# Bayesian Quantile Regression Using Flexible Likelihood Functions


Aziz Awadhallah S Aljuaid

School of mathematics

University of Leeds

Submitted in accordance with the requirements for the degree of

*Doctor of Philosophy*

October 2017

# Declaration

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

A Aljuaid , J Gosling and C Taylor developed the statistical methods. A Aljuaid derived the proof and investigated research-related issues, implemented the proposed approaches using the R programming and showed and discussed the findings. All of this manuscript is written by A Aljuaid. J Gosling and C Taylor gave comments that improved this manuscript.

Some sections from Chapters 2 & 3 were submitted for publication, in the Journal of Statistical Computation and Simulation, as given below:
Aljuaid, A., Gosling, J.P. & Taylor, C.C. (2017). Simultaneous Box-Cox Quantile Regression. In revision for *the Journal of Statistical Computation and Simulation*.

# Acknowledgements

I would like to express sincere appreciation and thanks to my supervisors Dr John Paul Gosling and Prof Charles C Taylor for their support, encouragement and patience with me over the last three and a half years; and for the vast reserve of valuable knowledge on programming, statistical methods and research skills that they have provided me.

I would like to express a sense of gratitude and thanks to my family, especially my father Awadhallah, my mother Awaidheh and my wife Hind for their support and encouragement, and for keeping my spirits high during my study. I am very grateful to my children Tulin, Mohammed and Aous. I would also like to thank my friends for their support and encouragement. Many thanks to Taif University that offered me the scholarship to study for a PhD.

# Abstract

Quantile regression is a statistical method used to investigate the full conditional distribution of a response variable. As such, it provides more information than ordinary least squares regression. To develop a novel Bayesian method to fit quantile regression, we need to deal with a number of issues such as likelihood choice, prior specification, posterior computation and the crossing of quantile functions. The main interest in this research is to find a flexible distribution that can provide a good approximation for the underlying distribution and then can be used to obtain accurate and reliable Bayesian inference for quantile functions.

In the context of Bayesian linear quantile regression based on the asymmetric Laplace likelihood, it is assumed that the relationship between the response variable and the covariates is linear with homoscedastic errors. To deal with some potential violations of these assumptions, the Box-Cox transformation can be used. In this research, we use a Box-Cox transformation to develop a Bayesian method based on a pseudo asymmetric Laplace likelihood to fit multiple quantile regressions simultaneously. We specify prior distributions for all parameters including the parameter of the Box-Cox transformation. Moreover, the issue of crossing quantile curves is investigated, and a solution based on prior constraints is considered. Also, we consider two extensions of Box-Cox regression: Box-Cox quantile regression with heteroscedastic errors and two-sided Box-Cox quantile regression.

In using the asymmetric Laplace distribution, it is assumed that the mode of data is represented by the quantile of interest. In addition, the quantile to be estimated and the skewness of the asymmetric Laplace distribution are determined by the same parameter. This implies that the asymmetric Laplace distribution is not flexible enough

to accommodate error distributions corresponding to different quantile functions. To overcome these drawbacks, in order to improve Bayesian inference for quantiles, we propose a new generalisation of the Gumbel distribution that can be used to provide good approximations of the error distributions for quantile regression. Then, we use this generalised Gumbel distribution to develop a Bayesian method to estimate linear quantile functions individually.

Although simultaneous estimation based on the pseudo asymmetric Laplace likelihood has the advantage of ensuring non-crossing of quantile regressions and improving the point estimation for quantiles (especially the extreme one), it can have a very low coverage probability. This is due to the poor approximation provided by the pseudo asymmetric Laplace distribution for the underlying distribution. Therefore, we approximate the joint distribution of quantiles using a weighted pseudo asymmetric Laplace distribution. The weights control the contribution of likelihood functions corresponding to individual quantiles in forming the joint distribution of multiple quantile functions included in the simultaneous estimation. We consider two types of weights which are fixed and estimated. This approximate joint distribution can provide a good representation of the true distribution of the data. Thus, the Bayesian quantile method based on this weighted pseudo asymmetric Laplace distribution, which is used to estimate multiple quantile functions simultaneously without crossing, shows significant improvement in the coverage probabilities approaching the nominal coverage probabilities.

To avoid the limitations of Bayesian quantile methods using the asymmetric Laplace likelihood and to provide a flexible distribution that can accommodate any number of quantiles, we propose a family of approximate distributions that can be used for the error distribution of quantile regression. This family of approximations shows a good ability to accommodate a variety of underlying distributions. Then, we apply these approximations to develop Bayesian method to estimate quantile curves for the linear models and heteroscedastic models.

In addition, to check the fit of quantile regression, we develop a new bootstrap test.

# Contents


**1 Introduction** **1**

1.1 Introduction . . . 1
1.2 The concept of quantiles . . . 3
1.3 The motivation of quantile regression . . . 6
1.3.1 Comprehensiveness . . . 6
1.3.2 Robustness . . . 7
1.3.3 Relative efficiency . . . 7
1.4 Simultaneous estimation of quantile functions . . . 9
1.5 Bayesian estimation of quantiles . . . 10
1.6 Bayesian quantile regression methods . . . 13
1.6.1 Quantile regression based on asymmetric Laplace Likelihood 13
1.6.2 Quantile approach using a representation of AL distribution based on a mixture distribution . . . 15
1.6.3 Quantile method using an infinite mixture of Gaussian distributions . . . 17
1.6.4 Multiple quantile regression based on the generalisation of Jeffreys's substitution likelihood . . . 18
1.6.5 Quantile regression based on an empirical likelihood . . . 19
1.6.6 Quantile regression based on an approximate likelihood . . . 19
1.7 The contributions of our project . . . 20
1.8 Thesis outline . . . 21
1.9 Conclusion . . . 23

**2 Simultaneous Box-Cox Quantile Regression** **25**

2.1 Introduction . . . 25
2.2 Box-Cox transformation . . . 26

2.3 Box-Cox quantile regression . . . . . 28
2.4 Bayesian model for Box-Cox quantile regression . . . . . 29
2.4.1 Likelihood construction . . . . . 29
2.4.2 The crossing issue and specification of prior distributions . 30
2.4.3 Posterior computation . . . . . 32
2.5 Simulations and applications . . . . . 35
2.5.1 Simulated Data . . . . . 36
2.5.2 Air Quality data . . . . . 40
2.6 Discussion . . . . . 44

**3 Extensions of Box-Cox regression 45**
3.1 Introduction . . . . . 45
3.2 Bayesian Box-Cox quantile regression with heteroscedastic errors . 47
3.3 Bayesian two-sided Box-Cox quantile regression . . . . . 49
3.4 Simulations and applications . . . . . 50
3.4.1 Simulated Data . . . . . 50
3.4.2 Housing values in Boston . . . . . 56
3.4.3 Air Quality data . . . . . 59
3.5 Discussion . . . . . 61

**4 On likelihood functions for Bayesian quantile estimation 69**
4.1 Introduction . . . . . 69
4.2 Review of asymmetric Laplace likelihood . . . . . 70
4.3 Generalised distributions for quantile estimation . . . . . 72
4.4 An approximate quantile estimation for linear models . . . . . 74
4.5 Univariate simulation study . . . . . 75
4.6 Linear models . . . . . 79
4.7 Discussion . . . . . 82

**5 On the approximation of the joint distribution of quantiles 87**
5.1 Introduction . . . . . 87
5.2 The pseudo asymmetric Laplace distribution . . . . . 88
5.3 The weighted pseudo asymmetric Laplace distribution . . . . . 89
5.4 Normalising term . . . . . 90
5.5 Maximising the weighted pseudo asymmetric Laplace likelihood . 91

5.6 The specification of the weight function . . . 92
5.6.1 Fixed weights . . . 93
5.6.2 Estimated weights . . . 93
5.7 Bayesian linear multiple quantile regressions . . . 94
5.7.1 Using equal fixed weights . . . 95
5.7.2 Using estimated weights . . . 95
5.7.3 Posterior computation . . . 96
5.8 Simulation studies and real data analysis . . . 97
5.8.1 Univariate models . . . 97
5.8.2 Linear models . . . 101
5.8.3 The Chatterjee-Price Attitude data . . . 107
5.9 Conclusion . . . 112

**6 A family of approximate likelihood functions for quantile regression 115**
6.1 Introduction . . . 115
6.2 Approximate likelihood functions for individual quantile estimation 116
6.2.1 The weighted normal distribution . . . 117
6.2.2 The weighted generalised normal distribution . . . 118
6.2.3 The weighted mixture distribution . . . 118
6.3 An approximate likelihood function for simultaneous estimation of quantiles . . . 119
6.4 Simultaneous estimation of linear quantile functions . . . 120
6.5 Simultaneous estimation of quantiles for heteroscedastic linear models . . . 124
6.6 Checking the goodness of fit . . . 125
6.7 Simulation studies and real data analysis . . . 127
6.7.1 Univariate models . . . 127
6.7.2 Linear models . . . 130
6.7.3 Heteroscedastic linear models . . . 133
6.7.4 Cats data . . . 135
6.8 Discussion . . . 137

**7 Comparison and Conclusion** **139**
7.1 Summary . . . 139
7.2 Comparison . . . 141
7.3 Future work . . . 146
7.3.1 The asymptotic properties of Bayesian estimators . . . 146
7.3.2 Simultaneous non-linear quantile regression . . . 147
7.3.3 Goodness of fit . . . 147
7.3.4 Prior distribution . . . 149
7.3.5 Prediction . . . 149

**A** **151**
A.1 The link between the pseudo AL likelihood and the weighted optimisation problem. . . . 151
A.2 Conditional distributions for parameters of Bayesian Box-Cox quantile model . . . 152
A.2.1 Conditional Distribution of $\alpha_{\tau_k}$ . . . 152
A.2.2 Conditional Distribution of $\beta$ . . . 153
A.2.3 Conditional Distribution of $w_{\tau_k,i}$ . . . 154
A.2.4 Conditional Distribution of $\sigma_{\tau_k}$ . . . 155

**B** **157**
B.1 Markov chain Monte Carlo (MCMC) . . . 157

**C** **167**
C.1 Properties of the generalised asymmetric Laplace distribution . . . 167
C.2 Properties of the generalised Gumbel distribution . . . 169

**D** **171**
D.1 Normalising term . . . 171
D.2 Maximizing the weighted pseudo asymmetric Laplace likelihood . . . 172

**E** **177**
E.1 The weighted mixture distribution . . . 177
E.2 The cumulative distribution function for the weighted mixture distribution . . . 179

**References** **186**

# List of Figures


1.1 The loss function $\rho(y-\mu_\tau)$ corresponding to $\tau = 0.25$ where $\mu_\tau = 0$. 4
1.2 AL densities corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) and $\mu_\tau = 0$. . . . . 5
1.3 The fitted quantile curves for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left) and ordinary least squares regression (right). . . . . 6
1.4 The estimated median in the presence of the outliers (green) is represented by red line and the estimated median after the outliers are removed is represented by blue line. . . . . 8
1.5 The asymptotic relative efficiency of the sample median to sample mean against the degrees of freedom $\nu$, where $F$ is Student $t$-distribution. The red and green lines represent $ARE(M, \overline{X}, t) =$ 1 and $2/\pi$ respectively. . . . . 9
1.6 The composite loss function (left) and the pseudo-asymmetric Laplace density (right) corresponding to $\tau = 0.05, 0.5, 0.75$ where $\mu_{\tau_{0.05}} = -2.5, \mu_{\tau_{0.5}} = 0$ and $\mu_{\tau_{0.75}} = 2$. . . . . 11

2.1 Box-Cox transformation corresponding to $\lambda = -1, 0, 0.5, 1, 2, 3$. . . 27
2.2 Scatter plot of the original response variable (left) and transformed response obtained using the PAL-BC method (right) against x. . . 36
2.3 Scatter plot of the transformed response obtained using the AL-BC method corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) against x. 37
2.4 The average of the posterior mean of quantile curves (red) obtained by applying the PAL-BC method (left) and the AL-BC method (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75. . . . . 38

2.5 The average of the 95% highest posterior density intervals (green) estimated by applying PAL-BC (top) and AL-BC (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75 (left to right). . . . . 39
2.6 The average of the posterior mean of quantile curves (red) obtained by applying the PAL-BC method (left) and the AL-BC method (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.05, 0.25, 0.5 0.75 and 0.95. . . . . 40
2.7 The average of the 95% highest posterior density intervals (green) estimated by applying the PAL-BC method (top) and the AL-BC method (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.05 and 0.95 (left to right). . . . . 41
2.8 Scatter plot of the original response variable (left) and transformed response obtained using the PAL-BC method (right) against *Temp*. 42
2.9 Scatter plot of the transformed response obtained using the AL-BC method corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) against *Temp*. . . . . 42
2.10 The quantile curve estimated by applying PAL-BC, AL-BC and local polynomial quantile regression (LPQR) are represented by the red, green and blue curves respectively for quantiles: 0.25, 0.5 and 0.75 (left to right). . . . . 43

3.1 Two-sided Box-Cox transformation for range values of transformation parameters $\lambda$ and $\eta$, where $y = x$. . . . . 46
3.2 Scatter plot of the original and transformed Engel data using the logarithmic transformation. . . . . 46
3.3 Scatter plot of the original response variable (left) and transformed response with $\lambda = 0.51$ estimated using PAL-HBC (right) against x. 51
3.4 Scatter plot of the transformed response obtained using the AL-HBC method corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) against x, where $\lambda_{0.25} = 0.77$, $\lambda_{0.5} = 0.52$ and $\lambda_{0.75} = 0.30$. . . . . 52
3.5 The average of the posterior mean of quantile curves (red) obtained by applying PAL-HBC (left) and AL-HBC (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75. . . . . 52

3.6 The average of the 95% highest posterior density intervals (green) estimated by applying PAL-HBC (top) and AL-HBC (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75 (left to right). . . . . 53
3.7 The average of the posterior mean of quantile curves (red) obtained by applying PAL-HBC (left) and AL-HBC (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.05, 0.25, 0.5, 0.75 and 0.95. . . . . 54
3.8 The average of the 95% highest posterior density intervals (green) estimated by applying PAL-HBC (top) and AL-HBC (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.05 and 0.95 (left to right). . . . . 55
3.9 Scatter plot of the original response variable against the original predictor (left) and the transformed response variable against the transformed predictor obtained using PAL-TBC (right). . . . . 56
3.10 Scatter plot of the transformed response variable against the transformed predictor obtained using AL-TBC for quantiles: 0.25, 0.5 and 0.75 (left to right). . . . . 57
3.11 Quantile curve estimated by applying PAL-TBC, AL-TBC and local polynomial quantile regression (LPQR) are represented by the red, green and blue curves respectively for quantiles: 0.25, 0.5 and 0.75 (left to right). . . . . 58
3.12 The scatter plot of original relationships between the variables. . . 60
3.13 The scatter plot of relationships between the transformed variables obtained using the PAL-TBC method. . . . . 62
3.14 The scatter plot of relationships between the transformed variables obtained using the AL-TBC method corresponding to $\tau = 0.25$. . 65
3.15 The scatter plot of relationships between the transformed variables obtained using the AL-TBC method corresponding to $\tau = 0.5$. . . 66
3.16 The scatter plot of relationships between the transformed variables obtained using the AL-TBC method corresponding to $\tau = 0.75$. . 67

4.1 The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for normal distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the asymmetric Laplace distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves. 76

4.2 The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for Student's $t$ distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the asymmetric Laplace distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves. 76

4.3 The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for gamma distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the asymmetric Laplace distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves. 77

4.4 The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for normal distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the generalised Gumbel distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves. 77

4.5 The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for Student's $t$ distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the generalised Gumbel distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves. 78

4.6 The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for gamma distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the generalised Gumbel distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves. 78

4.7 The box-plot of the maximum a posteriori estimators, obtained individually using Bayesian quantile method based on the asymmetric Laplace likelihood, over 1000 simulations. The red lines represent the true values. . . . . 84

4.8 The box-plot of the maximum a posteriori estimators, obtained individually using Bayesian quantile method based on the generalised Gumbel likelihood, over 1000 simulations. The red lines represent the true values. . . . . 85

5.1 The probability density function (left) and the cumulative distribution function (right) for the true distribution (black) which is N(0,1) and their approximations obtained using the PAL distribution based on 5 quantiles (red). . . . . 98

5.2 The average of Kullback–Leibler divergence, from the approximation obtained using the PAL distribution to N(0,1), over 500 simulations against the number of quantiles corresponding to Scheme 1 (left) and Scheme 2 (right). The dotted curves represent the 95% confidence intervals for the mean. . . . . 99

5.3 The probability density function (top) and the cumulative distribution function (bottom) for normal, gamma, Student's $t$ and mixture distributions (left to right). The true distributions are represented by black curves and their approximations obtained using the WPAL distribution based on 5 quantiles with equal fixed weights are represented by red curves. . . . . 100

