed et al. (2009) by assigning quantile dependent priors on the subset space and a quantile dependent prior distribution on the regression coefficients given a specific subset.

In order to perform the subset selection for the QR equation (1.2) we begin by defining an indicator vector $\boldsymbol{\gamma}' = (\gamma_1, ..., \gamma_k)$ with jth element γ_j such that $\gamma_j = 1$ if the jth independent variable (x_j) is active in the regression $(\beta_j \neq 0)$, and $\gamma_j = 0$ if the jth independent variable (x_j) is not active in the regression $(\beta_j = 0)$. Given the binary vector $\boldsymbol{\gamma}$, let $k_\gamma = \boldsymbol{\gamma}'1$ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and $\boldsymbol{x}_{i,\boldsymbol{\gamma}}$ are $k_\gamma \times 1$ vectors corresponding to all the components of $\boldsymbol{\beta}$ and $\boldsymbol{x}_i$ such that the corresponding γ_j's are equal to 1.

Given the quantile level p, we consider the following prior assumptions for $\boldsymbol{\beta}, \tau, \boldsymbol{v}$, and $\boldsymbol{\gamma}$:

1. $\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\tau, \boldsymbol{v} \sim N(\boldsymbol{\beta}_0, 2g\tau(\boldsymbol{X}'_{\boldsymbol{\gamma}}\boldsymbol{V}\boldsymbol{X}_{\boldsymbol{\gamma}})^{-1})$, where $\boldsymbol{\beta}_0 = 0$, $g > 0$, $p(\tau) \propto \tau^{-1}$ and each $v_i \sim Exp\{p(1-p)/\tau\}$.

2. Yuan and Lin (2005a) proposed a prior distribution over model space given by

$$p(\boldsymbol{\gamma}|\pi) \propto \pi^{k_\gamma}(1-\pi)^{k-k_\gamma}|\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^*}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^*}|^{1/2}, \quad 0 \leq \pi \leq 1, \tag{2.12}$$

where $\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^*}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^*}$ is the correlation matrix. Here, $\boldsymbol{X}_{\boldsymbol{\gamma}^*}$ have been standardized, referred to as "$\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^*}$", and $\boldsymbol{\gamma}^*$ is the model of $\boldsymbol{\gamma}$ without an intercept. We assume $k_\gamma = \boldsymbol{\gamma}^{*\prime}1 + 1$ if the intercept is included in the model $\boldsymbol{\gamma}$ and $k_\gamma = \boldsymbol{\gamma}^{*\prime}1$ otherwise. Under the prior (2.12), the prior odds ratio is given by

$$\frac{p(\gamma_j = 1|\boldsymbol{\gamma}_{-j})}{p(\gamma_j = 0|\boldsymbol{\gamma}_{-j})} = \frac{\pi}{(1-\pi)}r_0, \quad r_0 = \frac{|\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^{*}-j,\gamma_j=1}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^{*}-j,\gamma_j=1}|^{1/2}}{|\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^{*}-j,\gamma_j=0}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^{*}-j,\gamma_j=0}|^{1/2}}.$$

Here, $\boldsymbol{\gamma}^*_{-j}$ is the binary vector of $\boldsymbol{\gamma}^*$ without γ_j. As pointed out by Yuan and Lin (2005a), r_0 is small when x_j is strongly correlated with $\boldsymbol{X}^*{}_{\boldsymbol{\gamma}-j,\gamma_j=0}$ and consequently x_j can effectively be excluded from the whole model. However, it is known that the usual correlation coefficient is highly nonrobust (Wilcox, 1994). Additionally, with regards to QR, $|\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^*}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^*}|^{1/2}$ does not depend on the quantile level. To this end, we remedy these two undesirable properties by employing robust correlation coefficients as well as incorporating the quantile level into a prior of $\boldsymbol{\gamma}$ such that we have different priors for different quantiles over model space. In this respect, we suggest the following prior over subset space:

$$p(\boldsymbol{\gamma}|\pi) \propto \pi^{k_\gamma}(1-\pi)^{k-k_\gamma}|\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^*{}_{pb}}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^*{}_{pb}}|^{1/2},$$

where $\boldsymbol{X}^{*\prime}{}_{\boldsymbol{\gamma}^*{}_{pb}}\boldsymbol{X}^*{}_{\boldsymbol{\gamma}^*{}_{pb}}$ is the percentage bend correlation matrix. The percentage bend correlation, r_{pb}, between t_1 and t_2 is given by the following equation

(Wilcox, 1994)

$$r_{pb} = \frac{\sum A_i B_i}{\sqrt{\sum A_i^2 \sum B_i^2}}$$

where $A_i = \psi(U_{1i})$, $B_i = \psi(U_{2i})$, $\psi(x) = \max[-1, \min(1, x)]$, $U_{li} = (t_{li} - \phi_{pt_l})/h_{t_l}$, and $p(|t_l - \phi_{p_{t_l}}| < h_{t_l}) = 1 - \varphi$ for $l = 1, 2$. Here, $\phi_{p_{t_l}}$ and h_{t_l} represent the percentage measure of location and scale for the variable t_l, respectively. Following Wilcox (1994) and Shoemaker and Hettmansperger (1982), we set $\varphi = 0.1$.

3. $\pi \sim \text{Beta}(b_{01}, b_{02})$. We set $b_{01} = b_{02} = \frac{1}{2}$. In this case, $\mathbb{E}(\pi) = \frac{1}{2}$, and the anti-mode at the center of the distribution.

Following Smith and Kohn (1996), George and McCulloch (1993), we adopt an efficient MCMC-based computation technique for computing posterior model probabilities (PMP) in QR. Given $\boldsymbol{\gamma}, \boldsymbol{v}$, and $\boldsymbol{X}$, the marginal likelihood of $\boldsymbol{y}$ is given by

$$p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{v}, \boldsymbol{X}) \propto \int \left(\int p(\boldsymbol{y}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \tau, \boldsymbol{v}, \boldsymbol{X}) p(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \tau, \boldsymbol{v}) p(\boldsymbol{v}|\tau) d\boldsymbol{\beta}_{\boldsymbol{\gamma}} \right) p(\tau) d\tau.$$

Thus, we have

$$\begin{aligned}
&p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{v}, \boldsymbol{X}) \propto (1+g)^{-k\gamma/2} S(\boldsymbol{\gamma})^{-3n/2}, \\
&S(\boldsymbol{\gamma}) = \frac{1}{4}(\boldsymbol{y} - \xi\boldsymbol{v})' \boldsymbol{V} (\boldsymbol{y} - \xi\boldsymbol{v}) \\
&- \frac{g}{4(g+1)} (\boldsymbol{y} - \xi\boldsymbol{v})' \boldsymbol{V} \boldsymbol{X} (\boldsymbol{X}_{\boldsymbol{\gamma}}' \boldsymbol{V} \boldsymbol{X}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{X}_{\boldsymbol{\gamma}}' \boldsymbol{V} (\boldsymbol{y} - \xi\boldsymbol{v}) + p(1-p) \sum_{i=1}^{n} v_i.
\end{aligned}$$

Then, our MCMC-based computation method can be easily implemented to generate samples of

$$p(\boldsymbol{\gamma}|\boldsymbol{v}, \boldsymbol{y}, \boldsymbol{X}) \propto p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{v}, \boldsymbol{X}) p(\boldsymbol{\gamma}|\pi). \tag{2.13}$$

To build an efficient Gibbs sampler, as recommended by Lee et al. (2003), instead of updating $\boldsymbol{\gamma}$ as a vector from (2.13), we update an element γ_j from $p(\gamma_j|\boldsymbol{\gamma}_{-j}, \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{X})$. Hence, we have

$$p(\gamma_j|\boldsymbol{\gamma}_{-j}, \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{X}) \propto p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{v}, \boldsymbol{X})p(\gamma_j|\pi),$$

where

$$p(\gamma_j = 1|\boldsymbol{\gamma}_{-j}, \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{X}) = \frac{1}{1+h_j},$$
$$h_j = \sqrt{1+g}\left(\frac{S(\gamma_j=0)}{S(\gamma_j=1)}\right)^{-3n/2} \frac{(1-\pi)\ |\boldsymbol{X}^{*\prime}\boldsymbol{\gamma}^*_{pb-j,\gamma_j=0}\boldsymbol{X}^*\boldsymbol{\gamma}^*_{pb-j,\gamma_j=0}|^{1/2}}{\pi|\boldsymbol{X}^{*\prime}\boldsymbol{\gamma}^*_{pb-j,\gamma_j=1}\boldsymbol{X}^*\boldsymbol{\gamma}^*_{pb-j,\gamma_j=1}|^{1/2}}.$$

Since $\pi \sim \text{Beta}(b_{01}, b_{02})$, then, under model $\boldsymbol{\gamma}$ the full conditional distribution of π is $\text{Beta}(k_\gamma + b_{01}, k - k_\gamma + b_{02})$.

2.4 Simulations

2.4.1 Example 1 (*Inference*)

In this example, we consider our Bayesian QR approach (BQR_g) and semiparametric Bayesian approach (FBQR) assuming that the errors come from a mixture of Gaussian densities reported in Reich et al. (2010). The R code for FBQR can be obtained from the Web location "`http://www4.stat.ncsu.edu/~reich/Code/`". These approaches were compared with the standard QR approach (RQ) using the `rq()` function in the R package `quantreg` (Koenker, 2011). Our simulation design follows the setting of Reed and Yu (2009)

$$y_i = \beta_0 + \beta_1 x_i + (1 + \frac{x_i}{11})\varepsilon_i, \qquad i = 1, \cdots, 200,$$

where $x_i \sim \text{Uniform}(0, 10)$ and we set $\beta_0 = 10$, $\beta_1 = 5000$. The residuals ε_i are simulated from three distributions: N(0,1), a t_3 and a χ^2_3 distribution. Here, t_{ϱ_0} and

$\chi^2_{\varrho_0}$ denote the densities of a Student's t-distribution and a Chi-squared distribution respectively, where ϱ_0 denotes the number of degrees of freedom. In this example, we consider two choices of g: 100 and 10000.

For each residual distribution under consideration, 1000 replications are simulated assuming the number of observation is $n = 200$ and the models are fitted at three different quantiles $p = 0.50, 0.25$ and $p = 0.05$. Methods are evaluated based on the relative average bias

$$\widehat{\text{bias}}(\hat{\beta}_{m_0}) = \frac{1}{M}\sum_{r=1}^{M}\frac{\hat{\beta}^r_{m_0} - \beta_{m_0}}{\mid \beta_{m_0}\mid},$$

and the estimated relative efficiency

$$\widehat{\text{eff}}_{\text{model}}(\hat{\beta}_{m_0}) = \frac{S^2_{\text{model}}(\hat{\beta}_{m_0})}{S^2_{\text{BQR}g}(\hat{\beta}_{m_0})},$$

where M denotes the number of replications, $\hat{\beta}^r_{m_0}, m_0 = 1, 2$, is the QR coefficient estimate for the rth replication, β_{m_0} is the true value, $S^2(\hat{\beta}_{m_0}) = \frac{1}{M}\sum_{r=1}^{M}(\hat{\beta}^r_{m_0} - \bar{\beta}_{m_0})^2$ and $\bar{\beta}_{m_0} = \frac{1}{M}\sum_{r=1}^{M}\hat{\beta}^r_{m_0}$.

The simulation results for β_0 and β_1 are presented in Table 2.1, including the estimated relative bias and the efficiency. Across the three error distributions, the absolute bias obtained from our proposed method ($\text{BQR}_{g=10000}$) is much smaller at extreme quantiles than the competing approaches (RQ and FBQR). It can be observed that the semiparametric Bayesian model performs poorly for extreme quantiles, which is an undesirable situation when attention is focused on the extreme quantiles. In addition, as the quantiles become more extreme, the RQ and the semiparametric Bayesian approach yield high bias. Although, RQ and $\text{BQR}_{g=10000}$ perform better than $\text{BQR}_{g=100}$ in terms of bias, but $\text{BQR}_{g=100}$ is significantly better than RQ, $\text{BQR}_{g=10000}$ and FBQR in terms of efficiency. For instance, when the residual follows the normal distribution, the loss of efficiency for the RQ approach increased from 5.34% for β_0 when $p = 0.50$ to 14.72% when $p = 0.05$. It can also be observed that

Table 2.1: Estimated relative bias and relative efficiency for simulated data in Example 1 using three different error distributions. .

Error	Model	p	bias β_0	efficiency β_0	bias β_1	efficiency β_1
$\varepsilon \sim N(0,1)$	RQ	0.50	-0.0003	1.0534	0.0000	1.0018
	FBQR	0.50	-0.0168	1.1849	-0.0392	1.5223
	$\text{BQR}_{g=100}$	0.50	-0.0107	1.0000	-0.0099	1.0000
	$\text{BQR}_{g=10000}$	0.50	-0.0003	1.0144	-0.0001	1.0118
	RQ	0.25	0.0005	1.1197	0.0000	1.1233
	FBQR	0.25	-0.0328	2.3014	-0.0003	1.2336
	$\text{BQR}_{g=100}$	0.25	-0.0093	1.0000	-0.0097	1.0000
	$\text{BQR}_{g=10000}$	0.25	0.0005	1.0200	-0.0001	1.0150
	RQ	0.05	0.0009	1.1472	0.0000	1.2585
	FBQR	0.05	-0.0832	1.2137	-0.0098	2.7928
	$\text{BQR}_{g=100}$	0.05	-0.0096	1.0000	-0.0097	1.0000
	$\text{BQR}_{g=10000}$	0.05	0.0003	1.0036	-0.0001	1.0202
$\varepsilon \sim t_3$	RQ	0.50	0.0005	1.0458	0.0000	1.0863
	FBQR	0.50	-0.0167	1.5160	-0.0792	1.8483
	$\text{BQR}_{g=100}$	0.50	-0.0094	1.0000	-0.0099	1.0000
	$\text{BQR}_{g=10000}$	0.50	0.0004	1.0193	-0.0001	1.0325
	RQ	0.25	0.0007	1.0052	0.0000	1.0011
	FBQR	0.25	-0.0213	1.3118	-0.0095	2.0775
	$\text{BQR}_{g=100}$	0.25	-0.0097	1.0000	-0.0099	1.0000
	$\text{BQR}_{g=10000}$	0.25	0.0002	1.0097	-0.0001	1.0209
	RQ	0.05	-0.0050	1.0864	0.0000	1.0653
	FBQR	0.05	-0.0256	2.4654	-0.0188	2.4852
	$\text{BQR}_{g=100}$	0.05	-0.0167	1.0000	-0.0099	1.0000
	$\text{BQR}_{g=10000}$	0.05	-0.0042	1.0186	-0.0001	1.0286
$\varepsilon \sim \chi^2_3$	RQ	0.50	0.0000	1.0129	0.0000	1.0894
	FBQR	0.50	-0.0491	1.7054	-0.0082	2.4614
	$\text{BQR}_{g=100}$	0.50	-0.0101	1.0000	-0.0099	1.0000
	$\text{BQR}_{g=10000}$	0.50	-0.0001	1.0205	-0.0001	1.0207
	RQ	0.25	0.0024	1.0453	0.0000	1.0349
	FBQR	0.25	-0.0282	1.5351	-0.0072	1.4793
	$\text{BQR}_{g=100}$	0.25	-0.0073	1.0000	-0.0099	1.0000
	$\text{BQR}_{g=10000}$	0.25	0.0025	1.0195	-0.0001	1.0136
	RQ	0.05	0.0009	1.2059	0.0000	1.2684
	FBQR	0.05	-0.0511	2.1311	-0.0284	1.9357
	$\text{BQR}_{g=100}$	0.05	-0.0081	1.0000	-0.0095	1.0000
	$\text{BQR}_{g=10000}$	0.05	0.0007	1.0201	0.0000	1.0465

the loss of efficiency for the RQ approach when the error follows the t distribution increased from 4.58% for β_0 when $p = 0.50$ to 8.64% when $p = 0.05$.

The results of bias and relative efficiency clearly show that based on the simulated scenario BQR_g generally behaves well than the other methods (RQ and FBQR). As highlighted by Li et al. (2010), these results indicate that the reluctancy of the modeller in assuming an asymmetric Laplace residual distribution in the context of a nonparametric setting may be minimised and eliminated by the fact that the Bayesian techniques based on ALD are insensitive to the assumptions of the residual distribution. Furthermore, it can be argued that our simple Gibbs sampler via the modified g-prior contributed towards improvements of the QR results.

2.4.2 Example 2 *(Subset selection)*

In this example we compare our approach for Bayesian variable selection (BVSg) presented in Section 2.3 with the stochastic search variable selection reported in Reed et al. (2009) using the `SSVSquantreg()` function in the R package `MCMCpack` (Martin et al., 2011). The proposed method (BVSg) is also compared with the AIC and BIC methods using the least squares approximation (LSA) method reported in Wang and Leng (2007). The R code for the LSA method is available in the Web location "`http://www4.stat.ncsu.edu/~boos/var.select/LSA.R.txt`". Data are simulated from two model designs:

- Design I: $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)'$ and the rows of $\boldsymbol{X}$ follow a $N_k(0, \boldsymbol{\Sigma}_x)$ with $(\boldsymbol{\Sigma}_x)_{j_1 j_2} = 0.5^{|j_1 - j_2|}$, where $j_1 j_2$ refers to the $(j_1, j_2)^{th}$ entry of the matrix $\boldsymbol{\Sigma}_x$.
- Design II: Same as the first design except that $\boldsymbol{\beta}$=$(3, 1.5, 0, 0, 2, 0, 0, 0)'$.

In each design, we investigate four different distributions for the residuals ε_i, $i = 1, ..., n = 200$:

1. Normal distribution so that $\varepsilon_i \sim \text{N}(0, 9)$.

2. Mixture normal distributions so that $\varepsilon_i \sim 0.25\mathrm{N}(0,1) + 0.75\mathrm{N}(0,4)$.

3. Student's t-distribution so that $\varepsilon_i \sim t_{(3)}$.

4. Chi-squared distribution so that $\varepsilon_i \sim \chi^2_{(3)}$.

Subset selection is carried out at three different quantiles, $p \in \{0.50, 0.25, 0.05\}$, across the four error distributions. SSVSquantreg and BVSg are evaluated based on the average number of times for which the best candidate subset was chosen as the subset with highest Posterior Probability (PP) over 100 simulations, referred to as "%S". The results of SSVSquantreg and BVSg are presented in Tables 2.2 and 2.3. The tables also list the best subset selected, based on LSA. For each design and choice of the residual distribution, one can observe how often the proposed method selects the correct subset. Most noticeably, from Design II we observe that the SSVSquantreg tends to choose the whole model and attach very low PP to the correct subset. Alternatively, our proposed method BVSg almost always chose the true subset and gives high PP to the true subset. Additionally, we can observe that for Design II corresponding to the normal errors and $p = 0.50$, our proposed method and the LSA method using BIC and AIC criteria identify the correct subset $89\%, 61\%$ and 29% of the time, respectively. On the other hand, we can see that SSVSquantreg approach identifies the whole model 78% of the time.

Table 2.2: Comparing four approaches, corresponding to Design I: SSVSquantreg, BVSg, LSA.bic and LSA.aic based on the average posterior model probabilities (APMP) and the number of times each subset was chosen as the best subset over 100 simulations, referred to as "%S".

p	Error Distribution	SSVSquantreg subset	SSVSquantreg APMP (%S)	BVSg subset	BVSg APMP(%S)	LSA.bic subset	LSA.bic (%S)	LSA.aic subset	LSA.aic (%S)
0.50	normal	x_1	0.20 (60)	x_1	0.94 (98)	x_1	(82)	x_1	(36)
0.50	normal mixture	x_1	0.54 (99)	x_1	0.93 (99)	x_1	(71)	x_1	(26)
0.50	$t_{(3)}$	x_1	0.49 (99)	x_1	0.94 (99)	x_1	(80)	x_1	(41)
0.50	$\chi^2_{(3)}$	x_1	0.16 (41)	x_1	0.89 (81)	x_1	(70)	x_1	(24)
0.25	normal	x_1	0.20 (52)	x_1	0.93 (65)	x_1	(58)	x_1	(15)
0.25	normal mixture	x_1	0.56 (99)	x_1	0.93 (99)	x_1	(64)	x_1	(26)
0.25	$t_{(3)}$	x_1	0.45 (96)	x_1	0.90 (99)	x_1	(55)	x_1	(21)
0.25	$\chi^2_{(3)}$	x_1	0.23 (95)	x_1	0.92 (93)	x_1	(92)	x_1	(51)
0.05	normal	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.20 (59)	x_1	0.77 (72)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(6)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(13)
0.05	normal mixture	x_1	0.46 (95)	x_1	0.93 (93)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(7)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(18)
0.05	$t_{(3)}$	x_1	0.38 (88)	x_1	0.91 (99)	x_1	(4)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(14)
0.05	$\chi^2_{(3)}$	x_1	0.34 (99)	x_1	0.95 (99)	x_1	(43)	x_1	(8)

Table 2.3: Comparing four approaches, corresponding to Design II: SSVSquantreg, BVSg, LSA.bic and LSA.aic based on the APMP and the number of times each subset was chosen as the best subset over 100 simulations, referred to as "%S".

p	Error Distribution	SSVSquantreg subset	SSVSquantreg APMP (%S)	BVSg subset	BVSg APMP(%S)	LSA.bic subset	LSA.bic (%S)	LSA.aic subset	LSA.aic (%S)
0.50	normal	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.19 (78)	x_1, x_2, x_5	0.64 (89)	x_1, x_2, x_5	(61)	x_1, x_2, x_5	(29)
0.50	normal mixture	x_1, x_2, x_5	0.31 (99)	x_1, x_2, x_5	0.85 (100)	x_1, x_2, x_5	(76)	x_1, x_2, x_5	(42)
0.50	$t_{(3)}$	x_1, x_2, x_5	0.26 (96)	x_1, x_2, x_5	0.85 (99)	x_1, x_2, x_5	(82)	x_1, x_2, x_5	(48)
0.50	$\chi^2_{(3)}$	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.23 (96)	x_1, x_2, x_5	0.77 (56)	x_1, x_2, x_5	(43)	x_1, x_2, x_5	(24)
0.25	normal	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.25 (89)	x_1, x_2, x_5	0.59 (63)	x_1, x_2, x_5	(33)	x_1, x_2, x_5	(17)
0.25	normal mixture	x_1, x_2, x_5	0.28 (99)	x_1, x_2, x_5	0.83 (98)	x_1, x_2, x_5	(59)	x_1, x_2, x_5	(19)
0.25	$t_{(3)}$	x_1, x_2, x_5	0.22 (89)	x_1, x_2, x_5	0.84 (92)	x_1, x_2, x_5	(59)	x_1, x_2, x_4	(27)
0.25	$\chi^2_{(3)}$	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.14 (74)	x_1, x_2, x_5	0.78 (79)	x_1, x_2, x_5	(74)	x_1, x_2, x_5	(49)
0.05	normal	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.28 (96)	x_1, x_2, x_5	0.56 (27)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(14)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(25)
0.05	normal mixture	x_1, x_2, x_5	0.21 (77)	x_1, x_2, x_5	0.70 (82)	x_1, x_2, x_5	(7)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(17)
0.05	$t_{(3)}$	x_1, x_2, x_5	0.24 (67)	x_1, x_2, x_5	0.81 (90)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(10)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	(22)
0.05	$\chi^2_{(3)}$	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	0.13 (54)	x_1, x_2, x_5	0.80 (97)	x_1, x_2, x_5	(44)	x_1, x_2, x_5	(13)

2.5 Air pollution data

In this section, we consider the air pollution data which is available in the R package "truncSP" (Lindmark and Karlsson, 2012). There are 500 observations, 7 independent variables and one outcome variable, which is the log (concentration of NO_2 per hour). Independent variables include the log (number of cars per hour) (x_1), temperature at a height of two meters above the ground (x_2), wind speed in meters per second (x_3), the temperature difference between a height of 25 meters and a height of 2 meters above ground (x_4), wind direction (x_5), time of day in hours (x_6), and day number (x_7). We assume a QR model between the outcome log (concentration of NO_2 per hour) and the 7 independent variables, plus an intercept.

