Bayesian quantile regression for semiparametric models

Hu, Yuao

2013

Hu, Y. (2013). Bayesian quantile regression for semiparametric models. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/54830>

<https://doi.org/10.32657/10356/54830>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**BAYESIAN QUANTILE
REGRESSION FOR
SEMIPARAMETRIC MODELS**

HU YUAO

**Division of Mathematical Sciences
School of Physical and Mathematical Sciences**

2013

BAYESIAN QUANTILE REGRESSION FOR SEMIPARAMETRIC MODELS

HU YUAO

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

A thesis submitted to Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy in Mathematical Sciences

March 2013

Abstract

Quantile regression (Koenker and Bassett, 1978) has recently received a great deal of attention in both theoretical and empirical research. It can uncover different structural relationships between covariates and responses at the upper or lower tail, which is sometimes of significant interest in econometrics, educational and medicine applications. The methodologies of quantile regression for linear models have been well developed in both frequentist and Bayesian contexts. However, there has been relatively less work focusing on quantile regression for nonparametric models or semiparametric models, especially from a Bayesian perspective.

The principal goal of this work is to propose efficient approaches to implement Bayesian quantile regression with two kinds of semiparametric models, single-index models and partially linear additive models, using an asymmetric Laplace distribution which provides a mechanism for Bayesian inference of quantile regression. With carefully selected priors, we build hierarchical Bayesian models and design effective Markov chain Monte Carlo algorithms for posterior inference. We compare the proposed methods with some existing methods through simulation studies and real data applications.

Acknowledgements

First of all, I would like to express my heartfelt gratitude to my supervisor, Prof. Lian Heng, for his thoughtful guidance and warm encouragement during my Ph.D. study at Nanyang Technological University. His diligence, patience, and enthusiasm on research have been steadily influencing me, not to mention his knowledge in statistics. He has been a tremendous mentor for me. I hope that I can be a person like him some day.

I would like to thank the faculty of the Division of Mathematical Sciences at Nanyang Technological University, for providing all kinds of assistance during my Ph.D. journey. I would always remember my days here, with my teachers and friends.

On a personal note, I would like to thank my parents and husband. Thank you for loving me.

List of Works

Below is the list of work done, in chronological order, during my Ph.D. studies in NTU.

1. Hu, Y., Gramacy, R., and Lian, H., (2013). “Bayesian Quantile Regression for Single-Index Models”, *Statistics and Computing to appear*.
2. Hu, Y. and Lian, H., (2013). “Bayesian Additive Machine: Classification with a Semiparametric and Parsimonious Discriminant Function”, Submitted.
3. Hu, Y. and Lian, H., (2013). “Bayesian Quantile Regression for Partially Linear Additive Models”, Submitted.

Contents

Abstract	iii
Acknowledgements	iv
List of Works	v
1 Introduction	1
1.1 Two Semiparametric Models	2
1.1.1 Single-Index Models	2
1.1.2 Partially Linear Additive Models	4
1.2 Quantile Regression	7
1.2.1 The Frequentist Approaches	8
1.2.2 The Bayesian Approaches	9
1.3 Thesis Outline	12
2 Bayesian Quantile Regression for Single-Index Models	15
2.1 Hierarchical Bayesian Modelling	16
2.2 Numerical Illustrations	20
2.2.1 Simulations	21

2.2.2	Real Data Analysis	43
2.3	Summary of the Chapter	44
3	Bayesian Quantile Regression for Partially Linear Additive Models	49
3.1	Hierarchical Bayesian Modeling	50
3.2	Numerical Illustrations	55
3.2.1	Simulations	57
3.2.2	Real Data Analysis	68
3.2.2.1	GDP Growth Data	68
3.2.2.2	Housing Price Data	77
3.3	Summary of the Chapter	79
4	Discussion and Future Research	83
	Appendix	87
	Bibliography	95

List of Figures

2.1	The diagram of Bayesian hierarchical structure.	19
2.2	Trace plots of $\beta_1, \beta_2, \beta_3, \sigma, \gamma$ and λ at quantile 0.5 for simulation example 1, when $n = 100$. Two chains with different starting values are illustrated.	22
2.3	Summarizing estimators of β for $n = 100, 200$ in example 1. ‘B100’ (‘Q100’) denotes BQSIM (QSIM) with $n = 100$, for example.	27
2.4	Summarizing estimators of β for $n = 100, 200$ in example 1. ‘B100’ (‘Q100’) denotes BQSIM (QSIM) with $n = 100$, for example.	28
2.5	The autocorrelation plots for $\beta_1, \beta_2, \beta_3$. The left column represents the series produced by our partially collapsed sampler and the right column represents the uncollapsed sampler.	30

2.6	The estimated link functions. On the left panel, the values of estimated $\boldsymbol{\eta}_n$ are plotted against the true index, while for the right panel the values of estimated $\boldsymbol{\eta}_n$ are plotted against the estimated index. The true link function is also shown in the both figures (solid line).	31
2.7	Summarizing estimators of β for $n = 100, 200$ in example 2.	32
2.8	Summarizing estimators of β for $n = 100, 200$ in example 2.	33
2.9	Summarizing estimators of β for $n = 100, 200$ in example 3.	35
2.10	Summarizing estimators of β for $n = 100, 200$ in example 3.	36
2.11	The estimated link functions by BQSIM (left column) and QSIM (right column) for TC intensity at different quantiles with 95% credible/confidence intervals.	46
2.12	The boxplots on the left shows the posterior distribution of $\boldsymbol{\beta}$ (after normalization to unit norm). The histograms on the right shows the posterior distribution of $d = 1/\ \boldsymbol{\beta}\ ^2$	47
2.13	The boxplots on the left shows the posterior distribution of $\boldsymbol{\beta}$ (after normalization to unit norm). The histograms on the right shows the posterior distribution of $d = 1/\ \boldsymbol{\beta}\ ^2$	48
3.1	The diagram of Bayesian hierarchical structure.	54
3.2	Component selection results for one randomly selected replicate, when the distribution of errors is Gaussian and $p = 10$. From the left to the right, 10 bars in each subfigure represent function components 1 to 10. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively. . .	66

- 3.3 Component selection results for one randomly selected replicate, when the distribution of errors is Student's t and $p = 10$. From the left to the right, 10 bars in each subfigure represent function components 1 to 10. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively. 67
- 3.4 Component selection results for one randomly selected replicate with $p = 50$. The upper panels are results when error distribution is Gaussian, and the lower panels are results when the error distribution is Student's t . From the left to the right, 50 bars in each subfigure represent function components 1 to 50. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively. 69
- 3.5 Component selection results for one randomly selected replicate with $p = 100$. The upper panels are results when error distribution is Gaussian, and the lower panels are results when the error distribution is Student's t . From the left to the right, 100 bars in each subfigure represent function components 1 to 100. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively. 70
- 3.6 Component selection results for the GDP growth data. . . . 73

- 3.7 The fitted regression functions for GDP growth data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed) The red solid lines are the fitted curves of mean regression. 74
- 3.8 The fitted regression functions for GDP growth data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed) The red solid lines are the fitted curves of mean regression. 75
- 3.9 Trace plots of selected α , δ_0 , and μ of median regression for GDP growth data. 76
- 3.10 Component selection results for the housing price data. . . 79
- 3.11 The fitted regression functions for housing price data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed). The red solid lines are the fitted curves of mean regression. 80
- 3.12 The fitted regression functions for housing price data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed). The red solid lines are the fitted curves of mean regression. 81

List of Tables

2.1	Results of BQSIM for simulation example 1, $n = 100$	25
2.2	Results of BQSIM for simulation example 1, $n = 200$	26
2.3	Comparison of MSE for BQSIM (using posterior mean as the estimator) and QSIM based on 100 replications in each case for simulation example 1.	29
2.4	Comparison of MSE for BQSIM (using posterior mean as the estimator) and QSIM based on 100 replications in each case for simulation example 2.	34
2.5	Comparison of MSE for BQSIM (using posterior mean as the estimator) and QSIM based on 100 replications in each case for simulation example 3.	37
2.6	Results of BQSIM for simulation example 4, $n = 100$	39
2.7	Results of BQSIM for simulation example 4, $n = 200$	40
2.8	Comparison of MSE for BQSIM and QSIM based on 100 replications in each case for simulation example 5, where $\beta = (1, 1, 1, 0, \dots, 0)^T$	41

2.9	Comparison of MSE for BQSIM and QSIM based on 100 replications in each case for simulation example 5, where $\beta = \frac{1}{\sqrt{10}}(1, 1, 1, 1, \dots, 1)^T$	42
2.10	Estimates from BQSIM and QSIM for the TCs data.	45
3.1	Summary of $\sqrt{\widehat{ISE}}$ for each univariate function and $f = \sum_{j=1}^{10} f_j$ over 100 replicates, when the distribution of error is normal and $p = 10$. Standard errors based on simulations are shown as subscripts.	59
3.2	Summary of $\sqrt{\widehat{ISE}}$ for each univariate function and $f = \sum_{j=1}^{10} f_j$ over 100 replicates, when the distribution of error is Student's t and $p = 10$	60
3.3	Summary of testing errors for mean estimators and median estimators over 100 replicates, when $p = 10$. "Normal" and "Student's t" indicate the distribution of ϵ_i	63
3.4	Summary of testing errors for quantile estimators at different quantile levels over 100 replicates, when $p = 10$. "Normal" and "Student's t" indicate the distribution of ϵ_i	63
3.5	Summary of the component selection results for mean estimators and median estimators over 100 replicates, when $p = 10$. "Normal" and "Student's t" indicate the distribution of ϵ_i	64
3.6	Summary of the component selection results for quantile estimators at different quantile levels over 100 replicates, when the distribution of errors is Gaussian and $p = 10$	65
3.7	Summary of the component selection results for quantile estimators at different quantile levels over 100 replicates, when the distribution of errors is Student's t and $p = 10$	65

- 3.8 Summary of testing errors for mean and median estimators over 100 replicates, when $p = 50$ and 100. “Normal” and “Student’s t” indicate the distribution of ϵ_i 68
- 3.9 Summary of the component selection results for mean and median estimators over 100 replicates, when $p = 50$ and 100. “Normal” and “Student’s t” indicate the distribution of ϵ_i . . 71

Chapter 1

Introduction

The core inference procedure of Bayesian analysis is to update prior assumptions with data observations and to obtain posterior so that future observations can be predicted. In parametric Bayesian analysis, prior assumptions are viewed as probability distributions assigned to the parameters of the assumed model. For nonparametric Bayesian analysis, parametric assumptions are not kept so that more flexible description of data structure can be obtained. There is a spectrum of Bayesian regression models between stringent parametric models and full nonparametric ones. In this thesis, we focus on Bayesian inference for two kinds of semiparametric models in conditional quantile regression.

In this chapter, we first give an introduction to the two kinds of semiparametric models: single-index models and partially linear additive models in Section 1.1. Then we give a review to quantile regression in Section 1.2. We conclude this introduction with an outline of this thesis in Section 1.3.

1.1 Two Semiparametric Models

For data sets with high dimensional predictors, one unavoidable issue is the “curse of dimensionality”, which refers to the poor performance in convergence and other phenomena that will arise while analyzing and organizing data in high dimensional spaces. Two essential approaches can be used to circumvent the difficulty brought by this issue in the words of Xia et al. (2002): dimension reduction and function approximation. A favorite regression-type model for dimension reduction is single-index model. An attractive model used in function approximation is additive model. Here in this section, we briefly introduce these two models.

1.1.1 Single-Index Models

Single-index models (SIMs) provide an efficient way of coping with high-dimensional nonparametric estimation problems (Härdle et al., 1993; Yu and Ruppert, 2002) and avoid the “curse of dimensionality” (Bellman et al., 1966) in nonparametric problems by assuming that the response is only related to a single linear combination of the covariates. It is one of the most popular and important semiparametric models in statistics as well as applied quantitative sciences such as econometrics and psychometrics due to its dimension reduction capability (Ichimura, 1993). For example, Härdle et al. (1993) used SIMs to study the dependence of the severity of side impacts on the velocity and acceleration of the automobile in an accident, and Xia et al. (2004) demonstrated that SIMs provide a good fit in a study trying to identify causal factors associated with the prevalence and incidence of depression.

In the traditional mean regression, a single-index model is represented

by

$$E(y|\mathbf{x}) = \eta(\mathbf{x}^T \boldsymbol{\beta}), \quad (1.1.1)$$

where y is the response, $\mathbf{x} = (x_1, \dots, x_p)^T$ is the p -dimensional predictor vector, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the index vector, η is a univariate function, which is called “link function”, and $E(y|\mathbf{x})$ is the expectation of y conditioned on $\mathbf{x}$. Thus, the basic estimation of interest for SIMs includes two parts, the estimation of the coefficient vector $\boldsymbol{\beta}$ and the link function η .

From a frequentist perspective, numerous methods have been developed and discussed about SIMs, such as M -estimation approaches including semiparametric maximum likelihood estimator (Klein and Spady, 1993), semiparametric least squares estimator (Ichimura, 1993), and maximum rank correlation estimator (Han, 1987; Sherman, 1993). For M -estimation approaches, the unknown link function is considered as an infinite-dimensional nuisance parameter, while $\boldsymbol{\beta}$ is estimated by minimization of an M -functional with respect to $\boldsymbol{\beta}$ when η being replaced by its nonparametric estimator. M -estimators have favorable theoretical properties. However, to compute these estimators, one needs to solve optimization problems in a high-dimensional space which have been proved to be difficult. Alternative approaches include average derivative methods (ADE) (Stoker, 1986; Horowitz and Härdle, 1996; Hristache et al., 2001) and sliced inverse regression (SIR) methods (Li, 1991; Duan and Li, 1991). Using these methods, hard optimization problems can be avoided, but of course, both ADE and SIR have limitations. For example, ADE cannot be used when the link function is an even function and $\mathbf{x}$ is symmetrically distributed (Xia et al., 2002). SIR requires the distribution of covariates to be elliptically symmetric. Recently, Xia et al. (2002) proposed a new method named the (conditional) minimum average variance estimation (MAVE) which is inspired by the idea of ADE

methods, SIR methods and local linear smoothers (Fan and Gijbels, 1996). MAVE has been demonstrated to have favorable theoretical properties and good performance. Although frequentist methodologies about SIMs have been well developed, efficient and stable estimation of SIMs still remains challenging and has been inspiring many recent works in this area (Wang and Yang, 2009; Liang et al., 2010; Wang et al., 2010).

In comparison with the existing volume of frequentists' work, the Bayesian literature on SIMs regression is relatively limited. Antoniadis et al. (2004) and Wang (2009) proposed a Bayesian approach using polynomial splines to model the nonparametric link function, while Choi et al. (2011) and Gramacy and Lian (2012) used a Gaussian process (GP) prior. As noted in Gramacy and Lian (2012), one advantage of using GPs as the prior for the link function is that the index vector does not have to be normalized to have unit norm, which makes the choice of prior easier, and subsequently simplifies the sampling algorithm too.

Of course, single-index models are not restricted to mean regression. They can be combined with other regression models, such as binary response models and quantile regression models, see Klein and Spady (1993) and Wu et al. (2010). In this thesis, we will propose a Bayesian quantile regression approach for single-index models in Chapter 2.

1.1.2 Partially Linear Additive Models

Additive models are a kind of attractive (Hastie and Tibshirani, 1986) nonparametric regression models, which have drawn considerable attention over decades. They allow analysis of multidimensional problems which are difficult to fit with “full-dimensional” nonparametric methods, since they reduce multidimensional smoothing to a sequence of one-dimensional smoothing

(Opsomer and Ruppert, 1999). Besides, they provide enough flexibility by incorporating nonlinear relationships without making any specific assumptions about the functional form.

In mean regression, an additive model can be represented as

$$E(y|\mathbf{x}) = \mu + f_1(\mathbf{x}_1) + \cdots + f_p(\mathbf{x}_p), \quad (1.1.2)$$

where y is the response, $\mathbf{x} = (x_1, \dots, x_p)^T$ is the p -dimensional predictor vector, $f_j(\cdot), j = 1, \dots, p$, are unknown functions, μ is the intercept term, and $E(y|\mathbf{x})$ is the expectation of y conditioned on $\mathbf{x}$. One estimation of interest for the additive model is the component function estimation, where plenty of literature can be found from the frequentist point of view. The main methods include spline estimators (Stone, 1994; Huang et al., 2010), backfitting estimators (Mammen et al., 1999; Opsomer, 2000), and marginal integration estimators (Tjøstheim and Auestad, 1994; Linton and Nielsen, 1995; Fan et al., 1998). The oracle properties of these estimators have also been well developed.

Component selection is another interesting issue for additive models. Practically speaking, it is reasonable to believe that some components might be unrelated to responses, especially when the dimension of covariates is large. Statistically speaking, component selection reduces the complexity of the model and improves accuracy. Besides traditional criteria such as the Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978), penalization methods used for variable selection of linear regression models, such as lasso (Tibshirani, 1996) and SCAD (Fan and Li, 2001), are generalized to solve variable selection problems for additive models and they have become popular in recent years. For example, Zhang et al. (2004) used generalized lasso combined with

smoothing spline ANOVA for component selection. Lin and Zhang (2006) investigated the component selection and smoothing operator (COSSO). Xue (2009) proposed a nonparametric extension of the SCAD to additive model, Ravikumar et al. (2009) developed a L_2 penalized approach and Huang et al. (2010) proposed to use the adaptive group lasso for variable selection of additive models.

Besides, a number of existing papers are concerned with the estimation and selection of function components using Bayesian inference recently. In a Bayesian context, component functions are usually expressed as a linear combination of basis terms. One branch of literature focused on Bayesian model averaging method (Smith and Kohn, 1996; Luo and Wahba, 1997; Denison et al., 2002), for which, the basis terms can vary and the component functions are estimated by the average of varying dimension representations of the functions. A second branch of the literature (Wahba, 1990; Wong and Kohn, 1996; Lang and Brezger, 2004) focused on using fixed basis terms with informative priors on the coefficients of the basis functions. With priors corresponding to penalties in a frequentist context or the so-called “spike-and-slab” priors, this branch of approaches can be applied to select components for additive models with large number of predictors (Panagiotelis and Smith, 2008; Shively et al., 1999; Kneib et al., 2010).

An interesting special case of the additive model is the partially linear additive model (PLAM). PLAMs, consisting of both linear and nonlinear additive components, are more flexible than stringent linear models, and are more efficient than full nonparametric additive regression models. PLAMs have been widely applied in practice. For example, Liang et al. (2008) applied PLAMs to study the relationship between environmental chemical exposures and semen quality, while Panagiotelis and Smith (2008) applied

the models for intra-day electricity load analysis.

For a PLAM, one additional concern besides the component function estimation and significant component selection is the identification of linear components. Usually, pre-specification of linear components is required. Recently, several works considered to perform variable selection, parametric component identification, and parameter estimation all at the same time Zhang et al. (2011); Lian et al. (2012) from a frequentist perspective, for mean regression. We will propose a Bayesian quantile regression approach for partially linear additive models with the ability of both significant component selection and linear component identification in Chapter 3.

1.2 Quantile Regression

Quantile regression has drawn great attention both from theoretical and empirical point of view, since it is introduced by Koenker and Bassett (1978). It is desired when analysis of various quantiles of a conditional distribution is of interest. The traditional mean regression can be viewed as an average regression over all possible quantile levels. Hence, a simple analysis with mean regression can be not enough to capture the comprehensive relationship between responses and predictors. Besides, It is widely accepted that mean regression is not a robust estimate when the distribution of random errors is heavy-tailed, or when there are outliers in samples. As a useful supplement to mean regression, quantile regression produces a more complete description of the conditional response distribution and provides more robust analysis of data.

Quantile regression has been widely used in social sciences, economics, medicine, public health, financial return, environment and engineering. For

example, Buchinsky (1994) investigated the US wage structure from 1963 to 1987 with quantile regression. Korostelev et al. (1995) explored the image reconstruction and relevant problems where conditional quantiles near maximum or minimum are of interest. Abrevaya (2001) estimated the effects of demographics and maternal behavior during pregnancy at various quantiles of the birth weight distribution, and Cade and Noon (2003) investigated the interactions between different factors affecting organisms with data exhibiting heteroscedasticity at different quantile levels.