5.4 The probability density function (top) and the cumulative distribution function (bottom) for normal, gamma, Student's $t$ and mixture distributions (left to right). The true distributions are represented by black curves and their approximations obtained using the WPAL distribution based on 5 quantiles with estimated weights are represented by red curves. . . . . 101
5.5 The average of Kullback–Leibler divergence, from the approximation obtained using the WPAL distribution with equal fixed weights (red) and estimated weights (green) to the true distribution, over 500 simulations against the number of quantiles corresponding to Scheme 1. The dotted curves represent the 95% confidence intervals for the mean. . . . . 102
5.6 The average of Kullback–Leibler divergence, from the approximation obtained using the WPAL distribution with equal fixed weights (red) and estimated weights (green) to the true distribution, over 500 simulations against the number of quantiles corresponding to Scheme 2. The dotted curves represent the 95% confidence intervals for the mean. . . . . 103
5.7 The box-plot of the maximum a posteriori estimators, obtained using the Bayesian quantile method based on the WPAL likelihood with equal fixed weights, over 1000 simulations. The red lines represent the true values. . . . . 106
5.8 The box-plot of the maximum a posteriori estimators, obtained using the Bayesian quantile method based on the WPAL likelihood with estimated weights, over 1000 simulations. The red lines represent the true values. . . . . 107
5.9 The box-plot of the maximum a posteriori estimators, obtained using the Bayesian quantile method based on the WPAL likelihood with equal fixed weights, over 1000 simulations. The red lines represent the true values. . . . . 108
5.10 The box-plot of the maximum a posteriori estimators, obtained using Bayesian quantile method based the WPAL likelihood with estimated weights, over 1000 simulations. The red lines represent the true values. . . . . 109

5.11 The scatter plot of the overall rating against the handling of employee complaints (left). The estimated quantile functions, based on the maximum a posteriori estimators obtained individually using the asymmetric Laplace likelihood, for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (right). . . . . 110

5.12 The estimated quantile functions, based on the maximum a posteriori estimators obtained simultaneously using the WPAL likelihood with equal fixed weights and estimated weights, and the PAL likelihood (left to right), for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$. . . . . 110

5.13 Comparison of estimated conditional densities of $y|x = E(x)$ obtained using Bayesian quantile regression based on the WPAL distribution corresponding to 5 quantiles (left) and 10 quantiles (right) with equal fixed weights (green) and estimated weights (blue), and Bayesian regression based on normal distribution (black) and skew normal distribution (red). Overall rating is represented by $y$ and handling of employee complaints is represented by $x$. . . 112

6.1 The average of the estimated cumulative functions, obtained using the method developed by Nadaraya (1964) with a plug-in estimator of optimal bandwidth proposed by Altman & Leger (1995) for $\mathrm{N}(0, 1)$ and $\mathrm{Gamma}(3, 3)$, over 1000 simulations of size 150. . . . . 127

6.2 The probability density function (top) and the cumulative distribution function (bottom) for the normal, gamma and Student's $t$ distributions (left to right). The true distributions are represented by black curves and their approximations, obtained using the weighted normal distribution, are represented by red curves. . 128

6.3 The probability density function (top) and the cumulative distribution function (bottom) for the normal, gamma and Student's $t$ distributions (left to right). The true distributions are represented by black curves and their approximations, obtained using the weighted generalised normal distribution, are represented by red curves. . . . . 129

6.4 The probability density function (top) and the cumulative distribution function (bottom) for the normal, gamma and Student's $t$ distributions (left to right). The true distributions are represented by black curves and their approximations, obtained using the weighted mixture of skewed Laplace and skewed normal distributions, are represented by red curves. . . . . 130
6.5 The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted normal likelihood, over 1000 simulations. The red lines represent the true values. . . . . 131
6.6 The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted generalised normal likelihood, over 1000 simulations. The red lines represent the true values. . . . . 132
6.7 The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions, over 1000 simulations. The red lines represent the true values. . . . . 133
6.8 The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions with heteroscedastic scale, over 1000 simulations. The red lines represent the true values. . . . . 134
6.9 The estimated quantile functions based on the maximum a posteriori estimators obtained simultaneously using the weighted normal likelihood, the weighted generalised normal likelihood and the weighted mixture of skewed Laplace and skewed normal likelihoods (left to right), for $\tau = 0.25, 0.5, 0.75$. . . . . 135

6.10 The histogram of test statistics, corresponding to the 0.25, 0.5 and 0.75 quantile curves (left to right) estimated simultaneously using the weighted normal likelihood, the weighted generalised normal likelihood and the weighted mixture of skewed Laplace and skewed normal likelihoods (top to bottom), over 100,000 bootstrap samples. The red lines represent 95% percentile confidence interval. . 138

7.1 The colored lines represent the estimated quantile function, based on the maximum a posteriori estimators, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$. . . . . 144
7.2 The 95% highest posterior density intervals for intercept coefficients are represented by the colored lines. . . . . 145
7.3 The 95% highest posterior density intervals for slope coefficients are represented by the colored lines. . . . . 147
7.4 The joint 95% highest posterior density regions of the intercept and slope coefficients of quantile regression. . . . . 148
7.5 The conditional distribution of $y$=Heart Weight given $x$=Body Weight. The colored lines represent the estimated 0.25 and 0.5 quantiles. . . . . 148

B.1 The trace plots for the samples of the intercept coefficients obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom). . . . . 158
B.2 The trace plots for the samples of the slope coefficient obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right). 159
B.3 The trace plots for the samples of the scale parameters obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom). . . . . 159
B.4 The trace plots for the samples of the transformation parameter obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right). . . . . 160
B.5 The autocorrelation plot for the samples of the intercept coefficients obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom). . . . . 161

B.6 The autocorrelation plot for the samples of the slope coefficient obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right). . . . . 162
B.7 The autocorrelation plot for the samples of the scale parameters obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom). . . . . 162
B.8 The autocorrelation plot for the samples of the transformation parameter obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right). . . . . 163
B.9 The autocorrelation plot (after thinning) for the samples of the intercept coefficients obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom). . . . . 163
B.10 The autocorrelation plot (after thinning) for the samples of the slope coefficient obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right). . . . . 164
B.11 The autocorrelation plot (after thinning) for the samples of the scale parameters obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom). . . . . 164
B.12 The autocorrelation plot (after thinning) for the samples of the transformation parameter obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right). . . . . 165

# Chapter 1

# Introduction

## 1.1 Introduction

Since quantile regression was introduced by Koenker & Bassett (1978), it has been considered as a powerful statistical tool for investigating the relationships amongst variables. It offers an extension to ordinary least squares regression in that it describes the entire conditional distribution of the response variable. It allows estimating multiple quantile functions that can reveal some characteristics of the conditional distribution such as the shape, skewness and heteroscedasticity. Also, it is more robust to outliers and misspecification of the underlying distribution. Quantile regression has been applied in a variety of fields. For example, in health, Sherwood *et al.* (2013) used weighted quantile regression to describe the conditional distribution of health care cost data with missing covariates since quantile regression gives a more comprehensive image of how the health care cost can be affected by the covariates, and it can deal with skewed and heterogeneous data; in finance, since quantile regression can deal with extreme values and outliers, it was employed by Fin *et al.* (2009) to analyse Australian stocks data; in economics, to study determinants of the load cycle at upper quantiles of conditional distribution, Hendricks & Koenker (1992) applied hierarchical quantile regression to model household electricity demand; and in environmental science, Bel *et al.* (2015) employed quantile regression to investigate the impact of speed limits on air pollution.

There is extensive literature on frequentist estimation approaches for quantile regression. Since there is no requirement to specify the data's underlying distri-

bution, frequentist inference about quantile functions (e.g. confidence interval, and hypothesis test) is based on asymptotic theories. There are a number of methods developed to estimate quantile curves by minimising an asymmetric absolute loss function, by for example, Koenker & Hallock (2001), He & Zhu (2003) and Koenker (2005).

Research on Bayesian quantile regression is still limited. In contrast to classical estimation methods, Bayesian quantile regression requires the specification of the full distribution of residuals. Initial interest has been on developing Bayesian methods to fit median regression, such as the methods developed by Walker & Mallick (1999) and Kottas & Gelfand (2001). Yu & Moyeed (2001) proposed a Bayesian method to fit individual quantile regression. This was followed by the development of a number of Bayesian approaches such as the methods proposed by Tsionas (2003), Yu & Stander (2007), Geraci & Bottai (2007), Reich *et al.* (2010), Lancaster & Jae Jun (2010) and Kozumi & Kobayashi (2011).

Although simultaneous estimation of quantile curves has great advantages, developing simultaneous methods to fit quantile regression is a long standing issue. The key advantage of simultaneous estimation of quantiles over individual quantile estimation is that non-crossing of quantile functions can be achieved by ensuring the monotonic increase of quantile functions with respect to the quantile levels over the data space. Also, information over all quantiles of interest can be combined in order to provide better estimation accuracy and reliability. In frequentist statistics, there are a number of papers devoted to simultaneous quantile regression such as Zou & Yuan (2008), Wu & Liu (2009), Bondell *et al.* (2010), Jiang *et al.* (2012) and Xiong & Tian (2015). Moreover, there are a few examples of Bayesian multiple quantile approaches such as Dunson & Taylor (2005), Hahn & Burgette (2012) and Feng *et al.* (2015).

In this chapter, we introduce the concept of quantile regression. We highlight estimation of quantiles based on the asymmetric absolute function and asymmetric Laplace distribution. We discuss the motivation of quantile regression (e.g. Comprehensiveness and Robustness). We discuss the basic idea behind simultaneous estimation of quantile functions. In addition, we discuss Bayesian estimation of quantile curves. Finally, we briefly discuss some Bayesian quantile approaches. We show the main contributions of our project and the outline of this thesis.

## 1.2 The concept of quantiles

The $\tau$ sample quantile is the value that divides a sample of size $n$ into two subsets such that the lower subset includes almost $\tau n$ values (for more discussion on definitions of sample quantiles, see Hyndman & Fan, 1996). This definition of quantile can be extended to the entire population by considering $Y$, a random variable with cumulative distribution function:

$$F(y) = P(Y \leq y).$$

Here, for $\tau \in (0, 1)$, the $\tau$ quantile can be defined as

$$Q_y(\tau) = \inf \{y : F(y) \geq \tau\},$$

which is unique under the assumption of $F$ is continuous and strictly increasing. Koenker & Bassett (1978) proved that for a random sample $y_1, y_2, ..., y_n$ the $\tau$ quantile can be estimated by solving the optimisation problem:

$$\min_{\mu_\tau \in R} \sum_{i=1}^{n} \rho_\tau(y_i - \mu_\tau), \tag{1.1}$$

where $\mu_\tau$ is the $\tau$ quantile parameter and the loss function $\rho_\tau(\cdot)$ is given by

$$\begin{aligned} \rho_\tau(u) &= \frac{|u| + (2\tau - 1)u}{2} \\ &= \begin{cases} (1-\tau)|u|, & u \leq 0, \\ \tau|u|, & u > 0. \end{cases} \end{aligned} \tag{1.2}$$

This loss function is shown in Figure 1.1. The minimizer of the expected loss function written as

$$E\left[\rho_\tau(Y - \mu_\tau)\right] = (\tau - 1) \int_{-\infty}^{\mu_\tau} (y - \mu_\tau) dF(y) + \tau \int_{\mu_\tau}^{\infty} (y - \mu_\tau) dF(y),$$

is any element of $\{y : F(y) = \tau\}$, since $F$ is monotonic. In the case of the unique solution, the minimizer is $\mu_\tau = F^{-1}(\tau)$. Otherwise, the smallest value in the interval of $\tau$ quantiles must be the solution. For more discussion, see Koenker (2005, p.5).

By considering $\tau = 0.5$, the optimization problem given in (1.1) is just the

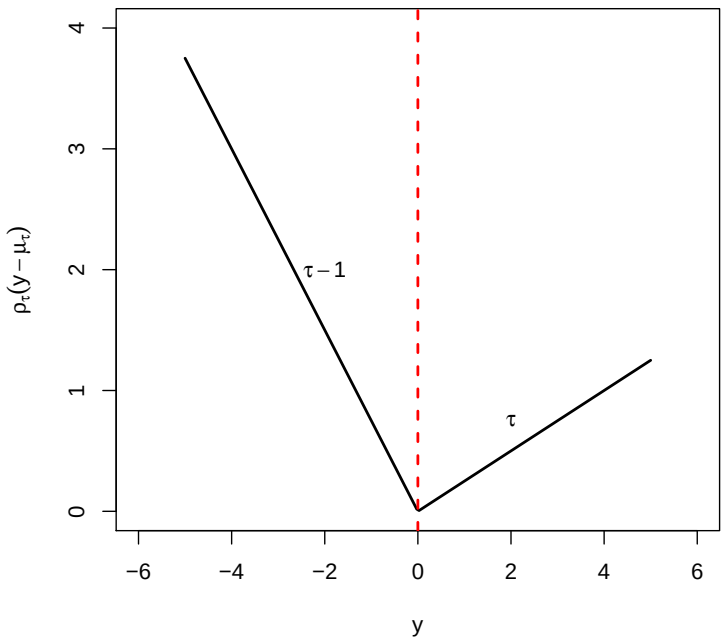



Figure 1.1: The loss function $\rho(y - \mu_\tau)$ corresponding to $\tau = 0.25$ where $\mu_\tau = 0$.

sum of the absolute values of the residuals $y_i - \mu_{0.5}$. Then, since the absolute loss function assigns equal weights to the spread of observations, its minimizer is the value $\mu_{0.5}$ that approximately equalises the number of negative and positive residuals. This implies that the number of observations above and below $\mu_{0.5}$ is approximately equal. Thus, by definition, the estimated $\mu_{0.5}$ is the sample median. For other quantile levels, the optimization problem given in (1.1) is the absolute loss function that is weighted asymmetrically to ensure that approximately $\tau n$ of observations are located below $\mu_\tau$ and $(1 - \tau)n$ is located above $\mu_\tau$ (see Figure 1.1).

Moreover, Yu & Moyeed (2001) considered estimating quantiles using the asymmetric Laplace (AL) distribution given by the density function

$$f_\tau(y|\mu_\tau) = \tau(1-\tau)\exp\left\{-\rho_\tau(y-\mu_\tau)\right\}, \tag{1.3}$$

rather than using the underlying distribution generating $y$, since the maximisation of this density is equivalent to the minimisation of the loss function given in (1.2).

As a special case, the standard symmetric Laplace distribution density is obtained by setting $\tau = 0.5$ :

$$f_\tau(y|\mu_\tau) = \frac{1}{4}\exp\left\{-\frac{|y-\mu_\tau|}{2}\right\},$$

which can be used to estimate the median.

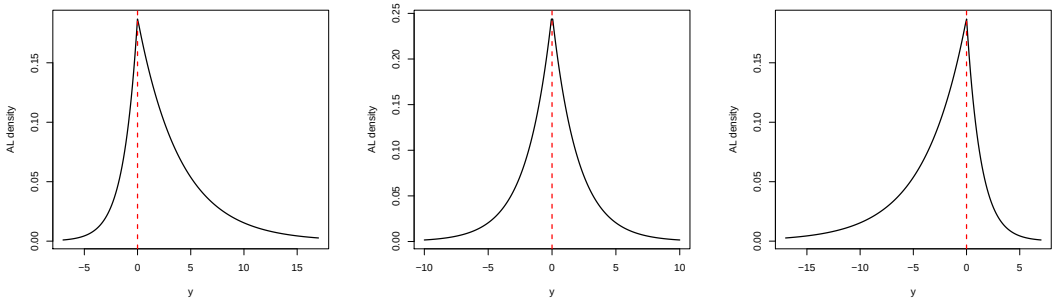



Figure 1.2: AL densities corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) and $\mu_\tau = 0$.

Figure 1.2, which shows the asymmetric Laplace densities corresponding to $\tau = 0.25, 0.5$ and $0.75$, illustrates the properties of the asymmetric Laplace distribution. It can be seen that the asymmetric Laplace density has long tails that may be able to accommodate any outlier observations. It is obvious that the asymmetric Laplace density is a non-differentiable function at its maximum. The location parameter $\mu_\tau$ is interpreted as $\tau$ quantile, since it satisfies the definition of quantile as follows:

$$\begin{aligned} F_\tau(Y \leq \mu_\tau) &= \int_{-\infty}^{\mu_\tau} f_\tau(y|\mu_\tau)dy \\ &= \int_{-\infty}^{\mu_\tau} \tau(1-\tau)\exp\left\{-(1-\tau)|y-\mu_\tau|\right\}dy \\ &= \tau. \end{aligned}$$

Yu & Zhang (2005) considered a quantile model depending on the asymmetric Laplace distribution with the density function given by

$$f_\tau(y|\mu_\tau, \sigma) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{y-\mu_\tau}{\sigma}\right)\right\}, \tag{1.4}$$

where $\mu_\tau$ and $\sigma$ are the location and the scale parameters respectively. They

also derived the properties of this class of asymmetric Laplace distributions. Although the scale parameter has no clear interpretation under non-Laplace error distributions, it could play an important role in improving the computation.

## 1.3 The motivation of quantile regression

To illustrate the motivation of quantile regression, we consider the following two properties.

### 1.3.1 Comprehensiveness

The ordinary least square estimator of the conditional mean function summarises the average relationship between the response variable and covariates. Therefore, in some cases, the least squares estimator is considered as a non-informative central tendency measure. In contrast, quantile regression allows for a more comprehensive analysis of the conditional distribution of the response variable given predictors. For more illustration, consider the example given in Figure 1.3 that shows the relationship between household food expenditure and income for 235 Belgian working class households (this dataset is now available publicly by Koenker, 2016).