In Table 2.4, we compare three methods: the standard frequentist QR using the `rq()` function (RQ) (Koenker, 2011), the semiparametric Bayesian model, and our approach. The approaches are assessed based on 95% intervals for three different choices of p, $0.50, 0.25$ and 0.05. It can be observed that our estimates are very close to the standard QR estimated and our credible intervals are much narrower than the intervals given by the standard frequentist QR and the semiparametric Bayesian model. This indicates that the model in (2.5) is a working model with artificial assumptions, employed on the outcome variable to achieve the equivalence between maximising ALD and minimising the check function of Koenker and Bassett (1978) (Yuan and Yin, 2010).

In Table 2.5 we compare the PMP using our method BVSg to those obtained using SSVSquantreg. Table 2.5 also reports the mean absolute deviation (MAD) along with the standard deviation of AD (SD) for the top subset based on SSVSquantreg, BVSg and LSA method using BIC and AIC criteria. From the table, in general one can observe that the best subset chosen by our approach has lower MAD and standard deviation than those of the models chosen by the SSVSquantreg, LSA.aic and LSA.bic. The comparison between the four methods indicates that our Bayesian subset selection BVSg produces promising subsets and behaves well compared to

Table 2.4: Estimates and 95% intervals for the 0.50, 0.25 and 0.05 QR parameters of the air pollution data. The proposed approach (BQR_g) is compared with two other approaches: the semiparametric Bayesian approach (FBQR) and the frequentist QR approach (RQ).

p	Parameters	RQ Mean (lower bd, upper bd)	FBQR Mean (95% CrI)	BQR_g Mean (95% CrI)
0.50				
	β_0	1.79755 (0.88164, 2.10382)	1.35451 (0.76992, 1.93767)	1.48084 (1.47788, 1.48376)
	β_1	0.25865 (0.19716, 0.34333)	0.31113 (0.22730, 0.39439)	0.28870 (0.28828, 0.28913)
	β_2	0.00393 (-0.00847, 0.01485)	-0.00616 (-0.01812, 0.00584)	0.00248 (0.00242, 0.00254)
	β_3	-0.11806 (-0.15096, -0.07518)	-0.12157 (-0.16645, -0.07898)	-0.11152 (-0.11171, -0.11131)
	β_4	0.01130 (-0.05877, 0.10029)	0.04645 (-0.04152, 0.13355)	0.03227 (0.03183, 0.03268)
	β_5	0.00002 (-0.00125, 0.00068)	0.00003 (-0.00085, 0.00091)	-0.00024 (-0.00024, -0.00024)
	β_6	-0.00176 (-0.01277, 0.01085)	0.00014 (-0.01294, 0.01318)	0.00202 (0.00195, 0.00208)
	β_7	0.00018 (-0.00013, 0.00052)	0.00036 (0.00000, 0.00073)	0.00022 (0.00022, 0.00022)
0.25				
	β_0	0.39446 (-0.48486, 1.12769)	0.62862 (-0.03797, 1.28096)	0.30376 (0.29993, 0.30758)
	β_1	0.37566 (0.29279, 0.49844)	0.35504 (0.26011, 0.44775)	0.36815 (0.36760, 0.36872)
	β_2	-0.00768 (-0.02370, 0.00898)	-0.01481 (-0.02916, 0.00044)	-0.01103 (-0.01111, -0.01095)
	β_3	-0.13303 (-0.15924, -0.05856)	-0.14073 (-0.19212, -0.09015)	-0.11623 (-0.11647, -0.11598)
	β_4	0.02011 (-0.13018, 0.12646)	0.05368 (-0.06065, 0.16697)	0.01743 (0.01688, 0.01798)
	β_5	-0.00007 (-0.00063, 0.00138)	0.00010 (-0.00088, 0.00102)	0.00014 (0.00013, 0.00015)
	β_6	0.00127 (-0.01474, 0.01031)	-0.00695 (-0.02136, 0.00772)	-0.00291 (-0.00300, -0.00282)
	β_7	0.00045 (-0.00004, 0.00082)	0.00048 (0.00007, 0.00087)	0.00049 (0.00048, 0.00049)
0.05				
	β_0	-0.69138 (-2.01104, -0.16466)	-0.17800 (-0.91368, 0.51518)	-0.85225 (-0.86316, -0.84179)
	β_1	0.49462 (0.36635, 0.60326)	0.37920 (0.27051, 0.48615)	0.45317 (0.45182, 0.45453)
	β_2	-0.03992 (-0.06891, -0.01210)	-0.02639 (-0.04685, -0.00386)	-0.03105 (-0.03134, -0.03075)
	β_3	-0.21045 (-0.24345, -0.08480)	-0.17865 (-0.26177, -0.10325)	-0.14944 (-0.15024, -0.14861)
	β_4	-0.08316 (-0.19286, 0.26049)	0.05641 (-0.09780, 0.21364)	0.02156 (0.01922, 0.02392)
	β_5	0.00060 (-0.00036, 0.00206)	0.00012 (-0.00130, 0.00146)	0.00070 (0.00069, 0.00072)
	β_6	-0.03201 (-0.05794, -0.00765)	-0.01688 (-0.03684, 0.00396)	-0.02612 (-0.02636, -0.02589)
	β_7	0.00064 (0.00001, 0.00149)	0.00070 (0.00008, 0.00129)	0.00060 (0.00059, 0.00061)

SSVSquantreg and LSA.

Table 2.5: Comparing MAD and SD for the best subset of the air pollution data. The PMP for the best subset of the air pollution data using SSVSquantreg and BVSg are also attached.

	SSVSquantreg			BVSg			LSA.aic		LSA.bic	
p	model	PMP	MAD (SD)	model	PMP	MAD (SD)	model	MAD (SD)	model	MAD (SD)
0.50	Inter., x_1, x_3	0.63	0.63 (0.51)	Inter., x_1, x_3	0.87	0.62 (0.51)	Inter., x_1, x_3	0.63 (0.51)	Inter., x_1, x_3	0.63 (0.51)
0.25	x_1	0.54	0.77 (0.58)	x_1, x_3	0.83	0.78 (0.58)	Inter., x_1, x_2, x_3, x_7	0.77 (0.58)	x_1, x_3, x_6	0.79 (0.59)
0.05	x_1	0.49	1.46 (0.74)	x_1, x_2, x_3	0.26	1.33 (0.71)	Inter., x_1, x_2, x_3, x_4, x_6, x_7	1.36 (0.81)	x_1, x_2, x_3, x_6	1.37 (0.77)

2.6 Chapter summary

In this chapter, we developed Bayesian techniques for subset selection and estimation of the independent variables coefficients in QR models based on conditional quantile dependent prior distributions. In particular, we assigned a quantile dependent prior distribution on the subset space and a quantile dependent prior distribution on the regression parameters given a specific subset. For regression coefficients, we developed a conditional conjugate prior distribution based on the familiar g-prior. In addition, the percentage bend correlation was used to find suitable prior distributions over subset space. MCMC-based computation algorithms are outlined based on the modified quantile dependent prior to generate samples from the posterior distributions over model space. Simulation studies and air pollution data show that, in comparison with existing Bayesian and non-Bayesian QR methods, the Bayesian QR method using a quantile dependent prior distribution generally perform better.

Chapter 3

Bayesian Tobit QR using g-prior distribution with ridge parameter

This chapter introduces the idea of the modified g-prior in the Tobit QR model. The prior is generalised by introducing a ridge parameter to address important challenges that may arise with left-censored data, such as multicollinearity and overfitting problems. Then, a simple MCMC-based computation technique is developed for Tobit QR based on the modified g-prior. We have developed an expression for the hyperparameter g to calibrate the modified g-prior with a ridge parameter to a corresponding g-prior. Some possible extensions of our approach are also presented, including the continuous and binary responses in QR. The techniques are illustrated using several simulation scenarios and the popular Leukemia data set.

3.1 Introduction

Tobit QR technique provides an active and crucial method of dealing with left-censored data and can be formulated as a QR model where the data on the outcome

of interest is not observed completely. A great body of work exists on Tobit QR methods and we refer to Powell (1986), Bilias et al. (2000), Yu and Stander (2007) and Wang and Fygenson (2009) for an overview. Consider the censoring model,

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \text{and} \quad y_i = \max\{y^0, y_i^*\}, \quad i = 1, \cdots, n, \tag{3.1}$$

where y_i is the outcome of interest, y^0 is a known fixed point, y_i^* is the corresponding latent unobserved outcome of the ith observation, $\boldsymbol{x}_i$ is a $k \times 1$ vector of predictors for the ith observation, $\boldsymbol{\beta}$ is a vector of unknown quantities of interest evaluated at pth quantile, and the residuals ε_i are restricted so that $\int_{-\infty}^{0} f_p(\varepsilon_i)d\varepsilon_i = p$. Following Powell (1986), it can be shown that the Tobit QR estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be estimated through the empirical check function

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \max\{y^0, y_i^*\}), \tag{3.2}$$

Yu and Stander (2007) observed that the posterior estimator of $\boldsymbol{\beta}$ obtained by assigning a likelihood that is based on the ALD-based working model at specific value of p, serves as the pth Tobit QR estimate. The authors assigned flat priors, independent of the value of p, for the Tobit QR coefficients vector and sampling $\boldsymbol{\beta}$ using the Metropolis-Hastings (MH) method. It is well known that flat priors could be useful for coefficient estimation in Tobit QR and other models but they cannot be used in subset selection techniques, owing to the fact that proper priors are needed to evaluate Bayes factors (Ibrahim and Chen, 2000). Yu and Stander (2007) also suggested families of symmetric prior distributions on the Tobit QR coefficients vector, such as normal and Laplace priors. Although these priors may lead to proper posterior, they are independent of the values of quantiles. That is, the prior is the same for modelling different order of quantiles. In this chapter, we use the modified g-prior to develop the Bayesian analysis of the Tobit QR model. Then, we generalised the g-prior by introducing a ridge parameter to address some issues that may arise with left-censored data such as, multicollinearity and overfitting problems. We also

developed an expression for the hyperparameter g to calibrate the modified g-prior with a ridge parameter to a corresponding g-prior.

The rest of this chapter is presented in the following way. Section 3.2 introduces our hierarchical Bayesian Tobit QR model, an extension of the modified g-prior is suggested and Bayesian MCMC steps are also outlined. Our Bayesian SSVS approach using the modified g-prior is presented in subsection 3.2.3. Section 3.3 extends the proposed technique to QR with continuous and binary outcome variables. Section 3.4 evaluates the methods using simulation examples and Section 3.5 applies the proposed technique to the popular Leukemia data set. A chapter summary follows in Section 3.6.

3.2 Methods

3.2.1 Hierarchical Bayesian modelling

At the pth quantile, we model conditional Tobit quantiles of the outcome y_i by assuming that $\varepsilon_i|v_i,\tau \sim \mathrm{N}((1-2p)v_i, 2\tau v_i)$ and $v_i|\tau \sim \mathrm{Exp}(p(1-p)/\tau)$, which is equivalent to assigning an ALD for ε_i, $i = 1, 2, ..., n$. To complete the prior specification, we assign a quantile dependent prior for $\boldsymbol{\beta}$ such that

$$\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N(0, 2g\tau(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}), \quad p(\tau) \propto \tau^{-1}. \tag{3.3}$$

In summary, our hierarchical Bayesian Tobit QR modelling is given by

$$\begin{aligned}
&y_i = \max\{y^0, y_i^*\}, \quad i = 1, \cdots, n,\\
&y_i^*|\boldsymbol{\beta}, \tau, v_i \sim N(\boldsymbol{x}_i'\boldsymbol{\beta} + \xi v_i, 2\tau v_i),\\
&\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N_k(0, 2\tau g(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}),\\
&v_i \sim \mathrm{Exp}(\frac{p(1-p)}{\tau}),\\
&p(\tau) \propto \tau^{-1}.
\end{aligned}$$

This hierarchical modelling produces an efficient MCMC algorithm by updating the latent variables y_i^* and v_i as well as the other parameters $\boldsymbol{\beta}$ and τ from their full conditional distributions.

- **Updating y_i^***

  Let $\Upsilon(.)$ denotes to a degenerate distribution, then the latent variable $y_i^*, i = 1, \cdots, n$, has a conditional distribution (CD) given by

$$y_i^*|y_i, \boldsymbol{\beta}, \tau, \boldsymbol{V} \sim \begin{cases} \Upsilon(y_i), & \text{if } y_i > y^0; \\ N(\boldsymbol{x}_i'\boldsymbol{\beta} + \xi v_i, 2\tau v_i) I(y_i^* \leq y^0), & \text{otherwise,} \end{cases} \tag{3.4}$$

- **Updating $\boldsymbol{\beta}$**

  The full CD of $\boldsymbol{\beta}$ is $\mathrm{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \frac{2\tau g}{g+1}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1} \text{ and } \boldsymbol{\mu} = \frac{g}{g+1}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{y}^* - \xi\boldsymbol{v}). \tag{3.5}$$

  Here, $\boldsymbol{y}^* = (y_1^*, \cdots, y_n^*)'$.

- **Updating τ**

$$\tau|\boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{v} \sim \mathrm{InvGa}((3n+k)/2, \frac{1}{4}(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta} - \xi\boldsymbol{v})'\boldsymbol{V}(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta} - \xi\boldsymbol{v}) + \frac{1}{4g}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})\boldsymbol{\beta} + p(1-p)\sum_{i=1}^{n} v_i).$$

- **Updating v_i**

  For $i = 1, ..., n$, each $v_i \sim \mathrm{GIG}(\nu, \varrho_1, \varrho_2)$, where $\nu = 0$, $\varrho_1^2 = ((y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \boldsymbol{\beta}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\beta}/g)/(2\tau)$ and $\varrho_2^2 = 1/(2\tau)$.

During MCMC iteration we sampled the latent variable $y_i^*, i = 1, ..., n$, using the `truncnorm` package (Trautmann et al., 2010) and we sampled $v_i, i = 1, ..., n$, using the `rgig()` function (Luethi and Breymann, 2012).

Since our target in the SSVS approach required computation of the marginal distribution of the data $p(\boldsymbol{y}^*|\tau, \boldsymbol{v})$, the following lemma gives the closed-form of

$p(\boldsymbol{y}^*|\tau, \boldsymbol{v})$ under the proposed prior.

Lemma 1. Under the quantile dependent prior (3.3), the marginal CD of the data $p(\boldsymbol{y}^*|\tau, \boldsymbol{v})$ is given by

$$p(\boldsymbol{y}^*|\tau, \boldsymbol{v}) = \frac{(1+g)^{-k/2}}{(4\pi)^{n/2}}(\prod_{i=1}^{n}(\tau v_i)^{-1/2}) \times \exp\{-(\boldsymbol{y}^* - \xi\boldsymbol{v})'(\frac{\boldsymbol{V}}{4\tau} - \frac{g\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}}{4(1+g)\tau})(\boldsymbol{y}^* - \xi\boldsymbol{v})\}. \tag{3.6}$$

The proof of Lemma 1 is straightforward and can be accomplished by integrating out the quantile coefficients vector $\boldsymbol{\beta}$ as in Smith and Kohn (1996).

3.2.2 Introducing a ridge parameter

In the original g-prior, the matrix $\boldsymbol{X}'\boldsymbol{X}$ suffers from singularity in case of multicollinearity or overfitting problems ($k >> n$). For this reason, Gupta and Ibrahim (2007) proposed a modification of the original Zellner's g-prior, motivated by the ridge parameter λ_0 which comes from ideas of *ridge regression* to deal with multicollinearity and overfitting problems. The authors showed that their technique allows consistent subset selection and coefficient estimation for overfitting problems. Baragatti and Pommeret (2012) considered the influence of λ_0 on the subset selection and suggested a technique to select the scaling factor. Similar to Gupta and Ibrahim (2007), in the situation of singularity of the matrix $\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$, we modified our prior with the ridge parameter ($\lambda_0 > 0$). More specifically, we propose the following prior for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N(0, 2\tau g_{\lambda_0}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} + 2\lambda_0\boldsymbol{I}_k)^{-1}), \tag{3.7}$$

where $g_{\lambda_0} > 0$ is a known scaling factor characterised by the parameter λ_0 and $\boldsymbol{I}_k$ is the $k \times k$ identity matrix. In this chapter, we assume $g_{\lambda_0} \neq g$.

Clearly in order for the conditional distribution (CD) of the quantile coefficients vector $\boldsymbol{\beta}$ under the prior (3.3) and the CD of $\boldsymbol{\beta}$ under the prior (3.7) to have identical CDs, we need $g(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1} = g_{\lambda_0}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} + 2\lambda_0\boldsymbol{I}_k)^{-1}$. The following lemma

characterises the relationship among the three parameters g, g_{λ_0} and λ_0.

Lemma 2. The conditional distribution (CD) of the quantile coefficients vector $\boldsymbol{\beta}$ under the prior (3.3) and the conditional distribution of $\boldsymbol{\beta}$ under the prior (3.7) are identical distributions if and only if

$$g_{\lambda_0}\boldsymbol{I}_k = g[\boldsymbol{I}_k + 2\lambda_0(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}]. \tag{3.8}$$

This lemma provides a technique to elicit g_{λ_0} and the proof of Lemma 2 is straightforward. By taking the expectation and trace of the second side of (3.8), we obtained

$$\hat{g}_{\lambda_0} = \frac{1}{k}\mathbb{E}[gk + \frac{2g\lambda_0}{\mathrm{tr}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})}], \tag{3.9}$$

where the expectation in Equation (3.9) is taken with respect to $\boldsymbol{V}$. In this chapter, similar to the previous chapter, we set the scaling factor to $g = 100$. We choose $\lambda_0 = 1/k$ as suggested by Baragatti and Pommeret (2012) which lies between 0 and 1 as recommended for Bayesian robustness (Gupta and Ibrahim, 2007). Under the prior (3.7), the full CD of $\boldsymbol{\beta}$ is $\mathrm{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = 2\tau[\frac{g_{\lambda_0}+1}{g_{\lambda_0}}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} + \frac{2\lambda_0}{g_{\lambda_0}}\boldsymbol{I}_k]^{-1} \quad \text{and} \quad \boldsymbol{\mu} = (2\tau)^{-1}\boldsymbol{\Sigma}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{y}^* - \xi\boldsymbol{v}). \tag{3.10}$$

During MCMC iteration, we updated g_{λ_0} using $g_{\lambda_0} = k^{-1}[gk + 2g\lambda_0/\mathrm{tr}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})]$ where $g = 100$ and $\lambda_0 = 1/k$. In the situation of nonsingularity of the matrix $\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$, we set $\lambda_0 = 0$ and $g_{\lambda_0} = g$.

3.2.3 Subset selection

This section extends the idea of Bayesian subset selection in Tobit QR reported in Ji et al. (2012) by using different priors for different quantiles. Given $p \in (0, 1)$ and $\tau = 1$, we consider the following prior distribution assumptions:

- The prior distribution of $\boldsymbol{\beta_\gamma}$ is taken as $\boldsymbol{\beta_\gamma}|\boldsymbol{V}, \boldsymbol{X_\gamma} \sim N(0, 2g_{\lambda_0}(\boldsymbol{X'_\gamma V X_\gamma} + 2\lambda_0 \boldsymbol{I}_{k_\gamma})^{-1})$, where $v_i \sim \text{Exp}(p(1-p))$ for $i = 1, ..., n$.

- The prior of $\boldsymbol{\gamma}$ is taken as $p(\boldsymbol{\gamma}|\pi) \propto \pi^{k_\gamma}(1-\pi)^{k-k_\gamma}$ (George and McCulloch, 1993, 1997), where $\pi \sim \text{Beta}(b_{01}, b_{02})$.

Under prior assumptions, we are able to use a MCMC based computation technique to update $\boldsymbol{y}^*, \boldsymbol{\beta_\gamma}, \boldsymbol{V}$ and π from the posterior:

- **Updating y_i^***

  Under $\boldsymbol{\gamma}$, the full CD of $y_i^*, i = 1, \cdots, n$, is reduced to

$$y_i^*|y_i, \boldsymbol{\beta_\gamma}, \boldsymbol{V} \sim \begin{cases} \Upsilon(y_i), & \text{if } y_i > y^0; \\ N(\boldsymbol{x}'_{i,\boldsymbol{\gamma}}\boldsymbol{\beta_\gamma} + \xi v_i, 2v_i) I(y_i^* \leq y^0), & \text{otherwise,} \end{cases} \tag{3.11}$$

- **Updating $\boldsymbol{\beta_\gamma}$**

  The full CD of $\boldsymbol{\beta_\gamma}$ is $\text{N}_{k_\gamma}(\boldsymbol{\mu_\gamma}, \boldsymbol{\Sigma_\gamma})$, where

$$\boldsymbol{\Sigma_\gamma} = 2[\frac{g_{\lambda_0}+1}{g_{\lambda_0}}\boldsymbol{X'_\gamma V X_\gamma} + \frac{2\lambda_0}{g_{\lambda_0}}\boldsymbol{I}_{k_\gamma}]^{-1},$$
$$\text{and} \quad \boldsymbol{\mu_\gamma} = 2^{-1}\boldsymbol{\Sigma_\gamma X'_\gamma V}(\boldsymbol{y}^* - \xi \boldsymbol{v}).$$

- **Updating $\boldsymbol{v}$**

  The full CD of each v_i can be obtained from the full CD of v_i in the subsection 3.2 by setting $\tau = 1$ and replacing $\boldsymbol{x}'_i$ and $\boldsymbol{\beta}$ everywhere with $\boldsymbol{x}'_{i,\boldsymbol{\gamma}}$ and $\boldsymbol{\beta_\gamma}$, respectively.

- **Updating γ_j**

  Each $\gamma_j, j = 1, \cdots, k$, has a full CD given by

$$p(\gamma_j = 1|\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{\beta_\gamma}, \boldsymbol{v}, \boldsymbol{\gamma}_{-j}) = \frac{1}{1+h_j},$$
$$h_j = \frac{p(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\beta_\gamma}, \boldsymbol{v}, \gamma_j = 0, \boldsymbol{\gamma}_{-j}) p(\boldsymbol{\beta_\gamma}|\gamma_j = 0, \boldsymbol{\gamma}_{-j}) p(\gamma_j = 0, \boldsymbol{\gamma}_{-j})}{p(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\beta_\gamma}, \boldsymbol{v}, \gamma_j = 1, \boldsymbol{\gamma}_{-j}) p(\boldsymbol{\beta_\gamma}|\gamma_j = 1, \boldsymbol{\gamma}_{-j}) p(\gamma_j = 1, \boldsymbol{\gamma}_{-j})}.$$

- **Updating π**

  The full CD of π is the same as in Section 2.3 of Chapter 2.