1.2.1 The Frequentist Approaches

The methodologies for linear quantile regression have been well developed. The main stream of approaches focuses on minimizing the check loss. Koenker and Bassett (1978) gave a detailed solution of the minimization problem for linear quantile regression by converting it into a linear program. Alternatively, Koenker (1994) suggested a method based on the theory of rank tests, which shows some more robust theoretical properties (Chen and Wei, 2005) than the first kind of approaches. However, the computation burden of the tests-based methods is heavy since the computing complexity is exponential in both the sample size and the dimension of predictors. For medium to large sized data sets, a better choice is bootstrap, see Parzen et al. (1994), He and Hu (2002) and Kocherginsky et al. (2005). To improve the prediction accuracy and select variables, regularization technique has been used in linear quantile regression. Koenker (2004) first used lasso penalty on random effects of a mixed-effect linear quantile regression model to shrink random effects towards zero. Li and Zhu (2008) and Belloni and Chernozhukov (2011) investigated an L_1 -norm (lasso) regularized regression approach to select variable in linear quantile regression. Wang et al. (2007)

and Bang and Jhun (2012) explored quantile regression estimates with adaptive lasso penalty. Wu and Liu (2009) developed SCAD and adaptive lasso regularized quantile regression approaches, and demonstrated their oracle properties.

Literature on nonparametric quantile regression is relative sparse in contrast to that on linear quantile regression. Yu and Lu (2004) developed nonparametric regression quantile regression estimation by kernel weighted local linear fitting. Koenker et al. (1994) and Koenker (2005) considered spline approaches, which is an alternative method for nonparametric quantile regression. Wu et al. (2010) investigated a single-index quantile regression model using kernel regression. De Gooijer and Zerom (2003), Horowitz and Lee (2005), and Yu and Lu (2004) all developed methodologies of nonparametric estimation for additive quantile regression.

1.2.2 The Bayesian Approaches

With the help of asymptotic theories, frequentist inference for quantile regression relies no more on a specific form of the residual distribution. However, for Bayesians, missing specification of the residual distributions could result in a problem that any estimation of unknown parameters is not possible because of the lack of specified likelihood. Yu and Moyeed (2001) suggested to use the asymmetric Laplace (AL) distribution with density

$$\pi(\epsilon|\tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\frac{1}{\sigma} \rho_{\tau}(\epsilon) \right\}, \quad (1.2.1)$$

to describe the distribution of the residuals ϵ , where $\rho_{\tau}(u) = u(\tau - I(u \leq 0))$ is the so-called *check function*, σ is the scale parameter, $\tau \in (0, 1)$ is the quantile level as well as the skewness of the AL distribution. With the assumption that $\epsilon_i, i = 1, \dots, n$ (n is the sample size) are i.i.d. from the AL

distribution (1.2.1), a “pseudo” likelihood can be employed in the form of

$$L(\mathbf{y}|\mathbf{x}, \sigma) = \frac{\tau^n(1-\tau)^n}{\sigma^n} \exp \left\{ - \sum_{i=1}^n \rho_\tau(y_i - f(\mathbf{x}_i)) \right\},$$

where $\mathbf{y} = (y_1, \dots, y_n)^T$ and $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ with $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$. $f(\cdot)$ varies with different assumptions of models. For example, for linear regression model,

$$f(\mathbf{x}_i) = \mu + \mathbf{x}_i^T \boldsymbol{\beta},$$

where μ is the intercept and $\boldsymbol{\beta}$ is the regression coefficient; for single-index regression model based on the model assumption (1.1.1),

$$f(\mathbf{x}_i) = \eta(\mathbf{x}_i^T \boldsymbol{\beta}),$$

and for additive model based on the model assumption (1.1.2)

$$f(\mathbf{x}_i) = \mu + \sum_{j=1}^p f(x_{ij}).$$

This likelihood is used to link Bayesian inference to frequentist inference instead of indicating the true data generating mechanism. Besides, The AL distribution is a member of tick exponential family, which leads to a favorable theoretical property that the posterior estimators are able to achieve posterior consistency for likelihood-based inference (Komunjer, 2005). The AL distribution has been widely used on Bayesian quantile regression, ever since Yu and Moyeed (2001) built a linear regression model based on it. For example, Geraci and Bottai (2007) focused on the Bayesian inference for longitudinal data with linear models based on AL likelihood. Yue and Rue (2011) proposed a Bayesian quantile regression approach for additive mixed models with AL distributed random errors. Chen and Yu (2009) and Thompson et al. (2010) used the AL likelihood to implement automatic Bayesian quantile curve fitting approaches on nonparametric regression models.

Kozumi and Kobayashi (2011) proved that AL distribution can be viewed as a mixture of an exponential and a scaled normal distribution called scale-mixture expression of AL. To be specific, the random variable

$$\epsilon = k_1 e + \sqrt{k_2 \sigma} e z \quad (1.2.2)$$

follows AL distribution (1.2.1) when $e \sim \exp(1/\sigma)$ is an exponential random variable with mean σ , z is a standard normal random variable which is independent of e , $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $k_2 = \frac{2}{\tau(1-\tau)}$. Reed and Yu (2009) considered a similar scaled-mixture expression as in Kozumi and Kobayashi (2011). Kottas and Krnjajić (2009) developed several error distributions based on Dirichlet process mixture models, including a scale mixture of AL densities.

Besides the AL likelihood, there are several other choices for Bayesian inference of quantile regression. For example, Dunson and Taylor (2005) used Jeffreys' (Jeffreys, 1961) substitution likelihood while Lancaster and Jun (2010) considered a Bayesian exponentially tilted empirical likelihood (Schemm, 2005).

In Bayesian quantile regression, variable selection methods are similar as those used in Bayesian mean regression. The priors of parameters might be used to impose regularization for likelihood-based inference. For example, in a linear regression model, posterior mode would correspond to the L_1 -norm quantile regression estimates (Li and Zhu, 2008) when the prior on coefficients β is set to be independent double exponential distributions. Li and Lin (2010) extended this idea to implement regularized Bayesian quantile regression based on penalties including lasso, group lasso and elastic net. Another technique of Bayesian variable selection used in mean regression is the so-called "slab-and-spike" technique. With the help

of indicator variables, Meligkotsidou et al. (2009), Reed et al. (2009) and Ji et al. (2011) generalized this methodology to Bayesian quantile regression. Alternatively, reversible jump Markov chain Monte Carlo (RJMCMC) of Green (1995) has also been used in quantile regression by Chen and Yu (2009) to select variables.

1.3 Thesis Outline

This thesis is organized into three chapters as follows.

In Chapter 2, we develop a fully Bayesian approach to fit single-index models in conditional quantile regression using an AL distribution. We use a Gaussian process prior for the unknown nonparametric link function and a Laplace distribution on the index vector, with the latter motivated by the recent popularity of the Bayesian lasso idea. We design a Markov chain Monte Carlo algorithm for posterior inference. Our simulations demonstrate the superior performance of the Bayesian method versus the frequentist approach. The method is further illustrated by an application to the hurricane data.

In Chapter 3, we develop a semiparametric Bayesian estimation and model selection approach for partially linear additive models in conditional quantile regression. The advantage of this new method is that nonlinear, linear and zero function components can be separated automatically and simultaneously during model fitting without the need of pre-specification or parameter tuning. Simulation studies are used to illustrate our algorithm. The proposed approach is further illustrated by application to two real data sets.

In Chapter 4, we conclude this thesis with a discussion. Also, some

directions for future work is briefly discussed.

All the MCMC algorithms mentioned in this thesis are implemented in R, and available upon request.

Chapter 2

Bayesian Quantile Regression for Single-Index Models

In this chapter, we consider a single-index quantile regression model. For a given quantile level $\tau \in (0, 1)$ and i.i.d. pairs $(\mathbf{x}_i, y_i)$, it is given by

$$Q_{y_i|\mathbf{x}_i}(\tau) = \eta(\mathbf{x}_i^T \boldsymbol{\beta}), \quad i = 1, 2, \dots, n.$$

Here y_i is the response, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ is the p -dimensional predictor vector, $Q_{y_i|\mathbf{x}_i}(\cdot) = F_{y_i|\mathbf{x}_i}^{-1}(\cdot)$ is the inverse cumulative distribution function of the response given the predictors, $\eta(\cdot)$ is the unknown univariate link function, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is the index which implicitly depends on the desired quantile τ . Dimension reduction is achieved specifically by the index vector so that η is a univariate function instead of p -variate one, as in fully nonparametric regression.

We propose a Bayesian treatment of the single-index quantile regression model. Recently, Wu et al. (2010) has considered a similar model using kernel regression, which will serve as a frequentist benchmark for our method. We establish a hierarchical Bayesian model by adopting the

asymmetric Laplace distribution and its location-scale mixture representation introduced in Chapter 1. Following Gramacy and Lian (2012), we assign a Gaussian process prior on the link function, to obtain a flexible nonparametric quantile regression model. The posterior inference of all parameters is performed via Markov chain Monte Carlo computations which automatically incorporate all sources of uncertainty.

The remainder of the chapter proceeds as follows. In Section 2.1, we describe the structure of our hierarchical Bayesian single-index quantile regression model and discuss our prior choices. We also consider the posterior sampling algorithm focusing on a more efficient partially collapsed sampler, where the link function is integrated out when drawing samples of the index vector. This is explained in detail in the Appendix. Then, numerical illustrations including simulation studies and a real data example are presented in Section 2.2. We summarize the chapter in Section 2.3.

2.1 Hierarchical Bayesian Modelling

At the τ -th quantile, we model the residual errors by the AL distribution. More specifically, the probability distribution of y given $\mu = \eta(\mathbf{x}^T \boldsymbol{\beta})$ is assumed to be

$$\pi(y|\mu, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\frac{1}{\sigma} \rho_{\tau}(y - \mu) \right\},$$

where $\rho_{\tau}(u) = u(\tau - I(u \leq 0))$, the quantile level τ is the skewness parameter in the distribution, μ is the location parameter, and σ is the scale parameter. In our context, with the setting $\mu_i = \eta(\mathbf{x}_i^T \boldsymbol{\beta})$, and $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, the conditional distribution for the observations is

$$\pi(\mathbf{y}|\boldsymbol{\beta}, \eta, \sigma) = \frac{\tau^n(1-\tau)^n}{\sigma^n} \exp \left\{ -\frac{1}{\sigma} \sum_{i=1}^n \rho_{\tau}(y_i - \eta(\mathbf{x}_i^T \boldsymbol{\beta})) \right\}. \quad (2.1.1)$$

With the location-scale mixture representation of the AL distribution (1.2.2), we can write the observations satisfying (2.1.1) alternatively as

$$y_i = \eta(\mathbf{x}_i^T \boldsymbol{\beta}) + k_1 e_i + \sqrt{k_2 \sigma e_i} z_i,$$

where $e_i \sim \exp(1/\sigma)$ is an exponential random variable with mean σ , z_i is a standard normal random variable and is independent of e_i , $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, and $k_2 = \frac{2}{\tau(1-\tau)}$. This suggests treating the e_i as latent variables, where the conditional distribution of $\mathbf{y}$ is rewritten as

$$\begin{aligned} \pi(\mathbf{y}|\boldsymbol{\beta}, \eta, \sigma, \mathbf{e}_n) &= \prod_{i=1}^n (2\pi k_2 \sigma e_i)^{-1/2} \exp \left\{ -\frac{1}{2k_2 \sigma e_i} (y_i - \eta(\mathbf{x}_i^T \boldsymbol{\beta}) - k_1 e_i)^2 \right\} \\ &\propto \exp \left\{ -\frac{(\mathbf{y} - \boldsymbol{\eta}_n - k_1 \mathbf{e}_n)^T \mathbf{E}^{-1} (\mathbf{y} - \boldsymbol{\eta}_n - k_1 \mathbf{e}_n)}{2} \right\} (\det[\mathbf{E}])^{-1/2}. \end{aligned}$$

Here $\mathbf{e}_n = (e_1, e_2, \dots, e_n)^T$, $\mathbf{E} = k_2 \sigma \text{diag}(e_1, e_2, \dots, e_n)$, and $\boldsymbol{\eta}_n = (\eta_1, \dots, \eta_n)^T = (\eta(\mathbf{x}_1^T \boldsymbol{\beta}), \dots, \eta(\mathbf{x}_n^T \boldsymbol{\beta}))^T$.

As in Choi et al. (2011) and Gramacy and Lian (2012), we model the link function by a Gaussian process prior distribution. More specifically, the prior distribution of η is a Gaussian process, with zero mean and a squared-exponential covariance function,

$$\eta \sim \text{GP}(\mathbf{0}, \mathbf{C}(\cdot, \cdot)), \quad \mathbf{C}(x, x') = \gamma \exp\{-(x - x')^2/d\}, \quad (2.1.2)$$

where γ and d are hyperparameters. Writing this out in the single-index model framework using the observed covariates $\mathbf{x}_i$, we have

$$\pi(\boldsymbol{\eta}_n|\boldsymbol{\beta}, \gamma) \propto \det[\mathbf{C}_n]^{-1/2} \exp \left\{ -\frac{\boldsymbol{\eta}_n^T \mathbf{C}_n^{-1} \boldsymbol{\eta}_n}{2} \right\},$$

where $\mathbf{C}_n$ is an $n \times n$ matrix with entries $\mathbf{C}(\mathbf{x}_i, \mathbf{x}_j) = \gamma \exp\{-(\mathbf{x}_i^T \boldsymbol{\beta} - \mathbf{x}_j^T \boldsymbol{\beta})^2/d\}$.

In the literature of single-index models, it is well-known that η and $\boldsymbol{\beta}$ are unidentifiable since $\eta(\mathbf{x}^T \boldsymbol{\beta}) = \eta_c(\mathbf{x}^T (c\boldsymbol{\beta}))$, $c \neq 0$, where $\eta_c(\cdot) = \eta(\cdot/c)$ and

thus β is only identifiable up to a constant scale. It is typically assumed that $\|\beta\| = 1$ so that β is identifiable up to sign (β and $-\beta$ leads to exactly the same model fit). Accordingly, Choi et al. (2011) restricted the support of the prior distribution for β to the unit sphere. On the other hand, Gramacy and Lian (2012) noted that without the constraint $\|\beta\| = 1$, only $\beta/\sqrt{d}$ is identifiable when using the Gaussian process prior, and thus one can remove the range parameter d and also remove the unit norm constraint on prior of β , which is mathematically equivalent to imposing $\|\beta\| = 1$ and keeping the range parameter d . With the latter approach, we have the advantage that the prior on β is more easily specified, and one fewer parameter (d) to worry about. Note that the model leaves the sign of β unidentified under either specification. In some cases when β is not of direct interest, it does not matter at all. When inference for β is a primary goal, some simple heuristics for reconciling the signs are provided by Gramacy and Lian (2012).

We therefore adopt the approach in Gramacy and Lian (2012), so that the entries of $\mathbf{C}_n$ are $\mathbf{C}(\mathbf{x}_i, \mathbf{x}_j) = \gamma \exp\{-(\mathbf{x}_i^T \beta - \mathbf{x}_j^T \beta)^2\}$. Since β is not constrained to have unit norm, we are free to choose any prior for β . A typical choice is to put an independent Gaussian prior on each component, which is sometimes called a *ridge prior*. One can also consider the popular g-prior (Zellner, 1986; George and Foster, 2000). A generalization of the ridge prior is the so-called Bayesian lasso, which has been of much interest in the recent literature (Park and Casella, 2008; Hans, 2009). Under this prior, β_j , $j = 1, \dots, p$ are independent and identically Laplace,

$$\pi(\beta|\sigma, \lambda) = \prod_{j=1}^p \frac{\lambda}{2\sigma} e^{-\lambda|\beta_j|/\sigma}, \quad \lambda > 0.$$

There are a suite of similar priors which share attractive properties and

yet further generalize the lasso. Examples include the normal–gamma prior (Griffin and Brown, 2010), the Bayesian elastic set (Li and Lin, 2010) and the horseshoe (Carvalho et al., 2010). Our reasons for calling these “generalizations” have to do with the form of their hierarchical latent variable representations, and corresponding data augmentation Gibbs samplers. We focus on the Bayesian lasso as a representative case in this paper. Simplifications (to the ridge) and further generalizations to the others are straightforward. Recognizing that estimators for β are not equivariant under such priors (Park and Casella, 2008), we take the common pre-processing step of scaling the inputs $\mathbf{x}_i$ to have a unit L_2 -norm. This also simplifies the choice of default priors for λ and σ .

To summarize, we provide the Bayesian hierarchical formulation and the diagram of hierarchical structure in Figure 2.1.

Figure 2.1: *The diagram of Bayesian hierarchical structure.*

$$\begin{aligned}
\mathbf{y}|\boldsymbol{\eta}_n, \mathbf{e}_n, \sigma &\sim N(\boldsymbol{\eta}_n + k_1 \mathbf{e}_n, \mathbf{E}), \\
\boldsymbol{\eta}_n|\mathbf{x}, \gamma, \boldsymbol{\beta} &\sim N(\mathbf{0}, \mathbf{C}_n), \\
\boldsymbol{\beta} &\sim \pi(\boldsymbol{\beta}|\sigma, \lambda), \quad e_i \stackrel{i.i.d.}{\sim} \exp(1/\sigma), \\
\sigma &\sim \pi(\sigma), \lambda \sim \pi(\lambda), \gamma \sim \pi(\gamma).
\end{aligned}$$

The hyperpriors for σ, λ, γ are set to be $IG(a_\sigma, b_\sigma)$, $Ga(a_\lambda, b_\lambda)$ and $IG(a_\gamma, b_\gamma)$, where IG denotes the inverse Gamma distribution and Ga denotes the Gamma distribution. All of the hyperparameters $a_\sigma, b_\sigma, a_\lambda, b_\lambda, a_\gamma, b_\gamma$ are set to be 0.5 in all our numerical experiments. Sensitivity analyses reveal that our results are not sensitive to these choices.

The posterior distribution of all parameters is found via MCMC, using a partially collapsed version integrating out $\boldsymbol{\eta}_n$. The details are left to the Appendix. For initial values, we set $\gamma = 0.5$, $\sigma = 0.5$, $\lambda = 0.5$, $e_i \sim \exp(1/\sigma)$ and $\beta_j \sim Unif(-1, 1)$. Convergence is generally very good, except for $\tau = 0.99$ for real data. In the latter case, the results presented are for one particular run where a different initialization value leads to different estimates.

2.2 Numerical Illustrations

We present three simulation examples and a real data application to illustrate the proposed method, the Bayesian quantile single-index regression model, which is denoted by BQSIM for short in the rest of the thesis.

2.2.1 Simulations

We illustrate the performance of the proposed method by comparing it with a non-Bayesian single-index quantile regression approach described by Wu et al. (2010), based on kernel estimation. This frequentist method is denoted by QSIM in the following. Since the frequentist approach requires the identifiability constraint $\|\beta\| = 1$, we also normalize the Bayesian estimate of the index vector to have unit norm and furthermore require the first component of the index vector to be positive to resolve the sign indeterminacy. The first three simulation examples are directly taken from Wu et al. (2010).

Example 1

Consider data generated from the following single-index model with homoscedastic errors,

$$y = \eta(\mathbf{x}^T \beta) + 0.1\mathcal{Z}, \quad \eta(t) = \sin\left(\frac{\pi(t - A)}{C - A}\right),$$

with $\beta = (\beta_1, \beta_2, \beta_3)^T = \frac{1}{\sqrt{3}}(1, 1, 1)^T$, $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $C = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\mathcal{Z} \sim N(0, 1)$. The predictors $\mathbf{x} = (x_1, x_2, x_3)^T$ are uniform in $[0, 1]^3$. We consider sample sizes $n = 100$ and $n = 200$. For each sample size, we fit the models at seven different quantiles $\tau = 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99$. In each case, the MCMC algorithm is run for 20000 iterations with a burn-in of 10000. For convergence diagnosis, we present trace plots of β , σ , λ and γ in Figure 2.2 with two different sets of initial values using $\tau = 0.5$. The plots suggest that the constructed chains mix quickly.

For a Bayesian point estimator we consider both the posterior mean and posterior median based on the sampled values after burn-in. The resulting estimates are summarized in Tables 2.1 and 2.2, together with the sample

Figure 2.2: Trace plots of β_1 , β_2 , β_3 , σ , γ and λ at quantile 0.5 for simulation example 1, when $n = 100$. Two chains with different starting values are illustrated.

standard deviation (S.D.) and 2.5% and 97.5% quantiles of the sampled values after burn-in. It is seen that both posterior mean and posterior median estimators work well and give similar estimates. Thus we only focus on posterior mean as our point estimate in the following.

Figures 2.3 and 2.4 display the boxplots of the estimated index vectors, comparing BQSIM and QSIM, with $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. To save space, cases with $\tau = 0.95$ and 0.99 are not presented. The plots are based on 100 independently generated datasets in each case and show that the Bayesian estimates have smaller bias and lower variance. These plots generally give the impression that BQSIM produces more precise and stable estimates than QSIM. Mean squared errors (MSE) of the estimates based on these 100 replications in each case are shown in Table 2.3 for both sample sizes and all seven quantile levels.