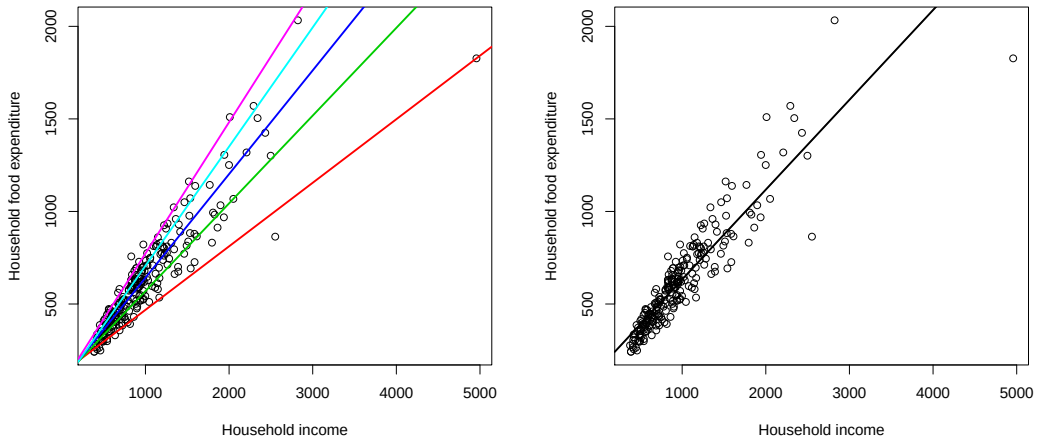



Figure 1.3: The fitted quantile curves for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left) and ordinary least squares regression (right).

In contrast to least squares regression, from Figure 1.3, quantile regression offers a complete picture of conditional distribution of household food expenditure. It can be seen that the estimated quantile curves suggest that the conditional distribution of household food expenditure is asymmetric around the centre with a long-lower tail and short-upper tail. Also, they can show the heteroscedastic error distribution (i.e. in this case, the scale of the conditional distribution is not a constant and increases with the predictors). The relationship is considerably different for households with a low expenditure from those with a medium or high expenditure.

### 1.3.2 Robustness

According to the properties of the estimation procedure based on the loss function given in (1.2), the quantiles are robust against the outliers. To illustrate this property, we estimate the median for data with some outliers and we do this estimation in the presence of the outliers and after they are removed. Then, the result, as in Figure 1.4, suggests that the median is not affected by the outliers. Also, they are considered powerful estimators that can deal with the distributions having long tails or being asymmetrical. The minimisation of this loss function corresponding to $\tau$ is equivalent to the maximisation of true data distribution under the error distribution being an asymmetric Laplace with skew parameter $\tau$. Moreover, the estimation procedure based on this loss function can perform well and provide reliable estimation under non-Laplace distributions of errors. Consequently, because of the theoretical link between loss function given in (1.2) and the asymmetric Laplace distribution defined in (1.3), the Bayesian methods based on the asymmetric Laplace distribution enjoy these attractive properties of estimation. Yu & Moyeed (2001) suggested that the Bayesian quantile method based on the asymmetric Laplace likelihood can work under different error distributions. Thus, the key motivations for using quantile regression technique are the robustness against outliers, and the fact that no assumption about the error distribution is required.

### 1.3.3 Relative efficiency

In this section, we illustrate the relative efficiency of the median as a spacial case

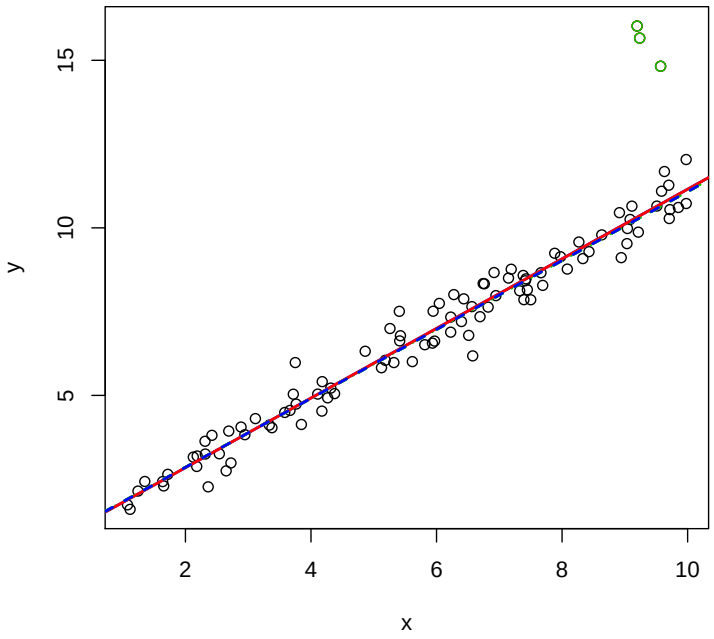



Figure 1.4: The estimated median in the presence of the outliers (green) is represented by red line and the estimated median after the outliers are removed is represented by blue line.

of the quantiles compared to the mean. Although the sample mean is considered sometimes to be more efficient, the sample median can be more robust. Serfling (2011) showed that the asymptotic relative efficiency (ARE) of sample median ($M$) to sample mean ($\overline{X}$) for a large sample from a distribution $F$ (e.g. with a density function $f(\theta)$) is given by

$$ARE(M, \overline{X}, F) = 4[f(\theta)]^2\sigma_F^2.$$

Then, in the case that $F$ is normal distribution $ARE(M, \overline{X}, F) = 2/\pi \approx 0.64$, which indicate that the sample mean is more efficient. However, in the case of $F$ is symmetric Laplace distribution $ARE(M, \overline{X}, F) = 2$. For Student $t$-distribution, the asymptotic relative efficiency of the sample median to sample mean is given by

$$ARE(M, \overline{X}, t) = \frac{4\nu}{\nu - 2}\left\{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\right\}^2,$$

which depend on the degrees of freedom $\nu$ (for more details, see Figure 1.5). Therefore, for the data from distributions having thick tails, the sample median

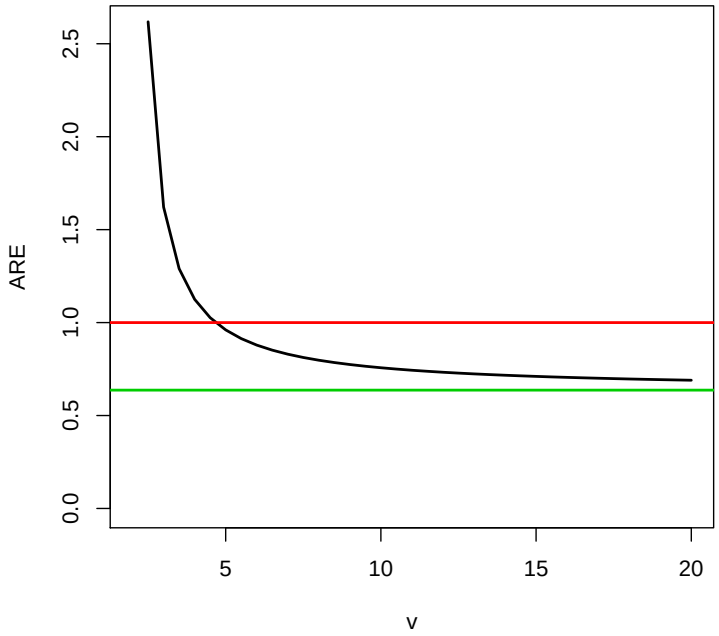



Figure 1.5: The asymptotic relative efficiency of the sample median to sample mean against the degrees of freedom $\nu$, where $F$ is Student $t$-distribution. The red and green lines represent $ARE(M, \overline{X}, t) = 1$ and $2/\pi$ respectively.

can is considered more efficient while for the data come from distributions having thin tails the sample mean is considered more efficient. This implies that the efficiency can changes significantly as the distribution of data changes. Thus, the efficiency may trade off against the robustness and this can be considered as a great motivation of robust estimator (e.g. quantiles).

## 1.4 Simultaneous estimation of quantile functions

The standard analysis of quantile regression discussed in the previous sections is based on the individual estimation of quantiles. To implement simultaneous

estimation of quantile functions, the optimisation problem given in (1.1) should be extended for a vector $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_m)$ by considering a sum of the check functions corresponding to different quantiles as follows:

$$\min_{\boldsymbol{\mu}_\tau} \sum_{k=1}^{m} \sum_{i=1}^{n} \rho_{\tau_k} \left( y_i - \mu_{\tau_k} \right), \tag{1.5}$$

where $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$. This is known as composite quantile regression (for more details, see Bondell *et al.*, 2010; Zou & Yuan, 2008). The key motivation for using this simultaneous method over the individual quantile estimation method given in (1.1) is that it can accommodate suitable constraints or assumptions to ensure that quantile functions monotonically increase in $\tau$ (i.e. $\mu_{\tau_{k-1}} < \mu_{\tau_k}$ for $k = 2, ..., m$). Simply, it can be shown that the minimisation of the composite optimisation problem given in (1.5) is equivalent to the maximisation of the pseudo asymmetric Laplace likelihood:

$$l\left(\boldsymbol{\mu}_\tau; \boldsymbol{y}\right) \propto \left( \prod_{k=1}^{m} \left[ \tau_k \left( 1 - \tau_k \right) \right]^n \right) \exp \left\{ - \sum_{k=1}^{m} \sum_{i=1}^{n} \rho_{\tau_k} \left( y_i - \mu_{\tau_k} \right) \right\}. \tag{1.6}$$

The comparison between the composite loss function $\sum_{k=1}^{m} \rho_{\tau_k}(y - \mu_{\tau_k})$ and the pseudo asymmetric Laplace density, related to the likelihood given in (1.6), is shown in Figure 1.6.

## 1.5 Bayesian estimation of quantiles

For a vector of observations $\boldsymbol{y}$, Bayesian inference about a set of unknown parameters $\Theta$ is made through the posterior distribution which is given by

$$p(\Theta|\boldsymbol{y}) \propto l(\Theta; \boldsymbol{y}) p(\Theta), \tag{1.7}$$

where $l(\Theta; \boldsymbol{y})$ is the likelihood function and $p(\Theta)$ is the prior density. Therefore, Bayesian models require a full specification of the underlying distribution. For Bayesian quantile regression, the distribution of the error $\epsilon$ should be with a density function $f$ such that

$$\int_{-\infty}^{0} f(\epsilon) dy = \tau, \tag{1.8}$$

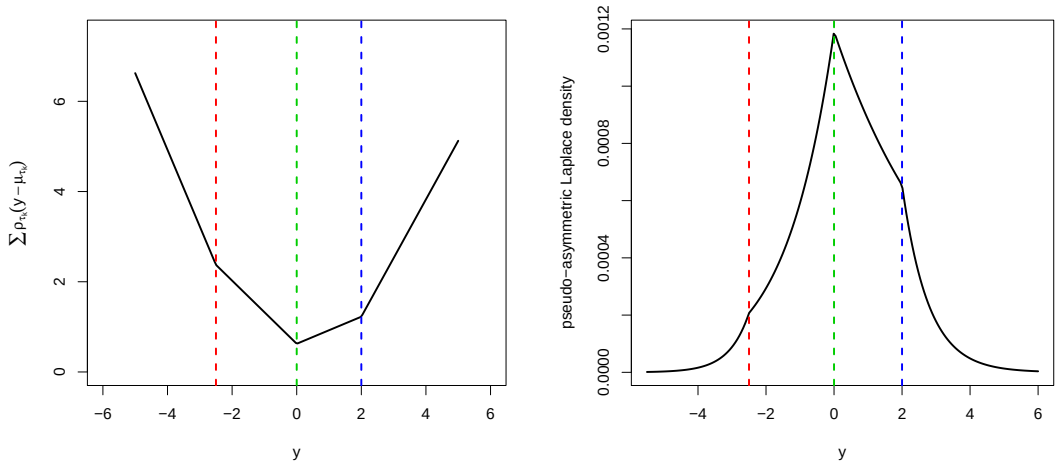



Figure 1.6: The composite loss function (left) and the pseudo-asymmetric Laplace density (right) corresponding to $\tau = 0.05, 0.5, 0.75$ where $\mu_{\tau_{0.05}} = -2.5, \mu_{\tau_{0.5}} = 0$ and $\mu_{\tau_{0.75}} = 2$.

so the error distribution is restricted to have the $\tau$ quantile of zero. In practice, it is very challenging to find the error distribution that satisfies the equation given in (1.8). As a result, a number of approaches were proposed to specify a suitable likelihood for Bayesian quantile regression. Yu & Moyeed (2001) were the first to develop a Bayesian method based on an asymmetric Laplace likelihood to fit quantile regression. Dunson & Taylor (2005) applied the substitution likelihood to construct a simultaneous quantile method to estimate multiple curves without crossing. Applying exponentially tilted empirical likelihood, Lancaster & Jae Jun (2010) developed a Bayesian approach to fit quantile regression (for more details, see Section 1.6).

In the context of Bayesian inference based on the posterior distribution given in (1.7), priors play a crucial role in inference. Non-informative and improper priors offer simplicity for some complex models and they can be useful in some situations, such as the representation of ignorance about the parameters. However, model selection can be impossible using Bayes factors with non-informative priors. Also, applying non-informative priors can lead to instability in the posterior estimates and cause convergence-related problems. In applied research, it is important to benefit from prior information that is available about the inves-

tigated problem. Thus, there is no doubt about the importance of specifying informative priors for all parameters included in the model. To specify priors for Bayesian model parameters, priors should be assigned over the model space. For example, in Bayesian quantile regression models that allow for simultaneous inference on multiple quantile curves, the priors for quantile coefficients should take into account the fact that quantile functions increase in $\tau$. Yu & Moyeed (2001) specified an improper prior for all quantile regression coefficients. Kozumi & Kobayashi (2011) utilised double exponential distributions as prior distributions for quantile regression coefficients. Alhamzawi & Yu (2013) introduced modified version of Zellner's g-prior for quantile regression.

The crossing of quantile functions can be a consequence of complex features of the conditional distribution of the response variable such as heteroscedastic variance. Also, it can be caused by a relatively small size of the dataset that is not sufficient to estimate quantile functions accurately. To address the problem of crossing quantile functions, a variety of approaches were proposed; see, for example, Dunson & Taylor (2005) and Rodrigues & Fan (2017).

Simultaneous quantile regression methods are sensitive to the number and location of quantile functions included in the estimation. The degree of sensitivity can be different due to the complexity of the underlying distribution and the available data. However, adding additional quantiles to the estimated model can increase stability of the quantile estimation. Bondell *et al.* (2010) suggested that including extra quantiles in a neighborhood of the quantile of interest to reach a desirable level of stability can improve the estimation of quantile functions, which is something we investigate later.

Analytical expressions for the posterior distribution and computing the summary statistics about the parameters of interest are unavailable in closed form whatever the prior distribution. Therefore, Markov Chain Monte Carlo methods (MCMC) are required to sample from the posterior distribution. Since Bayesian quantile regression may include a relatively large number of parameters, especially in the case of multiple quantile regression, it is difficult or sometimes even impossible to derive the full conditional distributions for all parameters of the model in order to apply the Gibbs sampler. Also, to improve the computational efficiency of the Metropolis-Hastings algorithm, the proposal distributions should

be constructed carefully. A number of Markov chain Monte Carlo methods similar to those considered by Tsionas (2003), Dunson & Taylor (2005) and Kozumi & Kobayashi (2011) can provide an efficient computation for the posterior distribution of quantile regression.

## 1.6 Bayesian quantile regression methods

To deal with the issues related to Bayesian estimation of quantile curves which are outlined in the previous section, a variety of Bayesian quantile regression methods have been developed. In this section, we discuss critically a number of novel Bayesian methods that have been developed to estimate quantile functions separately and simultaneously. We briefly describe these approaches and provide some of their key features and limitations.

### 1.6.1 Quantile regression based on asymmetric Laplace Likelihood

Yu & Moyeed (2001) considered the asymmetric Laplace Likelihood, which is unrelated to the assumed error distribution, to construct a Bayesian method to estimate quantile curves separately. This likelihood function is used because its maximisation is equivalent to the minimisation of the loss function given in (1.2). This leads to the likelihood function for the same model changing as the quantile of interest changes. To fit quantile regression corresponding to the linear model given by

$$y_i = \boldsymbol{x}_i^{'}\boldsymbol{\beta} + \epsilon_i, \quad \text{for } i = 1, ..., n \tag{1.9}$$

where $\boldsymbol{x_i^{'}}$ is the vector of covariates related to $i^{\text{th}}$ observations, $\boldsymbol{\beta}$ is a vector of unknown coefficients and $\epsilon_i$ is the error that follows an unknown distribution with mean zero and a constant variance, they employed the asymmetric Laplace likelihood given by

$$l_\tau(\boldsymbol{\beta}; \boldsymbol{y}) \propto \tau^n (1-\tau)^n \exp\left\{\sum_{i=1}^{n} -\rho_\tau(y_i - \boldsymbol{x}_i^{'}\boldsymbol{\beta})\right\}. \tag{1.10}$$

They also used the prior $p(\boldsymbol{\beta}) \propto 1$, and proved that using this prior with likelihood function given in (1.10) yields a proper posterior distribution. Beside

the simplicity and flexibility, this Bayesian model has a number of key features. Since the asymmetric Laplace distribution is an exponential of the negative asymmetric absolute function that is robust against outliers and can provide reliable estimation under different error distributions (for more details, see Section 1.3.2), this Bayesian quantile regression is robust against outliers and can perform well in terms of point estimation under a misspecification of the error distribution. However, this Bayesian quantile method has some limitations. Although this method assumes the desired quantile function represents the mode of the conditional distribution of the response variable, this is not true except for $\tau$ quantile function under an error distribution that is an asymmetric Laplace with skew parameter $\tau$. This implies that this Bayesian method can offer the target coverage probability for a particular quantile yet poor coverage probability for other quantile levels. Thus, the quality of Bayesian estimation based on the asymmetric Laplace distribution varies with respect to the quantile of interest and is determined by the similarity between the asymmetric Laplace distribution corresponding to the quantile of interest and the underlying distribution. In addition, the skewness and quantile of interest are controlled by the same parameter $\tau$. Using this approach, there is no guarantee that quantile functions will not cross.