3.3 Model extensions

3.3.1 Subset selection in QR

The proposed method in subsection 3.2.3 can be used, with some modifications, to find subset selection in QR with continuous outcome variable. By ignoring the link function $y_i = \max\{y^0, y_i^*\}$ and replacing y^* everywhere with y, our approach offers an alternative way for subset selection in QR model with continuous outcome to deal with multicollinearity and overfitting problems.

3.3.2 Subset selection in Binary QR

In this subsection, we show that our technique reported in subsection 3.2.3 can be extended to subset selection for binary QR model. Binary QR models have received considerable interest in the literature and we refer to Manski (1975, 1985), Kordas (2006) and Benoit and Poel (2011) for an overview. Suppose y_i is a binary outcome variable (e.g. normal and cancer), then the binary QR takes the form of (Manski, 1985)

$$\begin{aligned} y_i^* &= \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \\ y_i &= 1 \text{ if } \; y_i^* \geq 0, \quad y_i = 0 \;\; \text{otherwise.} \end{aligned} \tag{3.12}$$

Under the above model, the proposed method in subsection 3.2.3 can be used to find promising subset in binary QR by using the link function $y_i \sim 1(y_i^* \geq 0)$ and sampling

$y_i^*, i = 1, \cdots, n$, as follows

$$y_i^*|y_i = 1, v_i, \boldsymbol{\beta}_{\boldsymbol{\gamma}} \sim N(\boldsymbol{x}'_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \xi v_i, 2v_i) \text{ truncated at the left by 0,}$$
$$y_i^*|y_i = 0, v_i, \boldsymbol{\beta}_{\boldsymbol{\gamma}} \sim N(\boldsymbol{x}'_{i,\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \xi v_i, 2v_i) \text{ truncated at the right by 0.}$$

3.4 Simulations

3.4.1 Example 1 (*Inference*)

In this example, we consider our Bayesian Tobit quantile regression approach using g-prior (BTQRg) and Bayesian Tobit quantile regression approach (BTQ) using a symmetric prior distribution, $\boldsymbol{\beta} \sim \mathrm{N}_k(\mathbf{0}, 100\boldsymbol{I})$, as reported by Kozumi and Kobayashi (2011). These approaches were compared with the standard Tobit QR approach (crq) using the `crq()` function employing Powell's method in the R package `quantreg` (Koenker, 2011). Our simulation design follows the setting of Bilias et al. (2000) and Yu and Stander (2007), among others. We simulate data from the model

$$y_i = \max\{0, y_i^*\}, \ i = 1, ..., n,$$
$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

where $x_{1i} \sim$ Bernoulli(0.5) centered at zero, $x_{2i} \sim \mathrm{N}(0, 1)$ and $(\beta_0, \beta_1, \beta_2) = (1, 1, 1)$. The residuals ε_i are simulated from three distributions: N(0,1), $(1 + x_2)$N$(0, 1)$ and $0.75\mathrm{N}(0, 1) + 0.25\mathrm{N}(0, 4)$. For each residual distribution, 250 data sets are simulated assuming the number of observations are $n = 100$ and the models are fitted at three different quantiles, $p = 0.50, 0.75$ and $p = 0.95$. BTQRg and BTQ algorithms are run for 17000 iterations and the first 2000 were removed as burn in. Methods are evaluated based on the estimated relative average bias and efficiency which are defined in Section 2.4 of Chapter 2.

Clearly from Table 3.1, the biases due to BTQRg, BTQ and crq are more or less the same. However, BTQRg generally behaves much better than BTQ and crq in

Table 3.1: Estimated bias and relative efficiency for different error distributions. The proposed approach (BTQRg) is compared with two other approaches: the frequentist Tobit QR approach (crq) and the Bayesian approach using symmetric prior distribution for the regression coefficients (BTQ).

Model	p	bias (eff.) β_0	bias (eff.) β_1	bias (eff.) β_2
$\varepsilon \sim N(0,1)$				
crq	0.50	-0.01466 (2.15047)	0.03376 (2.18275)	0.01217 (2.06735)
BTQ	0.50	0.00597 (1.14267)	0.00996 (1.06784)	-0.00789 (1.01235)
BTQRg	0.50	0.00640 (1.00000)	-0.01036 (1.00000)	-0.05614 (1.00000)
crq	0.75	-0.01052 (1.72184)	-0.00405 (1.32100)	-0.01455 (1.59715)
BTQ	0.75	-0.02690 (1.35108)	0.06365 (1.57798)	0.09263 (1.85198)
BTQRg	0.75	0.00284 (1.00000)	-0.05996 (1.00000)	-0.01000 (1.00000)
crq	0.95	0.01208 (1.13496)	-0.02656 (1.21129)	-0.12986 (0.97514)
BTQ	0.95	0.09551 (0.91855)	0.05976 (1.03759)	0.07646 (1.09315)
BTQRg	0.95	0.00201 (1.00000)	0.00079 (1.00000)	-0.04465 (1.00000)
$\varepsilon \sim (1+x_2)N(0,1)$				
crq	0.50	-0.15563 (9.78132)	0.16054 (21.48298)	0.01826 (2.33432)
BTQ	0.50	-0.07266 (1.32119)	0.09796 (2.85943)	0.05578 (1.22709)
BTQRg	0.50	0.07632 (1.00000)	-0.01073 (1.00000)	-0.01023 (1.00000)
crq	0.75	-0.01138 (1.21019)	0.05568 (2.65549)	-0.07541 (1.62722)
BTQ	0.75	-0.05331 (1.30429)	0.28926 (2.96351)	0.03059 (1.67248)
BTQRg	0.75	0.04686 (1.00000)	-0.04863 (1.00000)	-0.01050 (1.00000)
crq	0.95	0.08300 (1.32362)	-0.01790 (1.18841)	-0.31592 (1.98608)
BTQ	0.95	0.18887 (0.89628)	0.27505 (1.01940)	-0.21792 (1.40077)
BTQRg	0.95	0.13293 (1.00000)	-0.00770 (1.00000)	-0.14665 (1.00000)
$\varepsilon \sim 0.75N(0,1)$ $+0.25N(0,4)$				
crq	0.50	-0.02559 (2.81012)	0.00989 (2.73070)	-0.00652 (1.82914)
BTQ	0.50	-0.01951 (1.49375)	0.00640 (1.36406)	-0.00642 (0.90888)
BTQRg	0.50	0.00207 (1.00000)	-0.00570 (1.00000)	-0.01228 (1.00000)
crq	0.75	-0.13335 (1.42499)	0.01562 (1.54906)	-0.01233 (1.78678)
BTQ	0.75	-0.14603 (1.86671)	0.08703 (1.85623)	0.10127 (1.89332)
BTQRg	0.75	-0.09073 (1.00000)	-0.04322 (1.00000)	-0.00878 (1.00000)
crq	0.95	-0.18130 (1.58685)	-0.05828 (1.15029)	-0.11391 (0.99865)
BTQ	0.95	-0.10930 (1.08060)	0.05055 (0.95292)	0.10552 (1.05016)
BTQRg	0.95	-0.10797 (1.00000)	-0.00999 (1.00000)	-0.02471 (1.00000)

terms of the absolute bias. Most noticeably, when $p = 0.95$ the absolute bias generated by BTQRg for all parameters is much smaller than the absolute bias generated by BTQ and crq. In addition, BTQRg appears more efficient than the BTQ and crq. For example, when the error is standard normal and estimating the median, the loss of efficiency of the standard Tobit QR (crq), with respect to BTQRg, was 107% for β_2 and larger for the other parameters. We may also investigate the estimation of β_0, β_1 and β_2 compared to the true QR coefficients $\boldsymbol{\beta}^{\boldsymbol{true}}$, which are presented in Table 3.2. The results suggest that BTQRg works well compared with the BTQ and crq. The posterior histograms of quantile coefficients β_0, β_1 and β_2 in Figure 3.1 also support this conclusion.

3.4.2 Example 2 (*Subset selection for left-censored response with* $k < n$)

Data are simulated from 2 model designs:

- Design I: $\boldsymbol{\beta} = (1, 5, 0, 0, 0, 0, 0, 0, 0)'$, including the intercept value, and the rows of $\boldsymbol{X}$ follow a $N_8(0, \boldsymbol{\Sigma}_x)$ with $(\boldsymbol{\Sigma}_x)_{j_1 j_2} = 0.5^{|j_1 - j_2|}$.
- Design II: Same as Design I except that $\boldsymbol{\beta}$=$(1, 3, 1.5, 0, 0, 2, 0, 0, 0)'$.

For Designs I and II, 250 datasets are generated each with $n = 200$ observations from the true censoring model

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \text{and} \quad y_i = \max\{0, y_i^*\}, \quad i = 1, \cdots, n, \tag{3.13}$$

The residuals $\varepsilon_i, i = 1, \cdots, 200$, are simulated from three distributions: N(0,1), t_3 and χ^2_3 distribution. In this example, we compare our Bayesian subset selection for Tobit QR using quantile dependent priors (BTQRg) with Bayesian variable selection in Tobit quantile regression (BVST) using a symmetric prior distribution as reported by Ji et al. (2012). The results of the standard Tobit QR approach (crq) are also reported. BTQRg, BVST and crq are evaluated based on median of mean absolute deviations, referred to as "MMAD". In other words, MMAD=median(mean($|\boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{x}_i'\boldsymbol{\beta}^{true}|$)),

Table 3.2: True parameter values and their estimates for Example 1. The results are averaged over 250 independent simulations.

Error	Method	p	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
$N(0,1)$	$\boldsymbol{\beta^{true}}$	0.50	1.0000	1.0000	1.0000
	crq	0.50	0.9831	1.0016	1.0201
	BTQ	0.50	1.0128	0.9826	0.9999
	BTQRg	0.50	1.0006	0.9897	1.0060
	$\boldsymbol{\beta^{true}}$	0.95	2.6449	1.0000	1.0000
	crq	0.95	2.6247	0.9548	0.8892
	BTQ	0.95	3.0503	1.0807	1.1221
	BTQRg	0.95	2.6503	1.0241	1.0896
$(1+x_2)N(0,1)$	$\boldsymbol{\beta^{true}}$	0.50	1.0000	1.0000	1.0000
	crq	0.50	0.9060	1.0847	1.0091
	BTQ	0.50	0.9392	1.0687	1.0404
	BTQRg	0.50	0.9889	1.0193	1.0034
	$\boldsymbol{\beta^{true}}$	0.95	2.6449	1.0000	2.6449
	crq	0.95	2.9249	1.0253	1.8588
	BTQ	0.95	3.3334	1.3619	2.1223
	BTQRg	0.95	2.7790	1.0191	2.1547
$0.75N(0,1)$ $+0.25N(0,4)$	$\boldsymbol{\beta^{true}}$	0.50	1.0000	1.0000	1.0000
	crq	0.50	0.9848	1.0253	1.0147
	BTQ	0.50	1.0032	1.0063	1.0052
	BTQRg	0.50	0.9993	1.0105	1.0014
	$\boldsymbol{\beta^{true}}$	0.95	3.0560	1.0000	1.0000
	crq	0.95	2.4880	0.9602	0.8669
	BTQ	0.95	2.9183	1.0863	1.1503
	BTQRg	0.95	2.9737	1.0107	1.0197

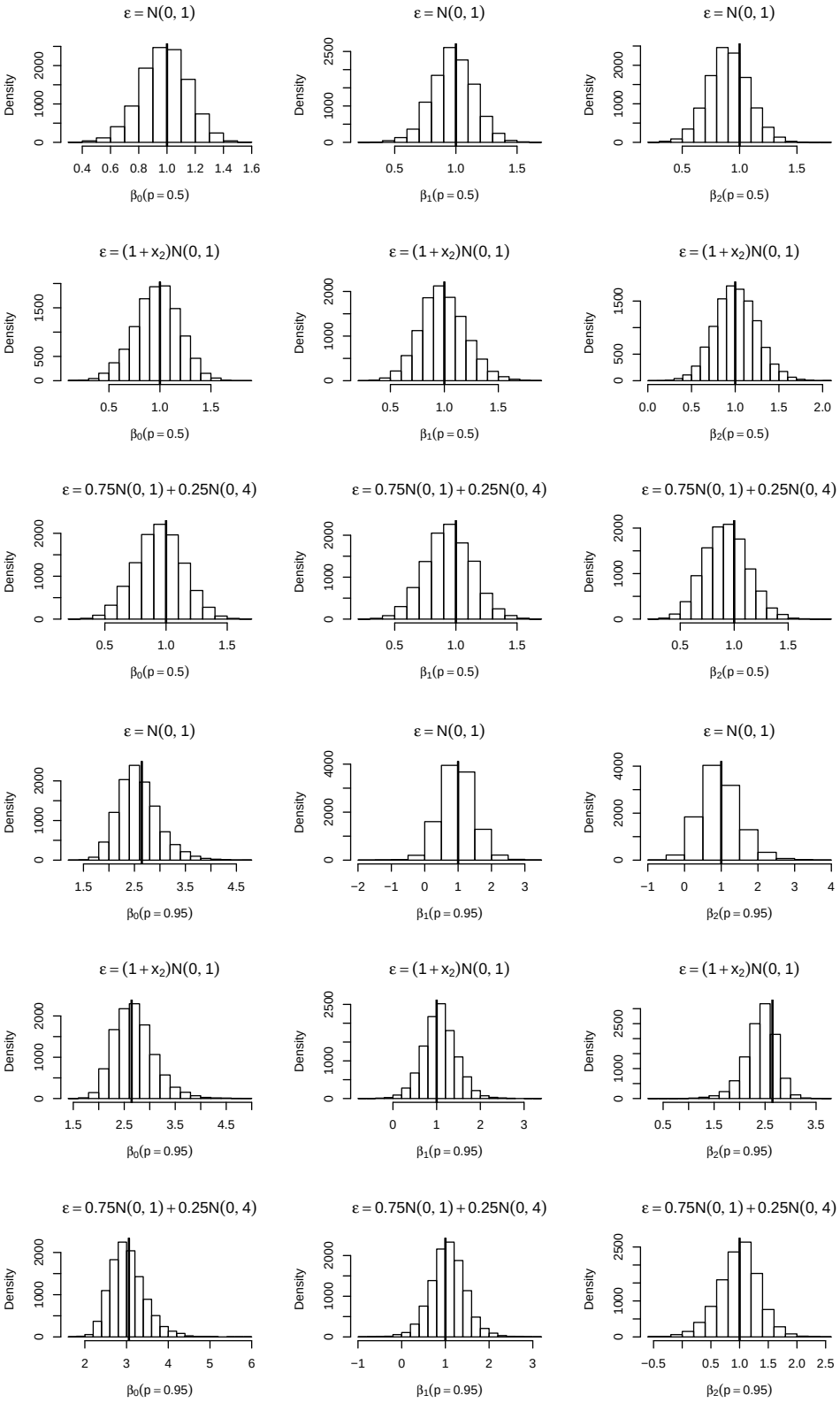

Figure 3.1: Posterior histograms of β_0, β_1 and β_2 at quantiles 0.50 and 0.95 for Example 1 using our Bayesian method.

where the median is calculated over the total number of replications. MMAD is a good technique of producing significant information about how well BTQRg, BVST and crq perform, where a lower MMAD suggests a better performance.

The results of the MMADs and SD of the MADs are listed in Table 3.3. For the MMADs and SD criteria, the proposed method (BTQRg) generally performs better than BVST and crq for all the distributions under consideration. In addition, BTQRg selects a highest average number of actual zeros than BVST. We can see that as the quantiles become more extreme, the BVST method yields a low average number of actual zeros compared with BTQRg, suggesting a good performance of BTQRg. Moreover, from Table 3.4, we observe that both methods BTQRg and BVST choose the true subset. However, we can observe that BTQRg tends to behave better in terms of average posterior model probability (APMP) for the correct subset than the BVST, especially for the most extreme quantile ($p = 0.95$). Hence, the modified g-prior plays a good role in finding the correct subset, even for extreme quantiles.

3.4.3 Example 3 (*Subset selection for left-censored response with $k > n$*)

The setup in Example 3 is the same as Example 2, except we set $k = 250$ (including the intercept value) and $\boldsymbol{\beta} = (1, 3, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 3, \underbrace{0, 0, \cdots, 0}_{236})'$ to investigate the performance of BTQRg and BVST in the case $k > n$. From Table 3.5, the performance of BTQRg appears quite well compared to the BVST. We observe that the MMAD produced using the BVST method is much higher than BTQRg. We also see that BTQRg tends to produce lower standard deviations than BVST, suggesting a good performance of the proposed method.

3.4.4 Example 4 (*Subset selection for continuous response with $k > n$*)

The setup in this example is the same as Example 3 but we ignore the link function (i.e., continuous response) and we set $n = 50$. In this example, we use one dataset to compare our approach reported in subsection 3.3.1, refereed to as "SSVSQ",

Table 3.3: MMADs, SD and the average number of actual zeros (correct) for the simulated data in Example 2, where $p = 0.50$, 0.75 and 0.95.

Model	p	$\varepsilon \sim N(0,1)$ MMAD (SD)	correct	$\varepsilon \sim t_3$ MMAD (SD)	correct	$\varepsilon \sim \chi^2_3$ MMAD (SD)	correct
Des. I							
crq	0.50	0.327 (0.112)	-	0.336 (0.191)	-	0.415 (0.289)	-
BVST	0.50	0.158 (0.083)	6.88	0.177 (0.105)	6.84	0.552 (0.386)	6.96
BTQRg	0.50	0.094 (0.082)	6.96	0.142 (0.094)	6.88	0.347 (0.264)	7.00
crq	0.75	0.311 (0.119)	-	0.363 (0.157)	-	0.444 (0.287)	-
BVST	0.75	0.267 (0.254)	6.84	0.322 (0.318)	6.80	0.357 (0.360)	6.68
BTQRg	0.75	0.146 (0.112)	7.00	0.184 (0.162)	6.96	0.297 (0.271)	6.92
crq	0.95	0.366 (0.189)	-	0.404 (0.280)	-	0.450 (0.520)	-
BVST	0.95	0.383 (0.318)	6.88	0.359 (0.393)	6.60	0.386 (0.388)	6.20
BTQRg	0.95	0.223 (0.168)	7.00	0.223 (0.217)	6.92	0.343 (0.290)	6.89
Des. II							
crq	0.50	0.296 (0.095)	-	0.343 (0.120)	-	0.408 (0.271)	-
BVST	0.50	0.159 (0.084)	4.84	0.315 (0.301)	4.64	0.371 (0.357)	4.46
BTQRg	0.50	0.149 (0.080)	4.96	0.221 (0.121)	4.84	0.296 (0.253)	4.99
crq	0.75	0.294 (0.092)	-	0.350 (0.121)	-	0.286 (0.263)	-
BVST	0.75	0.283 (0.211)	4.68	0.320 (0.299)	4.72	0.381 (0.338)	4.66
BTQRg	0.75	0.172 (0.081)	4.96	0.218 (0.111)	4.92	0.338 (0.257)	4.96
crq	0.95	0.345 (0.132)	-	0.386 (0.293)	-	0.477 (0.435)	-
BVST	0.95	0.384 (0.311)	4.76	0.390 (0.374)	4.66	0.437 (0.425)	4.36
BTQRg	0.95	0.234 (0.175)	5.00	0.258 (0.291)	4.96	0.381 (0.384)	4.96

Table 3.4: Top subsets in Example 2 for Tobit QR.

			$\varepsilon \sim N(0,1)$		$\varepsilon \sim t_3$		$\varepsilon \sim \chi^2_3$	
Des.	Model	p	Variables	APMP	Variables	APMP	Variables	APMP
I	BVST	0.50	Inter., x_1	0.90	Inter., x_1	0.87	Inter., x_1	0.85
	BTQRg	0.50	Inter., x_1	0.97	Inter., x_1	0.96	Inter., x_1	0.91
	BVST	0.75	Inter., x_1	0.85	Inter., x_1	0.84	Inter., x_1	0.74
	BTQRg	0.75	Inter., x_1	0.97	Inter., x_1	0.96	Inter., x_1	0.90
	BVST	0.95	Inter., x_1	0.82	Inter., x_1	0.76	Inter., x_1	0.60
	BTQRg	0.95	Inter., x_1	0.97	Inter., x_1	0.96	Inter., x_1	0.87
II	BVST	0.50	Inter., $x_1, x2, x5$	0.89	Inter., $x_1, x2, x5$	0.89	Inter., $x_1, x2, x5$	0.86
	BTQRg	0.50	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.91
	BVST	0.75	Inter., $x_1, x2, x5$	0.85	Inter., $x_1, x2, x5$	0.88	Inter., $x_1, x2, x5$	0.85
	BTQRg	0.75	Inter., $x_1, x2, x5$	0.92	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.91
	BVST	0.95	Inter., $x_1, x2, x5$	0.84	Inter., $x_1, x2, x5$	0.81	Inter., $x_1, x2, x5$	0.52
	BTQRg	0.95	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.91

Table 3.5: MMADs and SD for the simulated data in Example 3, where $p = 0.50$, 0.75 and 0.95.

		$\varepsilon \sim N(0,1)$	$\varepsilon \sim t_3$	$\varepsilon \sim \chi^2_3$
Model	p	MMAD (SD)	MMAD (SD)	MMAD (SD)
BVST	0.50	0.192 (0.089)	0.323 (0.235)	0.473 (0.358)
BTQRg	0.50	0.183 (0.093)	0.211 (0.155)	0.337 (0.294)
BVST	0.75	0.325 (0.339)	0.361 (0.289)	0.552 (0.403)
BTQRg	0.75	0.258 (0.278)	0.328 (0.293)	0.369 (0.413)
BVST	0.95	0.529 (0.363)	0.631 (0.346)	0.670 (0.619)
BTQRg	0.95	0.497 (0.321)	0.589 (0.334)	0.611(0.680)

with the stochastic search variable selection reported in Reed et al. (2009) using the `SSVSquantreg()` function (Martin et al., 2011). SSVSQ and SSVSquantreg were compared using marginal inclusion probabilities (MIP) at two quantiles, these were 0.50 and 0.95. We ran both algorithms SSVSQ and SSVSquantreg for 17000 iterations, removing the first 2000 as burn in. The results of the marginal inclusion probabilities are plotted in Figures 3.2, 3.3, 3.4 and 3.5. Clearly, one can observe that our approach (SSVSQ) tends to perform better than SSVSquantreg, especially for $p = 0.95$.