Figure 2.6 shows the fitted $\boldsymbol{\eta}_n$ values (posterior mean) by BQSIM at $\tau = 0.5$ on a typical run. In the left panel, the fitted $\boldsymbol{\eta}_n$ are plotted against the true index $\mathbf{x}_i^T \boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is the true value in the model. The true link function is also shown in the same figure. In the right panel, the fitted $\boldsymbol{\eta}_n$ are plotted against the fitted index $\mathbf{x}_i^T \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the posterior mean estimate, leading to visually smoother fitted values.

Finally, to demonstrate that partial collapsing can significantly improve mixing, we show the autocorrelation plots of the $\boldsymbol{\beta}$ series in Figure 2.5. The autocorrelations are calculated based on one random replicate at quantile 0.5 with $n = 100$. Observe that the autocorrelation of the series produced by the uncollapsed chain is much higher, implying the samples are much “stickier”. As discussed in the Appendix, $\mathbf{C}_n$ is nearly singular which caused numerical problems when using the uncollapsed Gibbs sampler. To avoid this numerical problem, a small artificial nugget effect is added to the matrix

(that is we use $\mathbf{C}_n + \epsilon I$ in place of $\mathbf{C}_n$ in evaluating the full conditional density, with $\epsilon = 10^{-5}$).

Example 2

Now consider data generated as follows:

$$y = \eta(\mathbf{x}^T \boldsymbol{\beta}) + \sqrt{(\sin(\mathbf{x}^T \boldsymbol{\beta}) + 1)} \mathcal{Z}, \text{ where } \eta(t) = 10 \sin(0.75t),$$

and $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = \frac{1}{\sqrt{5}}(1, 2)^T$, $\mathbf{x} = (x_1, x_2)^T$, and $\mathcal{Z}$ is a standard normal random variable. The x_j s, ($j = 1, 2$) are drawn identically and independently from a normal distribution with mean 0 and variance 0.25^2 . We conduct simulations at $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99\}$ with $n = 100$ and $n = 200$, each with 100 replications. Figures 2.7 and 2.8 show the boxplots for the estimated index vector and Table 2.4 reports the mean squared errors. With more accurate and stable estimates, BQSIM outperforms QSIM on each quantile level, when errors are heteroskedastic.

Example 3

Next, we consider a regression model with exponentially distributed errors,

$$y = \eta(\mathbf{x}^T \boldsymbol{\beta}) + \mathcal{E}, \quad \eta(t) = 5 \cos(t) + \exp(-t^2),$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = \frac{1}{\sqrt{5}}(1, 2)^T$, $\mathbf{x} = (x_1, x_2)^T$. $x_j \stackrel{i.i.d.}{\sim} N(0, 1)$, $j = 1, 2$, and $\mathcal{E} \sim \exp(1/2)$. Using the same sample sizes and quantile levels as for the previous two examples, the results are presented in Figures 2.9, 2.10, and Table 2.5, which again demonstrate the superiority of BQSIM.

Example 4

Here we follow a similar setup as in Example 1, except that the additive errors follow the AL distribution with $\sigma = 0.05$. That is we generate data

Table 2.1: *Results of BQSIM for simulation example 1, $n = 100$.*

$n = 100$		True	Mean	Median	S.D.	2.5%	97.5%
$\tau=0.10$	β_1	0.5774	0.5666	0.5817	0.0157	0.5354	0.5970
	β_2	0.5774	0.5739	0.5749	0.0159	0.5420	0.6044
	β_3	0.5774	0.5741	0.5770	0.0164	0.5410	0.6052
$\tau=0.25$	β_1	0.5774	0.5785	0.5766	0.0154	0.5481	0.6084
	β_2	0.5774	0.5748	0.5734	0.0153	0.5442	0.6047
	β_3	0.5774	0.5773	0.5783	0.0155	0.5466	0.6071
$\tau=0.50$	β_1	0.5774	0.5773	0.5775	0.0157	0.5465	0.6080
	β_2	0.5774	0.5732	0.5741	0.0158	0.5426	0.6043
	β_3	0.5774	0.5802	0.5815	0.0159	0.5486	0.6118
$\tau=0.75$	β_1	0.5774	0.5756	0.5755	0.0156	0.5455	0.6067
	β_2	0.5774	0.5738	0.5740	0.0156	0.5429	0.6046
	β_3	0.5774	0.5812	0.5802	0.0157	0.5503	0.6118
$\tau=0.90$	β_1	0.5774	0.5760	0.5751	0.0157	0.5453	0.6071
	β_2	0.5774	0.5751	0.5762	0.0152	0.5458	0.6053
	β_3	0.5774	0.5790	0.5811	0.0155	0.5490	0.6102
$\tau=0.95$	β_1	0.5774	0.5819	0.5823	0.0159	0.5511	0.6127
	β_2	0.5774	0.5745	0.5754	0.0154	0.5443	0.6047
	β_3	0.5774	0.5730	0.5735	0.0157	0.5432	0.6045
$\tau=0.99$	β_1	0.5774	0.5771	0.5786	0.0212	0.5338	0.6142
	β_2	0.5774	0.5710	0.5668	0.0184	0.5362	0.6077
	β_3	0.5774	0.5729	0.5802	0.0225	0.5304	0.6124

Table 2.2: *Results of BQSIM for simulation example 1, $n = 200$.*

$n = 200$		True	Mean	Median	S.D.	2.5%	97.5%
$\tau=0.10$	β_1	0.5774	0.5742	0.5745	0.0104	0.5543	0.5947
	β_2	0.5774	0.5749	0.5768	0.0010	0.5567	0.5959
	β_3	0.5774	0.5821	0.5789	0.0104	0.5592	0.6002
$\tau=0.25$	β_1	0.5774	0.5761	0.5756	0.0106	0.5552	0.5968
	β_2	0.5774	0.5764	0.5782	0.0105	0.5562	0.5972
	β_3	0.5774	0.5788	0.5771	0.0106	0.5575	0.5991
$\tau=0.50$	β_1	0.5774	0.5771	0.5779	0.0106	0.5561	0.5971
	β_2	0.5774	0.5760	0.5779	0.0106	0.5561	0.5976
	β_3	0.5774	0.5783	0.5766	0.0104	0.5571	0.5979
$\tau=0.75$	β_1	0.5774	0.5769	0.5752	0.0107	0.5567	0.5987
	β_2	0.5774	0.5754	0.5746	0.0109	0.5548	0.5973
	β_3	0.5774	0.5790	0.5803	0.0109	0.5568	0.5992
$\tau=0.90$	β_1	0.5774	0.5817	0.5802	0.0108	0.5600	0.6027
	β_2	0.5774	0.5816	0.5802	0.0098	0.5630	0.6017
	β_3	0.5774	0.5675	0.5644	0.0106	0.5471	0.5881
$\tau=0.95$	β_1	0.5774	0.5754	0.5766	0.0109	0.5546	0.5959
	β_2	0.5774	0.5810	0.5809	0.0114	0.5599	0.6037
	β_3	0.5774	0.5743	0.5738	0.0105	0.5543	0.5951
$\tau=0.99$	β_1	0.5774	0.5590	0.5749	0.0241	0.5244	0.6029
	β_2	0.5774	0.5700	0.5780	0.0155	0.5425	0.5983
	β_3	0.5774	0.5601	0.5722	0.0218	0.5277	0.5984

Figure 2.3: Summarizing estimators of β for $n = 100, 200$ in example 1. ‘ $B100$ ’ (‘ $Q100$ ’) denotes $BQSIM$ ($QSIM$) with $n = 100$, for example.

Figure 2.4: Summarizing estimators of β for $n = 100, 200$ in example 1. ‘B100’ (‘Q100’) denotes BQSIM (QSIM) with $n = 100$, for example.

Table 2.3: *Comparison of MSE for BQSIM (using posterior mean as the estimator) and QSIM based on 100 replications in each case for simulation example 1.*

		MSE ($n = 100$)			MSE ($n = 200$)		
		β_1	β_2	β_3	β_1	β_2	β_3
$\tau=0.10$	QSIM	0.00239	0.00447	0.00886	0.00346	0.00473	0.00052
	BQSIM	0.00041	0.00051	0.00045	0.00017	0.00023	0.00020
$\tau=0.25$	QSIM	0.00169	0.00312	0.00303	0.00276	0.00138	0.00310
	BQSIM	0.00029	0.00029	0.00030	0.00012	0.00014	0.00019
$\tau=0.50$	QSIM	0.00269	0.00154	0.00254	0.00058	0.00053	0.00089
	BQSIM	0.00023	0.00025	0.00029	0.00013	0.00013	0.00018
$\tau=0.75$	QSIM	0.00340	0.00601	0.00291	0.00039	0.00028	0.00048
	BQSIM	0.00026	0.00029	0.00038	0.00017	0.00016	0.00016
$\tau=0.90$	QSIM	0.00424	0.00794	0.00641	0.00139	0.00100	0.00192
	BQSIM	0.00040	0.00049	0.00059	0.00023	0.00028	0.00030
$\tau=0.95$	QSIM	0.00857	0.00739	0.00951	0.00157	0.00183	0.00245
	BQSIM	0.00051	0.00075	0.00094	0.00042	0.00029	0.00041
$\tau=0.99$	QSIM	0.05034	0.04259	0.06166	0.00494	0.09708	0.07046
	BQSIM	0.00083	0.00098	0.00125	0.00062	0.00192	0.00099

Figure 2.5: The autocorrelation plots for $\beta_1, \beta_2, \beta_3$. The left column represents the series produced by our partially collapsed sampler and the right column represents the uncollapsed sampler.

Figure 2.6: *The estimated link functions. On the left panel, the values of estimated η_n are plotted against the true index, while for the right panel the values of estimated η_n are plotted against the estimated index. The true link function is also shown in both figures (solid line).*

Figure 2.7: Summarizing estimators of β for $n = 100, 200$ in example 2.

Figure 2.8: *Summarizing estimators of β for $n = 100, 200$ in example 2.*

Table 2.4: *Comparison of MSE for BQSIM (using posterior mean as the estimator) and QSIM based on 100 replications in each case for simulation example 2.*

		MSE ($n = 100$)		MSE ($n = 200$)	
		β_1	β_2	β_1	β_2
$\tau=0.10$	QSIM	0.02510	0.02991	0.00744	0.00238
	BQSIM	0.00841	0.00197	0.00451	0.00115
$\tau=0.25$	QSIM	0.01997	0.00396	0.00823	0.00645
	BQSIM	0.00401	0.00095	0.00220	0.00059
$\tau=0.50$	QSIM	0.01098	0.00409	0.00214	0.00058
	BQSIM	0.00225	0.00056	0.00179	0.00041
$\tau=0.75$	QSIM	0.02025	0.00868	0.00345	0.00089
	BQSIM	0.00222	0.00059	0.00021	0.00056
$\tau=0.90$	QSIM	0.01983	0.01227	0.00584	0.00196
	BQSIM	0.00393	0.00104	0.00393	0.00104
$\tau=0.95$	QSIM	0.02048	0.00974	0.01044	0.01279
	BQSIM	0.00647	0.00150	0.00488	0.00126
$\tau=0.99$	QSIM	0.06640	0.04652	0.05911	0.03972
	BQSIM	0.00953	0.00226	0.00997	0.00210

Figure 2.9: Summarizing estimators of β for $n = 100, 200$ in example 3.

Figure 2.10: *Summarizing estimators of β for $n = 100, 200$ in example 3.*

Table 2.5: *Comparison of MSE for BQSIM (using posterior mean as the estimator) and QSIM based on 100 replications in each case for simulation example 3.*

		MSE ($n = 100$)		MSE ($n = 200$)	
		β_1	β_2	β_1	β_2
$\tau=0.10$	QSIM	0.00767	0.00447	0.00174	0.00114
	BQSIM	0.00033	0.00008	0.00018	0.00005
$\tau=0.25$	QSIM	0.00421	0.00175	0.00647	0.00384
	BQSIM	0.00074	0.00020	0.00035	0.00009
$\tau=0.50$	QSIM	0.00843	0.00306	0.00631	0.00351
	BQSIM	0.00173	0.00044	0.00090	0.00022
$\tau=0.75$	QSIM	0.03992	0.05156	0.01429	0.00690
	BQSIM	0.00221	0.00050	0.00573	0.00098
$\tau=0.90$	QSIM	0.06118	0.08449	0.06311	0.07486
	BQSIM	0.04856	0.01129	0.04571	0.00219
$\tau=0.95$	QSIM	0.07744	0.05669	0.09499	0.17752
	BQSIM	0.07094	0.05106	0.08097	0.05106
$\tau=0.99$	QSIM	0.10262	0.10510	0.12993	0.13203
	BQSIM	0.10064	0.05996	0.12071	0.06277

sets from the model (2.1.1), independently for each value of τ . This example mainly serves as an illustration that when the errors indeed follow AL distribution, so that estimation of σ becomes meaningful, we can indeed estimate its value satisfactorily. Since the frequentist approach QSIM has the distribution-free property on the error terms and the estimation of the error with QSIM is not meaningful for estimating σ , we only consider BQSIM in this example. These results are presented in Tables 2.6 and 2.7 for $n = 100$ and $n = 200$ respectively. We emphasize that the estimated σ is meaningful only if the true AL distribution is used in estimation. As an illustration of this point, we consider data generated from (2.1.1) using $\tau = 0.5$ and fitted using BQSIM at quantile level $\tau = 0.75$. In this case, the index vector β can still be estimated very close to the true values, while the estimated σ is about 0.01. In general, BQSIM cannot be used to estimate the error distribution directly. This limitation is discussed further in Chapter 4. In this example, the performance at $\tau = 0.99$ is less satisfactory than in previous examples.

Example 5

Finally, we increase the dimension in Example 1 to $p = 10$ to demonstrate the performance in higher dimensions. The only difference in the setup from Example 1 is that we now set $\beta = (1, 1, 1, 0, \dots, 0)^T$ and $\beta = \frac{1}{\sqrt{10}}(1, 1, 1, 1, \dots, 1)^T$, and only consider $n = 100$. For these two cases, the estimation results are shown in Tables 2.8 and 2.9 respectively. To save space, boxplots comparing BQSIM and QSIM are not shown now that we are estimating 10 coefficients. Our methods still perform much better than QSIM in these higher dimensional cases.

Table 2.6: *Results of BQSIM for simulation example 4, $n = 100$.*

$n = 100$		True	Mean	Median	S.D.	2.5%	97.5%	MSE
$\tau=0.10$	β_1	0.5774	0.5805	0.5818	0.0353	0.5084	0.6476	0.00117
	β_2	0.5774	0.5814	0.5771	0.0338	0.5123	0.6463	0.00069
	β_3	0.5774	0.5641	0.5615	0.0350	0.4923	0.6304	0.00132
	σ	0.0500	0.0521	0.0527	0.0054	0.0430	0.0637	0.00002
$\tau=0.25$	β_1	0.5774	0.5809	0.5823	0.0215	0.5377	0.6227	0.00055
	β_2	0.5774	0.5762	0.5760	0.0212	0.5341	0.6171	0.00043
	β_3	0.5774	0.5723	0.5704	0.0218	0.5290	0.6145	0.00063
	σ	0.0500	0.0520	0.0520	0.0054	0.0427	0.0638	0.00003
$\tau=0.50$	β_1	0.5774	0.5787	0.5786	0.0176	0.5432	0.6129	0.00030
	β_2	0.5774	0.5740	0.5759	0.0176	0.5394	0.6084	0.00026
	β_3	0.5774	0.5778	0.5783	0.0179	0.5425	0.6127	0.00030
	σ	0.0500	0.0514	0.0515	0.0053	0.0421	0.0631	0.00003
$\tau=0.75$	β_1	0.5774	0.5750	0.5765	0.0208	0.5346	0.6165	0.00041
	β_2	0.5774	0.5745	0.5750	0.0210	0.5330	0.6155	0.00047
	β_3	0.5774	0.5802	0.5780	0.0212	0.5384	0.6219	0.00043
	σ	0.0500	0.0519	0.0514	0.0053	0.0423	0.0634	0.00004
$\tau=0.90$	β_1	0.5774	0.5734	0.5682	0.0342	0.5055	0.6401	0.00083
	β_2	0.5774	0.5685	0.5672	0.0336	0.5020	0.6343	0.00094
	β_3	0.5774	0.5842	0.5829	0.0342	0.5178	0.6521	0.00131
	σ	0.0500	0.0514	0.0517	0.0054	0.0422	0.0633	0.00003
$\tau=0.95$	β_1	0.5774	0.5455	0.5652	0.0851	0.3827	0.6887	0.01151
	β_2	0.5774	0.5611	0.5658	0.0611	0.4412	0.6818	0.00352
	β_3	0.5774	0.5676	0.5912	0.0714	0.4175	0.6979	0.01261
	σ	0.0500	0.0518	0.0518	0.0053	0.0422	0.0640	0.00003
$\tau=0.99$	β_1	0.5774	0.1115	0.1589	0.4957	-0.7874	0.8895	0.28021
	β_2	0.5774	0.4797	0.4677	0.2595	0.0817	0.9325	0.02603
	β_3	0.5774	0.1051	0.1513	0.4771	-0.7418	0.8867	0.27783
	σ	0.0500	0.0529	0.0530	0.0054	0.0433	0.0646	0.00003

Table 2.7: *Results of BQSIM for simulation example 4, $n = 200$.*

$n = 200$		True	Mean	Median	S.D.	2.5%	97.5%	MSE
$\tau=0.10$	β_1	0.5774	0.5813	0.5797	0.0209	0.5389	0.6213	0.00042
	β_2	0.5774	0.5742	0.5736	0.0221	0.5299	0.6167	0.00041
	β_3	0.5774	0.5742	0.5753	0.0217	0.5301	0.6159	0.00041
	σ	0.0500	0.0521	0.0519	0.0037	0.0430	0.0637	0.00002
$\tau=0.25$	β_1	0.5774	0.5799	0.5798	0.0135	0.5534	0.6057	0.00020
	β_2	0.5774	0.5764	0.5806	0.0138	0.5493	0.6028	0.00025
	β_3	0.5774	0.5747	0.5754	0.0139	0.5476	0.6018	0.00025
	σ	0.0500	0.0515	0.0528	0.0037	0.0427	0.0638	0.00002
$\tau=0.50$	β_1	0.5774	0.5816	0.5830	0.0119	0.5584	0.6045	0.00030
	β_2	0.5774	0.5748	0.5738	0.0118	0.5516	0.5980	0.00026
	β_3	0.5774	0.5749	0.5752	0.0118	0.5517	0.5978	0.00030
	σ	0.0500	0.0522	0.0517	0.0038	0.0447	0.0592	0.00001
$\tau=0.75$	β_1	0.5774	0.5771	0.5789	0.0136	0.5511	0.6035	0.00027
	β_2	0.5774	0.5769	0.5774	0.0138	0.5503	0.6039	0.00023
	β_3	0.5774	0.5770	0.5751	0.0135	0.5505	0.6029	0.00022
	σ	0.0500	0.0512	0.0512	0.0037	0.0448	0.0594	0.00002
$\tau=0.90$	β_1	0.5774	0.5774	0.5760	0.0213	0.5363	0.6201	0.00057
	β_2	0.5774	0.5800	0.5788	0.0217	0.5370	0.6224	0.00044
	β_3	0.5774	0.5720	0.5697	0.0217	0.5295	0.6152	0.00047
	σ	0.0500	0.0511	0.0518	0.0037	0.0447	0.0593	0.00001
$\tau=0.95$	β_1	0.5774	0.5766	0.5739	0.0374	0.5009	0.6447	0.00197
	β_2	0.5774	0.5795	0.5841	0.0337	0.5128	0.6471	0.00099
	β_3	0.5774	0.5679	0.5688	0.0354	0.5003	0.6385	0.00145
	σ	0.0500	0.0518	0.0529	0.0038	0.0450	0.0649	0.00002
$\tau=0.99$	β_1	0.5774	0.3256	0.3507	0.3545	-0.3724	0.8083	0.12130
	β_2	0.5774	0.5350	0.5428	0.1791	0.2028	0.8372	0.02907
	β_3	0.5774	0.3318	0.3344	0.2977	-0.2720	0.7914	0.12535
	σ	0.0500	0.0524	0.0520	0.0038	0.0433	0.0646	0.00002

Table 2.8: Comparison of MSE for BQSIM and QSIM based on 100 replications in each case for simulation example 5, where $\beta = (1, 1, 1, 0, \dots, 0)^T$.