This approach was extended by Yu & Stander (2007) to fit Tobit quantile regression. Also, they discussed a variety of prior distributions that can be used with the asymmetric Laplace likelihood to provide a proper posterior distribution. They drew Bayesian inference based on the asymmetric Laplace distribution with the scale parameter given in (1.4). An advantage of this approach is to avoid solving a non-convex optimisation problem that is used to fit Tobit quantile regression. Moreover, Geraci & Bottai (2007) used an asymmetric Laplace likelihood to develop a Bayesian model including random effects, to fit linear quantile regression for longitudinal data.

Rodrigues & Fan (2017) developed a two-stage Bayesian method to fit multiple quantile regressions without crossing. In the first stage, all quantiles of interest are estimated separately using the Bayesian model suggested by Yu & Moyeed (2001). In the second stage, the initial estimates achieved in the first stage are adjusted by borrowing information from neighbor quantile functions using a Gaussian process. However, if the quantile functions based on initial estimates do not cross, the second stage has no effect on the estimation. This method can

be used to fit linear and non-linear quantile regression. However, in the case of poor initial estimates, this approach yields an unreasonably high bandwidth.

Bayesian quantile methods based on the asymmetric Laplace likelihood functions focus on the quality of the point estimation of parameters and ignore other Bayesian inference, such as the properties of posterior distribution and credible intervals. Most of these methods use asymmetric Laplace likelihood functions blindly to draw Bayesian inference under any underlying distribution without caring about the quality of the approximation achieved using these asymmetric Laplace distributions, which can be very poor in some cases. To deal with the estimation of quantile curves from a Bayesian viewpoint, more research is needed to decide if the asymmetric Laplace likelihood is an appropriate choice to construct a Bayesian model to fit quantile regression or is just a point estimation procedure that is not suited to make Bayesian inference on quantiles.

### 1.6.2 Quantile approach using a representation of AL distribution based on a mixture distribution

To estimate the quantile curves corresponding to the model given in (1.9) individually, Tsionas (2003) considered a representation of the asymmetric Laplace distribution based on a piecewise normal density given by

$$f_\tau(y_i|\boldsymbol{\beta},\sigma,w_i) \propto \frac{1}{\sigma\sqrt{w_i}} \exp\begin{cases} -\frac{\tau}{2\sigma^2 w_i}\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2, & y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}, \\ -\frac{1-\tau}{2\sigma^2 w_i}\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2, & y_i < \boldsymbol{x}_i'\boldsymbol{\beta}, \end{cases} \tag{1.11}$$

where $w_i$ follows the standard exponential distribution. He used the phrase "the normal mixture" density. Then he proved that the posterior distribution is given by

$$p(\boldsymbol{\beta},\sigma,\boldsymbol{w}|\boldsymbol{y}) \propto \sigma^{-(n+1)} \prod_{i=1}^{n} w_i^{-\frac{1}{2}} \exp\begin{cases} -\frac{(\tau-w_i)\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2 w_i}, & y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}, \\ -\frac{(1-\tau-w_i)\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2 w_i}, & y_i < \boldsymbol{x}_i'\boldsymbol{\beta}. \end{cases}$$

This representation allows the use of a Gibbs sampler with data augmentation to draw Bayesian inference from this posterior distribution. However, this algorithm would have a complexity of programming and would slow down as the number of observations increases. Also, the efficiency of this Gibbs sampler may be ques-

tionable, since each parameter $\beta_j \in \boldsymbol{\beta}$ is updated separately and this could result in highly correlated samples from the posterior distribution. To overcome such issues, Kozumi & Kobayashi (2011) considered the representation of the asymmetric Laplace distribution based on a mixture of normal distributions to develop a Gibbs sampler to fit Bayesian quantile regression. They showed that the asymmetric Laplace random variable can be represented as a mixture of location-scale normal distributions by considering $y_i \sim \mathrm{AL}(\mu_\tau, 1, \tau)$. Then, they showed

$$y_i \stackrel{d}{=} \mu_\tau + \theta_\tau \upsilon_{\tau,i} + z_i \sqrt{\phi_\tau \upsilon_{\tau,i}},$$

where $\upsilon_{\tau,i}$, which corresponds to the $\tau$ quantile and the $i^{\text{th}}$ observation, is an exponential random variable such that $\upsilon_{\tau,i} \sim \mathrm{Exp}(1)$ and $z_i$ is a normal random variable such that $z_i \sim \mathrm{N}(0,1)$. They did this by showing that the characteristic function of $y_i$ and $\mu_\tau + \theta_\tau \upsilon_{\tau,i} + z_i \sqrt{\phi_\tau \upsilon_{\tau,i}}$ are equivalent when

$$\theta_\tau = \tfrac{(1-2\tau)}{\tau(1-\tau)} \quad \text{and} \quad \phi_\tau = \tfrac{2}{\tau(1-\tau)}.$$

As an extension of this fact, $y_i \sim \mathrm{AL}(\mu, \sigma, \tau)$ can be represented by

$$y_i \stackrel{d}{=} \mu_\tau + \sigma\theta_\tau \upsilon_{\tau,i} + \sigma z_i \sqrt{\phi_\tau \upsilon_{\tau,i}},$$

which can re-parameterised as

$$y_i \stackrel{d}{=} \mu_\tau + \theta_\tau w_{\tau,i} + z_i \sqrt{\sigma\phi_\tau w_{\tau,i}},$$

where $w_{\tau,i}$ follows the exponential distribution with mean $\sigma$. We have $z_i \sim \mathrm{N}(0,1)$, so it follows that

$$\begin{aligned} z_i \sqrt{\sigma\phi_\tau w_{\tau,i}} &\sim \mathrm{N}(0, \sigma\phi_\tau w_{\tau,i}) \\ \mu_\tau + \theta_\tau w_{\tau,i} + z_i \sqrt{\sigma\phi_\tau w_{\tau,i}} &\sim \mathrm{N}(\mu_\tau + \theta_\tau w_{\tau,i}, \sigma\phi_\tau w_{\tau,i}) \end{aligned}$$

This implies that

$$y_i \sim \mathrm{N}\left(\mu_\tau + \theta_\tau w_{\tau,i}, \phi_\tau \sigma w_{\tau,i}\right). \tag{1.12}$$

Then, they applied this mixture of distributions, along with a prior distribution of double exponential for quantile coefficients, to fit linear quantile regression. Also, they extended the proposed method to draw inferences about Tobit quantile curves. By using this Gibbs sampler, all quantile coefficients are updated

jointly. Consequently, an important reduction in autocorrelation can be achieved compared to the Gibbs sampler proposed by Tsionas (2003). In addition, this Gibbs algorithm enjoys flexible and easy implementation. However, similarly to the Gibbs sampler developed by Tsionas (2003), this Gibbs algorithm could slow down as the size of sample becomes large. This Bayesian quantile method was extended by Lum & Gelfand (2012) to fit spatial quantile regression. Moreover, they proposed the asymmetric Laplace predictive process. Although they estimated quantile functions separately, they argued that the proposed conditional quantile model stochastically increases in $\tau$.

### 1.6.3 Quantile method using an infinite mixture of Gaussian distributions

Reich *et al.* (2010) developed a Bayesian quantile method based on an infinite mixture of Gaussian distributions. They assumed that the residual $\epsilon_i$ follows the infinite mixture of distributions given by

$$f(\epsilon_i|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{w}) = \sum_{l=1}^{\infty} w_l g(\epsilon_i|\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2, q_l),$$

where the $w_l$ is a weight such that $\sum_{l=1}^{\infty} w_l = 1$ and

$$g(\epsilon_i|\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2, q_l) = q_l \phi(\epsilon_i|\mu_{l1}, \sigma_{l1}^2) + (1 - q_l)\phi(\epsilon_i|\mu_{l2}, \sigma_{l2}^2),$$

where $\phi(.)$ is the normal density and $q_l$ is a weight calculated such that

$$\int_{-\infty}^{0} g(\epsilon_i|\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2, q_l) d\epsilon_i = \tau.$$

Also, they extended this approach to fit quantile regression for clustered data. They concluded that under a variety of distributions, the proposed method can perform better than traditional frequentist methods in terms of mean squared errors and coverage probabilities.

### 1.6.4 Multiple quantile regression based on the generalisation of Jeffreys's substitution likelihood

Lavine (1995) generalised the substitution likelihood proposed by Jeffreys (1961) to accommodate a vector of quantile functions given by $\boldsymbol{\mu}_\tau = (\mu_{\tau_1}, \mu_{\tau_2}, ..., \mu_{\tau_m})$ rather than just the median. This generalisation is given by

$$l_s(\mu_\tau; \boldsymbol{y}) \propto \binom{n}{\eta(\mu_{\tau_1})...\eta(\mu_{\tau_{m+1}})} \prod_{k=1}^{m+1} \Delta\tau_k^{\eta(\mu_{\tau_k})}, \tag{1.13}$$

where $n$ is the number of observations, $0 = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m < \tau_{m+1} = 1$, $\mu_{\tau_0} = -\infty, \mu_{\tau_{m+1}} = +\infty$, $\Delta\tau = (\tau_1, \tau_2 - \tau_1, ..., 1 - \tau_m)'$ and $\eta(\mu_{\tau_k}) = \sum_{i=1}^{n} 1_{(\mu_{\tau_{k-1}} < y_i \leq \mu_{\tau_k})}$. Jeffreys (1961) suggested that applying $l_s(\mu_\tau; \boldsymbol{y})$ instead of the likelihood function can yield a valid uncertainty. Also, the validity of Jeffreys's substitution likelihood was investigated by Monahan & Boos (1992) and they suggested that the posterior distributions formed using Jeffreys's substitution likelihood are not valid by coverage. However, they showed that by large-sample approximation of likelihood, Jeffreys's substitution likelihood is asymptotically valid. Moreover, Lavine (1995) suggested that $l_s(\mu_\tau; \boldsymbol{y})$ is asymptotically conservative at the truth. Dunson & Taylor (2005) applied this generalised substitution likelihood to construct a Bayesian quantile method that allows the drawing of Bayesian inference about multiple linear quantile curves simultaneously without crossing. The substitution likelihood $l_s(\mu_\tau; \boldsymbol{y})$ is a constant as $\mu_{\tau_1}$ and $\mu_{\tau_m}$ approach $-\infty$ and $\infty$ respectively. Therefore, to achieve a proper posterior distribution, Dunson & Taylor (2005) used a bounded prior distribution, which can ensure that quantile functions monotonically increase in $\tau$. Then, to simulate from the posterior distribution, they used a Metropolis-Hastings algorithm with a proposal distribution being a normal approximation for quantile regression coefficients. Under different distributions, Dunson & Taylor (2005) concluded that the proposed method shows a low bias of estimation and reasonable coverage probability. Also, they concluded that this method appeared to be non-sensitive to the number and locations of quantiles included in simultaneous estimation.

Dunson & Taylor (2005) implemented the investigation of the bias and the coverage probability for the method based on the generalised Jeffreys's substitution likelihood using simulations from univariate distributions. Therefore, further

work may be needed to check the performance of this method to handle linear regression or more complex models. In addition, more investigation may be required to examine the computational stability of the Bayesian method based on this generalised substitution likelihood.

### 1.6.5 Quantile regression based on an empirical likelihood

Lancaster & Jae Jun (2010) employed exponentially tilted empirical likelihood to develop a Bayesian method to draw inferences on quantile curves individually. They considered the posterior distribution given by

$$p(\mu_\tau|\boldsymbol{y}) \propto \pi(\mu_\tau) \prod_{i=1}^{n} \frac{\exp\{\lambda' g(y_i, \mu_\tau)\}}{\sum_{i=1}^{n} \exp\{\lambda' g(y_i, \mu_\tau)\}},$$

where $\mu_\tau$ is $\tau$ quantile, $\pi(.)$ is the prior distribution, and

$$g(y_i, \mu_\tau) = \begin{cases} 1-\tau & y_i \leq \mu_\tau, \\ -\tau & y_i > \mu_\tau. \end{cases}$$

Also, the vector $\lambda$ is the solution of the following optimisation problem:

$$\min_{\lambda} \frac{1}{n} \sum_{i=1}^{n} \exp\{\lambda' g(y_i, \mu_\tau)\}.$$

For quantile regression, they showed the asymptotic form of the posterior distribution. They concluded that inference from the posterior distributions based on the substitution likelihood proposed by Jeffreys (1961) and exponentially tilted empirical likelihood are similar, although the ratio of the two posterior distributions suggested that there is a difference asymptotically.

### 1.6.6 Quantile regression based on an approximate likelihood

Hahn & Burgette (2012) proposed an approximate likelihood to construct a Bayesian method to fit multiple quantiles corresponding to the vector $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$. They considered the approximate distribution that can be

written as

$$f_\tau\left(y_i|\mu_\tau\right)=\begin{cases}\frac{\tau_2-\tau_1}{\left(\mu_{\tau_2}-\mu_{\tau_1}\right)}\exp\left(-\frac{\tau_2-\tau_1}{\tau_1\left(\mu_{\tau_2}-\mu_{\tau_1}\right)}\left|y_i-\mu_{\tau_1}\right|\right), & y_i\leq\mu_{\tau_1},\\ \frac{\tau_k-\tau_{k-1}}{\mu_{\tau_k}-\mu_{\tau_{k-1}}}, & \mu_{\tau_{k-1}}<y_i\leq\mu_{\tau_k},\\ \frac{\tau_m-\tau_{m-1}}{\left(\mu_{\tau_m}-\mu_{\tau_{m-1}}\right)}\exp\left(-\frac{\tau_m-\tau_{m-1}}{(1-\tau_m)\left(\mu_{\tau_m}-\mu_{\tau_{m-1}}\right)}\left|y_i-\mu_{\tau_m}\right|\right), & y_i>\mu_{\tau_m}.\end{cases}$$

Then, they employed Gaussian process priors for quantile regression coefficients along with this approximate distribution to suggest a flexible solution for non-linear quantile regression. They applied a random walk Metropolis-within-Gibbs to sample from the posterior distribution.

To fit the linear model given in (1.9), Feng *et al.* (2015) developed a Bayesian quantile approach using an approximate distribution that is represented by the density function:

$$f_\tau\left(y_i|\boldsymbol{\beta}_\tau,\sigma^2\right)=\begin{cases}\tau_1\widehat{f}_L(y_i|\beta_{\tau_1},\sigma^2), & y_i\leq\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_1},\\ \sum_{k=1}^{m-1}\frac{\tau_{k+1}-\tau_k}{\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_{k+1}}-\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}}, & \boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}<y_i\leq\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_{k+1}},\\ (1-\tau_m)\,\widehat{f}_R(y_i|\beta_{\tau_m},\sigma^2), & y_i>\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_m},\end{cases}$$

where $\widehat{f}_L(y_i|\beta_{\tau_1},\sigma^2)$ is the left half of a normal distribution with mean $\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_1}$ and variance $\sigma^2$, and $\widehat{f}_R(y_i|\beta_{\tau_m},\sigma^2)$ is the right half of a normal distribution with mean $\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_m}$ and variance $\sigma^2$. Then, they developed a Bayesian quantile approach to estimate linear quantile functions simultaneously. A truncated normal distribution that ensures that quantile functions monotonically increase in $\tau$ at each observation is used as a prior for quantile linear coefficients. They concluded that their proposed MCMC converges to the target distribution as the number of quantiles $m$ approaches infinity. However, they mentioned that this method requires intensive computation. Also, they suggested that theoretical justifications are needed for some assumptions used by this method.

## 1.7 The contributions of our project

This research has several contributions. Firstly, we developed a Bayesian quantile method based on the pseudo asymmetric Laplace likelihood, that accommodates the Box-Cox transformation, to estimate quantile curves of non-linear

models with heteroscedastic errors without crossing. Then, this work extended to Box-Cox quantile regression with heteroscedastic error and two-sided Box-Cox quantile regression. Secondly, we proposed a new distribution called the generalised Gumbel distribution that shows more flexibility to approximate the underlying distribution than the asymmetric Laplace distribution. Then, the likelihood function corresponding to this proposed distribution was used to construct a Bayesian quantile approach to fit quantile regressions individually and provide better Bayesian inference. Thirdly, we developed a novel approximation of the joint distribution of multiple quantiles that was obtained by weighting the product of density functions corresponding to individual quantiles and can approximate the true distribution of the data well. Then, this approximation was used to estimate multiple quantile functions simultaneously without crossing, and obtain accurate and reliable Bayesian inference on quantiles' coefficients. Finally, to overcome the limitations of using the asymmetric Laplace distribution for error distribution, we considered a family of approximate distributions that can be used to model the errors in the context of quantile regression. This family was utilised to develop Bayesian quantile estimation for linear models with homoscedastic and heteroscedastic errors. In addition, we proposed a bootstrap test to check the fit of quantile regression.

## 1.8 Thesis outline

In Chapter 2, we use the Box-Cox transformation to handle nonlinearity and heteroscedasticity of the Bayesian quantile linear model considered by Yu & Moyeed (2001). To be able to estimate multiple quantiles simultaneously, we consider using the pseudo asymmetric Laplace likelihood which is formed by multiplying the asymmetric Laplace likelihoods corresponding to the individual quantiles. In addition, to ensure non-crossing of quantile functions, we specify suitable prior distributions for all parameters of the model. We propose a Gibbs algorithm with a Metropolis-Hastings step to sample more efficiently from the posterior distribution. Then, we use simulated and real data to compare simultaneous and individual estimates of quantile curves for non-linear models with heteroscedastic errors. To handle more complex models, this approach is extended in Chapter 3

to Box-Cox quantile regression with heteroscedastic errors and two-sided Box-Cox quantile regression.