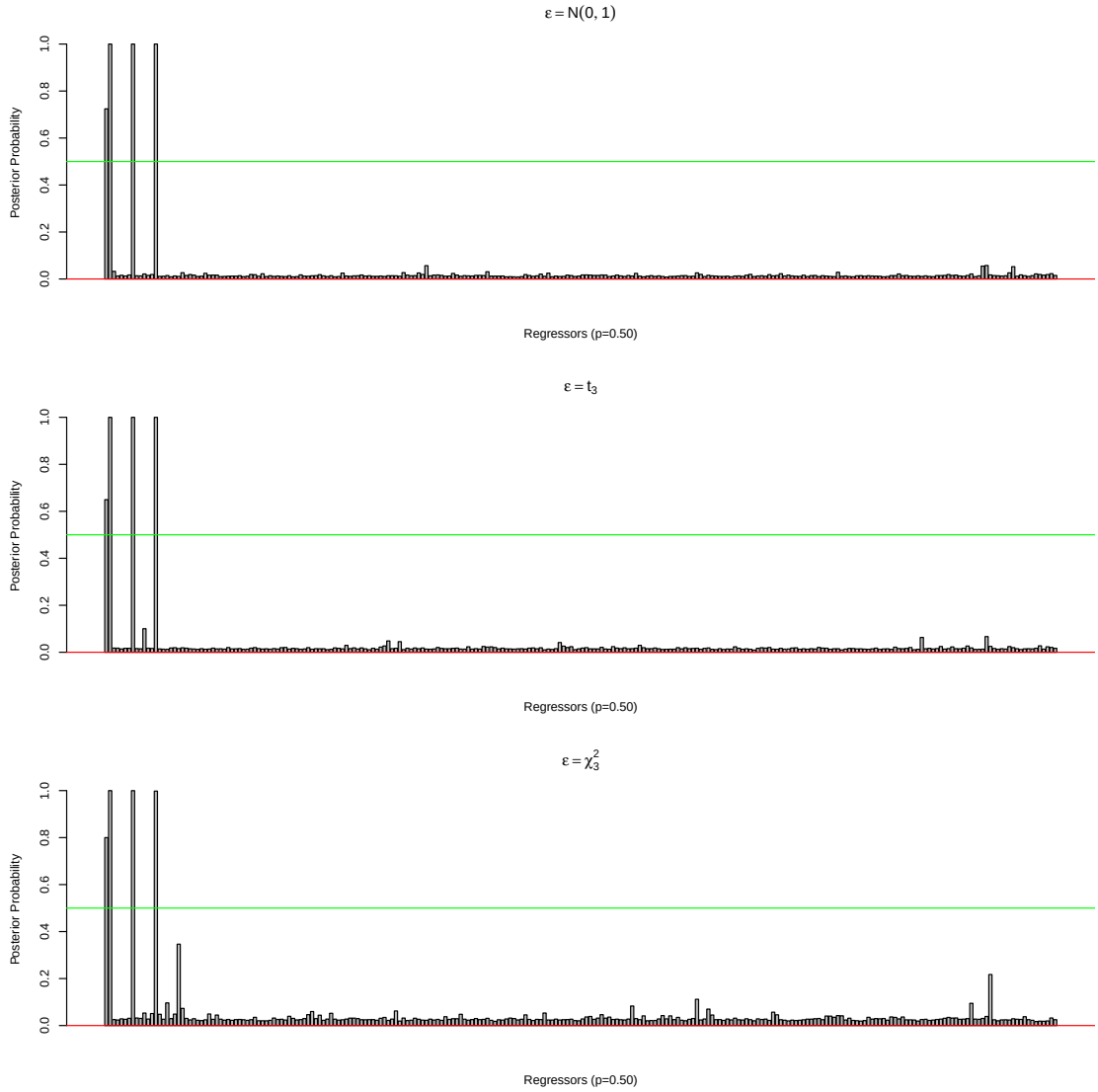

Figure 3.2: MIP for the simulated data in Example 4 at the median ($p = 0.50$) by using SSVSquantreg function.

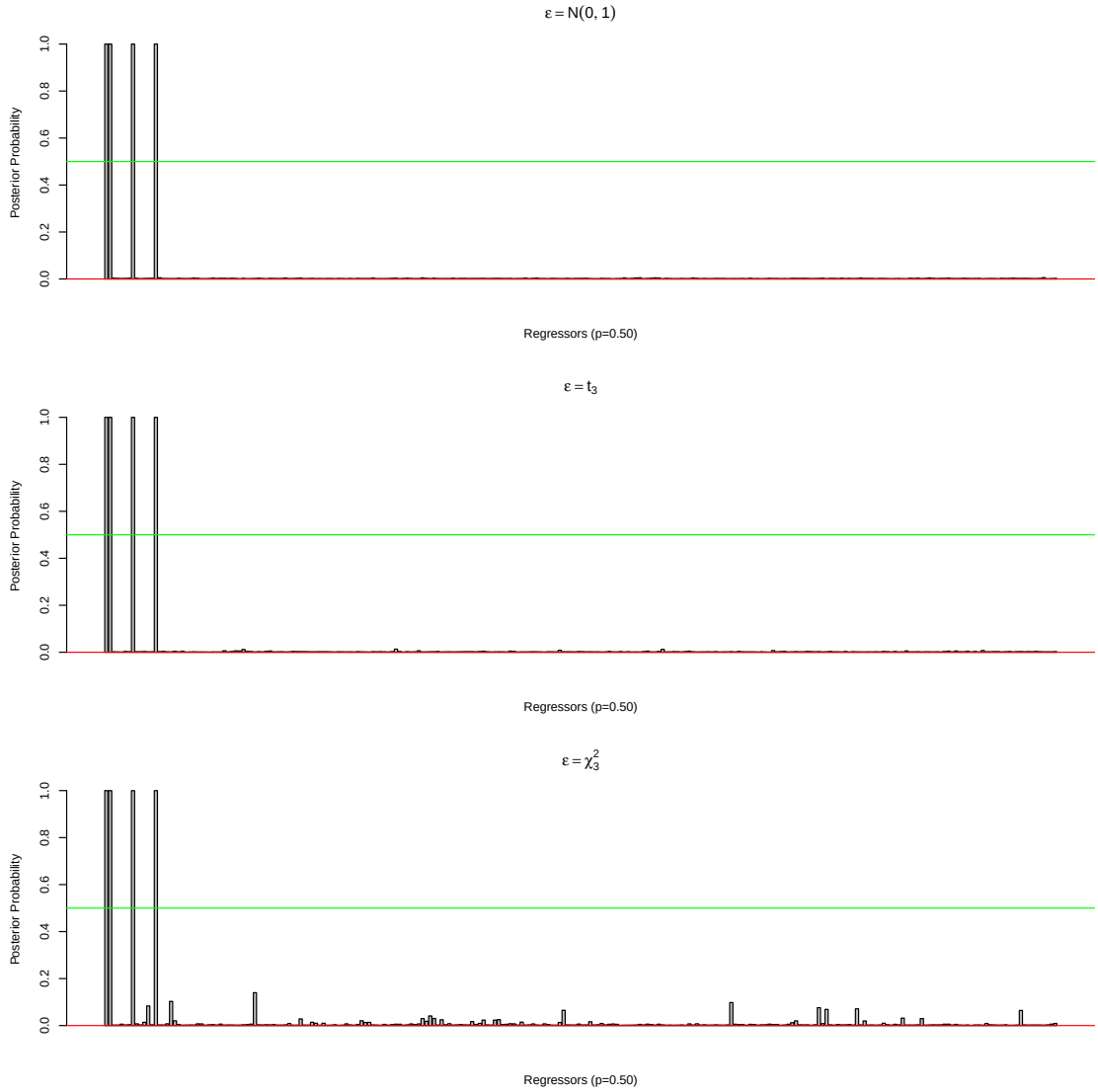

Figure 3.3: MIP for the simulated data in Example 4 at the median ($p = 0.50$) by using SSVSQ.

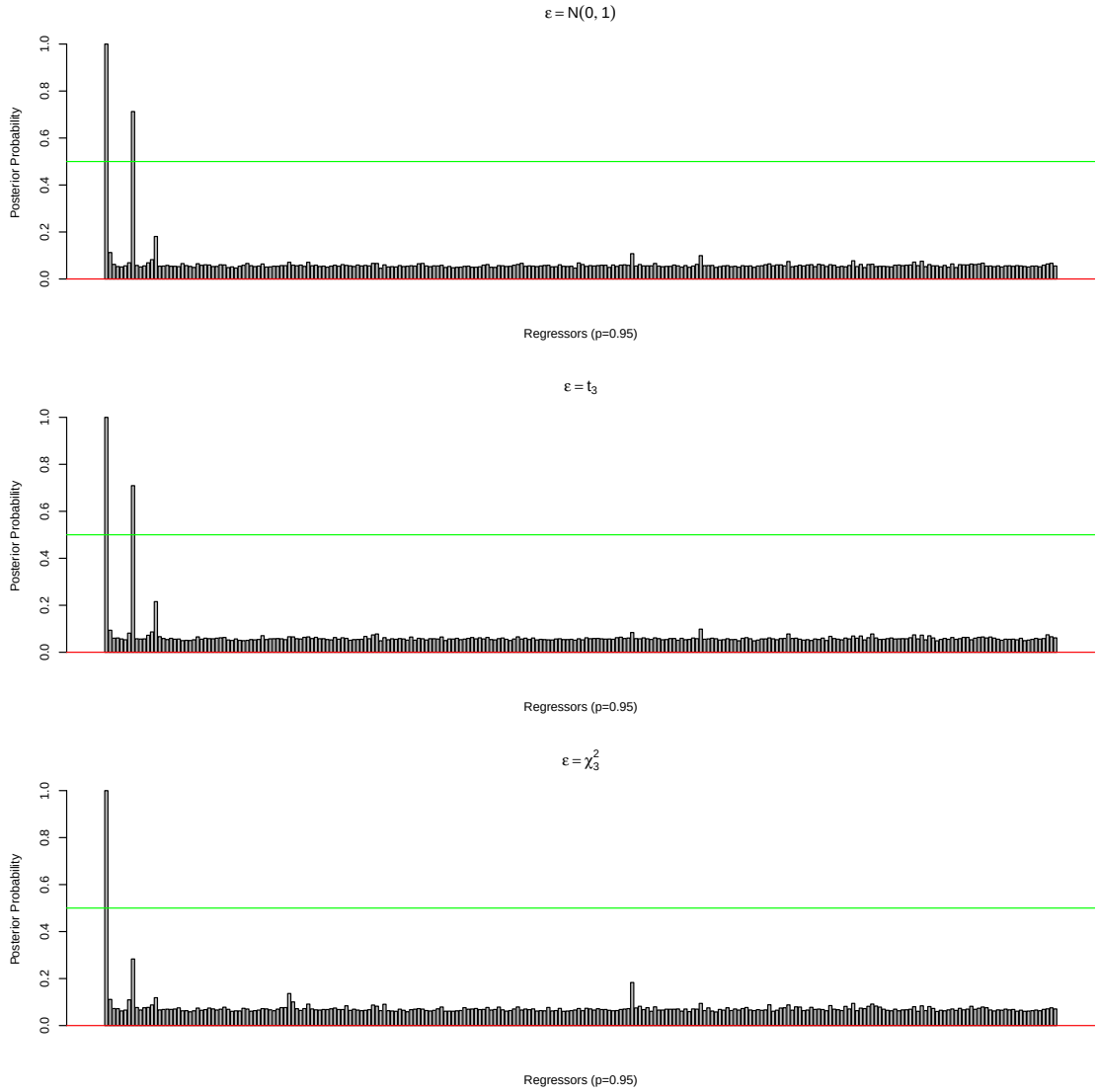

Figure 3.4: MIP for the simulated data in Example 4 when $p = 0.95$ by using SSVSquantreg function.

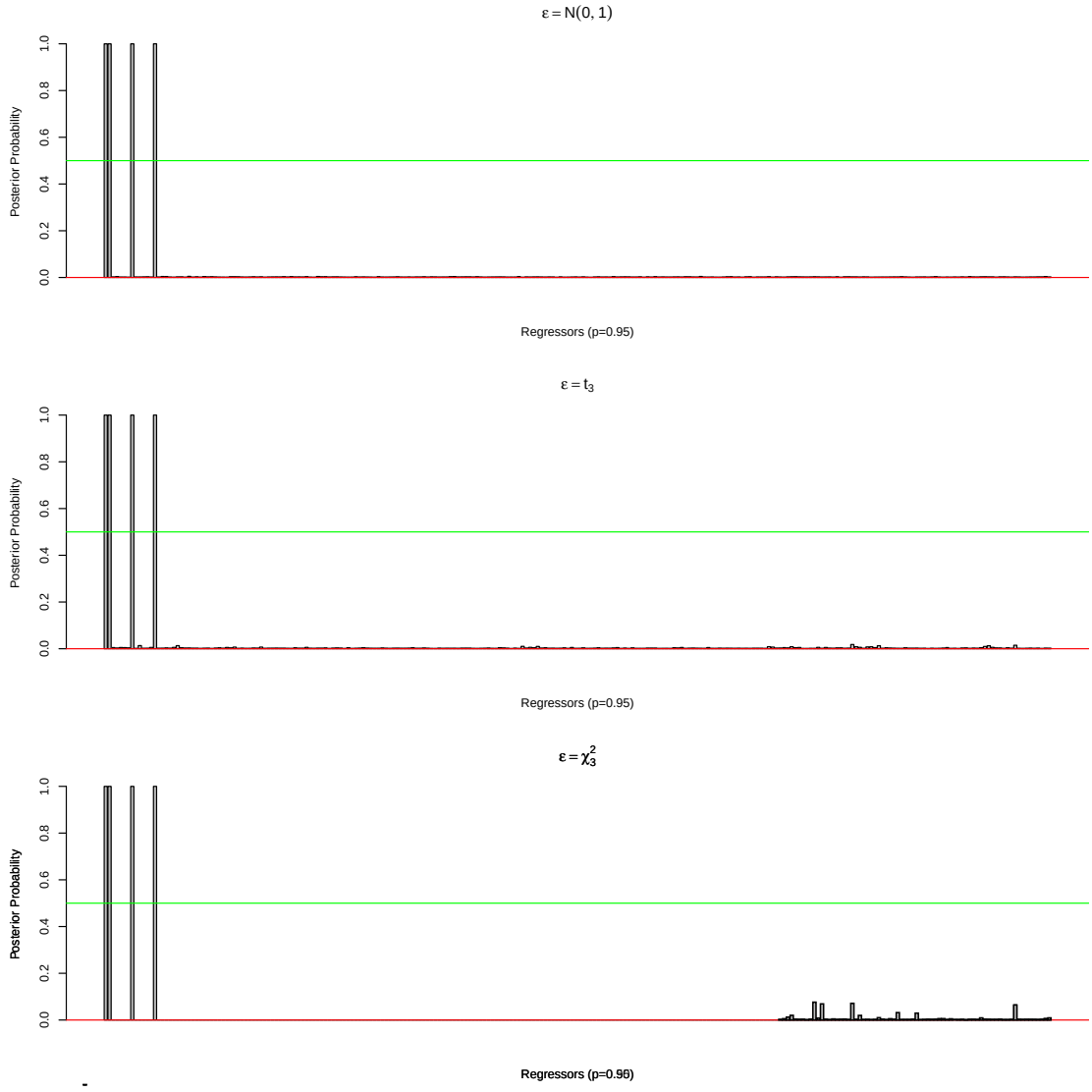

Figure 3.5: MIP for the simulated data in Example 4 when $p = 0.95$ by using SSVSQ.

3.5 Leukemia data set

The proposed technique in subsection 3.3.2 for subset selection in binary QR is illustrated using the popular leukemia dataset reported in Golub et al. (1999). This data describes 7,129 human genes in 72 patient samples labelled: ALL and AML. Here, ALL referred to as "*acute lymphoblastic leukemia*" which consists of 47 samples of 72 and the remaining are AML, which are referred to as "*acute myeloid leukemia*".

ALL further splits into 27 training and 11 testing samples, while AML splits into 20 training and 14 testing samples (Golub et al., 1999).

Table 3.6 lists the top 10 most significant genes selected by the proposed model in Subsection 3.3.2 for $p \in \{0.25, 0.50, 0.75\}$. Table 3.6 also shows the active genes chosen by Yang and Song (2010), Lee (2009) or Golub et al. (1999), when $p = 0.50$. The stronger gene is Zyxin which is also chosen as an active gene by Golub et al. (1999), Lee (2009), Bae and Mallick (2004), and Yang and Song (2010), among others. The crucial role of this gene in classification has been shown by Bae and Mallick (2004) who used only this gene for classification of testing dataset and got only three errors in classification, while Golub et al. (1999) applied 50 genes for classification and got five errors in classification. Our method identifies this gene (Zyxin) as the leading gene based on the posterior gene inclusion probabilities, which indicates that our method performs well.

A more complete view of gene effects can be supplied by the first and third quartiles, i.e, $p = 0.25$ and $p = 0.75$. From Table 3.6, it can be observed that the stronger gene is Macmarcks when $p = 0.25$, while the stronger gene is CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage) when $p = 0.75$. From Table 3.6, it can be seen that both genes are also identified by Golub et al. (1999), Lee (2009) and Yang and Song (2010). The simulation studies and the leukaemia dataset example indicate sturdy support for the apply of our method.

3.6 Chapter summary

In this chapter, we developed a method for Bayesian subset selection in Tobit QR based on a modification of Zellner's informative g-prior to take into account different priors for different quantile levels. The proposed prior is firstly developed for settings in which $k < n$, and then extended to deal with multicollinearity and over-fitting problems. Some extensions of our technique are also discussed, including the continuous and binary responses in quantile regression. We have also presented an expression for the hyperparameter g to calibrate the modified g-prior with a ridge parameter to a corresponding g-prior. Clear advantages over approaches proposed by Reed et al. (2009) and Ji et al. (2012) include quantile dependent priors and efficiency of Bayesian computation. The advantage of the method is that the prior distribution changes automatically when we change the quantile. Thus, we have different priors for different quantiles.

Table 3.6: The top 10 significant genes selected by the proposed model.

p	Rank	Index	Gene description
0.25	1	804	Macmarcks
	2	1685	Termianl transferase mRNA
	3	3847	HoxA9 mRNA
	4	2354	CCND3 Cyclin D3
	5	1779	MPO Myeloperoxidase
	6	4847	Zyxin
	7	2402	Azurocidin gene
	8	760	CYSTATIN A
	9	1882	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)
	10	6041	APLP2 Amyloid beta (A4) precursor-like protein 2
0.50	1	4847	Zyxin [a,b,c]
	2	760	CYSTATIN A[b,c]
	3	804	Macmarcks [a,b,c]
	4	4052	Catalase (EC 1.11.1.6) 5′flank and exon 1 mapping to chromosome 11, band p13 [a,c]
	7	1882	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)[a,b,c]
	6	1144	SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)[b]
	5	1745	LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
	8	1630	Inducible protein mRNA[b]
	9	2288	DF D component of complement (adipsin)[b]
	10	1953	Fc-epsilon-receptor gamma-chain mRNA
0.75	1	1882	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)
	2	4377	ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen
	3	1834	CD33 CD33 antigen (differentiation antigen)
	4	760	CYSTATIN A
	5	4336	ARHG Ras homolog gene family, member G (rho G)
	6	4847	Zyxin
	7	6041	APLP2 Amyloid beta (A4) precursor-like protein 2
	8	3847	HoxA9 mRNA
	9	1953	Fc-epsilon-receptor gamma-chain mRNA
	10	4328	PROTEASOME IOTA CHAIN

[a] Golub et al. (1999), [b] Lee (2009), [c] Yang and Song (2010)

Chapter 4

Bayesian subset selection for fixed and random effects in QR models

In many clustered applications, analysts are interested in identifying the coefficients of independent variables that may vary across a set of clusters to build good prediction models. This chapter considers the idea of Bayesian subset selection for both fixed and random coefficients in quantile mixed models (QMM) using an ALD-based working model. Some extensions are outlined and discussed, including the selection process in binary and Tobit quantile mixed-effects models. Illustrative examples involving age-related macular degeneration data are given to demonstrate the methodology.

4.1 Introduction

Clustered data is encountered in a wide variety of applications, including agriculture, education, finance, ecology, geology, medicine and social repeated measures studies. Since being introduced in Laird and Ware (1982), the mixed model with Random Effects (REs) has become a popular and effective technique to deal with clustered

data. This model consists of fixed and REs, the fixed coefficients of independent variables give the data intercept and slopes, while the REs account for the correlation and the heterogeneity among a set of clusters. One of the serious challenges in the linear-mixed model (LME) lies in selecting both fixed and REs. To solve this problem, AIC and BIC have been used over the years to select these effects. Recently, MCMC-based computation techniques have been proposed in traditional mean regression framework for selecting the fixed and REs (see, Chen and Dunson, 2003; Kinney and Dunson, 2007; Saville and Herring, 2009; Bondell et al., 2010; Ibrahim et al., 2011).

In this chapter a Bayesian approach for finding promising subsets of fixed and REs in the QR model is developed using a MCMC-based computation technique. This approach is related to the earlier approaches reported by Chen and Dunson (2003) and Kinney and Dunson (2007), but for mixed quantile regression (QR) models including: continuous, binary and left-censored responses. A key step in this approach is introducing a hierarchical Bayesian model to shrink the REs towards zero by proposing an l_1 penalty in the empirical check function of Koenker and Bassett (1978). This helps identify the exact prior distribution for the variances of the random effects (REs). Another key step is using a robust quantile dependent prior for subset selection and estimation in Bayesian QR. To author's knowledge, this is the first work discussing selecting both fixed and REs in QR models. Our motivating example is an analysis of age-related macular degeneration (ARMD) data, which was previously analysed by Chaili (2008). This study had a total of 203 patients who were randomly chosen from three cities (centres) including London, Belfast and Southampton. The goal of this study is to find the relationship between the distance visual acuity (DVA) and a subset of covariates. The change in DVA of each patient was measured four times over a 24 month period, where data was measured on the 3th, 6th, 12th and 24th months; see Chaili (2008) for more details. In this chapter we are interested in finding the most significant independent variables and random effects (REs) for the QR model, relating to the change in distance visual acuity (DVA).

The rest of this chapter is presented in the following way. In Section 4.2, the

re-parameterisation in linear mixed models are reviewed and an l_1-penalised check function for mixed QR models is proposed in Section 4.3. In Section 4.4, the prior specification and Bayesian MCMC technique are presented and Section 4.5 discusses some extensions of the method, including selection of both fixed and REs in binary and Tobit quantile mixed-effects models. The proposed methods are examined using simulation studies in Section 4.6 and using the ARMD data in Section 4.7. A chapter summary follows in Section 4.8.

4.2 Linear mixed models

Suppose y_{it} denotes the outcome variable for the ith cluster measured at the tth time point, $\boldsymbol{x}'_{it}$ is a $1 \times k$ vector of predictors, $\boldsymbol{z}'_{it}$ is a $1 \times q$ vector of predictors, and $\boldsymbol{x}'_{it}$ and $\boldsymbol{z}'_{it}$ are rows of the design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$, respectively, where $i = 1, ..., N$ and $t = 1, ..., n_i$. Then, according to Laird and Ware (1982), the LME model can be denoted as

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{\alpha}_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathrm{N}(0, \sigma^2), \tag{4.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$ are k and q-dimensional unknown fixed coefficients and REs, respectively, $\boldsymbol{X}_i$ is $n_i \times k$, $\boldsymbol{Z}_i$ is $n_i \times q$, $\boldsymbol{\alpha}_i \sim N(0, \boldsymbol{\Sigma}_\alpha)$, and ε_{it} is the residual term. We further assume that σ_{lr} denote the (l, r)th element of $\boldsymbol{\Sigma}_\alpha$, for $l, r = 1, \cdots, q$, where σ_{ll} is the lth RE variance.