$n = 100$		β_1	β_2	β_3	β_4	β_5
$\tau=0.10$	QSIM	0.03051	0.06458	0.09402	0.03894	0.03484
	BQSIM	0.01000	0.00251	0.03926	0.00068	0.00460
$\tau=0.25$	QSIM	0.00822	0.01876	0.04851	0.01342	0.00825
	BQSIM	0.00054	0.00046	0.00064	0.00045	0.00038
$\tau=0.50$	QSIM	0.01385	0.00483	0.00308	0.00617	0.00417
	BQSIM	0.00031	0.00039	0.00043	0.00033	0.00039
$\tau=0.75$	QSIM	0.00106	0.03489	0.03508	0.00102	0.00396
	BQSIM	0.00046	0.00039	0.00058	0.00063	0.00068
$\tau=0.90$	QSIM	0.00532	0.03687	0.03949	0.01752	0.00940
	BQSIM	0.00148	0.00106	0.00116	0.00156	0.00151
$\tau=0.95$	QSIM	0.03913	0.06571	0.10765	0.03061	0.030221
	BQSIM	0.00698	0.00368	0.00401	0.00178	0.00287
$\tau=0.99$	QSIM	0.04228	0.05799	0.06567	0.06315	0.06315
	BQSIM	0.02259	0.01966	0.03773	0.02252	0.01228
$n = 100$		β_6	β_7	β_8	β_9	β_{10}
$\tau=0.10$	QSIM	0.02318	0.02520	0.02052	0.03534	0.04236
	BQSIM	0.01572	0.00628	0.00214	0.00194	0.00432
$\tau=0.25$	QSIM	0.01096	0.01086	0.00622	0.00749	0.01302
	BQSIM	0.00056	0.00034	0.00038	0.00054	0.00036
$\tau=0.50$	QSIM	0.00717	0.01045	0.00452	0.00622	0.00974
	BQSIM	0.00043	0.00047	0.00029	0.00049	0.00032
$\tau=0.75$	QSIM	0.00117	0.00386	0.00157	0.01653	0.00803
	BQSIM	0.00035	0.00055	0.00049	0.00064	0.00046
$\tau=0.90$	QSIM	0.00766	0.04293	0.01118	0.01324	0.02452
	BQSIM	0.00118	0.00167	0.00065	0.00133	0.00181
$\tau=0.95$	QSIM	0.02806	0.03451	0.02941	0.03739	0.02923
	BQSIM	0.00288	0.00200	0.00451	0.00305	0.00263
$\tau=0.99$	QSIM	0.08010	0.06853	0.05461	0.06722	0.05835
	BQSIM	0.01346	0.01363	0.01514	0.01219	0.01140

Table 2.9: Comparison of MSE for BQSIM and QSIM based on 100 replications in each case for simulation example 5, where $\beta = \frac{1}{\sqrt{10}}(1, 1, 1, 1, \dots, 1)^T$.

$n = 100$		β_1	β_2	β_3	β_4	β_5
$\tau=0.10$	QSIM	0.00124	0.00149	0.00114	0.00133	0.00140
	BQSIM	0.00066	0.00085	0.00067	0.00079	0.00078
$\tau=0.25$	QSIM	0.00101	0.00077	0.00072	0.00089	0.00050
	BQSIM	0.00047	0.00035	0.00040	0.00039	0.00038
$\tau=0.50$	QSIM	0.00086	0.00062	0.00367	0.00061	0.00567
	BQSIM	0.00048	0.00032	0.00026	0.00048	0.00055
$\tau=0.75$	QSIM	0.00102	0.01325	0.00135	0.00122	0.00173
	BQSIM	0.00064	0.00048	0.00045	0.00053	0.00048
$\tau=0.90$	QSIM	0.00308	0.00256	0.01116	0.00666	0.00280
	BQSIM	0.00095	0.00082	0.00058	0.00081	0.00061
$\tau=0.95$	QSIM	0.01535	0.01194	0.03278	0.02291	0.02617
	BQSIM	0.00129	0.00144	0.00125	0.00162	0.00129
$\tau=0.99$	QSIM	0.01629	0.05267	0.02109	0.04479	0.04052
	BQSIM	0.01493	0.00967	0.01255	0.01236	0.01534
$n = 100$		β_6	β_7	β_8	β_9	β_{10}
$\tau=0.10$	QSIM	0.00123	0.00160	0.00141	0.00151	0.00100
	BQSIM	0.00096	0.00086	0.0065	0.00107	0.00066
$\tau=0.25$	QSIM	0.00074	0.00101	0.00069	0.00067	0.00084
	BQSIM	0.00054	0.00048	0.00052	0.00043	0.00033
$\tau=0.50$	QSIM	0.00098	0.00058	0.00180	0.00178	0.00064
	BQSIM	0.00037	0.00048	0.00033	0.00037	0.00029
$\tau=0.75$	QSIM	0.00283	0.00123	0.00428	0.00093	0.00123
	BQSIM	0.00057	0.00049	0.00057	0.00041	0.00041
$\tau=0.90$	QSIM	0.00385	0.00258	0.01294	0.00964	0.00341
	BQSIM	0.00070	0.00071	0.00079	0.00105	0.00064
$\tau=0.95$	QSIM	0.02083	0.02902	0.04447	0.02839	0.05003
	BQSIM	0.00132	0.00170	0.00077	0.00142	0.00104
$\tau=0.99$	QSIM	0.04999	0.04875	0.07079	0.06024	0.05711
	BQSIM	0.01191	0.01088	0.01134	0.02032	0.01921

2.2.2 Real Data Analysis

Finally, we apply the proposed method to the tropical cyclone (TC) data. Coastal tropical cyclones pose a serious threat to social and economic institutions. It is necessary and useful to provide a statistical way to analyze the TC data and evaluate the risk of the next catastrophic cyclone. Here we consider a dataset consisting of a sample of 422 TCs occurring near the US coastline over a 108-year period (1899-2006). Following Jagger and Elsner (2009) and Bondell et al. (2010), we model the wind speeds from TCs with four climate variables: the North Atlantic Oscillation Index (NAO), the Southern Oscillation Index (SOI), the Atlantic sea-surface temperature (SST) and the average sun spot number (SSN). The values of SOI, SST, and SSN are averaged values over the peak season of August through October and the values of NAO are averaged over the pre-season and early season months of May and June. Both Jagger and Elsner (2009) and Bondell et al. (2010) used linear quantile regression to analyze this data. Here we apply our BQSIM to analyze how these climate variables influence the wind speeds of TCs. We also fitted QSIM described in Wu et al. (2010) for comparison.

The particular focus for this type of data is on the upper quantiles, as these extreme hurricane-strength storms are of considerable importance. We consider five different quantile levels $\tau = (0.5, 0.75, 0.9, 0.95, 0.99)$. All covariates are scaled to have mean zero and standard deviation one.

Table 2.10 compares the obtained index vectors estimated by BQSIM and QSIM, and Figure 2.11 shows the estimated quantile curves of TC intensity at different levels. From the table, we can see that SOI is an important factor for TC intensity prediction, and the impact of SSN on TC intensity turns to be heavier as the quantile level increases. The index vectors ob-

tained by BQSIM and QSIM are qualitatively similar at lower quantiles, with more obvious deviations at levels above 0.9. This suggests that the estimates are not reliable for high quantile levels, especially for $\tau = 0.95$ and $\tau = 0.99$. The boxplots in Figures 2.12 and 2.13 (left columns) show the samples collected from the posterior distribution of β which are normalized to have unit norm. The histograms in Figures 2.12 and 2.13 show the implied distribution of d (see equation (2.1.2)). For Gaussian processes, larger values of d correspond to smoother functions. While it is generally hard to compare the performance of BQSIM and QSIM for real data, our previous simulations suggest that BQSIM is more trustworthy. We also performed model fitting on bootstrapped data and observed that the BQSIM estimates are more stable across bootstrap samples, except for $\tau = 0.99$ where estimates obtained from both BQSIM and QSIM are quite unstable.

2.3 Summary of the Chapter

In this chapter, we have proposed a Bayesian quantile regression method for single-index models based on a Gaussian process prior for the unknown link function. As detailed in the Appendix, we designed an efficient MCMC algorithm for posterior inference and demonstrated the superiority of the Bayesian approach to a modern non-Bayesian one. We carefully considered the possibility of marginalizing over the link function in some of the sampling steps, which leads to a partially collapsed sampler that balances sampling efficiency and implementation expediency. The performance of the proposed approach in our simulations is quite encouraging.

We used zero-mean Gaussian process with a squared exponential kernel. It is also possible to explicitly incorporate a mean function in the Gaussian

Table 2.10: *Estimates from BQSIM and QSIM for the TCs data.*

		BQSIM	QSIM
$\tau=0.50$	NAO(β_1)	0.46300	0.49633
	SOI(β_2)	-0.58446	-0.59011
	SST(β_3)	0.01665	-0.06932
	SSN(β_4)	0.51756	0.47965
$\tau=0.75$	NAO(β_1)	0.37840	0.51400
	SOI(β_2)	-0.87369	-0.73418
	SST(β_3)	-0.19373	-0.20965
	SSN(β_4)	0.07462	0.39093
$\tau=0.90$	NAO(β_1)	-0.13451	0.36110
	SOI(β_2)	-0.83952	-0.62212
	SST(β_3)	-0.40714	-0.58222
	SSN(β_4)	-0.18082	-0.37594
$\tau=0.95$	NAO(β_1)	-0.16689	0.10274
	SOI(β_2)	-0.90598	-0.11676
	SST(β_3)	-0.30985	-0.83128
	SSN(β_4)	-0.08797	-0.53364
$\tau=0.99$	NAO(β_1)	-0.03822	0.10860
	SOI(β_2)	-0.41609	-0.12104
	SST(β_3)	0.88491	-0.83293
	SSN(β_4)	0.13593	-0.52894

Figure 2.11: The estimated link functions by BQSIM (left column) and QSIM (right column) for TC intensity at different quantiles with 95% credible/confidence intervals.

Figure 2.12: The boxplots on the left shows the posterior distribution of β (after normalization to unit norm). The histograms on the right shows the posterior distribution of $d = 1/\|\beta\|^2$.

Figure 2.13: The boxplots on the left shows the posterior distribution of β (after normalization to unit norm). The histograms on the right shows the posterior distribution of $d = 1/\|\beta\|^2$.

process. This could add some flexibility to the model. For example, if a linear mean function is used, under independent zero-mean normal prior on the linear coefficients, the resulting process would be equivalent to a zero-mean Gaussian process with an additional quadratic term in the kernel function, as shown by MacKay (1998). Thus the consideration on whether to use a non-zero mean is similar to deciding what kind of kernel to use. We take the view that a zero-mean Gaussian process with the quadratic exponential kernel function is already flexible enough to model a variety of curves and thus do not consider these additional possibilities in modelling.

Chapter 3

Bayesian Quantile Regression for Partially Linear Additive Models

In this chapter, we propose a Bayesian quantile regression approach for partially linear additive models. At a given quantile level $\tau \in (0, 1)$, a partially linear additive model has the following form,

$$y_i = \mu_\tau + \sum_{j=1}^p f_{\tau,j}(x_{ij}) + \epsilon_{\tau,i}, i = 1, \dots, n, \quad (3.0.1)$$

where $(y_i, \mathbf{x}_i)$ are independent and identically distributed pairs. Here y_i is the response, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ is the p -dimensional predictor, μ_τ is the intercept, $\epsilon_i, i = 1, \dots, n$, are random errors with their τ th quantile equal to 0 and $f_{\tau,j}$ is a univariate component function which might be nonlinear, linear or zero.

The quantile regression approach we propose in this chapter has the ability of separating function components into those with nonlinear effects, those with linear effects, and those irrelevant to responses, in the context of quantile regression. In a Bayesian context, this separation of components is a soft decision based on the posterior probabilities of the components

being selected as nonlinear, linear, or zero. As in Chapter 2, we establish a hierarchical Bayesian model by adopting the AL distribution for errors. Then, extending the Bayesian variable selection approach, we introduce two sets of indicator variables in the spike-and-slab priors which make the separation of components possible.

The remainder of the chapter proceeds as follows. In Section 3.1, we describe our hierarchical Bayesian model for quantile regression based on the additive model structure. We also discuss our prior choices and introduce the posterior sampling algorithm focusing on an efficient partially collapsed sampler. The details of the algorithm are explained in the Appendix. In Section 3.2, we present numerical illustrations including simulation studies and two real data examples. In Section 3.3, we conclude this chapter.

3.1 Hierarchical Bayesian Modeling

In this chapter, we still implement quantile regression by solving a minimization problem based on the check loss function. With model (3.0.1), the specific problem is estimating μ_τ and the sequence of functions $f_{\tau,j}, j = 1, \dots, p$, by minimizing the following objective function,

$$L(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^n \rho_\tau \left(y_i - \mu_\tau - \sum_{j=1}^p f_{\tau,j}(x_{ij}) \right), \quad (3.1.1)$$

where $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$, and $\rho_\tau(u) = u(\tau - I(u \leq 0))$. We assume $\epsilon_{\tau,i}, i = 1, \dots, n$, are i.i.d. random variables from an AL distribution with density

$$\pi(\epsilon_{\tau,i}) = \frac{\tau(1-\tau)}{\delta_0} \exp \left\{ -\frac{1}{\delta_0} \rho_\tau(\epsilon_{\tau,i}) \right\},$$

where δ_0 is the scale parameter (In this chapter, we use δ_0 instead of σ to prevent confusion of parameters). Then the conditional distribution of $\mathbf{y}$ is

in the form of

$$\pi(\mathbf{y}|\mathbf{x}) = \frac{\tau^n(1-\tau)^n}{\delta_0^n} \exp \left\{ -\frac{1}{\delta_0} \sum_{i=1}^n \rho_\tau \left(y_i - \mu_\tau - \sum_{j=1}^p f_{\tau,j}(x_{ij}) \right) \right\}. \quad (3.1.2)$$

Hence, maximizing the likelihood (3.1.2) is equivalent to minimizing (3.1.1), giving (3.1.1) a likelihood-based interpretation. By introducing the location-scale mixture representation of the AL distribution, (3.1.2) can be equivalently written as

$$y_i = \mu_\tau + \sum_{j=1}^p f_{\tau,j}(x_{ij}) + k_1 e_i + \sqrt{k_2 \delta_0} e_i z_i, \quad (3.1.3)$$

where $e_i \sim \exp(1/\delta_0)$ follows an exponential distribution with mean δ_0 , z_i follows the standard normal distribution and is independent of e_i , $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, and $k_2 = \frac{2}{\tau(1-\tau)}$. For ease of notation, we will omit τ in the expressions in the following.

We assume the distribution of x_j , $j = 1, \dots, p$, is supported on $[0, 1]$ and also impose the condition $E f_j(x_j) = 0$ for identifiability. To model each unknown function f_j flexibly, we use the truncated power splines to approximate the functions in this chapter. Let $t_0 = 0 < t_1 < \dots < t_k < 1 = t_{k+1}$ partition $[0, 1]$ into subintervals $[t_i, t_{i+1})$, $i = 0, \dots, k$ with k internal knots. Here we focus on equally spaced knots although more complicated data-driven choice can be considered. For a given degree q , $B_0(x), \dots, B_K(x)$ are used to denote $K + 1 = q + k$ truncated power spline basis $x, x^2, \dots, x^q, (x - t_1)^q I(x > t_1), \dots, (x - t_k)^q I(x > t_k)$. Because of the centering constraint $E f_j(x_j) = 0$, we use the centered basis $\{B_{jk}(x) = B_k(x) - \sum_{i=1}^n B_k(x_{ij})/n, k = 0, \dots, K\}$ with $K = q + k - 1$. We separate linear basis (a single linear function) and nonlinear basis to enable the identification of linear and nonlinear components. Thus, with splines

approximation, we have

$$y_i = \mu + \sum_{j=1}^p \alpha_j B_{j0}(x_{ij}) + \sum_{j=1}^p \sum_{k=1}^K \beta_{jk} B_{jk}(x_{ij}) + k_1 e_i + \sqrt{k_2 \delta_0} e_i z_i, \quad i = 1, \dots, n.$$

We view e_i , $i = 1, \dots, n$ as latent variables and denote $\mathbf{e} = (e_1, \dots, e_n)^T$. By defining $\mathbf{E} = k_2 \delta_0 \text{diag}(e_1, \dots, e_n)$, $\mathbf{B}_0 = (\mathbf{B}_{10}, \dots, \mathbf{B}_{p0})$ with $\mathbf{B}_{j0} = (B_{j0}(x_{1j}), \dots, B_{j0}(x_{nj}))^T$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$, $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jK})^T$ and

$$\mathbf{B}_j = \begin{pmatrix} B_{j1}(x_{1j}) & B_{j2}(x_{1j}) & \dots & B_{jK}(x_{1j}) \\ B_{j1}(x_{2j}) & B_{j2}(x_{2j}) & \dots & B_{jK}(x_{2j}) \\ \vdots & \vdots & \vdots & \vdots \\ B_{j1}(x_{nj}) & B_{j2}(x_{nj}) & \dots & B_{jK}(x_{nj}) \end{pmatrix},$$

the full conditional distribution of $\mathbf{y}$ can be expressed as,

$$\begin{aligned} & \pi(\mathbf{y} | \boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \mathbf{e}, \delta_0, \mathbf{x}, \mu) \\ & \propto \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{f} - k_1 \mathbf{e})^T \mathbf{E}^{-1} (\mathbf{y} - \mathbf{f} - k_1 \mathbf{e}) \right\} (\det[\mathbf{E}])^{-1/2}. \end{aligned} \quad (3.1.4)$$

where

$$\mathbf{f} = \mu \mathbf{1}_n + \mathbf{B}_0 \boldsymbol{\alpha} + \sum_{j=1}^p \mathbf{B}_j \boldsymbol{\beta}_j,$$

and $\mathbf{1}_n = (1, \dots, 1)^T$ is a vector of dimension n with all components 1.

We choose spike-and-slab priors for α_j and $\boldsymbol{\beta}_j$ following George and McCulloch (1993), Panagiotelis and Smith (2008), and many others to enable variable selection and linear component selection. We introduce indicator variables $\boldsymbol{\gamma}^{(\nu)} = (\gamma_1^{(\nu)}, \dots, \gamma_p^{(\nu)})^T$, such that $\nu_j = 0$ if and only if $\gamma_j^{(\nu)} = 0$, with $\nu = \alpha$ or β . In other words, f_j is regarded as a nonlinear function, a linear function or a zero function under the situation that $(\gamma_j^{(\beta)} = 1)$, $(\gamma_j^{(\alpha)} = 1, \gamma_j^{(\beta)} = 0)$, or $(\gamma_j^{(\alpha)} = 0, \gamma_j^{(\beta)} = 0)$, respectively.

For α , we choose an independent Gaussian prior on each component, which is also called *ridge prior* (Goldstein and Smith, 1974),

$$\pi(\alpha|\gamma^{(\alpha)}, \sigma) \sim N(\mathbf{0}_p, \Sigma_\alpha),$$

$$\Sigma_\alpha = \text{diag}(\gamma_1^{(\alpha)}\sigma_1^2, \gamma_2^{(\alpha)}\sigma_2^2, \dots, \gamma_p^{(\alpha)}\sigma_p^2),$$

where $\mathbf{0}_p$ is a zero vector of dimension p and $\sigma = (\sigma_1, \dots, \sigma_p)$.

For β_j , we take the conjugate prior usually undertaken in a Bayesian context $\pi(\beta_j) \propto \exp\{-\frac{1}{2\tau_j^2}\beta_j^T \Omega_j \beta_j\}$ when $\gamma_j^{(\beta)} = 1$, such as Smith and Kohn (1996) and Chib and Jeliazkov (2006). Here the (k, k') entry of Ω_j is $\int_0^1 B_{jk}''(x)B_{jk'}''(x)dx$ (B_{jk}'' is the second derivative of B_{jk}). Since $\beta_j = \mathbf{0}_K$ when $\gamma_j^{(\beta)} = 0$, the prior of β_j can be described as,

$$\pi(\beta_j|\gamma_j^{(\beta)}, \tau_j) \sim N(\mathbf{0}_K, \gamma_j^{(\beta)}\tau_j^2\Omega_j^{-1}).$$

We use the same prior as in Cripps et al. (2005) on $\gamma^{(\nu)}$, $\nu = \alpha, \beta$,

$$\pi(\gamma^{(\nu)}) = \frac{1}{p+1} \begin{pmatrix} p \\ q_{\gamma^{(\nu)}} \end{pmatrix}^{-1},$$

where $q_{\gamma^{(\nu)}}$ is the number of non-zero ν_j . Under this prior, equal weights are placed on $\gamma^{(\nu)}$ with different numbers of non-zero ν_j . Alternatively, we can use $\pi(\gamma^{(\nu)}) = r^{q_{\gamma^{(\nu)}}}(1-r)^{p-q_{\gamma^{(\nu)}}}$ with r a constant. When $r = 1/2$, we have the uniform prior on $\gamma^{(\nu)}$. We find the results are not sensitive to this choice of prior.