In Chapter 4, we review the use of the asymmetric Laplace distribution for Bayesian estimation of quantile curves and illustrate its limitations as an error distribution. Also, we show that this Bayesian quantile approach can have poor coverage probabilities. Then, we suggest an alternative quantile model based on a generalisation of the Gumbel distribution that can approximate a variety of underlying distributions better than the asymmetric Laplace distribution. In addition, we use simulated data to implement a comparison study between the new quantile approach and that based on the asymmetric Laplace likelihood.

In Chapter 5, we discuss the quality of the approximation based on the pseudo asymmetric Laplace likelihood considered in Chapters 2 & 3 and explain how and why this approximation can be poor. Then, to improve the performance of this approximation, we develop a weighted pseudo asymmetric Laplace distribution, where the weights determine the contribution of each individual quantile-related asymmetric Laplace distribution in forming the underlying distribution. We consider two types of weights which are fixed weights and the estimated weights. Also, we consider an automated random walk Metropolis-Hastings based on a multivariate normal distribution with an updated covariance matrix to sample from the posterior distribution. We implement a simulation study to investigate the performance of Bayesian quantile approaches based on the weighted pseudo asymmetric Laplace likelihood and compare the obtained results with others given by the quantile method based on the pseudo asymmetric Laplace likelihood. Moreover, the proposed methods are used to fit multiple quantile regressions for real data.

In Chapter 6, we propose a family of approximate likelihood functions that can be used to construct Bayesian models to estimate quantile functions. We discuss the properties of these approximations and show how these approximations can improve the Bayesian inference for the quantile functions. Then, we use this family of approximations to develop Bayesian quantile approaches to fit homoscedastic and heteroscedastic linear regressions. In addition, we proposed a new bootstrap test to examine the goodness-of-fit of quantile regression.

In Chapter 7, we provide a summary of this research and we implement comparisons between the proposed methods in terms of the maximum a posteriori

(MAP) estimators and the 95% highest posterior density intervals (HPD). Also, we compare the proposed quantile approaches with the method considered by Yu & Moyeed (2001). Finally, we illustrate some potential future works.

## 1.9 Conclusion

The construction of efficient quantile regression models requires addressing issues related to the likelihood choice; specification of suitable priors over the model space for all components of the quantile regression coefficients; posterior computation; sensitivity to the number and locations of quantile functions included in the estimation; and crossing of quantile functions. However, the main aim in this research is to develop flexible likelihood functions that can be used to improve Bayesian inference for quantile functions.

# Chapter 2

# Simultaneous Box-Cox Quantile Regression

## 2.1 Introduction

The Bayesian linear quantile model considered by Yu & Moyeed (2001) assumes that the relationship between the response variable and the covariates is linear with homoscedastic errors. To make this model more flexible, transformation techniques can be used to address violations of these assumptions.

Transformation methods are widely considered as a useful statistical technique for addressing two violations of the linear regression model: non-linearity and heteroscedasticity. However, in order to benefit fully from the transformation, a suitable transformation for the response variable needs to be determined from the data. This can be achieved by applying the Box-Cox transformation proposed by Box & Cox (1964). The Box-Cox transformation includes as special cases the most popular transformations, such as linear, log, polynomial and square-root transformations. Box-Cox regression relaxes an assumption that the relationship between the response variable and the covariates is linear with the errors being independently normally distributed, with a mean equal to zero and constant variance. Thus, the Box-Cox regression model is considered as a great extension of linear regression.

In classical estimation, there are a number of methods developed to employ the Box-Cox transformation in the context of quantile regression. Powell (1991) used the Box-Cox transformation to develop an extension of linear quantile re-

gression. Chamberlain (1994) implemented Box-Cox quantile regression utilising an algorithm that includes linear programming and a one-dimensional search. Machado & Mata (2000) showed how Box-Cox quantile regression can be applied to describe the relationship between the industry attributes and the firm size. To overcome the problem of computing the inverse of the Box-Cox transformation which could be undefined for some observations, Fitzenberger *et al.* (2009) employed a modified version of the objective function considered by Buchinsky (1995) to suggest an estimator dealing with inadmissible observations. In addition, they generalised this result to include the case of multiple regression.

In this chapter, we employ the Box-Cox transformation to develop a Bayesian method based on a pseudo asymmetric Laplace likelihood to fit multiple quantile regression. Also, we specify convenient prior distributions for all parameters including the transformation parameter. Moreover, the issue of the crossing of quantile curves is discussed and a solution, based on prior distributions, is considered to ensure non-crossing of quantile functions over an unbounded space of covariates.

## 2.2 Box-Cox transformation

In linear regression analysis, it is often assumed that the relationship between the response variable and the covariates is linear with errors that are normally distributed with a constant variance. To make these assumptions more flexible, Box & Cox (1964) proposed the following family of power transformations:

$$\Lambda\left(y_i;\lambda\right)=\begin{cases}\frac{y_i^{\lambda}-1}{\lambda}, & \lambda\neq 0,\\ \log\left(y_i\right), & \lambda=0,\end{cases} \tag{2.1}$$

where $y_i>0$ for $i=1,...,n$ and $\lambda\in\mathbb{R}$ is the transformation parameter. To be able to compare the residual sum of squares obtained with different values of $\lambda$, Box & Cox (1964) suggested the standardized transformations:

$$\Lambda\left(y_i;\lambda\right)=\begin{cases}\frac{y_i^{\lambda}-1}{\lambda\widetilde{y}^{\lambda-1}}, & \lambda\neq 0,\\ \widetilde{y}\log\left(y_i\right), & \lambda=0,\end{cases} \tag{2.2}$$

where $\widetilde{y}$ is the geometric mean, given by $\widetilde{y} = \left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}}$. The Box-Cox transformation includes a variety of standard transformations as special cases. For example, $\lambda = -1, 0, 0.5$, implies reciprocal, logarithmic and square-root transformations respectively (for more examples, see Figure 2.1). Since the Box-Cox transformation is valid only for $y_i > 0$, for all $i = 1, 2, ..., n$, Box & Cox (1964) also proposed the shifted transformation:

$$\Lambda\left(y_i; \lambda, \phi\right) = \begin{cases} \frac{(y_i+\phi)^{\lambda}-1}{\lambda \widetilde{y}^{\lambda-1}}, & \lambda \neq 0, \\ \widetilde{y} \log\left(y_i + \phi\right), & \lambda = 0, \end{cases}$$

where $\phi$ is selected such that $y_i + \phi > 0$, for all $i = 1, 2, ..., n$ and $\widetilde{y}$ in this case is given by $\left(\prod_{i=1}^{n}(y_i + \phi)\right)^{\frac{1}{n}}$. For more details about alternative forms of the Box-Cox transformation, see Sakia (1992).

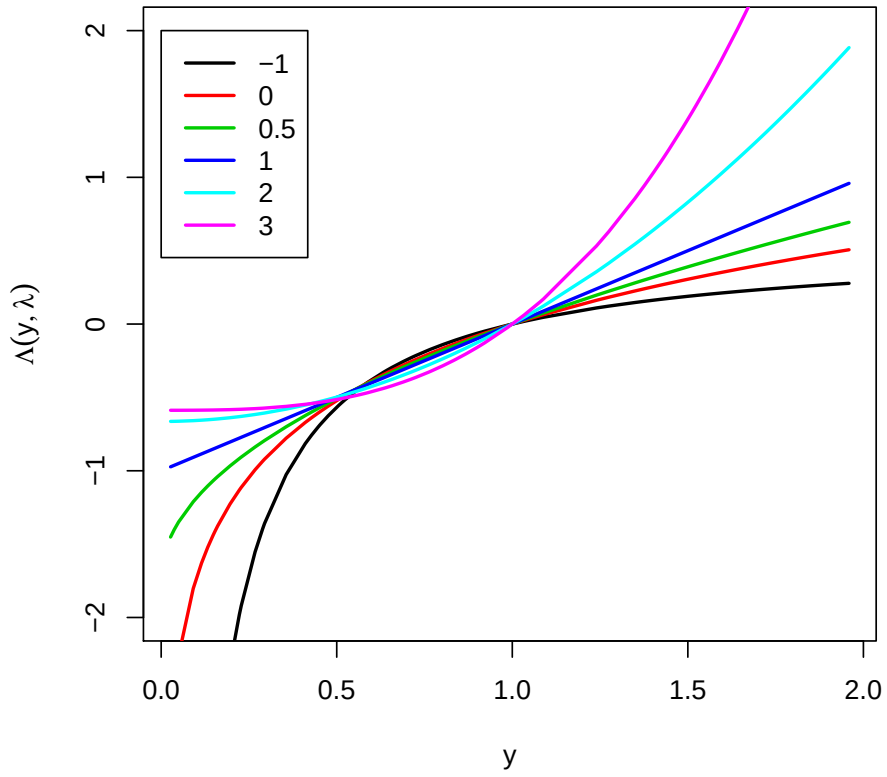



Figure 2.1: Box-Cox transformation corresponding to $\lambda = -1, 0, 0.5, 1, 2, 3$.

## 2.3 Box-Cox quantile regression

In the context of quantile regression, the Box-Cox transformation is applied in order to achieve the assumption of linearity and homoscedastic errors. After applying the Box-Cox transformation to the response variable, the regression model can be written as

$$\Lambda\left(y_i;\lambda\right)=\alpha+\boldsymbol{x}_i^{'}\boldsymbol{\beta}+\epsilon_i,\ \text{ for } i=1,2,...,n, \tag{2.3}$$

where $\epsilon_i$s are independently distributed with a mean equal to zero and constant variance; $\boldsymbol{x}_i$ is a $p\times 1$ vector of covariates for the $i^{\text{th}}$ observation; $\alpha$ is an unknown intercept coefficient and $\boldsymbol{\beta}$ is a $p\times 1$ vector of unknown slope coefficients. Then, the conditional quantile function is given by

$$Q(\tau_k|\boldsymbol{x}_i)=\Lambda^{-1}\left(\alpha_{\tau_k}+\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}|\lambda_{\tau_k}\right), \tag{2.4}$$

where $0<\tau_k<1$. Then, the $\tau_k$ quantile function can be estimated by solving the optimisation problem:

$$\min_{\alpha_{\tau_k},\boldsymbol{\beta}_{\tau_k},\lambda_{\tau_k}}\sum_{i=1}^{n}\rho_{\tau_k}\left(\Lambda\left(y_i;\lambda_{\tau_k}\right)-\alpha_{\tau_k}-\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}\right),$$

where $\rho_{\tau_k}$ is the loss function given in (1.2). This estimator is a consistent and asymptotically normal estimator (Powell, 1991). Moreover, Buchinsky (1995) proposed an iterative procedure, with two steps in each iteration, to estimate $\alpha_{\tau_k}$, $\boldsymbol{\beta}_{\tau_k}$ and $\lambda_{\tau_k}$:

1. Estimate $\alpha_{\tau_k}$ and $\boldsymbol{\beta}_{\tau_k}$ by solving

$$\min_{\alpha_{\tau_k},\boldsymbol{\beta}_{\tau_k}}\frac{1}{n}\sum_{i=1}^{n}\rho_{\tau_k}\left(\Lambda\left(y_i;\lambda_{\tau_k}\right)-\alpha_{\tau_k}-\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}\right),$$

   where $\Lambda\left(y_i;\lambda_{\tau_k}\right)$ is the Box-Cox transformation form given in (2.1).

2. Estimate $\lambda_{\tau_k}$ using

$$\lambda_{\tau_k}=\arg\min_{\lambda_{\tau_k}}\frac{1}{n}\sum_{i=1}^{n}\rho_{\tau}\left(y_i-\left[\lambda_{\tau_k}\left(\alpha_{\tau_k}+\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}\right)+1\right]^{1/\lambda_{\tau_k}}\right). \tag{2.5}$$

However, the optimization problem given in (2.5) can only be solved if

$$\lambda_{\tau_k}\left(\alpha_{\tau_k}+\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}\right)+1>0, \tag{2.6}$$

for each $i=1,2,...,n$ and $\lambda_{\tau_k}\in\mathbb{R}$. There are a number of suggestions to overcome this problem. Fitzenberger *et al.* (2009) used a defined set of admissible observations that satisfy the condition given in (2.6) for all $\lambda_{\tau_k}\in\mathbb{R}$, where this set of observations varies with different $\alpha_{\tau_k}$ and $\boldsymbol{\beta}_{\tau_k}$; for further discussions, see Manly (1976) and John & Draper (1980). However, this computational issue can be avoided by applying the proposed Bayesian Box-Cox quantile regression described in the next section, since there is no requirement to calculate the inverse of the Box-Cox transformation for the quantity given in (2.6) during the estimation process.

## 2.4 Bayesian model for Box-Cox quantile regression

### 2.4.1 Likelihood construction

To develop a Bayesian method to estimate an individual quantile curve, we can consider the asymmetric Laplace likelihood with the Box-Cox transformation (AL-BC) that is given by

$$\begin{aligned} l\left(\alpha_{\tau_k},\boldsymbol{\beta}_{\tau_k},\sigma_{\tau_k},\lambda_{\tau_k};\boldsymbol{y}\right) &\propto \left[\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right]^n \exp\left\{-\sum_{i=1}^{n}\rho_{\tau_k}\left(\frac{\Lambda\left(y_i;\lambda_{\tau_k}\right)-\alpha_{\tau_k}-\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}}{\sigma_{\tau_k}}\right)\right\} \\ &\propto \sigma_{\tau_k}^{-n}\exp\left\{-\sum_{i=1}^{n}\rho_{\tau_k}\left(\frac{\Lambda\left(y_i;\lambda_{\tau_k}\right)-\alpha_{\tau_k}-\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}}{\sigma_{\tau_k}}\right)\right\}, \end{aligned} \tag{2.7}$$

where $\Lambda\left(y_i;\lambda_{\tau_k}\right)$ is the Box-Cox transformation defined in (2.2). To understand the reason behind using the asymmetric Laplace distribution to construct a Bayesian model to fit quantile regression, see Sections 1.2 & 1.4. To improve the estimation and avoid the crossing of quantile functions, we also develop a Bayesian multiple quantile method using the pseudo asymmetric Laplace likeli-

hood with Box-Cox transformation (PAL-BC) that is given by

$$l\left(\boldsymbol{\alpha}_\tau, B_\tau, \boldsymbol{\sigma}_\tau, \boldsymbol{\lambda}_\tau; \boldsymbol{y}\right) \propto \left(\prod_{k=1}^{m} \sigma_{\tau_k}^{-n}\right) \exp\left\{-\sum_{k=1}^{m}\sum_{i=1}^{n} \rho_{\tau_k}\left(\frac{\Lambda\left(y_i;\lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}_{\tau_k}}{\sigma_{\tau_k}}\right)\right\}, \quad (2.8)$$

where $0 < \tau_1 < ... < \tau_m < 1$; $\boldsymbol{\alpha}_\tau = (\alpha_{\tau_1}, ..., \alpha_{\tau_m})$; $B_\tau = \left[\begin{array}{cccc} \boldsymbol{\beta}_{\tau_1} & \boldsymbol{\beta}_{\tau_2} & ... & \boldsymbol{\beta}_{\tau_m} \end{array}\right]_{p\times m}$ such that $\boldsymbol{\beta}_{\tau_k} = (\beta_{1,\tau_k}, ..., \beta_{p,\tau_k})'$ and $\boldsymbol{\lambda}_\tau = (\lambda_{\tau_1}, ..., \lambda_{\tau_m})$. Maximising this pseudo likelihood is equivalent to minimising a weighted version of the original optimisation problem given by

$$\min_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} \sum_{k=1}^{m}\sum_{i=1}^{n} \rho_{\tau_k}\left(\Lambda\left(y_i;\lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}_{\tau_k}\right), \quad (2.9)$$

with weights being $1/\sigma_{\tau_k}$ for $k = 1, .., m$ (for more details see Appendix A.1).