Chen and Dunson (2003) suggested a technique for finding active REs by re-parametrisation of the covariance matrix of the REs $\boldsymbol{\Sigma}_\alpha$ such that $\boldsymbol{\Sigma}_\alpha = \boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{\Delta}'\boldsymbol{A}$ and $\boldsymbol{z}'_{it}\boldsymbol{\alpha}_i = \boldsymbol{z}'_{it}\boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{h}_i$. Here, $\boldsymbol{A} = \mathrm{diag}(a_1, ..., a_q)'$ such that a_l is proportional to $\sqrt{\sigma_{ll}}$, $\boldsymbol{\Delta}$ is a $q \times q$ lower triangular matrix such that δ_{lr} is describing the correlation between the lth and rth random effects, $\mathrm{diag}(\boldsymbol{\Delta}) = (1, \ldots, 1)'$ and $\boldsymbol{h}_i \sim \mathrm{N}(0, I)$. The authors show that the parameters in $\boldsymbol{A}$ and $\boldsymbol{\Delta}$ have the conjugacy feature that allows to improve mixing and create an efficient MCMC-based computation technique for fitting the

LME (Chen and Dunson, 2003). Recently, Kinney and Dunson (2007) rewrite $\boldsymbol{\Sigma}_\alpha$ such that $\boldsymbol{\Sigma}_\alpha = \boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{\Omega}\boldsymbol{\Delta}'\boldsymbol{F}$, where $\boldsymbol{\Omega} = \text{diag}(\omega_1, ..., \omega_q)'$ is a diagonal matrix and $\boldsymbol{F} = \text{diag}(f_1, ..., f_q)'$ is a diagonal matrix, whose f_l element is proportional to a_l. Setting $f_l = 0$ is equivalent to reducing the model by removing the irrelavant lth random effect from the model. Thus, the authors replaced the model in (4.1) with

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim \text{N}(0, \sigma^2), \tag{4.2}$$

where $\boldsymbol{b}_i = (b_{i1}, ..., b_{iq})'$ and $\boldsymbol{b}_i \sim \text{N}(0, \boldsymbol{\Omega})$.

4.3 Linear mixed QR

Following the re-parametrisation given by Chen and Dunson (2003) and Kinney and Dunson (2007), we propose the following l_1-penalised check function,

$$\min_{\boldsymbol{\beta}_p, \boldsymbol{b}_p} \sum_{i=1}^{N}\sum_{t=1}^{n_i} \rho_p(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_p - \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_{ip}) + r_1 \sum_{i=1}^{N}\sum_{l=1}^{q} |b_{ilp}|, \tag{4.3}$$

where $\boldsymbol{b}_p = (\boldsymbol{b}_{1p}', ..., \boldsymbol{b}_{Np}')'$ and r_1 is a nonnegative regularisation parameter. In this chapter, r_1 is restricted to being 1. For simplicity of notation, we will omit the subscript p from $\boldsymbol{\beta}_p, \boldsymbol{b}_p, \boldsymbol{b}_{ip}$ and b_{ilp} in the remainder of the chapter.

If a Laplace prior $(1/2)\exp\{-|b_{il}|\}$ is employed on b_{il} and assumed that the residuals ε_{it} follow an ALD$(0, \tau, p)$, then the density of $(\boldsymbol{y}, \boldsymbol{b})$ is given by

$$f(\boldsymbol{y}, \boldsymbol{b}|\boldsymbol{\beta}, \tau, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{F}, \boldsymbol{\Delta}) \propto \exp\{-\sum_{i=1}^{N}(\sum_{t=1}^{n_i} \frac{|\varepsilon_{it}| + (2p-1)\varepsilon_{it}}{2\tau} + \sum_{l=1}^{q} |b_{il}|)\}, \tag{4.4}$$

where $\boldsymbol{y} = (y_{11}, \ldots, y_{N_{n_N}})'$, $\boldsymbol{b} = (\boldsymbol{b}_1', \ldots, \boldsymbol{b}_N')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_N')'$, $\boldsymbol{Z} = \text{diag}(\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_N')'$ and $\varepsilon_{it} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i$. In the clustered data, the mixture

representation of the ALD is given by

$$\tau^{-1}\exp\{-\frac{|\varepsilon_{it}|+(2p-1)\varepsilon_{it}}{2\tau}\} \propto \int \mathrm{N}(\varepsilon_{it};\xi v_{it},2\tau v_{it})\mathrm{Exp}(v_{it};\zeta)dv_{it}. \tag{4.5}$$

4.4 Priors specification and Bayesian sampler

4.4.1 Priors specification

In this chapter, the prior (2.10) is assigned for the fixed effects such that

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\tau,\boldsymbol{\gamma},\tilde{\boldsymbol{V}},\boldsymbol{X}_{\boldsymbol{\gamma}} \sim N_{k_\gamma}(0,2g\tau(\boldsymbol{X}'_{\boldsymbol{\gamma}}\tilde{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{\gamma}})^{-1}), \tag{4.6}$$

$$p(\tau) \propto \tau^{-1}, \tag{4.7}$$

$$v_{it}|\tau \sim \mathrm{Exp}(\zeta), \tag{4.8}$$

$$\boldsymbol{\gamma}|\pi \sim \pi^{k_\gamma}(1-\pi)^{k-k_\gamma}, \tag{4.9}$$

where $\tilde{\boldsymbol{V}}$=diag$({v_{11}}^{-1},...,{v_{N_{n_N}}}^{-1})$ and $\pi \in (0,1)$. According to Andrews and Mallows (1974), the Laplace prior $(1/2)\exp\{-|b_{il}|\}$ on $b_{il}, l=1,...,q$, can be written as:

$$\frac{1}{2}\exp\{-|b_{il}|\} = \int_0^\infty \mathrm{N}(b_{il};0,\omega_l)\mathrm{Exp}(\omega_l;\frac{1}{2})d\omega_l, \tag{4.10}$$

It should be noted that Kinney and Dunson (2007) assumed that b_{il} ~N$(0,\omega_l)$ and ω_l ~InvGa$(\omega_l;1/2,n/2)$, but from (4.10) it can be observed that the exact prior (knowledge) for ω_l is Gamma$(\omega_l;1,2)$. Although, in theory, any prior for ω_l could be used, it is crucial to elicit a prior for ω_l that is as informative as possible of the investigators idea.

The prior distributions for $f_l, l=1,...,q$, and $\boldsymbol{\delta}=(\delta_{lr}: l=2,\cdots,q; r=1,\cdots,l-1)'$ are specified in a same way as Chen and Dunson (2003) and Kinney and Dunson (2007). Thus, a standard half normal distribution truncated at the left by

zero ZI-N$^+(p_{l0}, 0, 1)$ is specified for f_l, where $p_{l0} = p(f_l = 0)$. For $\boldsymbol{\delta}$, it is assumed that $p(\boldsymbol{\delta}|\boldsymbol{f}) = \mathrm{N}(\mathbf{0}, \boldsymbol{R}_{\boldsymbol{\delta}}).1(\boldsymbol{\delta} \in \boldsymbol{R}_{\boldsymbol{f}})$, where 1(.) is an indicator function and $\boldsymbol{R}_{\boldsymbol{f}}$ restricted the components of $\boldsymbol{\delta}$ to be zero such that the corresponding f_l's, $l = 1, ..., q$, are equal to zero (Kinney and Dunson, 2007).

4.4.2 Bayesian sampler for variable selection

Under the specified prior distributions and following Chen and Dunson (2003) and Kinney and Dunson (2007), it is possible to simulate all parameters from their full conditional distributions. Let $\boldsymbol{y}_i = (y_{i1}, ..., y_{in_i})'$, $\tilde{\boldsymbol{v}} = (v_{11}, ..., v_{N_{n_N}})'$, $\tilde{\boldsymbol{V}}_i = diag(\tilde{\boldsymbol{v}}_i^{-1})$, $\tilde{\boldsymbol{v}}_i = (v_{i1}, ..., v_{in_i})'$, $T_{it} = \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i$, $\boldsymbol{T} = (T_{11}, \cdots, T_{N_{n_N}})'$, $\boldsymbol{o}_{1it} = (b_{il} f_m z_{itm} : l = 1, ..., q, m = l+1, ..., q)'$, $\boldsymbol{o}_{2it} = (z_{itl}(b_{il} + \sum_{m=1}^{l-1} b_{im}\delta_{ml}) : l = 1, ..., q)'$, $n = \sum_{i=1}^{N} n_i$ and k_γ denote the size of the $\boldsymbol{\gamma}$th subset model.

- **Updating $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$**

  The full conditional distribution (CD) of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is $\mathrm{N}_{k_\gamma}(\boldsymbol{\mu}_{\boldsymbol{\beta}_{\boldsymbol{\gamma}}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\boldsymbol{\gamma}}})$, where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_\gamma} = \frac{g}{g+1}(\boldsymbol{X}_{\boldsymbol{\gamma}}'\tilde{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{\gamma}})^{-1}\boldsymbol{X}_{\boldsymbol{\gamma}}'\tilde{\boldsymbol{V}}(\boldsymbol{y} - \boldsymbol{T} - \xi\tilde{\boldsymbol{v}}),$$

  and

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_\gamma} = \frac{2\tau g}{(g+1)}(\boldsymbol{X}_{\boldsymbol{\gamma}}'\tilde{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{\gamma}})^{-1}.$$

- **Updating $\boldsymbol{\gamma}$**

  The full CD of each γ_j, $p(\gamma_j|\boldsymbol{\gamma}_{-j}, \tilde{\boldsymbol{v}}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{T})$, is Bernoulli (π_1), where

$$\pi_1 = \left(1 + \frac{(1-\pi)S(\gamma_j = 0, \boldsymbol{\gamma}_{-j})\sqrt{1+g}}{\pi S(\gamma_j = 1, \boldsymbol{\gamma}_{-j})}\right)^{-1},$$

and

$$S(\boldsymbol{\gamma}) = (\frac{1}{4}(\boldsymbol{y}-\boldsymbol{T}-\xi\tilde{\boldsymbol{v}})'\tilde{\boldsymbol{V}}(\boldsymbol{y}-\boldsymbol{T}-\xi\tilde{\boldsymbol{v}}) - \frac{g}{4(g+1)}(\boldsymbol{y}-\boldsymbol{T}-\xi\tilde{\boldsymbol{v}})'\tilde{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{\gamma}}(\boldsymbol{X}'_{\boldsymbol{\gamma}}\tilde{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{\gamma}})^{-1} \times\boldsymbol{X}'_{\boldsymbol{\gamma}}\tilde{\boldsymbol{V}}(\boldsymbol{y}-\boldsymbol{T}-\xi\tilde{\boldsymbol{v}}) + p(1-p)\sum_{i=1}^{N}\sum_{t=1}^{n_i} v_{it})^{-3n/2}.$$

- **Updating $\boldsymbol{b}_i$**

  The full CD of $\boldsymbol{b}_i$, $p(\boldsymbol{b}_i|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\tau,\boldsymbol{\gamma},\tilde{\boldsymbol{v}},\boldsymbol{y},\boldsymbol{X}_{\boldsymbol{\gamma}i},\boldsymbol{Z}_i,\boldsymbol{F},\boldsymbol{\Delta})$, is $\mathrm{N}_q(\boldsymbol{\mu}_{\boldsymbol{b_i}},\boldsymbol{\Sigma}_{\boldsymbol{b_i}})$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{b_i}} = (\frac{\boldsymbol{\Delta}'\boldsymbol{F}\boldsymbol{Z}'_i\tilde{\boldsymbol{V}}_i\boldsymbol{Z}_i\boldsymbol{F}\boldsymbol{\Delta}}{2\tau}+\boldsymbol{\Omega}^{-1})^{-1},$$

  and

$$\boldsymbol{\mu}_{\boldsymbol{b_i}} = \frac{\boldsymbol{\Sigma}_{\boldsymbol{b_i}}\boldsymbol{\Delta}'\boldsymbol{F}\boldsymbol{Z}'_i\tilde{\boldsymbol{V}}_i}{2\tau}(\boldsymbol{y}_i-\boldsymbol{X}_{\boldsymbol{\gamma}i}\boldsymbol{\beta}_{\boldsymbol{\gamma}}-\xi\tilde{\boldsymbol{v}}_i).$$

- **Updating $\boldsymbol{\delta}$**

  The full CD of $\boldsymbol{\delta}$, $p(\boldsymbol{\delta}|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\tau,\tilde{\boldsymbol{v}},\boldsymbol{\gamma},\boldsymbol{y},\boldsymbol{X}_{\boldsymbol{\gamma}},\boldsymbol{Z},\boldsymbol{F},\boldsymbol{b})$, is given by $N(\boldsymbol{\mu}_{\boldsymbol{\delta}},\boldsymbol{\Sigma}_{\boldsymbol{\delta}}).1(\boldsymbol{\delta}\in\boldsymbol{R}_{\boldsymbol{f}})$ where

$$\boldsymbol{\Sigma}_{\boldsymbol{\delta}} = (\sum_{i=1}^{N}\sum_{t=1}^{n_i}\frac{\boldsymbol{o}_{1it}\boldsymbol{o}'_{1it}}{2\tau v_{it}}+\boldsymbol{R}_{\boldsymbol{\delta}}{}^{-1})^{-1},$$

  and

$$\boldsymbol{\mu}_{\boldsymbol{\delta}} = \boldsymbol{\Sigma}_{\boldsymbol{\delta}}(\sum_{i=1}^{N}\sum_{t=1}^{n_i}\frac{\boldsymbol{o}_{1it}(y_{it}-\boldsymbol{x}'_{\boldsymbol{\gamma}it}\boldsymbol{\beta}_{\boldsymbol{\gamma}}-\xi v_{it})}{2\tau v_{it}}).$$

- **Updating τ**

$$p(\tau|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\boldsymbol{\gamma},\tilde{\boldsymbol{v}},\boldsymbol{y},\boldsymbol{X}_{\boldsymbol{\gamma}},\boldsymbol{T}) = \mathrm{InvGa}(\frac{3n+k_\gamma}{2},\frac{1}{4}(\boldsymbol{y}-\boldsymbol{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}-\boldsymbol{T}-\xi\tilde{\boldsymbol{v}})'\tilde{\boldsymbol{V}}(\boldsymbol{y}-\boldsymbol{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}-\boldsymbol{T}-\xi\tilde{\boldsymbol{v}}) + \frac{1}{4g}\boldsymbol{\beta}'_{\boldsymbol{\gamma}}(\boldsymbol{X}'_{\boldsymbol{\gamma}}\tilde{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{\gamma}})\boldsymbol{\beta}_{\boldsymbol{\gamma}}+p(1-p)\sum_{i=1}^{N}\sum_{t=1}^{n_i}v_{it}).$$

- **Updating $\tilde{v}$**

  The full CD of each v_{it} is GIG($\nu = 0, \varrho_1, \varrho_2$), where

  $\varrho_1^2 = \left((y_{it} - \boldsymbol{x}'_{\boldsymbol{\gamma}it}\boldsymbol{\beta}_{\boldsymbol{\gamma}} - \boldsymbol{z}'_{it}\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i)^2 + \boldsymbol{\beta}'_{\boldsymbol{\gamma}}\boldsymbol{x}_{\boldsymbol{\gamma}it}\boldsymbol{x}'_{\boldsymbol{\gamma}it}\boldsymbol{\beta}_{\boldsymbol{\gamma}}/g\right)/(2\tau)$ and $\varrho_2^2 = 1/(2\tau)$. For $i = 1, ..., N$ and $t = 1, ..., n_i$, we update v_{it} using the `rgig()` function in the R package `ghyp` (Luethi and Breymann, 2012).

- **Updating $\omega_l, l = 1, ..., q$.**

  The full CD of each ω_l is GIG($\nu, \varrho_1, \varrho_2$), where $\nu = -(N+2)/2$, $\varrho_1^2 = \sum_{i=1}^N b_{il}^2$ and $\varrho_2^2 = N$. For $l = 1, ..., q$, we update ω_l using the `rgig()` function.

- **Updating $f_l, l = 1, ..., q$.**

  The full CD of f_l is given by $ZI - N^+(\hat{p}_l, \hat{f}_l, \tilde{\sigma}_{f_l}^2)$ where

$$\tilde{\sigma}_{f_l}^2 = (\sum_{i=1}^{N}\sum_{t=1}^{n_i}\frac{o_{2itl}^2}{2\tau v_{it}} + 1)^{-1},$$
$$\hat{f}_l = \tilde{\sigma}_{f_l}^2(\sum_{i=1}^{N}\sum_{t=1}^{n_i}\frac{o_{2itl}(y_{it} - \boldsymbol{x}'_{\boldsymbol{\gamma}it}\boldsymbol{\beta}_{\boldsymbol{\gamma}} - \sum_{s\neq l} o_{2its}f_s - \xi v_{it})}{2\tau v_{it}}),$$

  and

$$\hat{p}_l = \left(1 + \frac{(1-p_{l0})N(0;0,1)(1-\Phi(0;\hat{f}_l,\tilde{\sigma}_{f_l}^2))}{p_{l0}N(0;\hat{f}_l,\tilde{\sigma}_{f_l}^2)(1-\Phi(0;0,1))}\right)^{-1}.$$

  Here, $\Phi(.)$ is the normal cumulative distribution function.

4.5 Model extensions

The l_1 penalised check function with random effects in (4.3) can be extended in several ways. In this section, some extensions for binary and Tobit mixed-effects QR are described.

4.5.1 Binary mixed QR

Here, it is shown that the l_1 penalised check function with REs in (4.3) is directly extended to subset selection for fixed and REs in Binary mixed-effects QR models. Binary mixed-effects QR can be considered as a linear mixed-effects QR with a latent outcome that is not observed completely. We suppose that the outcome y_{it} is generated according to the link function

$$y_{it} = \begin{cases} 1, & \text{if} \quad y_{it}^* \geq 0, \\ 0, & \text{if} \quad y_{it}^* < 0, \end{cases} \tag{4.11}$$

where

$$y_{it}^* = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{\alpha}_i + \varepsilon_{it}. \tag{4.12}$$

Thus, the l_1 penalised check function in (4.3) can be written as

$$\min_{\boldsymbol{\beta},\boldsymbol{b}} \sum_{i=1}^{N}\sum_{t=1}^{n_i} \rho_p(y_{it} - \eta(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i)) + \sum_{i=1}^{N}\sum_{l=1}^{q} |b_{il}|, \tag{4.13}$$

where $\eta(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i) = I\{\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i \geq 0\}$. To conduct Bayesian subset selection in mixed effects binary QR models, it is supposed that the priors are the same as those given in subsection 4.4.1. Under $\boldsymbol{\gamma}$ and the mixture representation (4.5), the full CD of y_{it}^* is given by

$$\begin{aligned} y_{it}^*|y_{it} = 1, v_{it}, \tau, \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{x}_{\boldsymbol{\gamma}it}, \boldsymbol{z}_{it}, \boldsymbol{F}, \boldsymbol{\Delta}, \boldsymbol{b}_i \\ \sim N(\boldsymbol{x}_{\boldsymbol{\gamma}it}'\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i + \xi v_{it}, 2\tau v_{it}) \ \text{ left truncated at 0,} \\ y_{it}^*|y_{it} = 0, v_{it}, \tau, \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{x}_{\boldsymbol{\gamma}it}, \boldsymbol{z}_{it}, \boldsymbol{F}, \boldsymbol{\Delta}, \boldsymbol{b}_i \\ \sim N(\boldsymbol{x}_{\boldsymbol{\gamma}it}'\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{z}_{it}'\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i + \xi v_{it}, 2\tau v_{it}) \ \text{ right truncated at 0.} \end{aligned}$$

All parameters in binary mixed-effects QR can be easily obtained as in subsection 4.4.2 by replacing y_{it} everywhere with y_{it}^*.

4.5.2 Tobit mixed QR

Under the pth quantile, the penalised check function in (4.13) can be assumed as penalised check function for mixed-effects Tobit QR if it is assumed that the outcome y_{it} is generated according to

$$y_{it} = \max\{y^0, y^*_{it}\}, \tag{4.14}$$

and

$$y^*_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{\alpha}_i + \varepsilon_{it}, \tag{4.15}$$

where y^0 is a known fixed point, y^*_{it} denotes the left-censored jth outcome of the ith cluster and $\eta(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i) = \max\{y^0, y^*_{it}\}$. Under the specified priors in subsection 4.4.1, it is straightforward to find that the full CD of y^*_{it} is

$$\begin{aligned}
&y^*_{it}|y_{it}, v_{it}, \tau, \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{x}_{\boldsymbol{\gamma}it}, \boldsymbol{z}_{it}, \boldsymbol{F}, \boldsymbol{\Delta}, \boldsymbol{b}_i \\
&\qquad \sim \Upsilon(y_{it}), \qquad\qquad \text{if } y_{it} > y^0, \\
&y^*_{it}|y_{it}, v_{it}, \tau, \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{x}_{\boldsymbol{\gamma}it}, \boldsymbol{z}_{it}, \boldsymbol{F}, \boldsymbol{\Delta}, \boldsymbol{b}_i \\
&\qquad \sim N(\boldsymbol{x}'_{\boldsymbol{\gamma}it}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{z}'_{it}\boldsymbol{F}\boldsymbol{\Delta}\boldsymbol{b}_i + \xi v_{it}, 2\tau v_{it}) I(y^*_{it} \leq y^0) \quad \text{otherwise.}
\end{aligned}$$

Again, all parameters in mixed effects Tobit QR can be easily updated as in subsection 4.4.2 by replacing y_{it} everywhere with y^*_{it}.

4.6 Simulation study

To test the behavior of the proposed method in Section 4.3, we simulate data from the model

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it3} + \alpha_{i1} + \alpha_{i2} z_{it1} + \alpha_{i3} z_{it2} + \alpha_{i4} z_{it3} + \varepsilon_{it}, \tag{4.16}$$

where $x_{itj} \sim \text{Uniform}(-2, 2)$ for $j = 1, ..., 8$, and $z_{it\tilde{l}} \sim \text{Uniform}(-2, 2)$ for $\tilde{l} = 1, 2, 3$. We set $\boldsymbol{\beta} = (1, 1, 1, 1, 0, 0, 0, 0, 0)'$ and $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4})' \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$, where

$$\boldsymbol{\Sigma}_\alpha = \begin{pmatrix} 0.90 & 0.40 & 0.06 & 0 \\ 0.40 & 0.70 & 0.10 & 0 \\ 0.06 & 0.10 & 0.10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The residuals $\varepsilon_{it}, i = 1, \cdots, 30,\ t = 1, \ldots, 5$ are simulated from three distributions: N(0,1), a t_3 and a χ^2_3 distribution. For each residual distribution, 200 data sets were simulated from the model (4.16) assuming there were 30 clusters and 5 observations per cluster. Then subset selection in mixed effects QR was carried out at three quantiles, namely $0.50, 0.75$ and 0.95. Each generated data set is analysed using four methods: the proposed method (MQR) as described in subsection 4.4.2, stochastic search variable selection in QR (SSVSquantreg) for fixed effects in the R package `MCMCpack` (Martin et al., 2011), model selection for fixed effects in QR using AIC criteria in the R package `quantreg` (Koenker, 2011) and fixed and REs selection in linear regression model (FRES) as reported by Kinney and Dunson (2007). The R code for FRES can be obtained from the Web location "`http://www.stat.duke.edu/~sk11/pubs.html`". We set $\boldsymbol{R_\delta} = 0.5\boldsymbol{I}_q$, $\pi = 0.5$ and the same prior specifications as those in Section 4.4 is followed.