To summarize, we provide the diagram of hierarchical structure in Figure 3.1, and the Bayesian hierarchical formulation below,

Figure 3.1: The diagram of Bayesian hierarchical structure.

$$\begin{aligned}
 y|\alpha, \{\beta_j\}, e, \delta_0, \mathbf{x}, \mu &\sim N(\mathbf{f} + k_1 \mathbf{e}, \mathbf{E}), \\
 \alpha|\gamma^{(\alpha)}, \sigma &\sim N(\mathbf{0}_p, \Sigma_\alpha), \quad \gamma^{(\alpha)} \sim \pi(\gamma^{(\alpha)}), \\
 \beta_j|\gamma_j^{(\beta)}, \tau_j &\sim N(\mathbf{0}_K, \gamma_j^{(\beta)} \tau_j^2 \Omega_j^{-1}), \quad \gamma^{(\beta)} \sim \pi(\gamma^{(\beta)}), \\
 e_i &\stackrel{i.i.d.}{\sim} \exp(1/\delta_0), \quad \delta_0 \sim \pi(\delta_0), \\
 \sigma_j^2 &\sim \pi(\sigma_j^2), \quad \tau_j^2 \sim \pi(\tau_j^2), \quad \mu \sim \pi(\mu),
 \end{aligned}$$

where $\pi(\delta_0)$, $\pi(\sigma_j^2)$ and $\pi(\tau_j^2)$ represent the hyperpriors of δ_0 , σ_j^2 and τ_j^2 . They are set to be $IG(a_1, a_2)$, where IG denotes the inverse Gamma distribution, and a_1 and a_2 are set to be 0.5 in all our numerical experiments as an uninformative choice. Sensitivity analysis reveals that our results are not sensitive to these choices. Finally, we use an uninformative prior on μ , $\pi(\mu) \propto 1$.

We use the Metropolis-within-Gibbs algorithm to sample from the posterior distributions. To improve mixing of the Markov chains, we integrate out ν_j , $\nu = \alpha, \beta$ in some of the sampling steps, resulting in a partially

collapsed sampler (van Dyk and Park, 2008). The readers are referred to the Appendix for details.

3.2 Numerical Illustrations

We demonstrate the performance of our proposed quantile regression approach (denoted by BQPLAM) in terms of its estimation accuracy and model selection accuracy. For comparison, we consider the following 4 additional methods.

Method 1:

The hierarchical Bayesian model in section 2 can be easily revised to deal with partially linear additive mean regression model. Consider the following model,

$$y_i = \mu + \sum_{j=1}^p f_j(x_{ij}) + \epsilon_i, \quad i = 1, \dots, n,$$

where $\epsilon_i, i = 1, \dots, n$ are i.i.d. normally distributed with mean zero and variance δ_0^2 . The Bayesian hierarchical formulation will then be as follows,

$$\begin{aligned} \mathbf{y} | \boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \delta_0^2, \mathbf{x}, \mu &\sim N(\mathbf{f}, \delta_0^2 \mathbf{I}_{n \times n}), \\ \boldsymbol{\alpha} | \boldsymbol{\gamma}^{(\alpha)}, \boldsymbol{\sigma} &\sim N(\mathbf{0}_p, \boldsymbol{\Sigma}_\alpha), \quad \boldsymbol{\gamma}^{(\alpha)} \sim \pi(\boldsymbol{\gamma}^{(\alpha)}), \\ \boldsymbol{\beta}_j, \tau_j &\sim N(\mathbf{0}_K, \gamma_j^{(\beta)} \tau_j^2 \boldsymbol{\Omega}_j^{-1}), \quad \boldsymbol{\gamma}^{(\beta)} \sim \pi(\boldsymbol{\gamma}^{(\beta)}), \\ \delta_0^2 &\sim \pi(\delta_0^2), \quad \tau_j^2 \sim \pi(\tau_j^2), \quad \mu \sim \pi(\mu), \end{aligned}$$

where $\mathbf{I}_{n \times n}$ represents the $n \times n$ identity matrix. We denote this mean regression method to be BPLAM.

The other three are quantile regression approaches.

Method 2:

We consider a Bayesian additive model that combines nonlinear and linear terms together. This only performs variable selection but not linear component identification. To be specific, the Bayesian additive regression is based on model (3.1.4), with

$$\mathbf{f} = \mu \mathbf{1}_n + \sum_{j=1}^p \mathbf{B}_j \boldsymbol{\beta}_j,$$

where $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jK})^T$ and

$$\mathbf{B}_j = \begin{pmatrix} B_{j0}(x_{1j}) & B_{j1}(x_{1j}) & \dots & B_{jK}(x_{1j}) \\ B_{j0}(x_{2j}) & B_{j1}(x_{2j}) & \dots & B_{jK}(x_{2j}) \\ \vdots & \vdots & \vdots & \vdots \\ B_{j0}(x_{nj}) & B_{j1}(x_{nj}) & \dots & B_{jK}(x_{nj}) \end{pmatrix}.$$

With similar prior distributions and hyper-parameters as those described in Section 2, we obtain the following Bayesian hierarchical formulation,

$$\begin{aligned} \mathbf{y} | \boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \mathbf{e}, \delta_0, \mathbf{x}, \mu &\sim N(\mathbf{f} + k_1 \mathbf{e}, \mathbf{E}), \\ \boldsymbol{\beta}_j | \gamma_j^{(\boldsymbol{\beta})}, \tau_j &\sim N(\mathbf{0}_{K+1}, \gamma_j^{(\boldsymbol{\beta})} \tau_j^2 \boldsymbol{\Omega}_j^{-1}), \quad \gamma^{(\boldsymbol{\beta})} \sim \pi(\gamma^{(\boldsymbol{\beta})}), \\ e_i &\stackrel{i.i.d.}{\sim} \exp(1/\delta_0), \quad \delta_0 \sim \pi(\delta_0), \\ \tau_j^2 &\sim \pi(\tau_j^2), \quad \mu \sim \pi(\mu). \end{aligned}$$

This method is denoted by BQAM_V.

Method 3:

We consider Bayesian quantile linear regression based on model (3.1.4), with

$$\mathbf{f} = \mu \mathbf{1}_n + \mathbf{B}_0 \boldsymbol{\alpha},$$

in which all components are assumed to be linear. The Bayesian hierarchical structure is as follows,

$$\begin{aligned} \mathbf{y}|\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \mathbf{e}, \delta_0, \mathbf{x}, \mu &\sim N(\mathbf{f} + k_1\mathbf{e}, \mathbf{E}), \\ \boldsymbol{\alpha}|\boldsymbol{\gamma}^{(\alpha)}, \boldsymbol{\sigma} &\sim N(\mathbf{0}_p, \boldsymbol{\Sigma}_\alpha), \quad \boldsymbol{\gamma}^{(\alpha)} \sim \pi(\boldsymbol{\gamma}^{(\alpha)}), \\ e_i &\stackrel{i.i.d.}{\sim} \exp(1/\delta_0), \quad \delta_0 \sim \pi(\delta_0), \\ \sigma_j^2 &\sim \pi(\sigma_j^2), \quad \mu \sim \pi(\mu). \end{aligned}$$

The indicator variables $\boldsymbol{\gamma}^{(\alpha)}$ enable component selection. This method is denoted by BQLM_V.

Method 4:

The third quantile regression method for comparison is still based on the model (3.1.4), except that we fix all indicator variables $\gamma_j^{(\alpha)}$, $j = 1, \dots, p$, and $\gamma_j^{(\beta)}$, $j = 1, \dots, p$, to be 1. This method can only estimate component functions but cannot select variables. We denote it as BQAM_{NV}.

3.2.1 Simulations

We generate $n = 100$ observations $(\mathbf{x}_i, y_i)$, $i = 1, \dots, 100$, from the following heteroscedastic additive model

$$y_i = \sum_{j=1}^p f_j(x_{ij}) + (0.5 + x_{i2})\epsilon_i, \quad (3.2.1)$$

with $f_1(x) = \sin(2\pi x)/(2 - \sin(2\pi x))$, $f_2(x) = 5x(1-x)$, $f_3(x) = 2x$, $f_4(x) = x$, $f_5(x) = -x$ and $p = 10$. Thus, the first 2 components are nonlinear components, followed by 3 linear components and 5 zero components. The covariates x_{ij} are generated from the standard normal distribution with

correlations given by $Cov(x_{ij_1}, x_{ij_2}) = (1/2)^{|j_1 - j_2|}$, and then transformed to be marginally uniform on $[0, 1]$ by applying the cdf of the standard normal distribution. We consider two distributions of ϵ_i , a normal distribution with mean 0 and standard deviation 0.5, and a Student's t distribution with scale parameter 1/3 and degrees of freedom 2. We fit the mean regression model and the four quantile regression models at five different quantile levels $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. For each scenario, 100 data sets are generated and fitted. For each replicate, the MCMC algorithm is run for 20000 iterations with a burn-in of 10000. As stated in Chapter 2, we diagnose convergence by visually analyzing trace plots of parameters with two different sets of initial values for each setting. All the trace plots suggest that the constructed chains converge and mix quickly.

The performance was evaluated by the integrated squared error (ISE), for each component function, which is approximated over an equally spaced 1000 points $(t_1, \dots, t_T)$, $T = 1000$ on $[0, 1]$ by,

$$\widehat{ISE} = \frac{1}{T} \sum_{i=1}^T (\hat{f}_j(t_i) - f_j(t_i))^2, \quad (3.2.2)$$

where $f_j(t_i)$ is the true value of function f_j at t_i , and $\hat{f}_j(t_i)$ is the posterior mean of f_j at t_i based on the 10000 iterations after burn-in. Tables 3.1 and 3.2 summarize the average and standard deviation of $\sqrt{\widehat{ISE}}$ over 100 replicates for the first six components and for the regression function $f = \sum_{j=1}^{10} f_j$. From the results, we can see that BQPLAM is obviously more efficient than BQAM_V and BQAM_{NV} for the parametric components due to the separated linear basis and its associated indicator variables. BQLM_V performs poorly as expected, since it cannot capture nonlinear effects of the components. BQPLAM and BQAM_V outperform BQAM_{NV} for zero components (note that we only present results of f_6 among the five zero

Table 3.1: Summary of $\sqrt{\widehat{ISE}}$ for each univariate function and $f = \sum_{j=1}^{10} f_j$ over 100 replicates, when the distribution of error is normal and $p = 10$. Standard errors based on simulations are shown as subscripts.

		f_1	f_2	f_3	f_4	f_5	f_6	f
	BPLAM	0.165 _{0.049}	0.109 _{0.043}	0.069 _{0.051}	0.067 _{0.057}	0.061 _{0.061}	0.016 _{0.032}	0.219 _{0.047}
$\tau = 0.5$	BQPLAM	0.162 _{0.048}	0.112 _{0.049}	0.068 _{0.050}	0.073 _{0.076}	0.068 _{0.066}	0.010 _{0.029}	0.222 _{0.048}
	BQAM _v	0.156 _{0.046}	0.161 _{0.055}	0.156 _{0.054}	0.190 _{0.076}	0.176 _{0.080}	0.006 _{0.011}	0.224 _{0.072}
	BQLM _v	0.320 _{0.026}	0.364 _{0.015}	0.075 _{0.064}	0.112 _{0.091}	0.108 _{0.092}	0.015 _{0.034}	0.515 _{0.041}
	BQAM _{nv}	0.157 _{0.053}	0.127 _{0.051}	0.101 _{0.058}	0.104 _{0.059}	0.123 _{0.071}	0.114 _{0.073}	0.320 _{0.114}
$\tau = 0.1$	BQPLAM	0.195 _{0.066}	0.184 _{0.074}	0.105 _{0.068}	0.106 _{0.091}	0.110 _{0.092}	0.023 _{0.059}	0.287 _{0.084}
	BQAM _v	0.199 _{0.074}	0.222 _{0.073}	0.189 _{0.063}	0.233 _{0.064}	0.236 _{0.060}	0.007 _{0.016}	0.310 _{0.096}
	BQLM _v	0.359 _{0.063}	0.377 _{0.036}	0.107 _{0.093}	0.133 _{0.114}	0.147 _{0.099}	0.046 _{0.083}	0.526 _{0.053}
	BQAM _{nv}	0.148 _{0.045}	0.165 _{0.060}	0.158 _{0.062}	0.150 _{0.052}	0.146 _{0.054}	0.151 _{0.059}	0.396 _{0.139}
$\tau = 0.3$	BQPLAM	0.170 _{0.050}	0.136 _{0.054}	0.070 _{0.048}	0.072 _{0.066}	0.072 _{0.065}	0.014 _{0.039}	0.233 _{0.055}
	BQAM _v	0.164 _{0.050}	0.167 _{0.060}	0.157 _{0.050}	0.233 _{0.055}	0.221 _{0.061}	0.008 _{0.013}	0.262 _{0.089}
	BQLM _v	0.335 _{0.044}	0.362 _{0.048}	0.092 _{0.072}	0.134 _{0.098}	0.131 _{0.101}	0.020 _{0.046}	0.522 _{0.046}
	BQAM _{nv}	0.142 _{0.042}	0.158 _{0.064}	0.160 _{0.068}	0.145 _{0.053}	0.146 _{0.053}	0.147 _{0.076}	0.383 _{0.136}
$\tau = 0.7$	BQPLAM	0.166 _{0.045}	0.132 _{0.055}	0.070 _{0.052}	0.084 _{0.076}	0.073 _{0.71}	0.010 _{0.030}	0.239 _{0.064}
	BQAM _v	0.160 _{0.049}	0.179 _{0.065}	0.156 _{0.058}	0.240 _{0.055}	0.223 _{0.062}	0.007 _{0.010}	0.278 _{0.096}
	BQLM _v	0.359 _{0.063}	0.375 _{0.033}	0.077 _{0.068}	0.114 _{0.086}	0.114 _{0.090}	0.014 _{0.029}	0.517 _{0.043}
	BQAM _{nv}	0.150 _{0.056}	0.166 _{0.045}	0.153 _{0.056}	0.145 _{0.056}	0.144 _{0.049}	0.155 _{0.055}	0.379 _{0.126}
$\tau = 0.9$	BQPLAM	0.196 _{0.053}	0.204 _{0.078}	0.101 _{0.066}	0.115 _{0.101}	0.117 _{0.093}	0.026 _{0.058}	0.311 _{0.092}
	BQAM _v	0.181 _{0.055}	0.226 _{0.081}	0.179 _{0.069}	0.246 _{0.056}	0.234 _{0.057}	0.003 _{0.004}	0.357 _{0.108}
	BQLM _v	0.316 _{0.025}	0.405 _{0.075}	0.101 _{0.081}	0.119 _{0.092}	0.121 _{0.095}	0.028 _{0.064}	0.534 _{0.063}
	BQAM _{nv}	0.159 _{0.052}	0.181 _{0.060}	0.161 _{0.057}	0.155 _{0.059}	0.160 _{0.051}	0.164 _{0.059}	0.424 _{0.141}

components). The results show that the two sets of indicators can help reduce errors besides selecting components. The results of mean regression are similar to those of median regression when the errors follow a normal distribution. However, the advantage of median regression is more obvious when the errors follow a Student's t distribution.

To measure the prediction accuracy of our method, we generate $n' = 100,000$ independent test samples from the same generating model (3.2.1) as the training samples and present the test errors in Tables 3.3 and 3.4. For mean and median regression, we consider two error measures including the root mean squared errors (RMSE) and absolute deviation errors (AD).

Table 3.2: Summary of $\sqrt{\widehat{ISE}}$ for each univariate function and $f = \sum_{j=1}^{10} f_j$ over 100 replicates, when the distribution of error is Student's t and $p = 10$.

		f_1	f_2	f_3	f_4	f_5	f_6	f
	BPLAM	0.278 _{0.081}	0.232 _{0.104}	0.129 _{0.106}	0.134 _{0.095}	0.135 _{0.084}	0.032 _{0.043}	0.411 _{0.161}
$\tau = 0.5$	BQPLAM	0.176 _{0.055}	0.126 _{0.067}	0.072 _{0.050}	0.088 _{0.089}	0.087 _{0.089}	0.007 _{0.023}	0.244 _{0.073}
	BQAM _v	0.145 _{0.051}	0.154 _{0.068}	0.137 _{0.058}	0.193 _{0.074}	0.186 _{0.069}	0.032 _{0.053}	0.269 _{0.098}
	BQLM _v	0.319 _{0.029}	0.361 _{0.029}	0.075 _{0.064}	0.136 _{0.098}	0.121 _{0.099}	0.015 _{0.041}	0.517 _{0.041}
	BQAM _{nv}	0.160 _{0.057}	0.161 _{0.060}	0.158 _{0.065}	0.165 _{0.062}	0.161 _{0.058}	0.147 _{0.059}	0.439 _{0.160}
$\tau = 0.1$	BQPLAM	0.295 _{0.104}	0.274 _{0.129}	0.157 _{0.129}	0.176 _{0.099}	0.165 _{0.098}	0.027 _{0.059}	0.405 _{0.144}
	BQAM _v	0.230 _{0.089}	0.283 _{0.157}	0.244 _{0.155}	0.238 _{0.105}	0.241 _{0.124}	0.073 _{0.176}	0.492 _{0.301}
	BQLM _v	0.373 _{0.063}	0.370 _{0.027}	0.144 _{0.123}	0.191 _{0.105}	0.186 _{0.096}	0.048 _{0.088}	0.529 _{0.055}
	BQAM _{nv}	0.259 _{0.239}	0.251 _{0.178}	0.258 _{0.296}	0.261 _{0.244}	0.247 _{0.265}	0.231 _{0.235}	0.610 _{0.392}
$\tau = 0.3$	BQPLAM	0.201 _{0.080}	0.142 _{0.065}	0.076 _{0.064}	0.097 _{0.092}	0.093 _{0.086}	0.008 _{0.029}	0.268 _{0.087}
	BQAM _v	0.153 _{0.062}	0.166 _{0.072}	0.154 _{0.061}	0.193 _{0.076}	0.186 _{0.071}	0.037 _{0.058}	0.274 _{0.108}
	BQLM _v	0.339 _{0.049}	0.359 _{0.022}	0.098 _{0.077}	0.136 _{0.100}	0.120 _{0.103}	0.020 _{0.050}	0.523 _{0.041}
	BQAM _{nv}	0.183 _{0.082}	0.185 _{0.071}	0.178 _{0.081}	0.185 _{0.079}	0.170 _{0.072}	0.164 _{0.065}	0.464 _{0.173}
$\tau = 0.7$	BQPLAM	0.182 _{0.057}	0.146 _{0.085}	0.078 _{0.061}	0.098 _{0.094}	0.096 _{0.094}	0.006 _{0.021}	0.266 _{0.089}
	BQAM _v	0.163 _{0.054}	0.167 _{0.076}	0.154 _{0.073}	0.205 _{0.069}	0.195 _{0.067}	0.032 _{0.050}	0.308 _{0.129}
	BQLM _v	0.317 _{0.028}	0.369 _{0.027}	0.083 _{0.077}	0.130 _{0.093}	0.130 _{0.099}	0.013 _{0.037}	0.519 _{0.048}
	BQAM _{nv}	0.191 _{0.070}	0.205 _{0.110}	0.189 _{0.081}	0.175 _{0.080}	0.168 _{0.078}	0.178 _{0.076}	0.505 _{0.020}
$\tau = 0.9$	BQPLAM	0.279 _{0.100}	0.293 _{0.134}	0.146 _{0.112}	0.160 _{0.109}	0.156 _{0.116}	0.031 _{0.080}	0.408 _{0.155}
	BQAM _v	0.254 _{0.130}	0.264 _{0.134}	0.272 _{0.174}	0.247 _{0.091}	0.235 _{0.072}	0.083 _{0.131}	0.575 _{0.362}
	BQLM _v	0.345 _{0.054}	0.396 _{0.084}	0.130 _{0.107}	0.172 _{0.100}	0.175 _{0.105}	0.048 _{0.085}	0.576 _{0.111}
	BQAM _{nv}	0.253 _{0.130}	0.258 _{0.139}	0.259 _{0.165}	0.268 _{0.189}	0.220 _{0.097}	0.217 _{0.123}	0.679 _{0.387}

For quantile regressions at quantiles $\{0.1, 0.3, 0.7, 0.9\}$, the prediction error refers to the average check loss (ACL). The errors are defined by,

$$RMSE = \sqrt{\frac{1}{n'} \sum_{i=1}^{n'} (\hat{y}_i - y_i)^2},$$

$$AD = \frac{1}{n'} \sum_{i=1}^{n'} |\hat{y}_i - y_i|$$

and

$$ACL = \frac{1}{n'} \sum_{i=1}^{n'} \rho_\tau(\hat{y}_i - y_i),$$

where y_i , $i = 1, \dots, n'$, are the responses of the testing data and $\hat{y}_i$, $i = 1, \dots, n'$, are the predicted values estimated with

$$\hat{y}_i = \hat{\mu} + \sum_{j=1}^p \hat{f}_j(x_{ij}),$$

where $\hat{\mu}$ and $\hat{f}_j(x_{ij})$ are posterior mean based on the 10000 sampled values after burn-in. Our method BQPLAM results in smaller test errors and outperforms the other three quantile regression methods at all quantile levels. The median regression performs similarly as mean regression when the errors are Gaussian, while it outperforms mean regression when errors are Student's t as expected.