### 2.4.2 The crossing issue and specification of prior distributions

To overcome the crossing issue, the conditional quantile functions, for Box-Cox regression, given in (2.4) should satisfy

$$Q(\tau_k|\boldsymbol{x}_i) < Q(\tau_{k+1}|\boldsymbol{x}_i),$$

for all $k = 1, 2, ..., m-1$ and all $i = 1, ..., n$. Then, using Box-Cox transformation given in (2.2), it follows that

$$\left[\widetilde{y}^{\lambda_{\tau_k}-1}\lambda_{\tau_k}\left(\alpha_{\tau_k} + \boldsymbol{x}_i'\boldsymbol{\beta}_{\tau_k}\right) + 1\right]^{1/\lambda_{\tau_k}} < \left[\widetilde{y}^{\lambda_{\tau_{k+1}}-1}\lambda_{\tau_{k+1}}\left(\alpha_{\tau_{k+1}} + \boldsymbol{x}_i'\boldsymbol{\beta}_{\tau_{k+1}}\right) + 1\right]^{1/\lambda_{\tau_{k+1}}}, \quad (2.10)$$

It is obvious that the region of interest for non-crossing is bounded and depends on $\boldsymbol{x}_i$. Beside this disadvantage, applying this constraint can lead to unreliable estimation and computational issues as the number of observations satisfying this constraint is extremely low. However, the region of non-crossing can be extended to be the whole space of covariates, which is assumed to be unbounded, by specifying suitable prior distributions for *all* parameters rather than applying the constraint given in (2.10). To begin with the transformation parameter, allowing different transformations for different levels of quantiles implies each transformation parameter is estimated using the information conveyed through

its corresponding quantile function. Consequently, due to the lack of observed information at some quantile levels especially the extreme ones, the true transformation may be misspecified. Therefore, in order to combine information over multiple quantile functions to estimate the optimal transformation parameter, we assume a priori that all transformation parameters $\lambda_{\tau_k}$ are equal with common value $\lambda$. Then, we consider the following prior distribution:

$$p(\lambda) \propto \exp\left\{-\frac{1}{2v^2}\left(\lambda - \lambda^*\right)^2\right\}$$

where $\lambda^*, v$ are specified parameters. Also, by definition of linear quantile regression, the slope coefficients are identical over all quantile levels and represented by the vector $\boldsymbol{\beta}$. Hence, the prior distribution for the slope coefficients are specified as follows

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^*\right)' S^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^*\right)\right\}$$

where $\boldsymbol{\beta}^*, S$ are the specified mean vector and covariance matrix respectively. In addition, the quantile functions should increase in $\tau$. Therefore, conditional on the prior distributions for $\boldsymbol{\beta}$ and $\lambda$, the intercept coefficients should satisfy

$$\alpha_{\tau_k} < \alpha_{\tau_{k+1}} \text{ for } k = 1, 2, ..., m-1.$$

Hence, the prior distribution for intercept coefficients is specified as

$$p(\boldsymbol{\alpha}_\tau) \propto \exp\left\{-\frac{1}{2}\sum_{k=1}^{m}\frac{\left(\alpha_{\tau_k} - \alpha^*_{\tau_k}\right)^2}{c^2_{\tau_k}}\right\} 1_{\left(\alpha_{\tau_1} < \alpha_{\tau_2} < ... < \alpha_{\tau_m}\right)},$$

where $\alpha^*_{\tau_k}, c_{\tau_k}$are specified parameters. It is explicit that by using this specification of prior distributions, the solution of the inequality given in (2.10) does not depend on $\boldsymbol{x}_i$. Hence, the region of non-crossing determined by this condition is extended to be the whole space of covariates, which is assumed to be unbounded. The element of $\boldsymbol{\sigma}_\tau$ are assumed a priori independent with inverse-gamma distributions $\text{IG}(a_{\tau_k}, b_{\tau_k})$, so we can write

$$p(\boldsymbol{\sigma}_\tau) \quad \propto \quad \left(\prod_{k=1}^{m}\sigma_{\tau_k}^{-\left(a_{\tau_k}+1\right)}\right)\exp\left\{-\sum_{k=1}^{m}\frac{b_{\tau_k}}{\sigma_{\tau_k}}\right\},$$

where $a_{\tau_k}, b_{\tau_k}$ are specified parameters.

### 2.4.3 Posterior computation

The posterior distribution based on the pseudo asymmetric Laplace likelihood and the prior distributions given in the previous section is

$$p\left(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau, \lambda | \boldsymbol{y}\right) \propto \left(\prod_{k=1}^{m} \sigma_{\tau_k}^{-\left(n+a_{\tau_k}+1\right)}\right) \exp\left\{-\sum_{k=1}^{m}\sum_{i=1}^{n} \rho_{\tau_k}\left(\frac{\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma_{\tau_k}}\right)\right\}$$
$$\times \exp\left\{-\frac{\left(\lambda-\lambda^*\right)^2}{2v^2}\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)' S^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)\right\}$$
$$\times \exp\left\{-\sum_{k=1}^{m}\frac{b_{\tau_k}}{\sigma_{\tau_k}}\right\} \exp\left\{-\frac{1}{2}\sum_{k=1}^{m}\frac{\left(\alpha_{\tau_k}-\alpha_{\tau_k}^*\right)^2}{c_{\tau_k}^2}\right\} 1_{\left(\alpha_{\tau_1}<\ldots<\alpha_{\tau_m}\right)}.$$

To obtain a sample from this posterior distribution, we can apply the Metropolis-Hastings algorithm given in Algorithm 1. The hyperparameters $\Sigma_{\boldsymbol{\alpha}_\tau}$, $\Sigma_{\boldsymbol{\beta}}$, $\sigma_{\tau_k}$ and $\nu_\lambda$ are specified during the initial implementation when MCMC algorithm is monitored to ensure its convergence with an acceptance probability of approximately 0.234 (for more details, see Roberts *et al.*, 1997). Although the Metropolis-

**Algorithm 1** Metropolis-Hastings

**1. set initial values**
$r = 1$, $\boldsymbol{\alpha}_\tau^{(0)} = \hat{\boldsymbol{\alpha}}_\tau$, $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}$, $\boldsymbol{\sigma}_\tau^{(0)} = \hat{\boldsymbol{\sigma}}_\tau$, $\lambda^{(0)} = \widehat{\lambda}$.

**while** $r \leq R$

**2. generate candidate values from proposal distributions**
$\boldsymbol{\alpha}_\tau^\bullet \overset{ind}{\sim} \mathrm{N}_m(\boldsymbol{\alpha}_\tau^{(r-1)}, \Sigma_{\boldsymbol{\alpha}_\tau})$.
$\boldsymbol{\beta}^\bullet \overset{ind}{\sim} \mathrm{N}_p(\boldsymbol{\beta}^{(r-1)}, \Sigma_{\boldsymbol{\beta}})$.
$\sigma_{\tau_k}^\bullet \sim \mathrm{LN}(\sigma_{\tau_k}^{(r-1)}, \nu_{\sigma_{\tau_k}})$, for all $k$, where LN is the log-normal.
$\lambda^\bullet \sim \mathrm{N}(\lambda^{(r-1)}, \nu_\lambda)$.

**3. calculate**
$$\Lambda\left(y;\lambda\right) = \begin{cases} \frac{y^{\lambda^\bullet}-1}{\lambda^\bullet \widetilde{y}^{\lambda^\bullet-1}} & \lambda^\bullet \neq 0 \\ \widetilde{y}\log\left(y\right) & \lambda^\bullet = 0 \end{cases}$$

**4. generate**
$u \sim \mathrm{U}(0,1)$, where U is the uniform distribution.

**5. calculate**
$$\triangle = \frac{p(\boldsymbol{\alpha}_\tau^\bullet, \boldsymbol{\beta}^\bullet, \boldsymbol{\sigma}_\tau^\bullet, \lambda^\bullet | y)}{p\left(\boldsymbol{\alpha}_\tau^{(r-1)}, \boldsymbol{\beta}^{(r-1)}, \boldsymbol{\sigma}_\tau^{(r-1)}, \lambda^{(r-1)} \middle| y\right)} \frac{\prod_{k=1}^m \mathrm{LN}(\boldsymbol{\sigma}_{\tau_k}^{(r-1)} | \boldsymbol{\sigma}_{\tau_k}^\bullet)}{\prod_{k=1}^m \mathrm{LN}(\boldsymbol{\sigma}_{\tau_k}^\bullet | \boldsymbol{\sigma}_{\tau_k}^{(r-1)})}$$

**6. if** $u \leq \min(1, \triangle)$
**then** $\boldsymbol{\alpha}_\tau^{(r)} = \boldsymbol{\alpha}_\tau^\bullet$, $\boldsymbol{\beta}^{(r)} = \boldsymbol{\beta}^\bullet$, $\boldsymbol{\sigma}_\tau^{(r)} = \boldsymbol{\sigma}_\tau^\bullet$, $\lambda^{(r)} = \lambda^\bullet$.
**else** $\boldsymbol{\alpha}_\tau^{(r)} = \boldsymbol{\alpha}_\tau^{(r-1)}$, $\boldsymbol{\beta}^{(r)} = \boldsymbol{\beta}^{(r-1)}$, $\boldsymbol{\sigma}_\tau^{(r)} = \boldsymbol{\sigma}_\tau^{(r-1)}$, $\lambda^{(r)} = \lambda^{(r-1)}$.

**end**

Hastings algorithm offers a simple implementation, the specification of a suitable proposal distribution is very challenging due to the relatively large number of parameters in Bayesian multiple quantiles models. Also, the simulated samples obtained by applying the Metropolis-Hastings algorithm have high autocorrelation. This issue could be solved by choosing a better proposal though, and that is effectively what the Gibbs sampler is. Therefore, to improve the computational efficiency of the proposed method, we consider the Gibbs sampler, which requires deriving the full conditional distributions of all parameters included in the model. Since deriving the conditional distributions from the Bayesian model based on the asymmetric Laplace likelihood is intractable, the representation of the asymmetric Laplace distribution using a mixture of normal distributions considered by Kozumi & Kobayashi (2011), as shown in Section 1.6.2, is applied. Then, the likelihood function can be written as

$$l\left(\boldsymbol{\alpha}_{\tau}, B_{\tau}, W_{\tau}, \boldsymbol{\sigma}_{\tau}, \boldsymbol{\lambda}_{\tau}; \boldsymbol{y}\right) \propto \left(\prod_{k=1}^{m}\prod_{i=1}^{n} \sigma_{\tau_k}^{-\frac{1}{2}} w_{\tau_k,i}^{-\frac{1}{2}}\right) \exp\left\{-\sum_{k=1}^{m}\sum_{i=1}^{n} \frac{\left(\Lambda\left(y_i;\lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}_{\tau_k} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{2\phi_{\tau_k}\sigma_{\tau_k} w_{\tau_k,i}}\right\}.$$

where

$$\theta_{\tau_k} = \frac{(1-2\tau_k)}{\tau_k\left(1-\tau_k\right)}$$

and

$$\phi_{\tau_k} = \frac{2}{\tau_k(1-\tau_k)}.$$

Also, $w_{\tau_k,i}$ follows the exponential distribution with mean $\sigma_{\tau_k}$. Then, the posterior distribution, based on the mixture of normal likelihood functions and the specified prior distributions, is

$$\begin{aligned} p\left(\boldsymbol{\alpha}_{\tau}, \boldsymbol{\beta}, W_{\tau}, \boldsymbol{\sigma}_{\tau}, \lambda | \boldsymbol{y}\right) \propto & \left(\prod_{i=1}^{m}\prod_{i=1}^{n} \sigma_{\tau_k}^{-\frac{1}{2}} w_{\tau_k,i}^{-\frac{1}{2}}\right) \exp\left\{-\sum_{k=1}^{m}\sum_{i=1}^{n} \frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{2\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}}\right\} \\ & \times \left(\prod_{k=1}^{m} \sigma_{\tau_k}^{-\left(a_{\tau_k}+1\right)}\right) \exp\left\{-\sum_{k=1}^{m} \frac{b_{\tau_k}}{\sigma_{\tau_k}}\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)' S^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)\right\} \\ & \times \exp\left\{-\frac{\left(\lambda-\lambda^*\right)^2}{2v^2}\right\} \exp\left\{-\sum_{k=1}^{m} \frac{\left(\alpha_{\tau_k}-\alpha_{\tau_k}^*\right)^2}{2c_{\tau_k}^2}\right\} 1_{\left(\alpha_{\tau_1}<\alpha_{\tau_2}<\ldots<\alpha_{\tau_m}\right)}. \end{aligned}$$

After conditional distributions of all the parameters included in the model are derived as shown in Appendix A.2, the Gibbs sampler algorithm for the proposed

multiple Box-Cox quantile regressions is given in Algorithm 2. For more details about how the convergence of these MCMC methods are diagnosed see Appendix B.1.

**Algorithm 2** Gibbs sampler algorithm

**1. set initials**

$r = 1$, $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}$, $W_\tau^{(0)} = \widehat{W}_\tau$, $\boldsymbol{\sigma}_\tau^{(0)} = \hat{\boldsymbol{\sigma}}_\tau$, $\lambda^{(0)} = \widehat{\lambda}$.

**while** $r \leq R$

**2. generate** $\alpha_{\tau_k}^{(r)} | \boldsymbol{\beta}^{(r-1)}, \boldsymbol{w}_{\tau_k}^{(r-1)}, \sigma_{\tau_k}^{(r-1)}, \lambda^{(r-1)}$ using the conditional distribution: $\alpha_{\tau_k} | \boldsymbol{\beta}, \boldsymbol{w}_{\tau_k}, \sigma_{\tau_k}, \lambda \sim \mathrm{N}\left(\widetilde{\alpha}_{\tau_k}, \psi_{\tau_k}^2\right) 1_{\left(\alpha_{\tau_k} < \alpha_{\tau_{k+1}}\right)}$, where

$$\widetilde{\alpha}_{\tau_k} = \psi_{\tau_k}^2 \left( \frac{\alpha_{\tau_k}^*}{c_{\tau_k}^2} + \sum_{i=1}^{n} \frac{\left(\Lambda(y_i;\lambda) - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k,i}\right)}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right), \quad \psi_{\tau_k}^2 = \left( \frac{1}{c_{\tau_k}^2} + \sum_{i=1}^{n} \frac{1}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right)^{-1},$$

for all $k = 1, 2, ..., m$

**3. generate** $\boldsymbol{\beta}^{(r)} | \boldsymbol{\alpha}_\tau^{(r)}, W_\tau^{(r-1)}, \boldsymbol{\sigma}_\tau^{(r-1)}, \lambda^{(r-1)}$ using the conditional distribution: $\boldsymbol{\beta} | \boldsymbol{\alpha}_\tau, W_\tau, \boldsymbol{\sigma}_\tau, \lambda \sim \mathrm{N}\left(\widetilde{\boldsymbol{\beta}}, \Sigma\right)$, where

$$\widetilde{\boldsymbol{\beta}} = \Sigma \left( S^{-1}\boldsymbol{\beta}^* + \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\boldsymbol{x}_i\left(\Lambda(y_i;\lambda) - \alpha_{\tau_k} - \theta_{\tau_k} w_{\tau_k,i}\right)}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right), \quad \Sigma = \left( S^{-1} + \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\boldsymbol{x}_i\boldsymbol{x}_i'}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right)^{-1}.$$

**4. generate** $w_{\tau_k,i}^{(r)} | \alpha_{\tau_k}^{(r)}, \boldsymbol{\beta}^{(r)}, \sigma_{\tau_k}^{(r-1)}, \lambda^{(r-1)}$ using the conditional distribution:

$$w_{\tau_k,i} | \alpha_{\tau_k}, \boldsymbol{\beta}, \sigma_{\tau_k}, \lambda \sim \mathrm{GIG}\left( \frac{1}{2}, \sqrt{\frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k}}}, \sqrt{\frac{\theta_{\tau_k}^2}{\sigma_{\tau_k}\phi_{\tau_k}} + \frac{2}{\sigma_{\tau_k}}} \right),$$

where $GIG$ is the density of generalized inverse Gaussian distribution given in Appendix A.2.

**5. generate** $\sigma_{\tau_k}^{(r)} | \alpha_{\tau_k}^{(r)}, \boldsymbol{\beta}^{(r)}, \boldsymbol{w}_{\boldsymbol{\tau_k}}^{(r)}, \lambda^{(r-1)}$ using the conditional distribution: $\sigma_{\tau_k} | \alpha_{\tau_k}, \boldsymbol{\beta}, \boldsymbol{w}_{\boldsymbol{\tau_k}}, \lambda \sim \mathrm{IG}\left(\delta_0, \delta_1\right)$, where $IG$ is the density of the inverse-gamma distribution and

$$\delta_0 = \tfrac{3n}{2} + a_{\tau_k}, \quad \delta_1 = b_{\tau_k} + \sum_{i=1}^{n} \frac{\left(\Lambda(y_i;\lambda) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{2\phi_{\tau_k} w_{\tau_k,i}} + \sum_{i=1}^{n} w_{\tau_k,i}.$$

**6. update** $\lambda^{(r)} | \boldsymbol{\alpha}_\tau^{(r)}, \boldsymbol{\beta}^{(r)}, W_\tau^{(r)}, \boldsymbol{\sigma}_\tau^{(r)}$ using a one-step Metropolis-Hastings algorithm based on the conditional distribution:

$$p\left(\lambda | \boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, W_\tau, \boldsymbol{\sigma}_\tau\right) \propto \exp\left\{ -\sum_{i=1}^{n}\sum_{k=1}^{m} \left( \frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{2\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right) - \frac{\left(\lambda - \lambda^*\right)^2}{2v^2} \right\}.$$

## 2.5 Simulations and applications

In this section, we check and compare the performances of the individual quantile estimation method (AL-BC) and the simultaneous estimation method (PAL-BC) using simulations and real data. Mainly, we draw Bayesian inference on the first quartile, the median and the third quartile. In addition, we examine the proposed methods in terms of estimating the extreme quantiles corresponding to $\tau = 0.05$ and 0.95. To examine the proposed composite likelihood functions precisely, we may need to reduce the impact of prior distributions on the estimation. This can be achieved by considering weak prior distributions. Therefore, all approaches are constructed using the same diffuse proper prior distributions that lead to proper posterior distributions, each normal prior distribution has mean equal to zero and variance equal to $10^5$. Also, each multivariate normal distribution has a mean vector of zeros and a diagonal covariance matrix with entries equal to $10^5$. The inverse-gamma parameters are set equal to 0.01. To examine the performance of the proposed methods to handle the simulated and real datasets, we use a variety of goodness of fit measures. If the true quantile function $Q(\tau|.)$ is known, then the mean squared error ($\text{MSE}_\tau$) is given by

$$\text{MSE}_\tau = \frac{1}{n}\sum_{i=1}^{n}\left(Q(\tau|\boldsymbol{x}_i) - \overline{Q}(\tau|\boldsymbol{x}_i)\right)^2,$$

where $\Theta$ is a vector of the posterior parameters and $\overline{Q}$ is the estimated mean of the quantile function that is given by

$$\overline{Q}(\tau|\boldsymbol{x}_i) = \frac{1}{R}\sum_{r=1}^{R}\widehat{Q}(\tau|\boldsymbol{x}_i, \Theta_r),$$

where $R$ is the size of the sample simulated from the posterior distribution. Also, we use the mean absolute error ($\text{MAE}_\tau$)

$$\text{MAE}_\tau = \frac{1}{n}\sum_{i=1}^{n}\left|Q(\tau|\boldsymbol{x}_i) - \overline{Q}(\tau|\boldsymbol{x}_i)\right|.$$

For real data, we use the leave-one-out cross-validation estimation of the error,

which is given by

$$\text{QLCV}_\tau = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau \left( y_i - \overline{Q}^{-i}(\tau|\boldsymbol{x}_i) \right), \tag{2.11}$$

where $\overline{Q}^{-i}$ is the posterior mean of the quantile function that is estimated when the $i^{\text{th}}$ observation is left out. Since the loss function is more related to individual estimation, it could favour the individual estimation over the simultaneous estimation. Also, this loss function does not give the actual errors between the true observations and fitted values (that is, as happened in the case of ordinary least squares regression), it represents the errors weighted by either $\tau$ or $(1-\tau)$. However, it can be used to show the change in errors caused by using simultaneous estimation rather than the individual quantile estimations.