In binary and Tobit mixed effects QR, the model (4.16) was used to simulate $\boldsymbol{y}^*$, and using the link functions (4.11) and (4.14) observed outcomes $\boldsymbol{y}$, respectively. The proposed method for binary mixed effects QR (BMQR) was compared with the Bayesian variable selection in Binary quantile regression (BVSB) for fixed effects as reported in Ji et al. (2012), and fixed and the REs selection in logistic regression model (LFRES). The proposed method for Tobit mixed effects QR (TMQR) was compared with Bayesian variable selection in Tobit quantile regression (BVST) for fixed effects, reported in Ji et al. (2012).

Table 4.1: Comparing average numbers of actual and wrong zero fixed effects of independent variables in simulation study using three different distributions for the errors. The proposed method MQR is compared with 3 methods: the SSVS in QR for fixed effects (SSVSquantreg), AIC for fixed effects and fixed and REs selection in linear regression models (FRES).

p	Error Distribution		MQR	QR SSVSquantreg	AIC	Mean regression FRES
0.50	$N(0,1)$	correct	4.93	4.69	3.63	4.84
		wrong	(0.00)	(0.01)	(0.01)	(0.00)
0.75	$N(0,1)$	correct	4.86	4.65	3.42	-
		wrong	(0.00)	(0.00)	(0.00)	-
0.95	$N(0,1)$	correct	4.77	4.49	2.06	-
		wrong	(0.00)	(0.01)	(0.01)	-
0.50	t_3	correct	4.83	3.93	3.64	4.51
		wrong	(0.00)	(0.00)	(0.02)	(0.00)
0.75	t_3	correct	4.71	3.89	3.40	-
		wrong	(0.00)	(0.01)	(0.00)	-
0.95	t_3	correct	4.56	3.28	2.12	-
		wrong	(0.00)	(0.01)	(0.03)	-
0.50	χ^2_3	correct	4.52	1.94	3.57	4. 31
		wrong	(0.00)	(0.05)	(0.03)	(0.02)
0.75	χ^2_3	correct	4.61	1.79	3.36	-
		wrong	(0.01)	(0.07)	(0.01)	-
0.95	χ^2_3	correct	4.33	1.70	2.16	-
		wrong	(0.00)	(0.12)	(0.11)	-

4.6.1 Results

From Table 4.1, it can be observed that the Bayesian subset selection for fixed and REs (MQR) was more efficient compared to SSVSquantreg, AIC and FRES. As expected, the AIC and SSVSquantreg did not perform well because they ignore the random effects. It can be observed that MQR produced a higher average number of actual zeros compared with SSVSquantreg, AIC and FRES. Instead of looking at the average number of actual and wrong zeros, we may also look at the coefficients estimation and 95% intervals in Table 4.2. We can observe that the MQR performs well when comparing the estimates of $\beta_j, j = 1, ..., k$, and $\sigma_{ll}, l = 1, ..., q$, with the true values of β_j and σ_{ll}, respectively. It can be seen that the RQ procedure does not perform well for $p = 0.75$ and 0.95. It is easy to observe that the true value of β_0, for $p = 0.75$ and 0.95, lies outside the confidence interval (95% C.I) obtained using the RQ procedure because it ignores the random effects (REs) entirely. We also see that

the credible intervals (95% CrI) obtained using the MQR are generally shorter than the confidence intervals (95% C.I) obtained using the RQ procedure and the credible intervals obtained using FRES, suggesting an efficiency gain and stable estimation from the posterior distributions. For example, the interval width of β_0 using the MQR for $p = 0.50$ is 0.20 compared to the interval width of 0.55 for the RQ procedure and the interval width of 0.41 for the FRES procedure. These differences in the interval widths are more apparent for $\varepsilon_{it} \sim t_3$ and $\varepsilon_{it} \sim \chi^2_3$, the results for which are not shown here. Finally, given a coefficient of an independent variable, MQR and FRES give very similar marginal inclusion probabilities at the median. This indicates that the MQR performs well compared to the FRES.

From Table 4.3, we can see that the behavior of the BMQR, in terms of selecting actual zeros is better than the other approaches. The performance of BMQR is quite similar to logistic model, but the gap between BMQR and BVSB is very large especially in extreme quantiles. A similar conclusion can be observed in the Tobit mixed effects QR.

Table 4.2: Posterior mean (Pm), 95% intervals and marginal inclusion probabilities (MIP) for both the fixed and REs when the error is normal.

p	Parameter	True value	MQR Pm (95% CrI)	MIP	RQ Pm (95% C.I)	FRES Pm (95% CrI)	MIP
0.50	β_0	1.00	1.08 (0.91, 1.11)	1.00	1.00 (0.72, 1.27)	1.15 (0.94, 1.35)	1.00
	β_1	1.00	0.99 (0.88, 1.12)	1.00	0.99 (0.76, 1.24)	0.94 (0.85, 1.17)	1.00
	β_2	1.00	0.99 (0.94, 1.05)	1.00	1.00 (0.75, 1.23)	1.03 (0.96, 1.13)	1.00
	β_3	1.00	1.03 (0.98, 1.09)	1.00	0.98 (0.74, 1.22)	0.96 (0.81, 1.16)	1.00
	β_4	0.00	0.00 (0.00, 0.01)	0.09	0.01 (-0.23, 0.25)	0.01 (-0.05, 0.21)	0.05
	β_5	0.00	0.00 (-0.01, 0.00)	0.03	0.00 (-0.24, 0.24)	0.00 (-0.07, 0.10)	0.03
	β_6	0.00	0.00 (-0.01, 0.00)	0.09	0.00 (-0.25, 0.24)	0.09 (-0.13, 0.19)	0.07
	β_7	0.00	0.01 (-0.01, 0.02)	0.08	0.01 (-0.24, 0.25)	0.05 (-0.01, 0.25)	0.09
	β_8	0.00	0.00 (0.00, 0.01)	0.04	0.00 (-0.24, 0.24)	-0.08 (-0.14, 0.07)	0.04
	σ_{11}	0.90	0.92 (0.73, 1.17)	1.00	-	0.99 (0.65, 1.31)	1.00
	σ_{22}	0.70	0.67 (0.51, 0.88)	1.00	-	0.76 (0.43, 1.16)	1.00
	σ_{33}	0.10	0.18 (0.08, 0.31)	0.91	-	0.23 (0.06, 0.38)	0.78
	σ_{44}	0.00	0.00 (0.00, 0.00)	0.06	-	0.00 (0.00, 0.00)	0.02
0.75	β_0	1.67	1.66 (1.49, 1.83)	1.00	2.09 (1.81, 2.42)	-	
	β_1	1.00	1.07 (0.95, 1.17)	1.00	0.98 (0.74, 1.27)	-	
	β_2	1.00	0.99 (0.91, 1.08)	1.00	1.03 (0.73, 1.27)	-	
	β_3	1.00	0.98 (0.89, 1.12)	1.00	0.99 (0.73, 1.27)	-	
	β_4	0.00	0.00 (-0.08, 0.00)	0.05	0.00(-0.27, 0.26)	-	
	β_5	0.00	0.00 (-0.01, 0.02)	0.06	0.03 (-0.27, 0.27)	-	
	β_6	0.00	0.01 (-0.07, 0.24)	0.12	0.00(-0.26, 0.28)	-	
	β_7	0.00	0.00 (0.00, 0.03)	0.04	0.01 (-0.26, 0.27)	-	
	β_8	0.00	0.01 (-0.02, 0.05)	0.10	-0.01 (-0.27, 0.26)	-	
	σ_{11}	0.90	0.95 (0.61, 1.32)	1.00	-	-	
	σ_{22}	0.70	0.76 (0.41, 1.13)	1.00	-	-	
	σ_{33}	0.10	0.09 (0.05, 0.17)	0.84	-	-	
	σ_{44}	0.00	0.00 (0.00, 0.00)	0.08	-	-	
0.95	β_0	2.64	2.69 (2.41, 2.96)	1.00	3.74 (3.35, 4.35)	-	
	β_1	1.00	0.98 (0.91, 1.11)	1.00	0.99 (0.48 1.49)	-	
	β_2	1.00	1.07 (0.93, 1.22)	1.00	1.03 (0.49 1.51)	-	
	β_3	1.00	1.01 (0.91, 1.12)	1.00	1.01 (0.48 1.51)	-	
	β_4	0.00	0.00 (-0.11, 0.13)	0.07	0.00 (-0.51 0.51)	-	
	β_5	0.00	0.00 (0.00, 0.00)	0.02	0.01 (-0.50 0.52)	-	
	β_6	0.00	0.01 (0.00, 0.02)	0.08	0.01 (-0.50 0.50)	-	
	β_7	0.00	-0.08 (-0.17, -0.03)	0.04	0.00 (-0.51 0.52)	-	
	β_8	0.00	0.00 (0.00, 0.00)	0.09	0.01 (-0.51 0.51)	-	
	σ_{11}	0.90	1.01 (0.58, 1.35)	1.00	-	-	
	σ_{22}	0.70	0.78 (0.33, 1.27)	1.00	-	-	
	σ_{33}	0.10	0.14 (0.03, 0.26)	0.71	-	-	
	σ_{44}	0.00	0.00 (0.00, 0.00)	0.14	-	-	

Table 4.3: Comparing average numbers of actual and wrong zero fixed coefficients of independent variables in simulation study using three different distributions for the errors. The proposed method for binary data (BMQR) is compared with the SSVS in binary QR models for fixed effects (BVSB) and fixed and REs selection in logistic regression models (LFRES). Also, the proposed method for Tobit mixed effects quantile (TMQR) is compared with the SSVS in Tobit QR models for fixed effects (BVST).

p	Error Distribution		Binary QR		Logistic	Tobit QR	
			BMQR	BVSB	LFRES	TMQR	BVST
0.50	$N(0,1)$	correct	4.23	3.12	4.34	4.76	3.45
		wrong	(0.01)	(0.11)	(0.05)	(0.00)	(0.17)
0.75	$N(0,1)$	correct	4.15	3.04	-	4.39	3.62
		wrong	(0.01)	0.09	-	(0.0)	(0.09)
0.95	$N(0,1)$	correct	4.09	2.81	-	4.33	2.41
		wrong	(0.05)	0.12	-	(0.01)	(0.27)
0.50	t_3	correct	4.01	2.87	3.88	4.15	2.39
		wrong	(0.02)	(0.10)	(0.03)	(0.02)	(0.23)
0.75	t_3	correct	3.78	2.61	-	4.17	2.11
		wrong	(0.06)	(0.34)	-	(0.03)	(0.61)
0.95	t_3	correct	3.92	2.33	-	3.88	2.17
		wrong	(0.08)	(0.59)	-	(0.08)	(1.01)
0.50	χ^2_3	correct	3.82	1.98	3.81	3.97	2.21
		wrong	(0.04)	(0.73)	(0.12)	(0.05)	(0.37)
0.75	χ^2_3	correct	3.61	1.63	-	3.55	1.92
		wrong	(0.09)	(1.02)	-	(0.08)	(0.92)
0.95	χ^2_3	correct	3.45	1.35	-	3.61	1.05
		wrong	(0.13)	(1.72)	-	(0.09)	(1.45)

4.7 Analysis of ARMD data

In this section, the suggested methodology was applied to the ARMD data, previously analysed in Chaili (2008). This study had a total of 203 patients which were randomly chosen from the UK to investigate the treatment effects of teletherapy on the loss of vision. The sample consists of 70, 84 and 49 patients from London, Belfast and Southampton, respectively. The sample was divided into 2 groups with one group consisting of 101 patients randomly assigned to a treatment group and the remaining 102 assigned to a control group. Data was collected on the 3rd, 6th, 12th and 24th months from 203 patients and there is potential for heterogeneity across visits.

In this chapter, we fit a linear mixed QR model with a seven independent variables. The independent variables used were: x_1 = the actual time of the visits of each patient, x_2 = age, x_3 = sex, x_4 = centre (city), x_5 = whether or not the patient received teletherapy, x_6 = index eye, x_7 = both or one eye affected by the treatment. We set $\boldsymbol{Z} = \boldsymbol{X}$ and $\boldsymbol{b}_i \sim N_q(0, \boldsymbol{I})$. Similar to Section 4.6, the same priors were used and three choices of p are considered, $p = 0.50, 0.75$ and 0.95.

From Table 4.4, MQR appears quite good compared with SSVSquantreg and FRES. The results indicate that the SSVSquantreg method does not select the top model for $p = 0.50$ and it also selects the intercept only model as the top model for $p = 0.95$. Perhaps, this is due to the truth that SSVSquantreg does not consider the REs entirely and only focus on the fixed effects. On the other hand, it can be observed that MQR performs similar to FRES for $p = 0.50$ but with higher posterior model probability (PMP) for the top subset. It can be concluded from the analysis of the ARMD data that there are situations in which the MQR can perform well while the other criteria can perform poorly.

Table 4.4: Top two models selected for the age-related macular degeneration data.

	MQR		SSVSquantreg		FRES	
p	Model	PMP	Model	PMP	Model	PMP
0.50	Intercept, x_1, x_5, z_1	0.73	False choice	-	Intercept, x_1, x_5, z_1	0.64
	Intercept, x_1, x_5, z_1, z_5	0.08	False choice	-	Intercept, x_1, x_5, z_1, z_5	0.11
0.75	Intercept, x_1, x_2, x_5, z_1, z_5	0.63	x_1, x_5	0.41	-	
	Intercept, x_1, x_5, z_1	0.16	Intercept, x_1, x_5	0.23	-	
0.95	Intercept, x_1, x_3, x_5, x_6, z_1	0.43	Intercept	0.35	-	
	Intercept, x_1,,x_5, x_6, z_1	0.23	Intercept, x_1	0.22	-	

4.8 Chapter summary

We have considered a Bayesian framework for selecting the mixed effects in mixed effects QR models. We have used the idea of different priors for different quantiles to improve the Bayesian sampler. The approach has been extended for subset selection of mixed effects in binary and Tobit mixed effect QR models. Some extensions of the subset selection approach have been discussed, namely, model selection in binary mixed QR and Tobit mixed QR. The simulation studies and the age-related macular degeneration data have demonstrated the superiority of the methods for subset selection over the existing methods.

Chapter 5

Bayesian adaptive Lasso QR

Recently, variable selection and shrinkage of the coefficients of independent variables has attracted great interest in building good prediction models. In this chapter, we propose Bayesian adaptive Lasso quantile regression (BALQR). The method extends the Bayesian Lasso quantile regression (QR) reported in Li et al. (2010) by using different shrinkage weights for different quantile coefficients of independent variables. Inverse gamma priors with unknown hyperparameters are placed on the shrinkage weights, and then similar to Sun et al. (2010), the hyperparameters are considered as unknown quantities and estimated with other parameters. An MCMC-based computation technique with an additional MH update is developed to simulate the parameters of BALQR. Through simulation scenarios and analysis of a prostate cancer dataset, we compare the behavior of the BALQR with seven existing Bayesian and non-Bayesian methods. The simulation scenarios and the prostate cancer data analysis indicate that the BALQR method performs well in comparision to the other seven approaches.

5.1 Introduction

Lasso models (Tibshirani, 1996) are widely used regularisation and shrinkage models for coefficient estimation in regression problems. As we highlighted in Chapter 1, a flexible approach to the Lasso model has been proposed by Zou (2006), which is based on assigning different shrinkage weights for different regression coefficients, rather than a single one, as in the Lasso model. Further, Zou (2006) indicated that the flexibility version of Lasso, which is called adaptive Lasso, has oracle properties (OP) as reported in Fan and Li (2001), which the Lasso method estimators do not achieve. In the Cox model, Zhang and Lu (2007) considered a flexibility version of the Lasso estimator by penalising log partial likelihood and proved that their estimator has the OP. For clustered data, Bondell et al. (2010) suggested a penalised joint log-likelihood with adaptive shrinkage weights for subset selection and coefficient estimation in linear mixed-effects (LME) models and proved that their estimator enjoys the OP. Additionally, Huang et al. (2008) proved that the flexibility version of Lasso has the OP under some conditions in which some of the important and unimportant independent variables are weakly correlated. However, as pointed out by Sun et al. (2010), some of the important and unimportant independent variables are usually highly correlated, which is typical in areas such as chemometrics or bioinformatics.

In this chapter, our motivating example is prostate cancer data reported by Stamey et al. (1989) and analyzed by Yuan and Lin (2005a) and Tibshirani (1996), among others. The data set contains an outcome variable log(prostate specific antigen), which is used as a measure for testing prostate cancer in addition to 8 independent variables. Nowadays, significant effort is made in finding candidate independent variables that relate to prostate cancer. In this data set, certain correlations are present between the independent variables which are an argument to use the adaptive Lasso because the procedure deals with correlated independent variables by assigning adaptive shrinkage weights for the different coefficients of the independent variables. It could be expected that the conditional mean function is

inaccurate in representing the relationship between the independent variables and the level of prostate specific antigen.

In this chapter, we propose Bayesian adaptive Lasso QR (BALQR). In particular, we extend the Bayesian Lasso QR reported in Li et al. (2010) by allowing different shrinkage weights for different regression coefficients of independent variables. Inverse gamma priors with unknown hyperparameters are placed on the shrinkage weights, and then similar to Sun et al. (2010) and Yi and Xu (2008), the hyperparameters are considered as unknown quantities and estimated with other parameters. A Gibbs sampler with an additional MH update is developed to simulate the parameters of BALQR. Using both simulation studies and prostate cancer data we compared the behavior of the BALQR method with six existing Bayesian and non-Bayesian methods, which are already used in Li et al. (2010) to investigate the performance of the Bayesian regularised QR methods compared to other approaches. These methods encompass Bayesian Lasso QR (BLQR) and Bayesian elastic net QR (BQRnet). Also, non-Bayesian methods including the Lasso (Lasso), Lasso QR (RQL), the elastic net (Enet) and the frequentist QR (RQ) are used. Bayesian QR using the g prior approach (BQR$_g$), reported in subsection 2.2.4 of Chapter 2 is also included in the comparison. Both our simulation studies and prostate cancer dataset analysis indicate that BALQR performs well and this method may be preferred over most existing methods that it is compared against.

The rest of this chapter is presented in the following way. In Section 5.2, we review BLQR and propose the adaptive version. A Gibbs sampler with an additional MH update is outlined to simulate the parameters of BALQR in Section 5.3. In Section 5.4, we implement simulation scenarios to test the behavior of the BALQR and in Section 5.5, we illustrate the performance of BALQR via analysis of the prostate cancer data set. A chapter summary follows in Section 5.6.

5.2 Bayesian QR with adaptive Lasso penalty

Lasso QR (Li and Zhu, 2008) is a regularisation and shrinkage technique for quantile coefficients of independent variables. The Lasso QR (Li and Zhu, 2008) estimate is denoted as

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1, \tag{5.1}$$

for some $\lambda \geq 0$, where $\lambda\|\boldsymbol{\beta}\|_1$ is called the l_1 penalty which is used to impose sparsity and improve the efficiency in estimation of the coefficients of independent variables (Tibshirani, 1996). This shrinkage and selection penalty plays the most important role in the Lasso method (see, Tibshirani, 1996; Zou, 2006, among others). As the shrinkage weight λ in (5.1) increases, the l_1-penalised check function estimate is able to perform continuous variable selection and shrinkage for QR coefficients of independent variables towards zero. From a Bayesian framework, Li et al. (2010) employed a Laplace prior distribution $p(\beta_j|\tau,\lambda) = (\tau\lambda/2)\exp\{-\tau\lambda|\beta_j|\}$ on β_j, $\beta_j \in \boldsymbol{\beta}$ and assumed that the residuals ε_i follow the ALD. Specifically, Laplace prior distributions with a single shrinkage weight λ are assigned on the k quantile coefficients of independent variables.

In this chapter, we extend this idea to BALQR by assigning different shrinkage weights on the different coefficients of independent variables. Thus, we suggest a Laplace prior on β_j taking the form

$$p(\beta_j|\tau,\lambda_j) = \frac{1}{2\sqrt{\tau}\lambda_j}\exp\{-\frac{|\beta_j|}{\sqrt{\tau}\lambda_j}\}, \tag{5.2}$$

which can be represented as (Andrews and Mallows, 1974)

$$p(\beta_j|\tau,\lambda_j) = \int_0^\infty \frac{1}{\sqrt{2\pi s_j}}\exp\{-\frac{\beta_j^2}{2s_j}\}\frac{1}{2\tau\lambda_j^2}\exp\{-\frac{s_j}{2\tau\lambda_j^2}\}ds_j. \tag{5.3}$$

This two-level prior can provide flexible shrinkage weights for β_j, $j = 1,\cdots,k$, and represent an alternative model to the Bayesian Lasso model. Equation (5.3) motivates

us to assign an inverse gamma prior on the shrinkage weight $\lambda_j^2, j = 1, \cdots, k$, of the form

$$p(\lambda_j^2|\varsigma, \iota) = \frac{\iota^{\varsigma}}{\Gamma(\varsigma)}(\lambda_j^2)^{-\varsigma-1}\exp\{-\frac{\iota}{\lambda_j^2}\}, \tag{5.4}$$

where ς and ι are positive hyperparameters. These hyperparameters (ς and ι) determine how much shrinkage is needed in the prior and thus play a significant role in estimation of the coefficients of independent variables (Yi and Xu, 2008; Sun et al., 2010). Sun et al. (2010) suggest a joint improper prior on the parameters ι and ς of the form $p(\iota, \varsigma) \propto \iota^{-1}$, which is used in this chapter. The posterior density function of the shrinkage weight $\lambda_j^2,\ j = 1, \cdots, k$, combining the prior 5.4 with 5.3, is inverse gamma distribution (InvGa) with shape $1 + \varsigma$ and scale $s_j/(2\tau) + \iota$. We also assume that the prior of τ takes the form of $p(\tau) = \tau^{-a_{01}-1}\exp\{-a_{02}/\tau\}$, where $a_{01} = a_{01} = 0.1$. The procedure of BALQR is quite different from Bayesian Lasso QR reported in Li et al. (2010), in the sense that each quantile coefficient has a Lasso-type of a positive shrinkage weight which controls the complexity of the model. In other words, we added flexibility due to employment of multiple positive shrinkage weights rather than a single one as in Li et al. (2010).