The results of zero and linear component selection are based on the estimation of posterior distribution of $\gamma^{(\alpha)}$ and $\gamma^{(\beta)}$. To be specific, we compare the estimated values of the posterior probabilities $\pi(\gamma_j^{(\beta)} = 1 | \mathbf{y}, \mathbf{x})$, $\pi(\gamma_j^{(\alpha)} = 1, \gamma_j^{(\beta)} = 0 | \mathbf{y}, \mathbf{x})$ and $\pi(\gamma_j^{(\alpha)} = 0, \gamma_j^{(\beta)} = 0 | \mathbf{y}, \mathbf{x})$. The component will be identified as nonlinear, linear or zero according to which of the three probabilities is the largest (this rule is used only for illustration and Bayesians usually do not perform this step of making hard decision). In Tables 3.5, 3.6, and 3.7, we report the number of nonzero components

selected, number of nonzero components selected that are nonzero in the true model, number of linear components selected, and the number of linear components selected that are linear in the true model for BQPLAM and BPLAM. We report the number of variables selected and the number of variables selected that are truly nonzero for BQAM_v and BLAM_v, since these two methods can only select nonzero components. Figures 3.2 and 3.3 display estimates of the three posterior probabilities for each component based on one randomly selected replicate among the 100. The black, dark grey, and light grey areas represent the percentage of times the component is selected as nonlinear, linear and zero, respectively. It is seen from the figures that BQPLAM can detect linear and nonlinear components quite accurately.

Finally, we increase the dimension in the example (by adding zero components) to demonstrate the performance of our proposed method in higher dimensions. We consider the dimension $p = 50$ and 100 . To save space and time, we only consider BQPLAM at $\tau = 0.5$ and BPLAM. Test errors and variable selection results are presented in Tables 3.8 and 3.9. The estimates of three posterior probabilities for each component (for one data set) are displayed in Figures 3.4 and 3.5. Even with large p , BQPLAM can still select nonlinear and linear components accurately. BPLAM performs poorly when the error is heavy tailed as expected.

Table 3.3: Summary of testing errors for mean estimators and median estimators over 100 replicates, when $p = 10$. “Normal” and “Student’s t ” indicate the distribution of ϵ_i .

		RMSE	AD
Normal	BPLAM	0.586 _{0.030}	0.457 _{0.026}
	BQPLAM	0.587 _{0.031}	0.459 _{0.026}
	BQAM _v	0.646 _{0.032}	0.505 _{0.026}
	BQLM _v	0.762 _{0.030}	0.604 _{0.025}
	BQAM _{nv}	0.696 _{0.251}	0.540 _{0.165}
Student’s t	BPLAM	1.224 _{0.079}	0.648 _{0.101}
	BQPLAM	1.160 _{0.032}	0.551 _{0.054}
	BQAM _v	1.350 _{0.022}	0.589 _{0.031}
	BQLM _v	1.354 _{0.015}	0.679 _{0.022}
	BQAM _{nv}	1.236 _{0.043}	0.666 _{0.058}

Table 3.4: Summary of testing errors for quantile estimators at different quantile levels over 100 replicates, when $p = 10$. “Normal” and “Student’s t ” indicate the distribution of ϵ_i .

		$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.7$	$\tau = 0.9$
Normal	BQPLAM	0.634 _{0.100}	0.311 _{0.041}	0.322 _{0.046}	0.660 _{0.128}
	BQAM _v	0.652 _{0.083}	0.338 _{0.045}	0.351 _{0.048}	0.669 _{0.096}
	BQLM _v	0.868 _{0.119}	0.415 _{0.058}	0.424 _{0.064}	0.846 _{0.113}
	BQAM _{nv}	0.684 _{0.084}	0.356 _{0.065}	0.355 _{0.048}	0.664 _{0.089}
Student’s t	BQPLAM	0.819 _{0.219}	0.360 _{0.061}	0.365 _{0.063}	0.843 _{0.182}
	BQAM _v	0.831 _{0.261}	0.393 _{0.063}	0.391 _{0.058}	0.853 _{0.228}
	BQLM _v	0.988 _{0.206}	0.464 _{0.060}	0.464 _{0.566}	0.967 _{0.146}
	BQAM _{nv}	0.835 _{0.514}	0.445 _{0.077}	0.456 _{0.075}	0.850 _{0.309}

Table 3.5: *Summary of the component selection results for mean estimators and median estimators over 100 replicates, when $p = 10$. “Normal” and “Student’s t ” indicate the distribution of ϵ_i .*

		Normal	Student’s t
BPLAM	#of Nonzero Variables	5.21 _{0.67}	4.69 _{1.89}
	#of Correct Nonzero Variables	4.91 _{0.32}	4.00 _{1.21}
	#of Linear Variables	3.19 _{0.82}	3.22 _{1.99}
	#of Correct Linear Variables	2.81 _{0.46}	2.16 _{1.03}
BQPLAM	#of Nonzero Variables	5.14 _{0.69}	4.69 _{1.01}
	#of Correct Nonzero Variables	4.88 _{0.39}	4.54 _{0.93}
	#of Linear Variables	2.90 _{0.94}	2.58 _{0.98}
	#of Correct Linear Variables	2.56 _{0.69}	2.35 _{0.92}
BQAM _v	#of Nonzero Variables	4.51 _{0.93}	5.66 _{2.23}
	#of Correct Nonzero Variables	4.43 _{0.87}	4.31 _{0.85}
BQLM _v	#of Nonzero Variables	4.39 _{1.18}	3.82 _{1.26}
	#of Correct Nonzero Variables	3.90 _{0.89}	3.49 _{0.97}

Table 3.6: *Summary of the component selection results for quantile estimators at different quantile levels over 100 replicates, when the distribution of errors is Gaussian and $p = 10$.*

		$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.7$	$\tau = 0.9$
BQPLAM	#of Nonzero Variables	5.00 _{1.00}	5.01 _{0.77}	4.92 _{0.73}	4.92 _{1.05}
	#of Correct Nonzero Variables	4.56 _{0.75}	4.77 _{0.52}	4.73 _{0.61}	4.53 _{0.79}
	#of Linear Variables	2.40 _{1.39}	2.67 _{1.05}	2.58 _{1.05}	2.45 _{1.39}
	#of Correct Linear Variables	1.97 _{1.04}	2.40 _{0.84}	2.35 _{0.88}	2.00 _{1.00}
BQAM _v	#of Nonzero Variables	4.17 _{0.91}	4.37 _{0.91}	4.28 _{0.81}	4.21 _{0.83}
	#of Correct Nonzero Variables	4.12 _{0.90}	4.24 _{0.89}	4.26 _{0.81}	4.17 _{0.83}
BQLM _v	#of Nonzero Variables	4.44 _{1.92}	3.88 _{1.91}	4.19 _{1.18}	4.88 _{1.54}
	#of Correct Nonzero Variables	3.45 _{1.07}	3.49 _{1.05}	3.75 _{0.90}	3.52 _{0.96}

Table 3.7: *Summary of the component selection results for quantile estimators at different quantile levels over 100 replicates, when the distribution of errors is Student's t and $p = 10$.*

		$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.7$	$\tau = 0.9$
BQPLAM	#of Nonzero Variables	4.15 _{1.53}	4.58 _{0.96}	4.57 _{0.94}	4.31 _{1.66}
	#of Correct Nonzero Variables	3.72 _{1.36}	4.51 _{0.93}	4.49 _{0.87}	3.76 _{1.07}
	#of Linear Variables	2.31 _{1.38}	2.49 _{1.11}	2.49 _{1.12}	2.63 _{1.96}
	#of Correct Linear Variables	1.76 _{1.03}	2.28 _{0.92}	2.29 _{1.10}	1.65 _{1.02}
BQAM _v	#of Nonzero Variables	4.20 _{2.07}	5.45 _{2.23}	5.18 _{2.07}	4.22 _{1.97}
	#of Correct Nonzero Variables	3.69 _{0.93}	4.21 _{0.82}	4.13 _{0.83}	3.66 _{0.85}
BQLM _v	#of Nonzero Variables	3.68 _{1.93}	3.78 _{1.56}	3.94 _{1.39}	4.07 _{2.08}
	#of Correct Nonzero Variables	2.81 _{1.26}	3.35 _{1.08}	3.58 _{1.05}	3.02 _{1.29}

Figure 3.2: Component selection results for one randomly selected replicate, when the distribution of errors is Gaussian and $p = 10$. From the left to the right, 10 bars in each subfigure represent function components 1 to 10. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively.

Figure 3.3: *Component selection results for one randomly selected replicate, when the distribution of errors is Student's t and $p = 10$. From the left to the right, 10 bars in each subfigure represent function components 1 to 10. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively.*

Table 3.8: Summary of testing errors for mean and median estimators over 100 replicates, when $p = 50$ and 100. “Normal” and “Student’s t ” indicate the distribution of ϵ_i .

		$p = 50$		$p = 100$	
		RMSE	AD	RMSE	AD
Normal	BPLAM	0.625 _{0.055}	0.483 _{0.046}	0.643 _{0.060}	0.506 _{0.050}
	BQPLAM	0.624 _{0.057}	0.482 _{0.048}	0.638 _{0.054}	0.497 _{0.045}
Student’s t	BPLAM	1.251 _{0.069}	0.685 _{0.088}	1.563 _{0.657}	0.709 _{0.108}
	BQPLAM	1.176 _{0.039}	0.569 _{0.065}	1.185 _{0.052}	0.588 _{0.081}

Table 3.9: Summary of the component selection results for mean and median estimators over 100 replicates, when $p = 50$ and 100. “Normal” and “Student’s t ” indicate the distribution of ϵ_i .

		$p = 50$		$p = 100$	
		Normal	Student’s t	Normal	Student’s t
BPLAM	#of Nonzero Variables	4.32 _{1.12}	2.73 _{1.45}	4.82 _{1.11}	2.42 _{1.73}
	#of Correct Nonzero Variables	4.19 _{0.99}	2.57 _{1.38}	4.10 _{1.11}	2.21 _{1.49}
	#of Linear Variables	2.38 _{1.30}	1.92 _{1.21}	2.26 _{1.04}	1.61 _{1.40}
	#of Correct Linear Variables	2.10 _{1.07}	1.45 _{0.93}	1.96 _{0.99}	1.31 _{1.32}
BQPLAM	#of Nonzero Variables	4.41 _{0.17}	4.05 _{1.17}	4.29 _{1.07}	3.72 _{1.41}
	#of Correct Nonzero Variables	4.30 _{1.05}	4.02 _{1.16}	4.20 _{1.04}	3.72 _{1.43}
	#of Linear Variables	2.24 _{1.26}	2.13 _{0.97}	2.16 _{1.07}	2.06 _{1.10}
	#of Correct Linear Variables	2.06 _{1.03}	1.91 _{1.05}	1.98 _{1.06}	1.81 _{1.01}

Figure 3.4: *Component selection results for one randomly selected replicate with $p = 50$. The upper panels are results when error distribution is Gaussian, and the lower panels are results when the error distribution is Student's t . From the left to the right, 50 bars in each subfigure represent function components 1 to 50. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively.*

Figure 3.5: *Component selection results for one randomly selected replicate with $p = 100$. The upper panels are results when error distribution is Gaussian, and the lower panels are results when the error distribution is Student's t . From the left to the right, 100 bars in each subfigure represent function components 1 to 100. The black, dark grey, and light grey areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively.*

3.2.2 Real Data Analysis

Now we illustrate the methodology with two data sets. In the following examples, we fix $\gamma^{(\beta)}$ associated with the discrete dummy variables to be zero all the time, which means that dummy variables can only be modeled as linear or zero based on the posterior estimate of $\gamma^{(\alpha)}$. All covariates are standardized by a linear transformation to lie in $[0, 1]$ before analysis.

3.2.2.1 GDP Growth Data

We apply our method to a GDP growth data. This data set includes data for 74 countries over two 20-year time intervals: 1960-1979 and 1980-1999, with sample size $n = 147$. This data set is used by Tan (2010) to uncover the interplay between geography, institutions, and fractionalization in economic development by employing a regression tree analysis. Here we use our method to detect the determinants of economic growth. The dependent variable is the difference between the logarithmic values of real per capita GDP for the start and end years of each 20-year time interval. We consider four categories of independent variables, a total of 13. The first covariate we consider is a variable measuring the quality of expropriation risk of government across the years 1984-1997, which is denoted by EXPROP8497. Tan (2010) also considers another variable ICRG8497 to measure the quality of institution. Since the correlation between the two variables is very high, at over 0.8, we drop the second one and only keep EXPROP8497. The second category of independent variables includes: proportion of a country's land area that experiences more than 5 frost days per month in winter (FROST5), percentage of a country's land area that is classified as a tropical eco-zone (ZTROPICS), and percentage of a country's land area within

100 km of an ice-free coast (LCR100KM). This category is used as proxy for climate. The third category measures degree of ethnic fractionalization, including ELF60, ETHNIC, LANG and RELIG. ELF60 measures the probability in 1960 that two randomly selected people from a given country will not belong to the same ethno-linguistic subgroup. ETHNIC combines racial and linguistic characteristics, LANG is based on data for shares of languages spoken as ‘mother tongues’, and RELIG describes differences in religion. The last category of independent variables includes several familiar neoclassical determinants: log net depreciation rate (MNGD), log investment share (MINV), log schooling (MSCH15), a dummy variable for the period 1960-1979 (DUM6079), and log initial per capita income (MGDP0).

In Figure 3.6, we show the variable selection results of BPLAM and BQPLAM. For mean regression, we identified 6 nonzero components, all of which are linear components. For quantile regression, we identified 4, 3, 4, 5 and 6 nonzero components, of which 1, 0, 0, 1, 2 components are nonlinear at 5 quantile levels respectively. The barplots show that some covariates only have effects at lower or upper quantiles. For example, MSCH15 and ELF60 only have effects at upper quantiles, and ETHNIC only affects lower quantiles. In Figures 3.7 and 3.8, we show the fitted curves of the estimated regression functions (posterior mean based on the 10000 iterations after burn-in) as one covariate varies while others are fixed at 0.5. For convergence diagnosis, we present trace plots of α (we only present α of those covariates detected to be nonzero components), δ_0 , and μ for the median regression in Figure 3.9 (we plot one sampled point in every ten to reduce the image size).

To summarize, besides those neoclassical determinants, which are fundamental determinants used in the production function (Mankiw et al.,

Figure 3.6: *Component selection results for the GDP growth data.*

Figure 3.7: The fitted regression functions for GDP growth data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed). The red solid lines are the fitted curves of mean regression.

Figure 3.8: *The fitted regression functions for GDP growth data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed). The red solid lines are the fitted curves of mean regression.*

Figure 3.9: Trace plots of selected α , δ_0 , and μ of median regression for GDP growth data.

1992), both institutions and ethnic fractionalization are found to have effects on economical development while geography is found to play no roles. These results are consistent with the results of Tan (2010), and lead to the same conclusion that more attention should be paid on the impact of quality of institutions and degree of ethnic fractionalization on the development of economics.

3.2.2.2 Housing Price Data

Another data set we use is a housing price data. This data set is used to investigate the relationship between house price and several features, including physical attributes of the residential houses, characteristics of the surrounding rural district and environmental nuisances. This application concerns transaction prices ($n = 2053$) for the residential houses sold during 1996 and 1997 in Brittany. We consider four physical characteristics of a house and its adjacent lot: age of the house (AGE), state of repair (REPAIR)(dummy), number of rooms (ROOMS), and lot size (LOT). We use another four indicators on the characteristics of the surrounding rural district: the population of the district (POP), average taxable family income (AVINC), proportion of vacant houses (VACANT), and a dummy variable indicating whether or not the surrounding district is located in the department of Ille et Vilaine (COUNTY). The environmental nuisances are expressed by two indicators, amount of nitrogen emissions from livestock farming per hectare of arable land in the rural district where the residential house is located (NITRO), and proportion of permanent grassland converted into cultivated grassland (TMEAD). This data set was used by Bontemps et al. (2008) to investigate the effects of agricultural pollution on house prices. Bontemps et al. (2008) pre-specifies all the explanatory

variables (AGE, REPAIR, ROOMS, LOT, COUNTY, VACANT, POP and AVINC) except the two environment indicators (TMEAD and NITRO) to be linear components in the semiparametric models, and incorporated the two environmental indicators in a nonparametric way.

We display the barplots in Figure 3.10 and the fitted regression functions in Figures 3.11 and 3.12. The results show that all the components are identified as nonzero components and 4 of them are nonlinear for mean regression. For quantile regression at the 5 quantile levels, we identified 9, 10, 10, 10 and 9 nonzero components, and 1, 1, 1, 0, 0 nonlinear components respectively. Examining the influence of the physical characteristics of houses on prices at different levels, we find that a larger number of rooms, a bigger lot size and the fact that a house has been repaired are factors contributing to an increase in the price of a house, while an older age has a negative impact on the price. Similarly, the characteristics of the surrounding districts show their effects on house prices that conform to our expectations. For example, the price of houses located in the districts of the most urbanized country of Brittany (Ille-et-Vilaine) is higher, and the price of residential houses located in districts with lower housing vacancy rates is higher. For the two environmental indicators, we reveal their negative effects on house prices. The effects of covariates on responses at different quantile levels show some degree of heterogeneity. For example, the effect of AGE is linear at quantile 0.9, but is nonlinear at lower quantiles, which shows that mid-low price markets have more fluctuated depreciation rates than high price markets over years. Another example is that the negative effect of NITRO is not obvious at quantile 0.1, but more obvious at upper quantiles. This phenomena shows that environmental factors impact high price markets more, which is also reasonable.

Figure 3.10: *Component selection results for the housing price data.*

3.3 Summary of the Chapter

In this chapter, we have proposed a Bayesian quantile regression method for partially linear additive models, which explicitly models components that have linear and nonlinear effects. As detailed in the Appendix, we designed an efficient MCMC algorithm for posterior inferences. With sim-

Figure 3.11: The fitted regression functions for housing price data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed). The red solid lines are the fitted curves of mean regression.

Figure 3.12: *The fitted regression functions for housing price data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1$ (dashed), $\tau = 0.3$ (dotted), $\tau = 0.5$ (solid), $\tau = 0.7$ (dot dashed), and $\tau = 0.9$ (long dashed). The red solid lines are the fitted curves of mean regression.*

ulation studies, we illustrated the empirical performances of our proposed quantile regression approach and demonstrated the efficacy of the two sets of indicator variables. The method can automatically determine the type of component effects. The performance of the proposed approach in our simulations is quite encouraging, even when p is large.

Note that the error distributions we use in the simulation examples are very different from the asymmetric Laplace distribution. Our simulation results show that our methods perform well even when the errors are not generated based on their assumed generating mechanism. Theoretically it has been recently demonstrated that Bayesian quantile regression based on the asymmetric Laplace distribution is consistent even when the error distribution is misspecified Sriram et al. (2013). Li and Lin (2010) have showed that Bayesian methods are not sensitive to this assumption. We will give a more detailed discussion about this problem in Chapter 4.

Chapter 4

Discussion and Future Research

The principal aim of this work is to implement Bayesian inference for semi-parametric quantile regression models. We have established hierarchical Bayesian models and designed efficient MCMC algorithms for posterior inference. Extensive simulation studies and real data examples demonstrated our approaches in term of both accuracy and stability. Our approaches showed superiority over mean regression and frequentist approaches. Specifically speaking, we illustrated the satisfactory performance of GP prior utilised for link function estimation of single index models in the nonparametric Bayesian framework. Other parameters including index vector, noise variance, and hyperparameters were also estimated accurately. For the BQ-PLAM approach, the classic stochastic search variable selection performed very well in both selecting non-zero components and linear components, and led to an efficient estimation of component functions for Bayesian additive models. The asymmetric Laplace distribution (AL distribution) was proved to provide a handy solution for Bayesian quantile regression. From the computational perspective, the usage of partially collapsed MCMC algorithms were demonstrated to be computational effective. It is a goal that

the methodologies contained herein would find usage in a wide variety of data analysis application.