### 2.5.1 Simulated Data

In this section, we consider 200 observations from the model:

$$y_i = \log(1 + x_i + \epsilon_i), \tag{2.12}$$

where $x_i$ is generated from the uniform distribution U(1,6) and $\epsilon_i$ is simulated from the normal distribution $\text{N}(0, 0.7)$. The simulated data are shown in Figure 2.2.

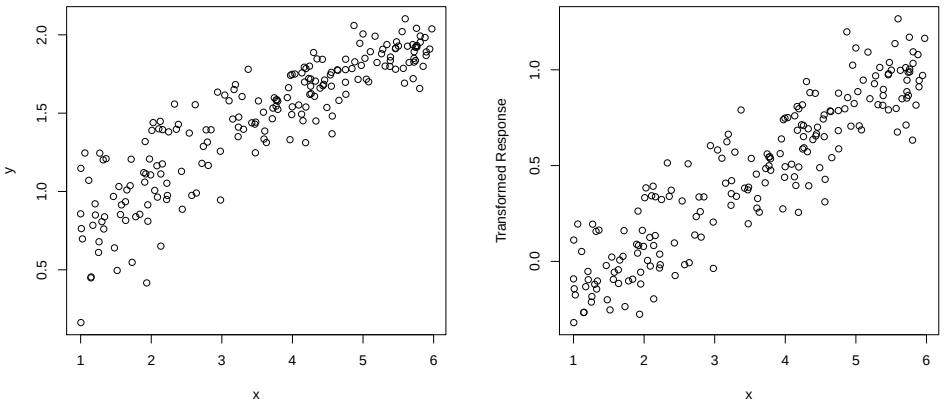



Figure 2.2: Scatter plot of the original response variable (left) and transformed response obtained using the PAL-BC method (right) against x.

Figure 2.2 shows the relationship between the response variable and the covariate to be non-linear with heteroscedastic errors. However, it also shows that after applying the PAL-BC method, the relationship between the transformed response variable $\Lambda(y;\lambda)$ with $\lambda = 2.13$ (that is, the posterior mean) and the covariate is linear with a constant variance. This suggests that the PAL-BC method can handle the violation of linearity and homoscedastic errors for this particular example. Figure 2.3 shows that the AL-BC method suggests similar transformations to that obtained by the PAL-BC method in the case of $\tau = 0.25, 0.5$ with $\lambda_{0.25} = 2.25$ and $\lambda_{0.5} = 2.05$, and a slightly different transformation in the case $\tau = 0.75$ with $\lambda_{0.75} = 1.72$. To provide more explanation about this difference in estimating the transformations and its effect on the estimation of quantile functions, we implement a simulation study using 300 samples, that just differ in the errors $\epsilon_i$, from the model given in (2.12).

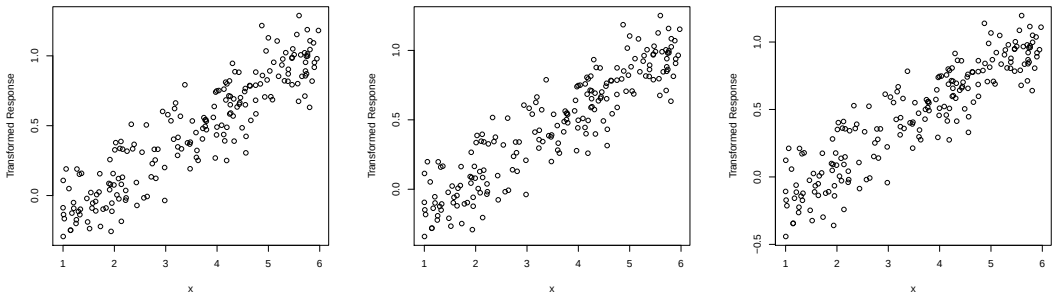



Figure 2.3: Scatter plot of the transformed response obtained using the AL-BC method corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) against x.

Figure 2.4 shows a comparison between the estimated quantile curves using the two approaches and the true quantile functions. For the 0.5 quantile functions, we can say that the two approaches share similar results, while for the 0.25 and 0.75 quantile functions, the PAL-BC approach outperforms the other approach that fails to estimate the correct transformation. This is because that although the observed information at the 0.25 and 0.75 quantile levels is not sufficient to estimate the transformation parameter and the quantile coefficients accurately, the PAL-BC method combines the observed information over multiple quantile levels to give estimates close to the true values for the parameters related to the

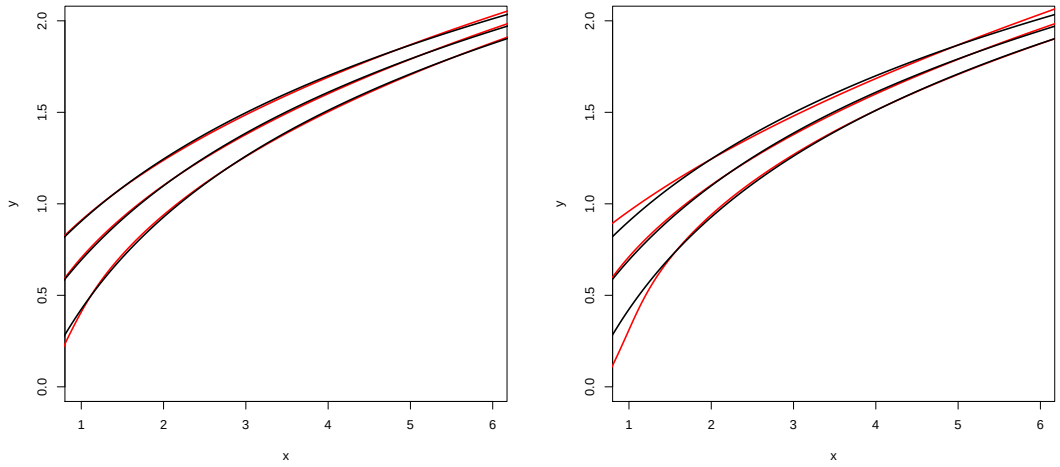



Figure 2.4: The average of the posterior mean of quantile curves (red) obtained by applying the PAL-BC method (left) and the AL-BC method (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75.

0.25 and 0.75 quantile functions. This conclusion is supported by Table 2.1 which suggests that the simultaneous estimation using the PAL-BC method succeeds in reducing $\text{MSE}_\tau$ and $\text{MAE}_\tau$ for the 0.25 and 0.75 quantile functions more than individual estimation using the AL-BC method. Table 2.1 also shows that the AL-BC approach performs better when the quantile of interest is close to the true mode of the data. From Figure 2.5, the two approaches suggest the averages of the 95% highest posterior density intervals for quantile functions containing the true quantile curves. The PAL-BC approach shows a better precision around the true quantile functions especially in the case of the 0.75 quantile.

| $\tau$ | $\text{MSE}_\tau$ | | $\text{MAE}_\tau$ | |
|---|---|---|---|---|
| | PAL-BC | AL-BC | PAL-BC | AL-BC |
| 0.25 | 0.13 (0.0076) | 0.19 (0.0135) | 2.38 (0.0643) | 2.53 (0.0667) |
| 0.5 | 0.07 (0.0039) | 0.07 (0.0042) | 1.87 (0.0501) | 1.91 (0.0539) |
| 0.75 | 0.05 (0.0029) | 0.13 (0.0219) | 1.66 (0.0494) | 2.43 (0.1008) |

Table 2.1: The averages of mean squared error ($\text{MSE}_\tau \times 10^{-2}$), mean absolute error ($\text{MAE}_\tau \times 10^{-2}$) and their standard errors (s.e. $\times 10^{-2}$) in the brackets, based on 300 samples.

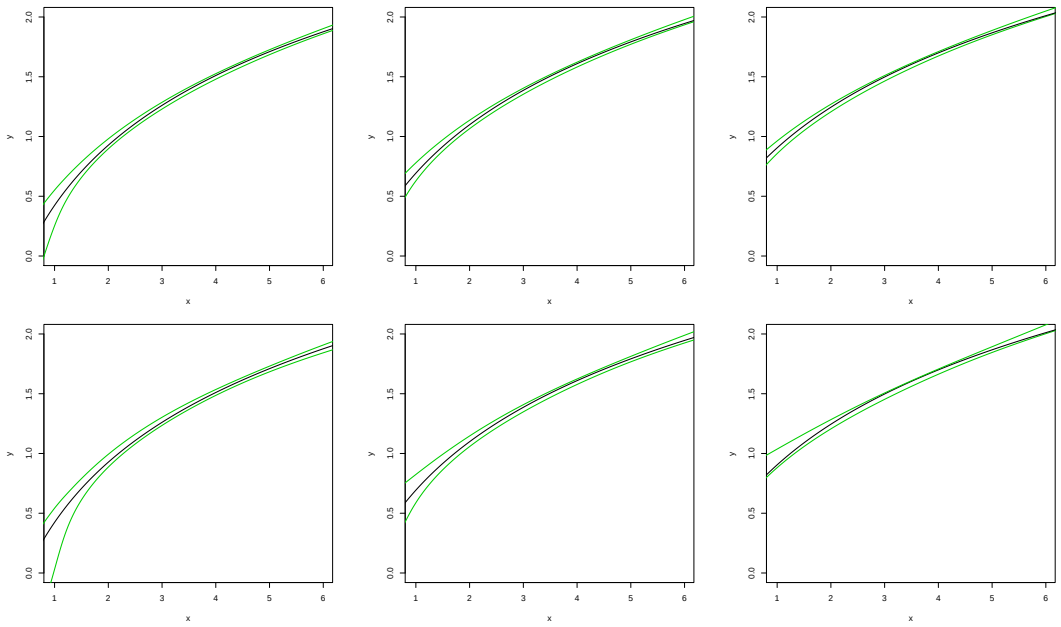


Figure 2.5: The average of the 95% highest posterior density intervals (green) estimated by applying PAL-BC (top) and AL-BC (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75 (left to right).

To investigate how increasing the number of quantiles affects simultaneous estimation and compare the proposed approaches in terms of estimating the extreme quantile functions, we implement individual estimations for the 0.05 and 0.95 quantiles, and simultaneous estimations for the 0.05 and 0.95 quantiles along with the 0.25, 0.5 and 0.75 quantiles. Figure 2.6 shows that the PAL-BC approach catches the quantiles of the underlying distribution well while the AL-BC method performs badly because the quantile of interest comes away from the median which is the mode of the data. In addition, Figures 2.4 & 2.6 suggest that PAL-BC method provides almost the same estimates under different number of quantiles. From Figure 2.7 the averages of the 95% highest posterior density intervals estimated using the PAL-BC method for the 0.95 quantile function includes the true quantile curve while the averages of the 95% highest posterior density intervals estimated using the AL-BC method does not. Therefore, it can be concluded that applying the PAL-BC method reduces the uncertainty shown by the other method about the location of the true 0.95 quantile function.

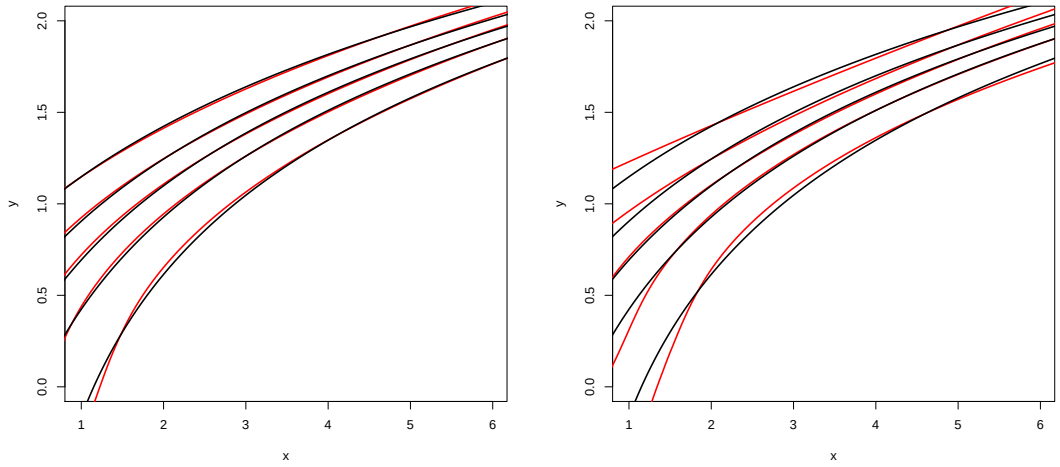



Figure 2.6: The average of the posterior mean of quantile curves (red) obtained by applying the PAL-BC method (left) and the AL-BC method (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.05, 0.25, 0.5 0.75 and 0.95.

### 2.5.2 Air Quality data

This dataset contains daily measurements of air quality, in New York city over the period from May to September 1973, on six variables which are *Ozone*: the mean of ozone in parts per billion, *Solar.R*: Solar radiation in the frequency band 4000 - 7700, *Wind*: the average of wind speed in mile per hour, *Temp*: the maximum of daily temperature in Fahrenheit degree, *Month* and *Day* (these data are available publicly, R Core Team, 2016). In this section, we are interested in analysing the relationship between *Ozone* and *Temp* that is described in Figure 2.8.

Figure 2.8 shows the nonlinear relationship between *Ozone* and *Temp* with heteroscedastic errors. Also, it shows how the PAL-BC method can deal with these violations of assumptions. Figure 2.9 shows how the AL-BC method suggests different transformations for the quantile levels. Table 2.2 gives more details about this difference in transformations and its effect on the estimates.

Figure 2.10 shows the comparison between quantile functions estimated using the PAL-BC method, the AL-BC method and local polynomial quantile regression (LPQR) proposed by Koenker (2005). For illustration, LPQR method estimates

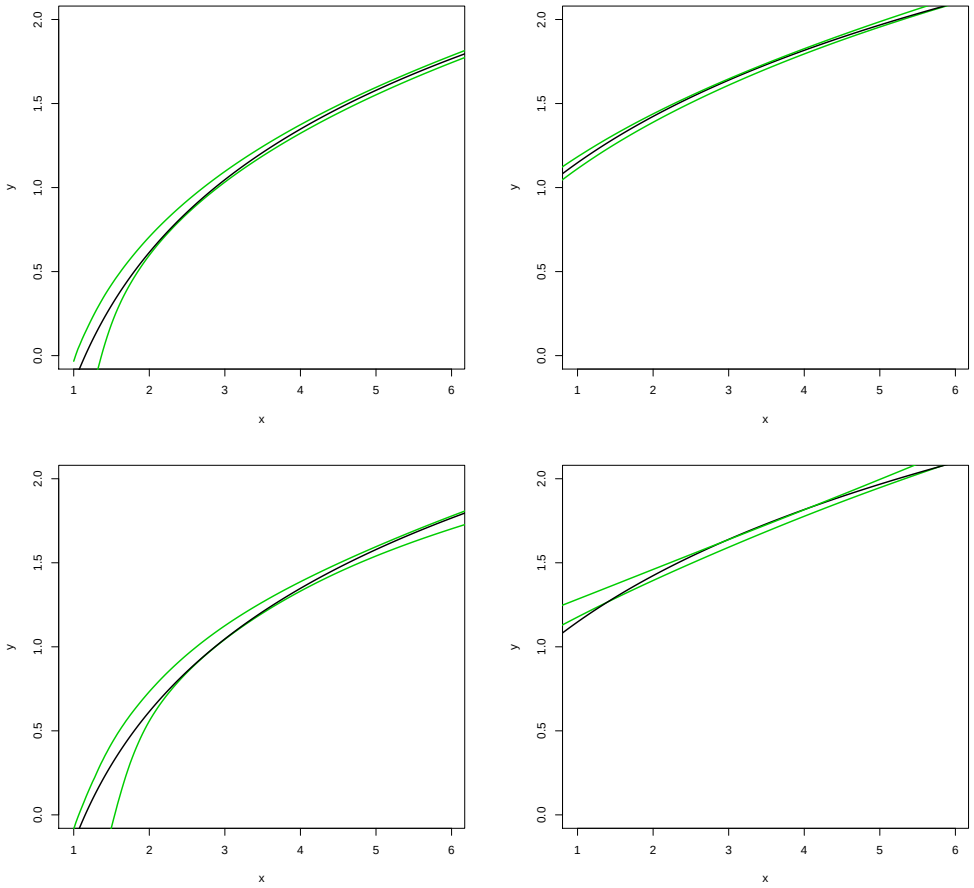



Figure 2.7: The average of the 95% highest posterior density intervals (green) estimated by applying the PAL-BC method (top) and the AL-BC method (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.05 and 0.95 (left to right).

the quantile functions by solving the optimisation problem given by

$$\min_{\alpha_{\tau_k},\beta_{\tau_k}} \sum_{i=1}^{n} w_i(x)\rho_{\tau_k}\Big(y_i - \alpha_{\tau_k} - \beta_{\tau_k}(x_i - x)\Big),$$

where $w_i(x) = K\Big((x_i - x)/h\Big)/h$ for a positive symmetric unimodal kernel function $K$ and a bandwidth parameter $h$. It suggests all methods behave in a similar way with some differences in estimating quantile functions. Table 2.2 and Figure 2.10 suggest that there is some agreement in estimating median functions between

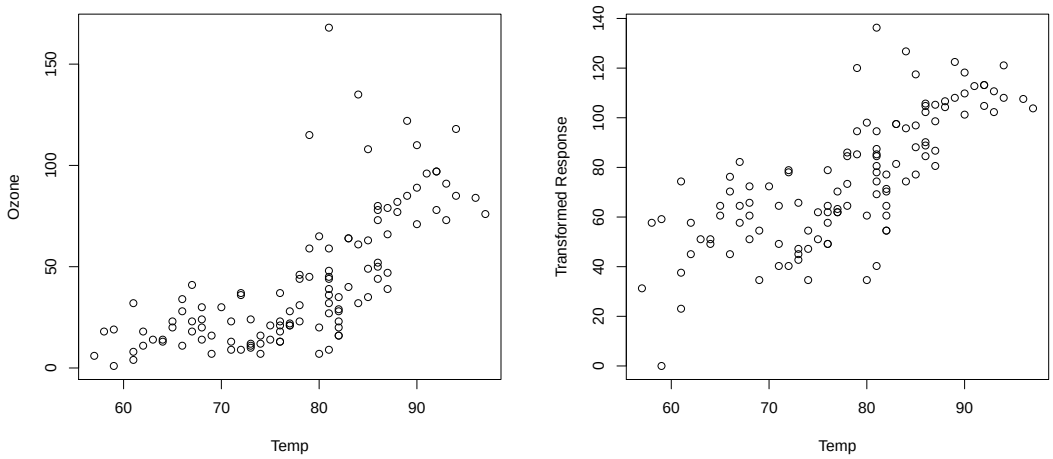



Figure 2.8: Scatter plot of the original response variable (left) and transformed response obtained using the PAL-BC method (right) against *Temp*.