To summarise, BALQR is a Bayesian hierarchical model given by

$$p(y_i, v_i|\boldsymbol{\beta}, \tau) \propto \frac{1}{\tau\sqrt{\tau v_i}}\exp\{-\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \xi v_i)^2}{4\tau v_i} - \zeta v_i\}, \tag{5.5}$$

$$p(\beta_j, s_j|\tau, \lambda_j^2) = \frac{1}{\sqrt{2\pi s_j}}\exp\{-\frac{\beta_j^2}{2s_j}\}\frac{1}{2\tau\lambda_j^2}\exp\{-\frac{s_j}{2\tau\lambda_j^2}\}, \tag{5.6}$$

$$p(\lambda_j^2|\varsigma, \iota) = \frac{\iota^{\varsigma}}{\Gamma(\varsigma)}(\lambda_j^2)^{-\varsigma-1}\exp\{-\frac{\iota}{\lambda_j^2}\}, \tag{5.7}$$

$$p(\tau) \propto \tau^{-a_{01}-1}\exp\{-\frac{a_{02}}{\tau}\}, \tag{5.8}$$

$$p(\iota, \varsigma) \propto \iota^{-1}. \tag{5.9}$$

Then, the joint posterior distribution for $\boldsymbol{\beta}, \tau, \boldsymbol{v}$, $\boldsymbol{s} = (s_1, ..., s_k)'$ and $\boldsymbol{\lambda}^2 = (\lambda_1^2, ..., \lambda_k^2)'$

is given by

$$
\begin{aligned}
&p(\boldsymbol{\beta}, \tau, \boldsymbol{v}, \boldsymbol{s}, \boldsymbol{\lambda}^2 | \boldsymbol{y}, \boldsymbol{X}) \\
&\quad \propto \prod_{i=1}^{n} \frac{1}{\tau\sqrt{\tau v_i}} \exp\{-\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \xi v_i)^2}{4\tau v_i} - \zeta v_i\} \\
&\quad \times \prod_{j=1}^{k} \frac{1}{\sqrt{s_j}} \exp\{-\frac{\beta_j^2}{2s_j}\} \frac{1}{2\tau\lambda_j^2} \exp\{-\frac{s_j}{2\tau\lambda_j^2}\}, \\
&\quad \times \prod_{j=1}^{k} \frac{\iota^{\varsigma}}{\Gamma(\varsigma)} (\lambda_j^2)^{-\varsigma-1} \exp\{-\frac{\iota}{\lambda_j^2}\}, \\
&\quad \times \tau^{-a_{01}-1} \exp\{-\frac{a_{02}}{\tau}\} \iota^{-1}.
\end{aligned} \tag{5.10}
$$

5.3 Posterior inference

The joint posterior distribution (5.10) allows for improved mixing and creates an efficient MCMC-based computation algorithm that works as follows:

- **Updating $\boldsymbol{v}^{-1}$**

  The full conditional distribution (CD) of each v_i^{-1} for $i = 1, \ldots, n$, is IG(μ_i', λ'), where $\mu_i' = 1/\sqrt{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}$ and $\lambda' = 1/(2\tau)$. Here, IG refereed to the Inverse Gaussian density which is given by (Chhikara and Folks, 1989)

$$
f(x|\lambda', \mu') = \sqrt{\frac{\lambda'}{2\pi}} x^{-3/2} \exp\{\frac{-\lambda'(x-\mu')^2}{2(\mu')^2 x}\}, \quad x > 0. \tag{5.11}
$$

  We use the `rinvGauss()` function in the R package `SuppDists` (Wheeler, 2009) to sample from the inverse Gaussian distribution.

- **Updating β_j**

  The full CD of each β_j for $j = 1, \ldots, k$, is N$(\tilde{\beta}_j, \tilde{\sigma}_j^2)$, where $\tilde{\sigma}_j^2 = (\sum_{i=1}^{n} x_{ij}^2/(2\tau v_i) + s_j^{-1})^{-1}$, and $\tilde{\beta}_j = \tilde{\sigma}_j^2 \sum_{i=1}^{n} x_{ij}(y_i - \sum_{l \neq j} x_{il}\beta_l - \xi v_i)/(2\tau v_i)$

- **Updating s_j^{-1}**

The full CD of each s_j^{-1} for $j = 1, \ldots, k$, is IG (μ_j', λ_j'), where $\mu_j' = \sqrt{1/(\tau\lambda_j^2\beta_j^2)}$ and $\lambda_j' = 1/(\tau\lambda_j^2)$.

- **Updating τ**

  The full CD of τ is InvGa(G_1, G_2), where $G_1 = 3n/2 + k + a_{01}$, and

$$G_2 = \sum_{i=1}^{n}\left(\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \xi v_i)^2}{4v_i} + p(1-p)v_i\right) + \sum_{j=1}^{k}\frac{s_j}{2\lambda_j^2} + a_{02}.$$

- **Updating λ_j^2**

  The full CD of λ_j^2 is InvGa$(1+\varsigma,\ s_j/(2\tau) + \iota)$.

- **Updating ι**

  The full CD of ι is Gamma $(k\varsigma,\ \sum_{j=1}^{k}\lambda_j^{-2})$.

- **Updating ς**

  Because the full CD of ς is $p(\varsigma|\ \lambda_j^2, \iota) \propto (\Gamma(\varsigma))^{-k}\iota^{k\varsigma}\prod_{j=1}^{k}\lambda_j^{-2\varsigma}$, there is no closed form solution for ς. Since $p(\varsigma|\ \lambda_j^2, \iota)$ is log-concave (Sun et al., 2010), the adaptive rejection sampling algorithm Gilks (1992) is used to update ς.

5.4 Simulation studies

We compare the performance of BALQR with six existing methods, including BLQR, BQRnet, Lasso, RQL, Enet and RQ. These six methods have been compared and evaluated in (Li et al., 2010) who showed that the Bayesian methods BLQR and BQRnet often outperform the other methods. Our method BQR$_g$ in subsection 2.2.4 of Chapter 2 is also added to the comparison. The simulation setup is same to the simulation studies 1, 2 and 3 in Li et al. (2010) with different parameter values for the error distributions. In addition, we further test the methodology of the eight methods with two alternative error distributions. Specifically, we simulate 20 training observations, 20 validation observations and 200 testing observations from

$y_i = x_i'\boldsymbol{\beta} + \varepsilon_i$ where the true values for the $\boldsymbol{\beta}$'s are set as follows:

Design I: $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)'$,

Design II: $\boldsymbol{\beta} = (\underbrace{0.85, 0.85, \cdots, 0.85}_{8})'$,

Design III: $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)'$,

Design IV: $\boldsymbol{\beta} = (5, 5, 5, \underbrace{0, 0, \cdots, 0}_{15})'$.

We have set up Design IV as a sparse recovery problem in which $k = 18$, with most coefficients of independent variables being set to zero, except $\beta_j = 5$, where $j = 1, 2, 3$. We fit the linear QR model using the simulated dataset and select the shrinkage weights in the non-Bayesian methods (Lasso, Enet and RQL) via an independent validation set. The rows of X follow a $N_k(0, \boldsymbol{\Sigma}_x)$ with $(\boldsymbol{\Sigma}_x)_{j_1 j_2} = 0.5^{|j_1 - j_2|}$, where the outcome variable is centered around zero and the columns of design matrix X have been standardised. In each simulation study and for each $p \in \{0.50, 0.75, 0.95\}$, the residuals $\varepsilon_i,\ i = 1, \cdots, n$ are simulated from the following six distributions, where the parameter μ in the normal distributions and Laplace distributions is selected so that the pth quantile is zero:

1. $N(\mu, 1)$.
2. Mixture of 2 normal distributions: $0.1N(\mu, 1) + 0.9N(\mu, 9)$.
3. Laplace distribution: Laplace$(\mu, 1)$.
4. Mixture of 2 Laplace distributions: $0.1\text{Laplace}(\mu, 1) + 0.9\text{Laplace}(\mu, 3)$.
5. $t_{(3)}$.
6. $\chi^2_{(3)}$.

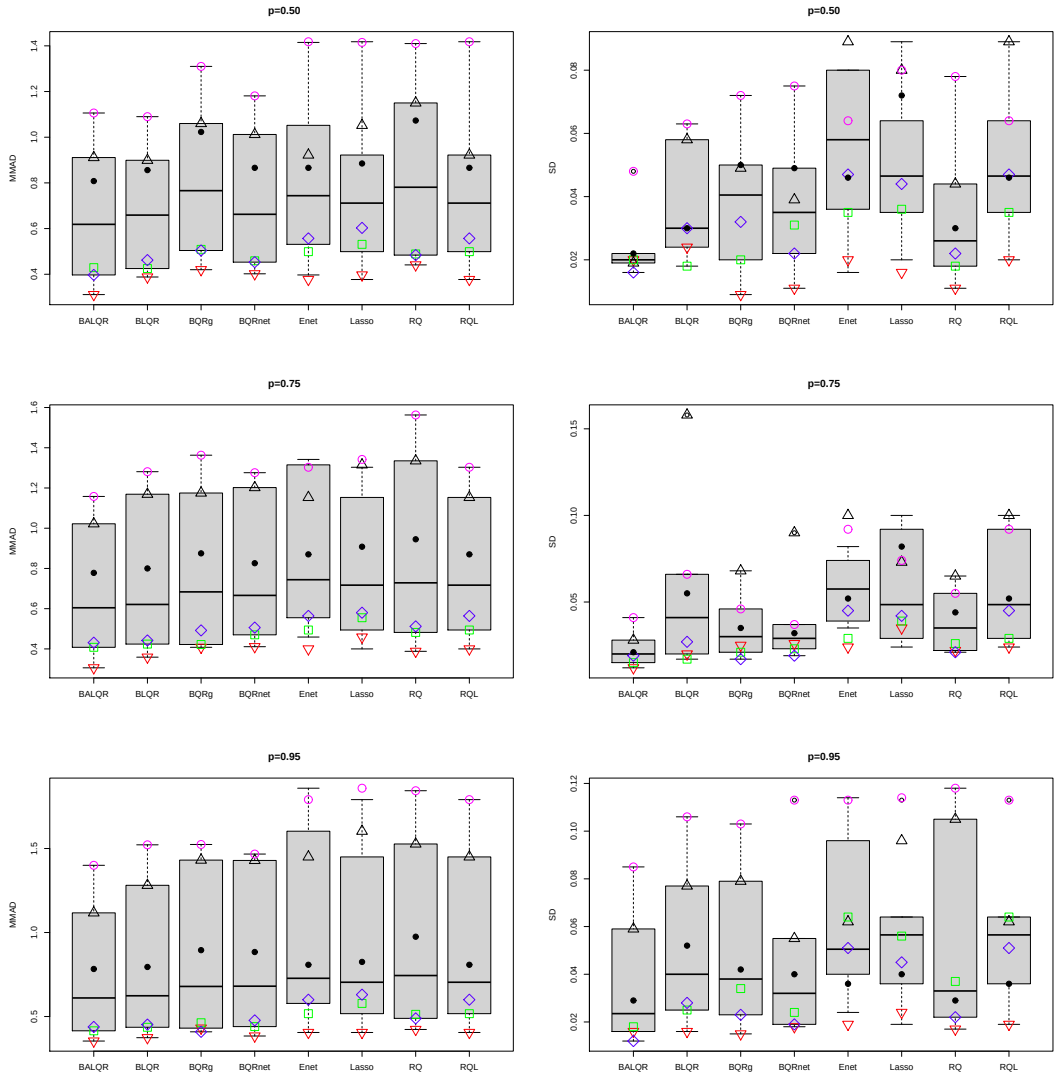

Figure 5.1: Boxplots summarising the MMADs and the standard deviations of MADs (SD) for the eight methods using the six error distributions in Design I. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

For each Design and quantile level $p \in \{0.50, 0.75, 0.95\}$ a total of 150 replications are considered. A number of observations can be considered from Figures 5.1, 5.2, 5.3, and 5.4. For the MMAD and the standard deviation criteria (SD), BALQR generally performs better than the other seven methods for all the distributions under consideration. Most noticeably, when $p = 0.75$ and $p = 0.95$, BALQR was significantly more efficient than the other seven methods. Secondly, from Table 5.1 we see that BALQR performs well when comparing the estimates of

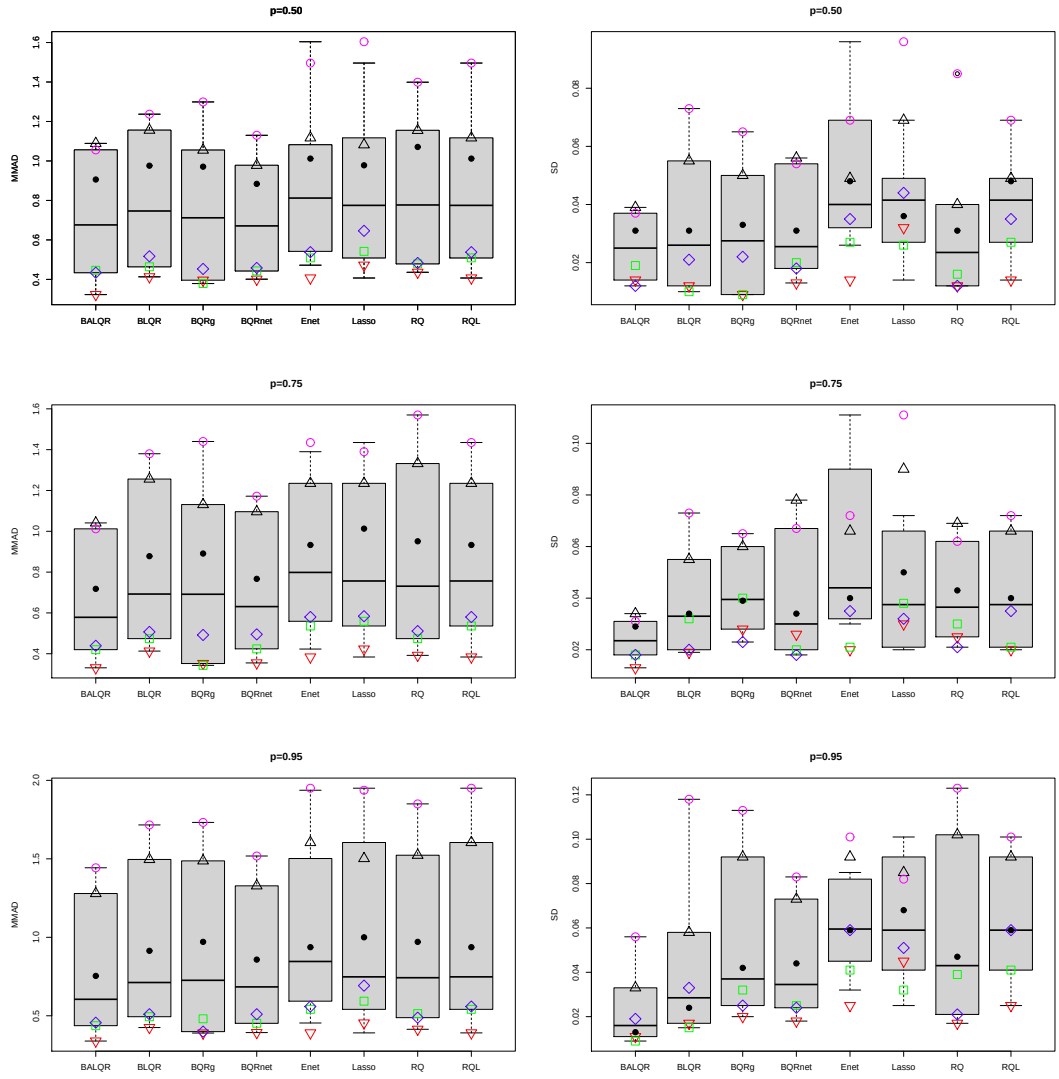

Figure 5.2: Boxplots summarising the MMADs and the standard deviations of MADs (SD) for the eight methods using the six error distributions in Design II. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

$\beta_j, j = 1, \cdots, 8$, with the true values of β_j (β_j^{true}).

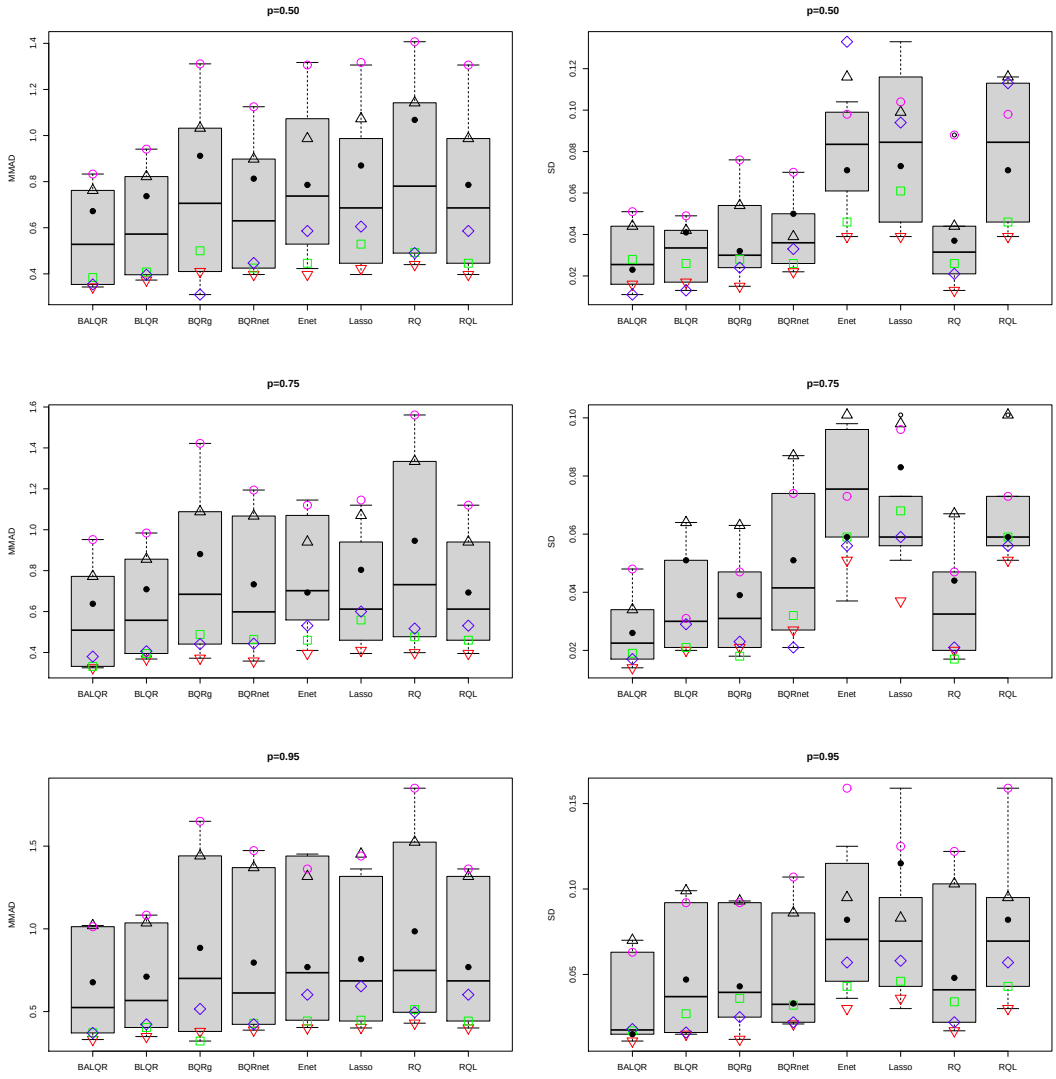

Figure 5.3: Boxplots summarising the MMADs and the standard deviations of MADs (SD) for the eight methods using the six error distributions in Design III. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

5.5 Prostate cancer data (PCD) analysis

This section considers the performance of the BALQR in the PCD reported by Stamey et al. (1989) and analysed by Tibshirani (1996) and Yuan and Lin (2005a), among others. This study had a total of 97 male patients who suffer from prostate cancer and is available in the R package "`bayesQR`" (Benoit et al., 2011). The outcome of interest is the level of prostate antigen (lpsa). The dataset consists of eight independent

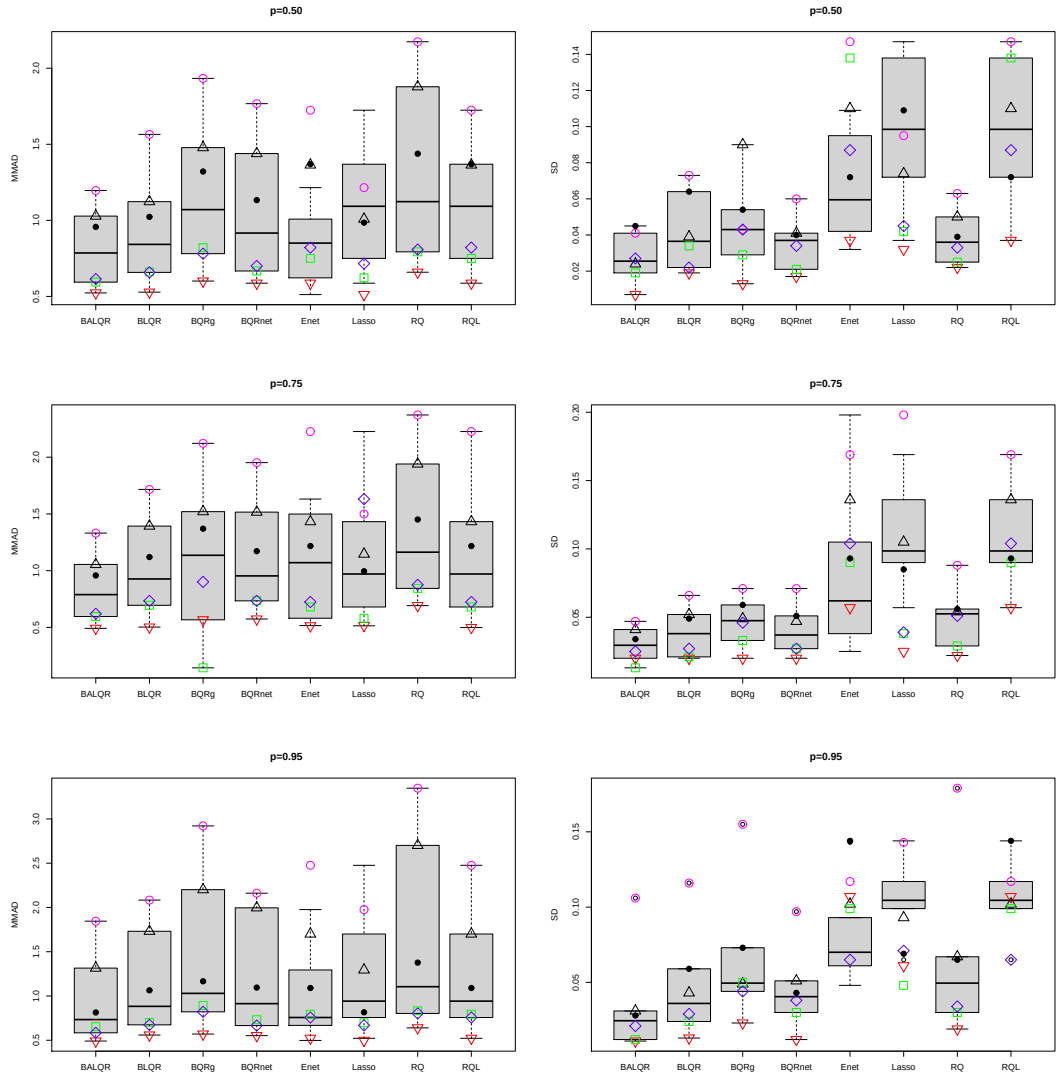

Figure 5.4: Boxplots summarising the MMADs and the standard deviations of MADs (SD) for the eight methods using the six error distributions in Simulation IV. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

variables. These independent variables are the logarithm of cancer amount (x_1), logarithm of the weight of prostate (x_2), age of male patient (x_3), logarithm of the volume of benign enlargement of the prostate (x_4), vesicular glands invasion (x_5), logarithm of Capsular penetration in prostate cancer (x_6), Gleason score in male patient (x_7) and percentage of Gleason scores 4 or 5 (x_8). We estimate a QR model between the response lpsa and the 8 independent variables without an intercept. Here, the outcome variable is centered around zero and the columns of design matrix X

have been standardised. In this section, we set $p \in \{0.50, 0.75, 0.95\}$ and the shrinkage weights in the non-Bayesian regularised methods are tuned by 5-fold cross-validation.