It is worth noting that we mainly regard AL distribution as a tool for estimating the conditional quantile, much like in the frequentist approach. By assuming the errors follow AL distribution, the posterior can give spurious confidence if the underlying data come from a different model than AL distribution. This is like using a quasi-likelihood to replace the true likelihood in frequentist estimation. In particular, the error distribution may not be accurately estimated by our approach. As we demonstrated, both single-index models and partially linear additive models are accurately estimated. Finally we remind the readers that AL distribution can lead to incoherent inferences in the sense that quantile curves are permitted to intersect each other. This is because the inference for distinct quantiles would proceed separately/independently and there is nothing to prevent them from overlapping. Other approaches may be able to address this issue, but we choose to use asymmetric Laplace distribution for its simplicity and computational convenience.

One disadvantage of our proposed methods is the computational burden. For example, on an ordinary PC, fitting the BQSIM on a single generated data set under our simulation setup would take about 10-20 minutes. This is less of a problem for our real data but is quite a burden for simulations. The time will increase as the number of samples and dimension of predictors increase. For this reason, we can only deal with data sets with sample size and dimensionality in the order of a couple of hundreds. Due to the relatively slow computational speed which is a common problem that plagues MCMC algorithms, some approximation methods such as variational Bayes might be desirable. For inference of partially linear ad-

ditive models, another possibility is to parallelize the sampling of different components, which would be helpful if p is large.

Our proposed approaches can be extended to other models. For example, the quantile regression method for single-index models, which we introduced in Chapter 2, can be extended to deal with multiple-index models in quantile regression. Xia (2008) introduced a frequentist solution for multiple-index models, which can be the benchmark of our future extension. However, sampling the index matrix poses some serious challenges. Similarly, corresponding frameworks based on latent variable representation of regularized support vector machines(SVM's) of Polson and Scott (2011) present a possible extension from quantile regression for partially linear additive models in Chapter 3 to binary regression models. In a same spirit, varying coefficient models and semiparametric mixed models for longitudinal data (Yue and Rue, 2011; Fahrmeir and Lang, 2001) can also be considered.

Appendix

Appendix A: MCMC algorithm details of Bayesian single-index quantile regression models

The posterior distribution for all of the unknown parameters and latent variables is proportional to the joint distribution, given by

$$\begin{aligned} & \pi(\boldsymbol{\beta}, \boldsymbol{\eta}_n, \mathbf{e}_n, \sigma, \lambda, \gamma | \mathbf{y}) \\ \propto & \exp \left\{ -\frac{(\mathbf{y} - \boldsymbol{\eta}_n - k_1 \mathbf{e}_n)^T \mathbf{E}^{-1} (\mathbf{y} - \boldsymbol{\eta}_n - k_1 \mathbf{e}_n)}{2} \right\} \\ & \times (\det[\mathbf{E}])^{-1/2} \det[\mathbf{C}_n]^{-1/2} \exp \left\{ -\frac{\boldsymbol{\eta}_n^T \mathbf{C}_n^{-1} \boldsymbol{\eta}_n}{2} \right\} \\ & \times \prod_{j=1}^p \frac{\lambda}{2\sigma} e^{-\lambda|\beta_j|/\sigma} \prod_{i=1}^n \frac{1}{\sigma} \exp \left\{ -\frac{e_i}{\sigma} \right\} \\ & \times \left(\frac{1}{\sigma} \right)^{a_\sigma+1} \exp \left\{ -\frac{b_\sigma}{\sigma} \right\} \left(\frac{1}{\gamma} \right)^{a_\gamma+1} \exp \left\{ -\frac{b_\gamma}{\gamma} \right\} \lambda^{a_\lambda-1} \exp\{-b_\lambda \lambda\}. \end{aligned}$$

The Metropolis-within-Gibbs algorithm may be used to sample from the posterior distribution. Mathematically speaking, it is possible to integrate out $\boldsymbol{\eta}_n$ before sampling, and it is well-known that marginalization can improve the mixing of the chain (Liu, 2008). However, if $\boldsymbol{\eta}_n$ is not sampled,

then the full conditional distributions for e_i and σ are no longer well-known distributions which lead to extra difficulty in sampling. On the other hand, we note the full conditional distribution of β is

$$\pi(\beta|\eta_n, \lambda, \sigma, \mathbf{y}) \propto \det[\mathbf{C}_n]^{-1/2} \exp \left\{ -\frac{\eta_n^T \mathbf{C}_n^{-1} \eta_n}{2} \right\} \times \prod_{j=1}^p \frac{\lambda}{2\sigma} e^{-\lambda|\beta_j|/\sigma},$$

and thus the evaluation of the density involves the inverse of $\mathbf{C}_n$. Unfortunately, since $\mathbf{C}_n$ is a kernel matrix, in many simulations we found it is nearly singular. When η_n is integrated out, this singularity problem is avoided since we only have to compute the inverse of the matrix $\mathbf{C}_n + \mathbf{E}$, where $\mathbf{E}$ is a diagonal matrix (see (A.1) below).

The conditional posterior densities of all the parameters and variables, except for β and γ , are common distributions. The conditional distributions used in the sampling are presented below. η_n is marginalized out when considering the posterior conditional distribution of β and γ .

$$\begin{aligned} \pi(\beta|\mathbf{e}_n, \sigma, \lambda, \gamma, \mathbf{y}) &\propto \int \pi(\mathbf{y}|\mathbf{e}_n, \sigma, \eta_n) \pi(\eta_n|\beta, \gamma) d\eta_n \times \pi(\beta|\sigma, \lambda) \\ &\propto \exp \left\{ -\frac{(\mathbf{y} - k_1 \mathbf{e}_n)^T (\mathbf{E} + \mathbf{C}_n)^{-1} (\mathbf{y} - k_1 \mathbf{e}_n)}{2} \right\} \quad (\text{A.1}) \\ &\times (\det[\mathbf{C}_n + \mathbf{E}])^{-1/2} \times \prod_{j=1}^p e^{-\lambda|\beta_j|/\sigma}, \end{aligned}$$

$$\begin{aligned} \pi(\gamma|\beta, \mathbf{e}_n, \sigma, \lambda, \mathbf{y}) &= \pi(\gamma|\beta, \mathbf{e}_n, \sigma, \mathbf{y}) \\ &\propto \int \pi(\mathbf{y}|\mathbf{e}_n, \sigma, \eta_n) \pi(\eta_n|\beta, \gamma) d\eta_n \times \pi(\gamma) \\ &\propto \exp \left\{ -\frac{(\mathbf{y} - k_1 \mathbf{e}_n)^T (\mathbf{E} + \mathbf{C}_n)^{-1} (\mathbf{y} - k_1 \mathbf{e}_n)}{2} \right\} \quad (\text{A.2}) \\ &\times (\det[\mathbf{C}_n + \mathbf{E}])^{-1/2} \left(\frac{1}{\gamma} \right)^{a_\gamma+1} \exp \left\{ -\frac{b_\gamma}{\gamma} \right\}, \end{aligned}$$

$$\begin{aligned}
\pi(\boldsymbol{\eta}_n | \boldsymbol{\beta}, \mathbf{e}_n, \sigma, \lambda, \gamma, \mathbf{y}) &= \pi(\boldsymbol{\eta}_n | \boldsymbol{\beta}, \mathbf{e}_n, \sigma, \gamma, \mathbf{y}) \sim N(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n), \\
\boldsymbol{\mu}_n &= \mathbf{C}_n(\mathbf{C}_n + \mathbf{E})^{-1}(\mathbf{y} - k_1 \mathbf{e}_n), \\
\boldsymbol{\Sigma}_n &= \mathbf{C}_n(\mathbf{C}_n + \mathbf{E})^{-1} \mathbf{E},
\end{aligned} \tag{A.3}$$

$$\begin{aligned}
\pi(\sigma | \boldsymbol{\beta}, \boldsymbol{\eta}_n, \mathbf{e}_n, \lambda, \gamma, \mathbf{y}) &= \pi(\sigma | \boldsymbol{\eta}_n, \boldsymbol{\beta}_n, \mathbf{e}_n, \lambda, \mathbf{y}) \sim IG(\alpha_\sigma, \nu_\sigma), \\
\alpha_\sigma &= \frac{3n}{2} + p + a_\sigma, \\
\nu_\sigma &= \sum_{i=1}^n \left(\frac{(y_i - \eta(\mathbf{x}_i^T \boldsymbol{\beta}) - k_1 e_i)^2}{2k_2 e_i} + e_i \right) + \sum_{j=1}^p \lambda |\beta_j| + b_\sigma,
\end{aligned}$$

$$\pi(\lambda | \sigma, \boldsymbol{\beta}, \boldsymbol{\eta}_n, \mathbf{e}_n, \gamma, \mathbf{y}) = \pi(\lambda | \sigma, \boldsymbol{\beta}) \sim Ga(a_\gamma + p, b_\gamma + \sum_{j=1}^p |\beta_j| / \sigma^2).$$

The full conditional distribution for e_i is a generalized inverse Gaussian distribution (*GIG*),

$$\pi(e_i | \sigma, \boldsymbol{\beta}, \boldsymbol{\eta}_n, \lambda, \gamma, \mathbf{y}) = \pi(e_i | \sigma, \boldsymbol{\eta}_n, \mathbf{y}) \sim GIG \left(\frac{1}{2}, \sqrt{\frac{(y_i - \eta(\mathbf{x}_i^T \boldsymbol{\beta}))^2}{k_2 \sigma}}, \sqrt{\frac{k_1^2}{k_2 \sigma} + \frac{2}{\sigma}} \right),$$

where the probability density function of $GIG(\rho, m, n)$ is

$$\begin{aligned}
f(x | \rho, m, n) &= \frac{(n/m)^\rho}{2K_\rho(mn)} x^{\rho-1} \exp \left\{ -\frac{1}{2}(m^2 x^{-1} + n^2 x) \right\}, \\
x &> 0, -\infty < \rho < \infty, m \geq 0, n \geq 0,
\end{aligned}$$

and K_ρ is the modified Bessel function of the third kind (Barndorff-Nielsen and Shephard, 2001).

We use a superscript $(.)^{(t)}$ to denote the sampled values of different quantities at iteration t . The variables $\boldsymbol{\eta}_n^{(t)}, \lambda^{(t)}, \gamma^{(t)}, \mathbf{e}_n^{(t)}$ can be directly

generated in R based on the respective full conditional distributions. For $\boldsymbol{\beta}^{(t)}$, we use a Metropolis step with proposal distribution $N(\boldsymbol{\beta}^{(t-1)}, \sigma_\beta^2 I)$, and for $\gamma^{(t)}$, we propose the new value from $\log \gamma^{(t)} \sim N(\log \gamma^{(t-1)}, \sigma_\gamma^2)$. In practice, σ_β and σ_γ are manually tuned to ensure the acceptance rate to be within 10% $\sim$ 30%. This manual tuning is simplified by transforming all predictors and responses to have mean 0 and variance 1 before running the MCMC algorithm.

Our sampling strategy is “partially collapsed” in the sense of van Dyk and Park (2008), and in particular is similar to Sampler 7 in that paper. It is easy to see the validity of the constructed sampler (that is, it does not change the stationary distribution). More specifically, to obtain this partially collapsed sampler by modifying the Gibbs sampler, we first marginalize the full conditional distributions for $\boldsymbol{\beta}$ and γ to get $\pi(\boldsymbol{\beta}, \boldsymbol{\eta}_n | \mathbf{e}_n, \sigma, \lambda, \gamma, \mathbf{y})$ and $\pi(\gamma, \boldsymbol{\eta}_n | \boldsymbol{\beta}, \mathbf{e}_n, \sigma, \lambda, \mathbf{y})$, by moving $\boldsymbol{\eta}_n$ from being conditioned to being sampled. Since $\boldsymbol{\eta}_n$ is sampled again in (A.3) immediately following $\boldsymbol{\beta}$ and γ , the two intermediate $\boldsymbol{\eta}_n$ ’s are redundant and thus can be trimmed, resulting in (A.1) and (A.2) respectively. It was shown in Theorem 1 of van Dyk and Park (2008) that the marginalization step can only improve the autocorrelation of the chain. As shown in our simulation examples, this improvement is dramatic for our specific problem.

Appendix B: MCMC algorithm details of Bayesian quantile regression for partially linear quantile regression models

The joint distribution of all the variables is

$$\begin{aligned} & \pi(\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \mathbf{E}, \mu, \delta_0 | \mathbf{y}, \mathbf{x}) \\ & \propto \exp \left\{ -\frac{1}{2} \left(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{B}_0 \boldsymbol{\alpha} - \sum_{j=1}^p \mathbf{B}_j \boldsymbol{\beta}_j - k_1 \mathbf{e} \right)^T \mathbf{E}^{-1} \left(\mathbf{y} - \mu \mathbf{1}_n - \mathbf{B}_0 \boldsymbol{\alpha} - \sum_{j=1}^p \mathbf{B}_j \boldsymbol{\beta}_j - k_1 \mathbf{e} \right) \right\} \\ & \quad \times \det[\mathbf{E}]^{-1/2} \times \pi(\boldsymbol{\alpha}) \times \prod_{j=1}^p \pi(\boldsymbol{\beta}_j) \times \pi(\delta_0) \times \prod_{i=1}^n \pi(e_i) \times \pi(\mu), \end{aligned}$$

where $\pi(\boldsymbol{\beta}_j)$, $\pi(\boldsymbol{\alpha})$, $\pi(e_i)$, $\pi(\delta_0)$ and $\pi(\mu)$ are the prior distributions of $\boldsymbol{\beta}_j$, $\boldsymbol{\alpha}$, e_i , δ_0 , and μ respectively.

We use the Metropolis-within-Gibbs algorithm to sample from the posterior distribution. We integrate out α_j in step 3 and $\boldsymbol{\beta}_j$ in step 5 to improve mixing of the Markov chain. The posterior distribution of each variable is as follows ($\sim$ denotes all variables except the one to be sampled):

1. Sample $\pi(\alpha_j | \sim) = p(\alpha_j | \mathbf{y}^*, \mathbf{e}, \delta_0, \sigma_j^2, \gamma_j^{(\alpha)})$, $j = 1, \dots, p$, from the conditional distribution of α_j ,

$$\begin{aligned} \pi(\alpha_j | \mathbf{y}^*, \mathbf{e}, \delta_0, \sigma_j^2, \gamma_j^{(\alpha)} = 1) & \sim N(\mu_j, \xi_j^2), \\ \pi(\alpha_j | \mathbf{y}^*, \mathbf{e}, \delta_0, \sigma_j^2, \gamma_j^{(\alpha)} = 0) & = 0, \end{aligned}$$

where $\mathbf{y}^* = \mathbf{y} - \mu \mathbf{1}_n - \sum_{i \neq j}^p \alpha_i \mathbf{B}_{i0} - \sum_{i=1}^p \mathbf{B}_i \boldsymbol{\beta}_i - k_1 \mathbf{e}$, $\xi_j^2 = (\mathbf{B}_{j0}^T \mathbf{E}^{-1} \mathbf{B}_{j0} + \frac{1}{\sigma_j^2})^{-1}$ and $\mu_j = \xi_j^2 \mathbf{B}_{j0}^T \mathbf{E}^{-1} \mathbf{y}^*$.

2. Sample $\pi(\mu | \sim) = \pi(\mu | \mathbf{y}^*, \mathbf{e}, \delta_0)$, from the conditional distribution of

μ ,

$$\pi(\mu|\mathbf{y}^*, \mathbf{e}, \delta_0) \sim N(\mu_0, \xi_0^2),$$

where $\xi_0^2 = k_2 \delta_0 (\sum_{i=1}^n e_i^{-1})^{-1}$, $\mu_0 = \xi_0^2 \mathbf{1}_n^T \mathbf{E}^{-1} \mathbf{y}^*$, and $\mathbf{y}^* = \mathbf{y} - \sum_{i=1}^p \alpha_i \mathbf{B}_{i0} - \sum_{i=1}^p \mathbf{B}_i \beta_i - k_1 \mathbf{e}$.

3. Sample $\pi(\gamma_j^{(\alpha)} | \sim) = \pi(\gamma_j^{(\alpha)} | \mathbf{y}^*, \mathbf{e}, \delta_0, \sigma_j^2)$, $j = 1, \dots, p$, from its conditional posterior after integrating over α_j , $\pi(\gamma_j^{(\alpha)} = 1 | \mathbf{y}^*, \mathbf{e}, \delta_0, \sigma_j^2) = \frac{1}{1+h}$, with

$$h = h_1 \frac{\pi(\gamma_j^{(\alpha)} = 0 | \gamma_{i \neq j}^{(\alpha)})}{\pi(\gamma_j^{(\alpha)} = 1 | \gamma_{i \neq j}^{(\alpha)})},$$

$$h_1 = \exp \left\{ -\frac{1}{2} (\mathbf{B}_{j0}^T \mathbf{E}^{-1} \mathbf{y}^*)^2 \left(\mathbf{B}_{j0}^T \mathbf{E}^{-1} \mathbf{B}_{j0} + \frac{1}{\sigma_j^2} \right)^{-1} \right\}$$

$$\times (\sigma_j^2 \mathbf{B}_{j0}^T \mathbf{E}^{-1} \mathbf{B}_{j0} + 1)^{\frac{1}{2}},$$

where $\pi(\gamma_j^{(\alpha)} = 0 | \gamma_{i \neq j}^{(\alpha)}) = (p - q_{\gamma_0^{(\alpha)}})/(p + 1)$ and $p(\gamma_j^{(\alpha)} = 1 | \gamma_{i \neq j}^{(\alpha)}) = (1 + q_{\gamma_0^{(\alpha)}})/(p + 1)$ with $\gamma_0^{(\alpha)} = (\gamma_1^{(\alpha)}, \dots, \gamma_{j-1}^{(\alpha)}, 0, \gamma_{j+1}^{(\alpha)}, \dots, \gamma_p^{(\alpha)})^T$, and $\mathbf{y}^* = \mathbf{y} - \mu \mathbf{1}_n - \sum_{i \neq j}^p \alpha_i \mathbf{B}_{i0} - \sum_{i=1}^p \mathbf{B}_i \beta_i - k_1 \mathbf{e}$.

4. Sample $\pi(\beta_j | \sim) = \pi(\beta_j | \mathbf{y}^*, \mathbf{e}, \delta_0, \tau_j^2, \gamma_j^{(\beta)})$, $j = 1, \dots, p$,

$$\pi(\beta_j | \mathbf{y}^*, \mathbf{e}, \delta_0, \tau_j^2, \gamma_j^{(\beta)} = 1) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j),$$

$$\pi(\beta_j | \mathbf{y}^*, \mathbf{e}, \delta_0, \tau_j^2, \gamma_j^{(\beta)} = 0) = 0,$$

where $\mathbf{y}^* = \mathbf{y} - \mu \mathbf{1}_n - \sum_{i=1}^p \alpha_i \mathbf{B}_{i0} - \sum_{i \neq j}^p \mathbf{B}_i \beta_i - k_1 \mathbf{e}$, $\boldsymbol{\mu}_j = \boldsymbol{\Sigma}_j \mathbf{B}_j^T \mathbf{E}^{-1} \mathbf{y}^*$, $\boldsymbol{\Sigma}_j = (\mathbf{B}_j^T \mathbf{E}^{-1} \mathbf{B}_j + \frac{1}{\tau_j^2} \boldsymbol{\Omega}_j)^{-1}$.