Table 5.2 summarises the results of the 5-fold cross-validation technique for BALQR and the other seven methods. The results in Table 5.2 show that the BQRnet outperforms the other seven methods when $p = 0.50$. However, the performance of the BALQR is very close to the performance of the BQRnet method. Moreover, for the quantiles $p = 0.75$ and $p = 0.95$, the BALQR performs better than the other seven methods. Also, the results show that BLQR has a poor performance when p= 0.50 and $p = 0.95$ due to the high pairwise correlations between some of variables. Thus, the proposed method attempts to remedy the shortcomings of BLQR by using adaptive weights for different quantile coefficients of independent variables.

Table 5.3 summarizes the posterior estimates for the prostate cancer data set using the Bayesian regularised quantile regression methods (BALQR, BLQR and BQRnet) for $p = 0.50$ and $p = 0.75$. We can see that our method gives very similar posterior mean estimates compared to the other Bayesian methods. However, more importantly, it can be observed that the credible intervals for our approach are narrower than the alternative Bayesian methods. Hence, the analysis shows strong support for the use of the proposed method to inference for quantile regression.

5.6 Chapter summary

In this chapter, we proposed Bayesian adaptive Lasso QR (BALQR) for subset selection and quantile regression coefficient estimation. This method extends Bayesian QR with Lasso penalty by allowing different shrinkage weights for different coefficients of independent variables. Independent inverse gamma priors with unknown hyperparameters are assigned on the shrinkage weights of Laplace priors. We developed Bayesian hierarchical model for BALQR as well as a Gibbs sampler with an additional MH update to simulate the parameters of BALQR. The simulation studies and prostate cancer data (PCD) analysis both indicate that the BALQR

behaves quite well and perhaps preferred over current existing Bayesian and non-Bayesian approaches.

Table 5.1: Posterior means for the simulated data in Designs I-III when the error is normal and p=0.95.

Design	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
I	$\boldsymbol{\beta}^{true}$	3.000	1.500	0.000	0.000	2.000	0.000	0.000	0.000
	BALQR	2.988	1.469	0.002	0.013	1.994	0.001	-0.006	0.011
	BLQR	2.937	1.446	-0.024	0.046	1.964	0.054	-0.019	0.037
	BQR_g	2.931	1.461	-0.042	0.031	1.966	0.036	-0.021	0.048
	BQRnet	2.922	1.472	-0.030	0.049	1.957	0.040	-0.033	0.065
	Lasso	2.811	1.357	0.000	0.000	1.782	0.000	0.000	0.000
	Enet	2.796	1.453	0.000	0.000	1.774	0.000	0.000	0.000
	RQ	2.960	1.462	-0.053	0.038	2.000	0.034	-0.049	0.057
	RQL	2.915	1.392	0.000	0.000	1.810	0.000	-0.001	0.000
II	$\boldsymbol{\beta}^{true}$	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850
	BALQR	0.835	0.852	0.849	0.865	0.858	0.860	0.846	0.863
	BLQR	0.811	0.820	0.774	0.865	0.851	0.877	0.790	0.878
	BQR_g	0.812	0.822	0.798	0.836	0.867	0.834	0.813	0.862
	BQRnet	0.800	0.831	0.775	0.872	0.854	0.879	0.783	0.887
	Lasso	0.805	0.836	0.788	0.834	0.849	0.837	0.747	0.873
	Enet	0.710	0.814	0.897	0.889	0.915	0.913	0.808	0.731
	RQ	0.838	0.820	0.786	0.857	0.865	0.853	0.790	0.897
	RQL	0.463	0.753	0.691	0.522	0.765	0.429	0.784	0.555
III	$\boldsymbol{\beta}^{true}$	5.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	BALQR	4.971	0.004	-0.024	0.008	0.012	0.007	-0.010	0.056
	BLQR	4.883	0.010	-0.047	0.033	0.035	0.035	-0.031	0.058
	BQR_g	4.871	0.017	-0.041	0.023	0.028	0.043	-0.033	0.054
	BQRnet	4.869	0.020	-0.055	0.045	0.042	0.048	-0.025	0.062
	Lasso	4.591	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Enet	4.614	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	RQ	4.934	-0.019	-0.053	0.018	0.026	0.054	-0.049	0.067
	RQL	4.936	0.003	-0.020	0.000	0.000	0.000	0.000	0.001

Table 5.2: Cross validation results with standard errors in parentheses for the prostate cancer data analysis.

Method	Test error $p = 0.50$	$p = 0.75$	$p = 0.95$
BALQR	0.26754 (0.05448)	0.26722 (0.04798)	0.26743 (0.04857)
BLQR	0.29061 (0.05952)	0.26979 (0.05803)	0.28289 (0.07249)
BQR_g	0.27353 (0.05222)	0.27618 (0.05232)	0.27534 (0.05137)
BQRnet	0.26416 (0.05214)	0.28537 (0.07039)	0.27455 (0.05701)
Lasso	0.27990 (0.05902)	0.27719 (0.06380)	0.27719 (0.06380)
Enet	0.27938 (0.05897)	0.27876 (0.06002)	0.27876 (0.06002)
RQ	0.27618 (0.05218)	0.27618 (0.05218)	0.27618 (0.05218)
RQL	0.30146 (0.06471)	0.28493 (0.07208)	0.29032 (0.07216)

Table 5.3: Estimates and 95% intervals for the 0.50 and 0.75 QR parameters of the prostate cancer data.

Parameter	BALQR Mean	BALQR (95% CrI)	BLQR Mean	BLQR (95% CrI)	BQRnet Mean	BQRnet (95% CrI)
$p = 0.50$						
β_1	0.53078	(0.38274, 0.66123)	0.52309	(0.37570, 0.68599)	0.51705	(0.37220, 0.67497)
β_2	0.21302	(0.06615, 0.34062)	0.20503	(0.06285 0.35521)	0.21070	(0.07437, 0.35645)
β_3	-0.13750	(-0.25283, -0.01722)	-0.12493	(-0.23960, 0.00256)	-0.13886	(-0.25110, -0.01036)
β_4	0.17268	(0.03419, 0.28174)	0.15847	(0.01500, 0.29417)	0.17282	(0.02901, 0.30591)
β_5	0.26972	(0.10117, 0.39785)	0.24593	(0.07156, 0.40071)	0.26862	(0.09171, 0.41762)
β_6	-0.10503	(-0.28424, 0.01720)	-0.07359	(-0.27000, 0.08943)	-0.09756	(-0.26894, 0.07100)
β_7	0.07309	(-0.07209, 0.19336)	0.06678	(-0.06616, 0.20894)	0.07767	(-0.06654, 0.21280)
β_8	0.09697	(-0.04403, 0.24801)	0.08070	(-0.04848, 0.26103)	0.09526	(-0.04671, 0.26796)
$p = 0.75$						
β_1	0.52711	(0.38137, 0.65933)	0.52205	(0.37392, 0.68092)	0.51980	(0.37012, 0.67187)
β_2	0.21207	(0.06682, 0.33181)	0.20270	(0.05545, 0.35561)	0.21098	(0.08045, 0.35721)
β_3	-0.13779	(-0.25122, -0.01309)	-0.12092	(-0.24227, 0.00577)	-0.14071	(-0.25083, -0.01436)
β_4	0.17069	(0.03506, 0.29005)	0.15440	(0.01508, 0.29312)	0.17244	(0.02852, 0.30042)
β_5	0.26928	(0.09577, 0.41174)	0.24605	(0.06719, 0.39534)	0.26999	(0.08897, 0.41257)
β_6	-0.09946	(-0.28306, 0.06293)	-0.07532	(-0.24954, 0.07862)	-0.09781	(-0.28021, 0.07059)
β_7	0.07525	(-0.07353, 0.21173)	0.06542	(-0.07091, 0.20112)	0.08085	(-0.06850, 0.22461)
β_8	0.09332	(-0.05461, 0.24219)	0.08338	(-0.04382, 0.26263)	0.09457	(-0.05122, 0.26255)

Chapter 6

Conclusions and Future Research

This thesis has proposed several Bayesian hierarchical models for subset selection and coefficient estimation in QR models. Clear advantages over existing methods include a quantile dependent prior, efficient MCMC-based computation techniques and use of data augmentation to allow binary and left-censored outcome variables. The main contributions and future research topics are listed below.

6.1 Main Contributions

Bayesian QR methods for subset selection and coefficient estimation are proposed in chapter 2. These approaches rely on quantile dependent priors for regression coefficients and over model space. For regression coefficients, an extension of the familiar g-prior distribution is suggested to allow a quantile dependent prior. For the model space, novel priors based on percentage bend correlation are used. Our proposed approaches are advantageous in that different quantiles have different priors, which are automatically selected. In particular, the quantile dependent priors and the proposed MCMC algorithm represent a quite useful alternative to existing methods.

In chapter 3, the modified g-prior is generalised by introducing a ridge parameter to address important challenges that may arise in many applications, such as multicollinearity and overfitting problems. An expression for the hyperparameter g to calibrate the modified g-prior with a ridge parameter to a corresponding g-prior is proposed. Possible extensions to the proposed approach are also discussed including the continuous and binary responses in QR. Then, MCMC based computation techniques are proposed based on g-prior to facilitate the computation of the posterior.

In chapter 4, Bayesian subset selection method for fixed and REs in QR mixed effects model is proposed. This approach is related to earlier approaches reported in (Kinney and Dunson, 2007) and (Chen and Dunson, 2003) for linear mixed models. Some possible extensions of the proposed approach are also presented, including binary and left-censored outcome variables. Several advantages of the proposed approach over existing methods are discussed.

In chapter 5, Bayesian adaptive Lasso QR (BALQR) is proposed for subset selection and estimation. The method allows different shrinkage weights for different regression coefficients of independent variables. An MCMC-based computation technique with an additional MH update is developed to simulate the parameters of BALQR.

6.2 Recommendations for Future Research

The work considered in chapter 2 opens the door to new research directions for subset selection and coefficient estimation in QR models by using the modified g-prior. One of these directions has already been studied by Dortet-Bernadet and Fan (2012) who adapts an auxiliary variable approach to fitting QR curves using the modified g-prior. There are many other possible extensions such as using the modified g-prior in Bayesian single index QR or Bayesian nonparametric QR. The idea of Bayesian model selection for fixed and REs reported in Chapter 4 can be extended to Bayesian QR with single index. One can also extend the idea of Bayesian

adaptive Lasso QR in Chapter 5 to other models such as Bayesian adaptive Lasso Tobit QR, Bayesian adaptive Lasso binary QR, Bayesian adaptive Lasso single index QR, and many others.

All of the approaches proposed in this thesis can be extended to the Bayesian QR models with right-censored or interval censored responses.

Bibliography

Agliari, A. and C. C. Parisetti (1988). A-g reference informative prior: a note on Zeliner's g-prior. *The Statistician 37*, 271–275.

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.

Alhamzawi, R. and K. Yu (2012a). Bayesian Lasso mixed quantile regression. *Journal of Statistical Computation and Simulation*, in press.

Alhamzawi, R. and K. Yu (2012b). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics & Data Analysis*, in press.

Alhamzawi, R. and K. Yu (2013). Bayesian tobit quantile regression using *g*-prior distribution with ridge parameter. *Computational Statistics & Data Analysis*, revised manuscript.

Alhamzawi, R., K. Yu, and D. F. Benoit (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling 12*, 279–297.

Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B 36*, 99–102.

Bae, K. and B. Mallick (2004). Gene selection using a two-level hierarchical Bayesian model. *Bioinformatics 20*, 3423–3430.

Baragatti, M. C. and D. Pommeret (2012). A study of variable selection using g-prior distribution with ridge parameter. *Computational Statistics & Data Analysis 56*, 1920–1934.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Benoit, D. F. and D. V. D. Poel (2011). Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. *Journal of Applied Econometrics 27*, 1174–1188.

Benoit, F., R. Alhamzawi, K. Yu, and D. V. P. Poel (2011). `bayesQR`: Bayesian quantile regression. R package version 2.1.

Bilias, Y., S. Chen, and Z. Ying (2000). Simple resampling methods for censored regression quantiles. *Journal of Econometrics 68*, 303–338.

Bondell, H. D., A. Krishna, and S. K. Ghosh (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics 66*, 1069–1077.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics 24*, 2350–2383.

Burgette, L. F. and J. P. Reiter (2012). Modeling adverse birth outcomes via confirmatory factor quantile regression. *Biometrics 68*, 92–100.

Cade, B. S. and N. Barry, R. (2003). A gentle introduction to quantile regression for ecologistis. *Frontiers in Ecology and the Environment 1*, 412–420.

Celeux, G., M. El-Anbari, J. Marin, and C. Robert (2012). Regularization in regression: comparing Bayesian and frequentist methods in a poorly informative situation. *Bayesian Analysis 7*, 477–502.

Chaili, P. (2008). Statistical modelling of age-related macular degeneration. A dissertation submitted to the University of Manchester for the degree of Master of Science in the Faculty of Engineering and Physical Sciences.

Chen, M.-H., J. Ibrahim, and C. Yiannoutsos (1999). Prior elicitation, variable selection, and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society, Series B 61*, 223–242.

Chen, M.-H. and J. G. Ibrahim (2003). Conjugate priors for generalized linear models. *Statistica Sinica 13*, 461–476.

Chen, R.-B., C.-H. Chu, T.-Y. Lai, and Y. N. Wu (2011). Stochastic matching pursuit for Bayesian variable selection. *Statistics and Computing 21*, 247–259.

Chen, Z. and D. Dunson (2003). Random effects selection in linear mixed models. *Biometrics 59*, 762–769.

Chhikara, R. and L. Folks (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, New York.

Clyde, M. and E. I. George (2004). Model uncertainty. *Statistical Science 19*, 81–94.

Cole, T. J. (1988). Fitting smoothed centile curves to reference data (with discussion). *Journal of the Royal Statistical Society A 151*, 385–418.

Dortet-Bernadet, J. and Y. Fan (2012). On Bayesian quantile regression curve fitting via auxiliary variables. Technical report.

Fahrmeir, L., T. Kneib, and S. Konrath (2010). Bayesian regularisation in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing 20*, 203–219.

Fan, J. and R. Z. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*, 881–889.

George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*, 339–373.

Geraci, M. and M. Bottai (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics 8*, 140–154.

Gilks, W. R. (1992). *Derivative-free adaptive rejection sampling for Gibbs sampling*. In Bernardo, J., Berger, J., Dawid, A. P., and Smith, A. F. M., editors, Bayesian Statistics 4, pages 641-649. Oxford University Press.

Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caliguri, C. D. Bloomfield, and E. S. Lander (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science 286*, 531–537.

Griffin, J. E. and P. J. Brown (2010). Bayesian adaptive Lasso with non-convex penalization. Technical report, Institute of Mathematics, Statistics and Actuarial Science, University of Kent.

Gupta, M. and J. G. Ibrahim (2007). Variable selection in regression mixture modeling for the discovery of gene regulatory networks. *Journal of the American Statistical Association 102*, 867–880.

Gupta, M., P. Qu, and J. Ibrahim (2007). A temporal hidden markov regression model for the analysis of gene regulatory networks. *Biostatistics 8*, 805–820.

Härdle, W. K., V. Spokoiny, and W. Wang (2011). Local quantile regression. *Journal of Statistical Planning and Inference*, in press.

Huang, J., S. Ma, and C.-H. Zhang (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica 18*, 1603–1618.

Ibrahim, J. G. and M.-H. Chen (2000). Power prior distributions for regression models. *Statistical Science 15*, 46–60.

Ibrahim, J. G., H. Zhu, R. I. Garcia, and R. Guo (2011). Fixed and random effects selection in mixed effects models. *Biometrics 67*, 495–503.

Ji, Y., N. Lin, and B. Zhang (2012). Model selection in binary and Tobit quantile regression using the Gibbs sampler. *Computational Statistics & Data Analysis 56*, 827–839.

Kass, R. E. and L. Wasserman (1995). A reference Bayesian test for nested hypothesis and its relationship to the Schwarz criterion. *Journal of the American Statistical Association 90*, 928–934.

Khare, K. and J. P. Hobert (2012). Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression. *Journal of Multivariate Analysis 112*, 108–116.

Kinney, S. K. and D. B. Dunson (2007). Fixed and random effects selection in linear and logistic models. *Biometrics 63*, 690–698.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate 91*, 74–89.

Koenker, R. (2005). *Quantile Regression*. Cambridge Books. Cambridge University Press.

Koenker, R. (2011). `quantreg`: Quantile regression. R package version 4.71.

Koenker, R. and G. J. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

Koenker, R. and V. D'Orey (1987). Algorithm AS 229: Computing regression quantiles. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 36*, 383–393.

Koenker, R. and J. A. F. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*, 1296–1310.

Kordas, G. (2006). Smoothed binary regression quantiles. *Journal of Applied Econometrics 21*, 387–407.

Kottas, A. and M. Krnjajić (2009). Bayesian semiparametric modeling in quantile regression. *Scandinavian Journal of Statistics 36*, 297–319.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation 81*, 1565–1578.

Krishna, A., H. D. Bondell, and S. K. Ghosh (2008). Bayesian variable selection using an adaptive powered correlation prior. *Journal of Statistical Planning and Inference 139*, 2665–2674.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics 38*, 963–974.

Lee, D. and T. Neocleous (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 59*, 905–920.

Lee, D. J. (2009). Testing parameter stability in quantile models: An application to the U.S. inflation process. Working papers 2009-26, University of Connecticut, Department of Economics.

Lee, K., N. Sha, E. R. Dougherty, M. Vannucci, and B. K. Mallik (2003). Gene selection: a Bayesian variable selection approach. *Bioinformatics. 19*, 90–97.

Li, Q., R. Xi, and N. Lin (2010). Bayesian regularized quantile regression. *Bayesian Analysis 5*, 533–556.

Li, Y. and J. Zhu (2008). l_1-norm quantile regressions. *Journal of Computational and Graphical Statistics 17*, 163–185.

Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association 103*, 410–423.

Lindmark, A. and M. Karlsson (2012). truncsp: Semi-parametric estimators of truncated regression models. R package version 11.1.

Luethi, D. and W. Breymann (2012). `ghyp`: A package on the generalized hyperbolic distribution and its special cases. R package version 1.5.5.

Lum, K. and A. Gelfand (2012). Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Analysis 7*, 235–258.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics 15*, 661–675.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: asymptotic properties ofthe maximum score estimator. *Journal of Econometrics 27*, 313–333.

Martin, A., K. Quinn, and J. Park (2011). `MCMCpack`: Markov chain Monte Carlo. R package version 1.0-10.

Meligkotsidou, L., I. D. Vrontos, and S. D. Vrontos (2009). Quantile regression analysis of hedge fund strategies. *Journal of Empirical Finance 16*, 264–279.

Park, T. and G. Casella (2008). The Bayesian Lasso. *Journal of the American Statistical Association 103*, 681–686.

Polsen, N. G. and J. G. Scott (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *In Bayesian Statistics 9 (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West). Oxford: Oxford University Press*.

Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics 32*, 143–155.

Reed, C., D. Dunson, and K. Yu (2009). Bayesian variable selection in quantile regression. Technical report, Brunel University, Department of Mathematical Sciences.

Reed, C. and K. Yu (2009). A partially collapsed Gibbs sampler for Bayesian quantile regression. Technical report, Brunel University, Department of Mathematical Sciences.

Reich, B. J., H. D. Bondell, and H. J. Wang (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics 11*, 337–352.

Saville, B. R. and A. H. Herring (2009). Testing random effects in the linear mixed model using approximate Bayes factors. *Biometrics 65*, 369–376.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Shoemaker, L. H. and T. P. Hettmansperger (1982). Robust estimates and tests for the one-and two-sample scale models. *Biometrika 69*, 47–53.

Smith, M. and R. Kohn (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics 75*, 317–343.

Stamey, T., J. Kabalin, J. McNeal, I. Johnstone, F. Freiha, E. Redwine, and N. Yang (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.

Sun, W., J. G. Ibrahim, and F. Zou (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. *Genetics 185*, 349–359.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B 58*, 267–288.

Tibshirani, R., S. Rosset, J. Zhu, and K. Knight (2005). Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society, Series B 67*, 91–108.

Trautmann, H., D. Steyer, and O. Meramann (2010). `truncnorm`: Truncated normal distribution. R package version 1.0-4.

Wang, H. and C. Leng (2007). Unified Lasso estimation by least squares approximation. *Journal of the American Statistical Association 102*, 1039–1048.

Wang, H., G. Li, and G. Jiang (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics 25*, 347–355.

Wang, H. J. and M. Fygenson (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics 37*, 756–781.

Wheeler, B. (2009). `SuppDists`: Supplementary distributions. R package version 1.1-8.

Wilcox, R. R. (1994). The percentage bend correlation coefficient. *Psychometrica 59*, 601–616.

Wu, Y. and Y. Liu (2009). Variable selection in quantile regression. *Statistica Sinica 19*, 801–817.

Yang, A. J. and X. Y. Song (2010). Bayesian variable selection for disease classification using gene expression data. *Bioinformatics 26*, 215–222.

Yi, N. and S. Xu (2008). Bayesian Lasso for quantitative trait loci mapping. *Genetics 179*, 1045–1055.

Yu, K. and M. C. Jones (1998). Local linear quantile regression. *Journal of the American Statistical Association 93*, 228–237.

Yu, K., Z. Lu, and J. Stander (2003). Quantile regression: Applications and current research areas. *The Statistician 52*, 331–350.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*, 437–447.

Yu, K. and J. Stander (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics 137*, 260–276.

Yu, K. and J. Zhang (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics - Theory and Methods 34*, 1867–1879.

Yuan, M. and Y. Lin (2005a). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association 100*, 1215–1225.

Yuan, M. and Y. Lin (2005b). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B 68*, 49–67.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika 94*, 19–35.

Yuan, Y. and G. Yin (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics 66*, 105–114.

Yue, Y. R. and H. Rue (2011). Bayesian inference for additive mixed quantile regression models. *Computational Statistics & Data Analysis 55*, 84–96.

Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *The Statistician 32*, 23–34.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Bayesian inference and decision techniques. *Stud. Bayesian Econometrics Statist. 6*, 233–243.

Zellner, A. and A. Siow (1980). Posterior odds ratios for selected regression hypotheses. *In Bayesian Statistics. Proceedings of the First Valencia International Meeting Held in Valencia (Spain) (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 585-603. Valencia Univ. Press*.

Zhang, H. H. and W. Lu (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika 94*, 691–703.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B 67*, 301–320.

Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics 36*, 1108–1126.