5. Sample $\pi(\gamma_j^{(\beta)} | \sim) = p(\gamma_j^{(\beta)} | \mathbf{y}^*, \mathbf{e}, \delta_0, \tau_j^2)$, $j = 1, \dots, p$, from its conditional posterior after integrating over $\boldsymbol{\beta}_j$, $\pi(\gamma_j^{(\beta)} = 1 | \mathbf{y}^*, \mathbf{e}, \delta_0, \tau_j^2) = \frac{1}{1+h}$, with

$$h = h_1 \frac{\pi(\gamma_j^{(\beta)} = 0 | \gamma_{i \neq j}^{(\beta)})}{\pi(\gamma_j^{(\beta)} = 1 | \gamma_{i \neq j}^{(\beta)})},$$

$$h_1 = \exp \left\{ -\frac{1}{2} \left[\mathbf{y}^{*T} \mathbf{E}^{-1} \mathbf{B}_j \left(\mathbf{B}_j^T \mathbf{E}^{-1} \mathbf{B}_j + \frac{1}{\tau_j^2} \boldsymbol{\Omega}_j \right)^{-1} \mathbf{B}_j^T \mathbf{E}^{-1} \mathbf{y}^* \right] \right\}$$

$$\times \det \left[\frac{1}{\tau_j^2} \boldsymbol{\Omega}_j \right]^{-\frac{1}{2}} \times \det \left[\mathbf{B}_j^T \mathbf{E}^{-1} \mathbf{B}_j + \frac{1}{\tau_j^2} \boldsymbol{\Omega}_j \right]^{\frac{1}{2}},$$

where $\mathbf{y}^* = \mathbf{y} - \mu \mathbf{1}_n - \sum_{i=1}^p \alpha_i \mathbf{B}_{i0} - \sum_{i \neq j}^p \mathbf{B}_i \boldsymbol{\beta}_i - k_1 \mathbf{e}$, $\pi(\gamma_j^{(\beta)} = 0 | \gamma_{i \neq j}^{(\beta)}) = (p - q_{\gamma_0^{(\beta)}})/(p + 1)$ and $\pi(\gamma_j^{(\beta)} = 1 | \gamma_{i \neq j}^{(\beta)}) = (1 + q_{\gamma_0^{(\beta)}})/(p + 1)$ with $\gamma_0^{(\beta)} = (\gamma_1^{(\beta)}, \dots, \gamma_{j-1}^{(\beta)}, 0, \gamma_{j+1}^{(\beta)}, \dots, \gamma_p^{(\beta)})^T$.

6. Sample δ_0 ,

$$\delta_0 \sim IG(a_1 + 3n/2, \nu),$$

$$\nu = a_2 + \left[(2k_2 e_i)^{-1} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \alpha_j B_{j0}(x_{ij}) - \sum_{j=1}^p \sum_{k=1}^K \beta_{jk} B_{jk}(x_{ij}) - k_1 e_i \right)^2 + e_i \right].$$

7. Sample σ_j^2 , $j = 1, \dots, p$, and τ_j^2 , $j = 1, \dots, p$, from their conditional posterior distributions if $\gamma_j^{(\alpha)}, \gamma_j^{(\beta)} \neq 0$,

$$\sigma_j^2 \sim IG(a_1 + 1/2, a_2 + (\alpha_j^2/2)),$$

$$\tau_j^2 \sim IG(a_1 + K/2, a_2 + (\boldsymbol{\beta}_j^T \boldsymbol{\Omega}_j \boldsymbol{\beta}_j/2)).$$

Otherwise they are generated from their priors.

8. The full conditional distribution of e_i , $i = 1, \dots, n$ is a generalized

inverse Gaussian distribution (*GIG*),

$$\pi(e_i|\delta_0, \nu_i) \sim GIG \left(\frac{1}{2}, \sqrt{\frac{(y_i - \nu_i)^2}{k_2 \delta_0}}, \sqrt{\frac{k_1^2}{k_2 \delta_0} + \frac{2}{\delta_0}} \right),$$

$$\nu_i = y_i - \mu - \sum_{j=1}^p \alpha_j B_{j0}(x_{ij}) - \sum_{j=1}^p \sum_{k=1}^K \beta_{jk} B_{jk}(x_{ij}),$$

where the probability density function of $GIG(\rho, m, n)$ is

$$f(x|\rho, m, n) = \frac{(n/m)^\rho}{2K_\rho(mn)} x^{\rho-1} \exp \left\{ -\frac{1}{2}(m^2 x^{-1} + n^2 x) \right\},$$

$$x > 0, -\infty < \rho < \infty, m \geq 0, n \geq 0,$$

and K_ρ is the modified Bessel function of the third kind (Barndorff-Nielsen and Shephard, 2001).

References

- Abrevaya, J. “The effects of demographics and maternal behavior on the distribution of birth outcomes.” *Empirical Economics*, 26(1):247–257 (2001).
- Akaike, H. “Maximum likelihood identification of Gaussian autoregressive moving average models.” *Biometrika*, 60(2):255–265 (1973).
- Antoniadis, A., Grégoire, G., and McKeague, I. “Bayesian estimation in single-index models.” *Statistica Sinica*, 14(4):1147–1164 (2004).
- Bang, S. and Jhun, M. “Simultaneous estimation and factor selection in quantile regression via adaptive sup-norm regularization.” *Computational Statistics and Data Analysis*, 56(4):813–826 (2012).
- Barndorff-Nielsen, O. and Shephard, N. “Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241 (2001).
- Bellman, R., Bellman, R., and Bellman, R. *Adaptive Control Processes: A Guided Tour*. Princeton University Press (1966).

- Belloni, A. and Chernozhukov, V. “ L_1 -penalized quantile regression in high-dimensional sparse models.” *The Annals of Statistics*, 39(1):82–130 (2011).
- Bondell, H., Reich, B., and Wang, H. “Noncrossing quantile regression curve estimation.” *Biometrika*, 97(4):825–838 (2010).
- Bontemps, C., Simioni, M., and Surry, Y. “Semiparametric hedonic price models: assessing the effects of agricultural nonpoint source pollution.” *Journal of Applied Econometrics*, 23(6):825–842 (2008).
- Buchinsky, M. “Changes in the US wage structure 1963-1987: application of quantile regression.” *Econometrica: Journal of the Econometric Society*, 62(2):405–458 (1994).
- Cade, B. and Noon, B. “A gentle introduction to quantile regression for ecologists.” *Frontiers in Ecology and the Environment*, 1(8):412–420 (2003).
- Carvalho, C., Polson, N., and Scott, J. “The horseshoe estimator for sparse signals.” *Biometrika*, 97(2):465–480 (2010).
- Chen, C. and Wei, Y. “Computational issues for quantile regression.” *Sankhyā: The Indian Journal of Statistics*, 67(Part 2):399–417 (2005).
- Chen, C. and Yu, K. “Automatic Bayesian quantile regression curve fitting.” *Statistics and Computing*, 19(3):271–281 (2009).
- Chib, S. and Jeliazkov, I. “Inference in semiparametric dynamic models for binary longitudinal data.” *Journal of the American Statistical Association*, 101(474):685–700 (2006).

- Choi, T., Shi, J., and Wang, B. “A Gaussian process regression approach to a single-index model.” *Journal of Nonparametric Statistics*, 23(1):21–36 (2011).
- Cripps, E., Carter, C., and Kohn, R. “Variable selection and covariance selection in multivariate regression models.” *Handbook of Statistics*, 25(2):519–552 (2005).
- De Gooijer, J. and Zerom, D. “On additive conditional quantiles with high-dimensional covariates.” *Journal of the American Statistical Association*, 98(461):135–146 (2003).
- Denison, D., Mallick, B., and Smith, A. “Automatic Bayesian curve fitting.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):333–350 (2002).
- Duan, N. and Li, K. “A bias bound for least squares linear regression.” *Statistica Sinica*, 1:127–136 (1991).
- Dunson, D. and Taylor, J. “Approximate Bayesian inference for quantiles.” *Nonparametric Statistics*, 17(3):385–400 (2005).
- Fahrmeir, L. and Lang, S. “Bayesian inference for generalized additive mixed models based on Markov random field priors.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2):201–220 (2001).
- Fan, J. and Gijbels, I. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*, volume 66. Chapman and Hall/CRC (1996).

- Fan, J., Härdle, W., and Mammen, E. “Direct estimation of low-dimensional components in additive models.” *The Annals of Statistics*, 26(3):943–971 (1998).
- Fan, J. and Li, R. “Variable selection via nonconcave penalized likelihood and its oracle properties.” *Journal of the American Statistical Association*, 96(456):1348–1360 (2001).
- George, E. and Foster, D. “Calibration and empirical Bayes variable selection.” *Biometrika*, 87(4):731–747 (2000).
- George, E. and McCulloch, R. “Variable selection via Gibbs sampling.” *Journal of the American Statistical Association*, 88(423):881–889 (1993).
- Geraci, M. and Bottai, M. “Quantile regression for longitudinal data using the asymmetric Laplace distribution.” *Biostatistics*, 8(1):140–154 (2007).
- Goldstein, M. and Smith, A. “Ridge-type estimators for regression analysis.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 36(2):284–291 (1974).
- Gramacy, R. and Lian, H. “Gaussian process single-index models as emulators for computer experiments.” *Technometrics*, 54(1):30–41 (2012).
- Green, P. “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.” *Biometrika*, 82(4):711–732 (1995).
- Griffin, J. and Brown, P. “Inference with Normal–Gamma prior distributions in regression problems.” *Bayesian Analysis*, 5(1):171–188 (2010).
- Han, A. “Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator.” *Journal of Econometrics*, 35(2):303–316 (1987).

- Hans, C. “Bayesian lasso regression.” *Biometrika*, 96(4):835–845 (2009).
- Härdle, W., Hall, P., and Ichimura, H. “Optimal smoothing in single-index models.” *The Annals of Statistics*, 21(1):157–178 (1993).
- Hastie, T. and Tibshirani, R. “Generalized additive models.” *Statistical Science*, 1(3):297–310 (1986).
- He, X. and Hu, F. “Markov chain marginal bootstrap.” *Journal of the American Statistical Association*, 97(459):783–795 (2002).
- Horowitz, J. and Härdle, W. “Direct semiparametric estimation of single-index models with discrete covariates.” *Journal of the American Statistical Association*, 91(436):1632–1640 (1996).
- Horowitz, J. and Lee, S. “Nonparametric estimation of an additive quantile regression model.” *Journal of the American Statistical Association*, 100(472):1238–1249 (2005).
- Hristache, M., Juditsky, A., and Spokoiny, V. “Direct estimation of the index coefficient in a single-index model.” *The Annals of Statistics*, 29(3):593–623 (2001).
- Huang, J., Horowitz, J., and Wei, F. “Variable selection in nonparametric additive models.” *The Annals of Statistics*, 38(4):2282–2313 (2010).
- Ichimura, H. “Semiparametric least squares (SLS) and weighted SLS estimation of single-index models.” *Journal of Econometrics*, 58(1):71–120 (1993).
- Jagger, T. and Elsner, J. “Modeling tropical cyclone intensity with quantile regression.” *International Journal of Climatology*, 29(10):1351–1361 (2009).

- Jeffreys, H. *Theory of Probability*. Clarendon Press Oxford (1961).
- Ji, Y., Lin, N., and Zhang, B. “Model selection in binary and tobit quantile regression using the Gibbs sampler.” *Computational Statistics and Data Analysis*, 64(4):827–839 (2011).
- Klein, R. and Spady, R. “An efficient semiparametric estimator for binary response models.” *Econometrica: Journal of the Econometric Society*, 61(2):387–421 (1993).
- Kneib, T., Konrath, S., and Fahrmeir, L. “High dimensional structured additive regression models: Bayesian regularization, smoothing and predictive performance.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(1):51–70 (2010).
- Kocherginsky, M., He, X., and Mu, Y. “Practical confidence intervals for regression quantiles.” *Journal of Computational and Graphical Statistics*, 14(1):41–55 (2005).
- Koenker, R. “Confidence intervals for regression quantiles.” In Mandl, P. and Huskova, M. (eds.), *Asymptotic Statistics*, 349–359. Confidence intervals for regression quantiles (1994).
- . “Quantile regression for longitudinal data.” *Journal of Multivariate Analysis*, 91(1):74–89 (2004).
- . *Quantile Regression*, volume 38. Cambridge University Press (2005).
- Koenker, R. and Bassett, G. “Regression quantiles.” *Econometrica: Journal of the Econometric Society*, 46(1):33–50 (1978).

- Koenker, R., Ng, P., and Portnoy, S. “Quantile smoothing splines.” *Biometrika*, 81(4):673–680 (1994).
- Komunjer, I. “Quasi-maximum likelihood estimation for conditional quantiles.” *Journal of Econometrics*, 128(1):137–164 (2005).
- Korostelev, A., Simar, L., and Tsybakov, A. “Efficient estimation of monotone boundaries.” *The Annals of Statistics*, 23(2):476–489 (1995).
- Kottas, A. and Krnjajić, M. “Bayesian semiparametric modelling in quantile regression.” *Scandinavian Journal of Statistics*, 36(2):297–319 (2009).
- Kozumi, H. and Kobayashi, G. “Gibbs sampling methods for Bayesian quantile regression.” *Journal of Statistical Computation and Simulation*, 81(11):1565–1578 (2011).
- Lancaster, T. and Jun, S. “Bayesian quantile regression methods.” *Journal of Applied Econometrics*, 25(2):287–307 (2010).
- Lang, S. and Brezger, A. “Bayesian P-splines.” *Journal of Computational and Graphical Statistics*, 13(1):183–212 (2004).
- Li, K. “Sliced inverse regression for dimension reduction.” *Journal of the American Statistical Association*, 86(414):316–327 (1991).
- Li, Q. and Lin, N. “The Bayesian elastic net.” *Bayesian Analysis*, 5(1):151–170 (2010).
- Li, Y. and Zhu, J. “ L_1 -norm quantile regression.” *Journal of Computational and Graphical Statistics*, 17(1):163–185 (2008).

- Lian, H., Chen, X., and Yang, J. “Identification of partially linear structure in additive models with an application to gene expression prediction from sequences.” *Biometrics*, 68(2):437–445 (2012).
- Liang, H., Liu, X., Li, R., and Tsai, C. “Estimation and testing for partially linear single-index models.” *The Annals of Statistics*, 38(6):3811–3836 (2010).
- Liang, H., Thurston, S. W., Ruppert, D., Apanasovich, T., and Hauser, R. “Additive partial linear models with measurement errors.” *Biometrika*, 95(3):667–678 (2008).
- Lin, Y. and Zhang, H. “Component selection and smoothing in multivariate nonparametric regression.” *The Annals of Statistics*, 34(5):2272–2297 (2006).
- Linton, O. and Nielsen, J. “A kernel method of estimating structured nonparametric regression based on marginal integration.” *Biometrika*, 82(1):93–100 (1995).
- Liu, J. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer (2008).
- Luo, Z. and Wahba, G. “Hybrid adaptive splines.” *Journal of the American Statistical Association*, 92(437):107–116 (1997).
- MacKay, D. “Introduction to Gaussian processes.” *Nato Asi Series F. Computer and Systems Sciences*, 168:133–166 (1998).
- Mammen, E., Linton, O., and Nielsen, J. “The existence and asymptotic properties of a backfitting projection algorithm under weak conditions.” *The Annals of Statistics*, 27(5):1443–1490 (1999).

- Meligkotsidou, L., Vrontos, I., and Vrontos, S. “Quantile regression analysis of hedge fund strategies.” *Journal of Empirical Finance*, 16(2):264–279 (2009).
- Opsomer, J. “Asymptotic properties of backfitting estimators.” *Journal of Multivariate Analysis*, 73(2):166–179 (2000).
- Opsomer, J. and Ruppert, D. “A root-n consistent backfitting estimator for semiparametric additive modeling.” *Journal of Computational and Graphical Statistics*, 8(4):715–732 (1999).
- Panagiotelis, A. and Smith, M. “Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models.” *Journal of Econometrics*, 143(2):291–316 (2008).
- Park, T. and Casella, G. “The Bayesian lasso.” *Journal of the American Statistical Association*, 103(482):681–686 (2008).
- Parzen, M., Wei, L., and Ying, Z. “A resampling method based on pivotal estimating functions.” *Biometrika*, 81(2):341–350 (1994).
- Polson, N. and Scott, S. “Data augmentation for support vector machines.” *Bayesian Analysis*, 6(1):1–23 (2011).
- Ravikumar, P., Lafferty, J., Liu, H., and Wasserman, L. “Sparse additive models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):1009–1030 (2009).
- Reed, C., Dunson, D., and Yu, K. “Bayesian variable selection in quantile regression.” *Department of Mathematical Sciences, Brunel University. Technical Report* (2009).

- Reed, C. and Yu, K. “An efficient Gibbs sampler for Bayesian quantile regression.” *Department of Mathematical Sciences, Brunel University. Technical report* (2009).
- Schennach, M. “Bayesian exponentially tilted empirical likelihood.” *Biometrika*, 92(1):31–46 (2005).
- Schwarz, G. “Estimating the dimension of a model.” *The Annals of Statistics*, 6(2):461–464 (1978).
- Sherman, R. “The limiting distribution of the maximum rank correlation estimator.” *Econometrica: Journal of the Econometric Society*, 61(1):123–137 (1993).
- Shively, T., Kohn, R., and Wood, S. “Variable selection and function estimation in additive nonparametric regression using a data-based prior.” *Journal of the American Statistical Association*, 94(447):777–794 (1999).
- Smith, M. and Kohn, R. “Nonparametric regression using Bayesian variable selection.” *Journal of Econometrics*, 75(2):317–343 (1996).
- Sriram, K., Ramamoorthi, R., and Ghosh, P. “Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density.” *Bayesian Analysis*, 8(2):1–26 (2013).
- Stoker, T. “Consistent estimation of scaled coefficients.” *Econometrica: Journal of the Econometric Society*, 54(6):1461–1481 (1986).
- Stone, C. “The use of polynomial splines and their tensor products in multivariate function estimation.” *The Annals of Statistics*, 22(1):118–171 (1994).

- Tan, C. “No one true path: uncovering the interplay between geography, institutions, and fractionalization in economic development.” *Journal of Applied Econometrics*, 25(7):1100–1127 (2010).
- Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. “Bayesian nonparametric quantile regression using splines.” *Computational Statistics and Data Analysis*, 54(4):1138–1150 (2010).
- Tibshirani, R. “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 58(1):267–288 (1996).
- Tjøstheim, D. and Auestad, B. “Nonparametric identification of nonlinear time series: projections.” *Journal of the American Statistical Association*, 89(428):1398–1409 (1994).
- van Dyk, D. and Park, T. “Partially collapsed Gibbs samplers.” *Journal of the American Statistical Association*, 103(482):790–796 (2008).
- Wahba, G. *Spline Models for Observational Data*, volume 59. Society for Industrial Mathematics (1990).
- Wang, H. “Bayesian estimation and variable selection for single index models.” *Computational Statistics and Data Analysis*, 53(7):2617–2627 (2009).
- Wang, H., Li, G., and Jiang, G. “Robust regression shrinkage and consistent variable selection through the LAD-lasso.” *Journal of Business and Economic Statistics*, 25(3):347–355 (2007).
- Wang, J., Xue, L., Zhu, L., and Chong, Y. “Estimation for a partial-linear single-index model.” *The Annals of Statistics*, 38(1):246–274 (2010).

- Wang, L. and Yang, L. “Spline estimation of single-index models.” *Statistica Sinica*, 19(2):765–783 (2009).
- Wong, C. and Kohn, R. “A Bayesian approach to additive semiparametric regression.” *Journal of Econometrics*, 74(2):209–235 (1996).
- Wu, T., Yu, K., and Yu, Y. “Single-index quantile regression.” *Journal of Multivariate Analysis*, 101(7):1607–1621 (2010).
- Wu, Y. and Liu, Y. “Variable selection in quantile regression.” *Statistica Sinica*, 19(2):801–817 (2009).
- Xia, Y. “A multiple-index model and dimension reduction.” *Journal of the American Statistical Association*, 103(484):1631–1640 (2008).
- Xia, Y., Li, W., Tong, H., and Zhang, D. “A goodness-of-fit test for single-index models.” *Statistica Sinica*, 14(1):1–39 (2004).
- Xia, Y., Tong, H., Li, W., and Zhu, L. “An adaptive estimation of dimension reduction space.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):363–410 (2002).
- Xue, L. “Consistent variable selection in additive models.” *Statistica Sinica*, 19(3):1281–1296 (2009).
- Yu, K. and Lu, Z. “Local linear additive quantile regression.” *Scandinavian Journal of Statistics*, 31(3):333–346 (2004).
- Yu, K. and Moyeed, R. “Bayesian quantile regression.” *Statistics and Probability Letters*, 54(4):437–447 (2001).

- Yu, Y. and Ruppert, D. “Penalized spline estimation for partially linear single-index models.” *Journal of the American Statistical Association*, 97(460):1042–1054 (2002).
- Yue, Y. R. and Rue, H. “Bayesian inference for additive mixed quantile regression models.” *Computational Statistics and Data Analysis*, 55(1):84–96 (2011).
- Zellner, A. “On assessing prior distributions and Bayesian regression analysis with g-prior distributions.” In Goel, P. and Zellner, A. (eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. North-Holland/Elsevier (1986).
- Zhang, H., Cheng, G., and Liu, Y. “Linear or nonlinear? Automatic structure discovery for partially linear models.” *Journal of the American Statistical Association*, 106(495):1099–1112 (2011).
- Zhang, H., Wahba, G., Lin, Y., Voelker, M., Ferris, M., Klein, R., and Klein, B. “Variable selection and model building via likelihood basis pursuit.” *Journal of the American Statistical Association*, 99(467):659–672 (2004).