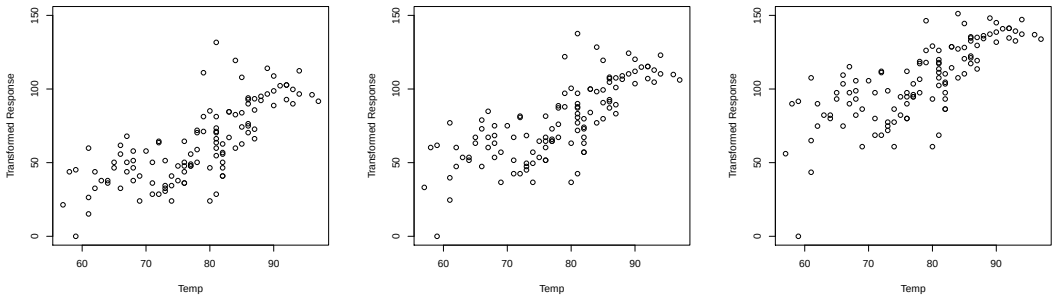



Figure 2.9: Scatter plot of the transformed response obtained using the AL-BC method corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) against *Temp*.

the PAL-BC and AL-BC methods. Table 2.3 shows that the means of errors estimated using the leave-one-out cross-validation evidences this agreement. For the 0.25 and 0.75 quantiles, there are slight differences in the means of errors obtained using the proposed methods.

| PAL-BC | | | AL-BC | | |
|---|---|---|---|---|---|
| Coefficients | P.M. | 95% HPD | Coefficients | P.M. | 95% HPD |
| $\alpha_{0.25}$ | $-92.54$ | $(-114.09, -70.34)$ | $\alpha_{0.25}$ | $-125.90$ | $(-162.48, -94.05)$ |
| $\alpha_{0.5}$ | $-80.35$ | $(-100.80, -58.25)$ | $\alpha_{0.5}$ | $-76.98$ | $(-108.12, -44.98)$ |
| $\alpha_{0.75}$ | $-69.74$ | $(-91.06, -48.27)$ | $\alpha_{0.75}$ | $-12.52$ | $(-47.20, 30.66)$ |
| $\beta$ | 2.01 | (1.84, 2.20) | $\beta_{0.25}$ | 2.27 | (1.85, 2.72) |
| | | | $\beta_{0.5}$ | 2.02 | (1.70, 2.32) |
| | | | $\beta_{0.75}$ | 1.70 | (1.47, 1.92) |
| $\lambda$ | 0.22 | (0.11, 0.33) | $\lambda_{0.25}$ | 0.38 | (0.18, 0.55) |
| | | | $\lambda_{0.5}$ | 0.20 | (0.01, 0.38) |
| | | | $\lambda_{0.75}$ | $-0.01$ | $(-0.17, 0.13)$ |
| $\sigma_{0.25}$ | 5.47 | (4.48, 6.49) | $\sigma_{0.25}$ | 5.38 | (4.42, 6.39) |
| $\sigma_{0.5}$ | 6.83 | (5.60, 8.09) | $\sigma_{0.5}$ | 6.86 | (5.62, 8.13) |
| $\sigma_{0.75}$ | 5.18 | (4.25, 6.16) | $\sigma_{0.75}$ | 5.03 | (4.15, 5.99) |

Table 2.2: The posterior mean (P.M.) and the 95% highest posterior density intervals (95% HPD).

| $\tau$ | PAL-BC | AL-BC |
|---|---|---|
| 0.25 | 5.49 | 5.42 |
| 0.5 | 7.83 | 7.83 |
| 0.75 | 6.71 | 6.69 |

Table 2.3: The mean of errors estimated using the leave-one-out cross-validation for the different quantile functions.

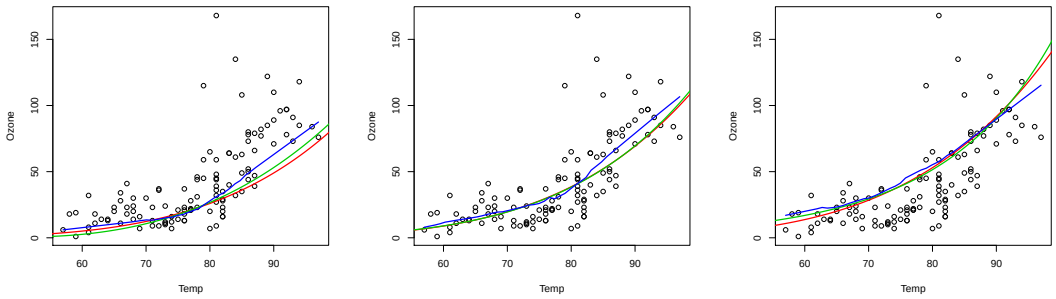



Figure 2.10: The quantile curve estimated by applying PAL-BC, AL-BC and local polynomial quantile regression (LPQR) are represented by the red, green and blue curves respectively for quantiles: 0.25, 0.5 and 0.75 (left to right).

## 2.6 Discussion

Along with handling the violations of the linearity and homoscedasticity efficiently, the PAL-BC method can ensure non-crossing of quantile functions. The PAL-BC method shows a great ability to estimate the suitable transformation for the response variable by combining the information over multiple quantile functions, while the AL-BC method suggests different transformation for each quantile and this could lead to misspecifying the true transformation, especially in the case of extreme quantiles.

Simultaneous estimation of quantile curves can outperform the individual estimation. Thus, combining observed information over multiple-levels of quantiles can enhance the estimation of quantile curves and ensure that quantile functions are monotonically increasing in quantile level. The key features of the proposed methods are the easy implementation and the flexible assumptions. Also, they do not require intensive computations and can provide smooth and reliable estimation of quantile functions for complex non-linear models with heteroscedastic errors. Thus, the proposed methods can be considered as a useful extension of Bayesian linear quantile regression.

However, the PAL-BC and AL-BC methods have some limitations. Due to the asymmetry property of the Box-Cox transformation, the PAL-BC and AL-BC approaches cannot offer a suitable transformation for the response variable in the cases where covariates should be transformed. Moreover, they are impractical in the context of multiple regressions where transformation of the response variable is not enough to fix the violations of linearity over all covariates.

# Chapter 3

# Extensions of Box-Cox regression

## 3.1 Introduction

Despite the flexibility of the Box-Cox transformation for the response variable, it has some limitations. In some situations, it can handle non-linearity, but heteroscedasticity may still exist. Therefore, to improve the estimation, the heteroscedastic variance of the conditional distribution of the transformed response should be taken into account. Also, since transforming the response variable is occasionally not enough to deal with the violation of linearity especially in the case of multiple covariates, a two-sided transformation, that allows for transforming each covariate along with the response variable may be needed.

To handle these issues, we consider two extensions of Box-Cox regression. Firstly, we consider a more general form of Box-Cox regression for heteroscedastic errors. Secondly, we consider a generalisation of Box-Cox regression that, in addition to transforming the response variable, allows different transformations for different predictors. Two-sided Box-Cox transformation includes a wide range of transformations that are symmetric around the linear transformation. Thus, the two-sided Box-Cox regression can handle more complex nonlinear models effectively. Figure 3.1 illustrates the properties of the two-sided Box-Cox transformation.

The two-sided Box-Cox transformation has a number of motivations. For example, in some situations where it is required to deal with the heteroscedasticity of linear models, the Box-Cox transformation of the response variable is not able to handle this problem. Therefore, the two-sided Box-Cox transformation is needed

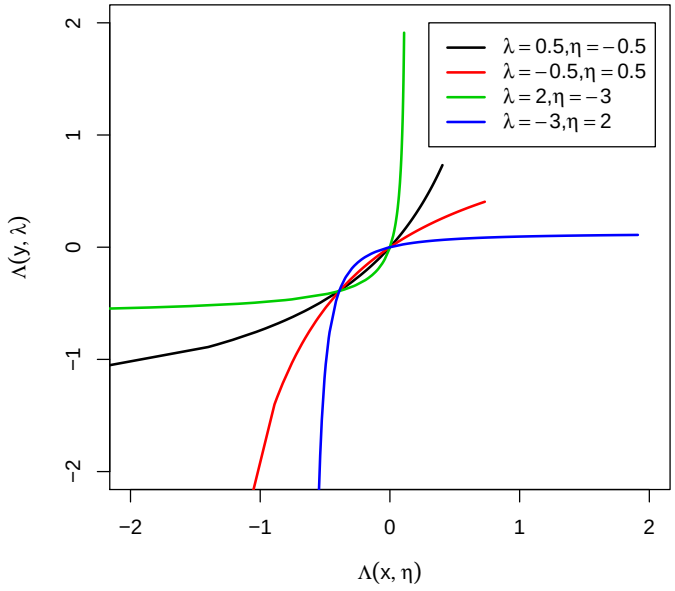



Figure 3.1: Two-sided Box-Cox transformation for range values of transformation parameters $\lambda$ and $\eta$, where $y = x$.

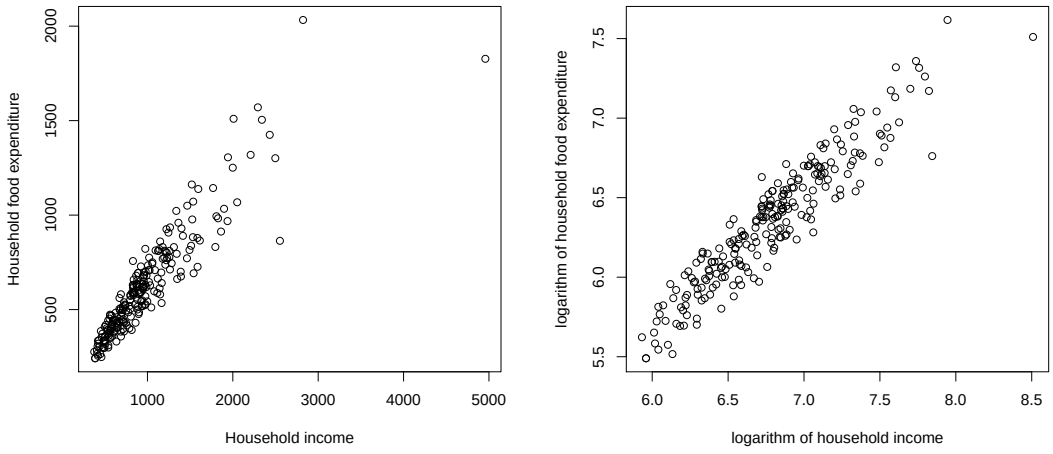



Figure 3.2: Scatter plot of the original and transformed Engel data using the logarithmic transformation.

to address the violation of homoscedasticity. To elaborate, we consider the Engel dataset that is now available publicly by Koenker (2016) and includes 235 observations on household food expenditure and income. To treat heteroscedasticity in the Engel data, the logarithmic transformation of both sides of the regression can be used as in Figure 3.2. In addition, the standard Box-Cox transformation of the response variable may not be practical in the case of multiple regression, since it can not fix the violation of linearity for all covariates. Hence, the two sided Box-Cox transformation can be considered as an alternative.

In this chapter, we propose two Bayesian methods to estimate quantile functions for these two extensions of Box-Cox regression simultaneously and individually. Then, we use simulations and real datasets to investigate the performance of the proposed methods and to implement a comparison study.

## 3.2 Bayesian Box-Cox quantile regression with heteroscedastic errors

To take into account heteroscedasticity that is not eliminated by Box-Cox transformation, we consider a more general form of Box-Cox regression given by

$$\Lambda\left(y_i;\lambda\right)=\alpha+\boldsymbol{x}_i^{'}\boldsymbol{\beta}+\sigma\left(\boldsymbol{x}_i\right)\epsilon_i,$$

where $\epsilon_i$ is independently distributed with a mean equal to zero and constant variance; $\boldsymbol{\sigma}\left(\boldsymbol{x}_i\right)$ is the standard deviation of the heteroscedastic error depending on $\boldsymbol{x}_i$; $\boldsymbol{x}_i$ is a $p\times 1$ vector of covariates for the $i^{\text{th}}$ observation; $\alpha$ is a unknown intercept coefficient and $\boldsymbol{\beta}$ is a $p\times 1$ vector of unknown slope coefficients. Then, the conditional quantile function is given by

$$Q(\tau_k|\boldsymbol{x}_i)=\Lambda^{-1}\left(\alpha_{\tau_k}+\boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}|\lambda_{\tau_k}\right).$$

Thus, the Bayesian quantile model for this regression is based on the assumption of a linear relationship between the transformed response variable and covariates. To accommodate heteroscedastic scale in the Bayesian quantile model, we assume that

$$\sigma\left(x_i\right)=\exp\left(\psi_0+\boldsymbol{x}_i^{'}\boldsymbol{\psi}\right), \tag{3.1}$$

where

$$\boldsymbol{\psi} = (\psi_1, \psi_2, ..., \psi_p)\,.$$

Then, the multiple quantile functions of Box-Cox regression with heteroscedastic error can be estimated by maximising the pseudo asymmetric Laplace likelihood function (PAL-HBC) that is obtained by replacing the scale parameters in the formula (2.8) with the heteroscedastic scale given in (3.1) and can be written as follows:

$$\begin{aligned} l\left(\boldsymbol{\alpha}_\tau, B_\tau, \boldsymbol{\lambda}_\tau, \boldsymbol{\psi}_{0,\tau}, \Psi_\tau; \boldsymbol{y}\right) \;\propto\; & \exp\left\{ -\sum_{k=1}^{m}\sum_{i=1}^{n} \rho_{\tau_k}\left( \frac{\Lambda\left(y_i;\lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}}{\exp\left(\psi_{0,\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\psi}_{\tau_k}\right)} \right) \right. \\ & \left. - \sum_{k=1}^{m}\sum_{i=1}^{n} \left(\psi_{0,\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\psi}_{\tau_k}\right) \right\}. \end{aligned} \tag{3.2}$$

where $\boldsymbol{\psi}_{0,\tau} = (\psi_{0,\tau_1}, ..., \psi_{0,\tau_m})$ and $\Psi_\tau$ is a $p \times m$ matrix whose $k^{\text{th}}$ column is given by $\boldsymbol{\psi}_{\tau_k} = (\psi_{1,\tau_k}, ..., \psi_{p,\tau_k})$. Also, the quantile functions can be estimated individually by maximising the asymmetric Laplace likelihood function (AL-HBC), which is obtained by setting $m = 1$ in the likelihood given in (3.2). The prior distributions for the coefficients of quantile regression are specified as follows:

$$p(\boldsymbol{\alpha}_\tau) \;\propto\; \exp\left\{ -\frac{1}{2}\sum_{k=1}^{m} \frac{\left(\alpha_{\tau_k} - \alpha^*_{\tau_k}\right)^2}{c^2_{\tau_k}} \right\},$$

$$p(B_\tau) \;\propto\; \exp\left\{ -\frac{1}{2}\sum_{k=1}^{m} \left(\boldsymbol{\beta}_{\tau_k} - \boldsymbol{\beta}^*_{\tau_k}\right)^{'} S^{-1}_{\tau_k} \left(\boldsymbol{\beta}_{\tau_k} - \boldsymbol{\beta}^*_{\tau_k}\right) \right\},$$

and for the vector of the Box-Cox transformation parameters $\boldsymbol{\lambda}_\tau$ are as given in Section 2.4.2. The prior distributions for all variance parameters are given by

$$p(\boldsymbol{\psi}_{0,\tau}) \propto \exp\left\{ -\sum_{k=1}^{m} \frac{\left(\psi_{0,\tau_k} - \psi^*_{0,\tau_k}\right)^2}{2d^2_{\tau_k}} \right\},$$

$$p(\Psi_\tau) \propto \exp\left\{ -\frac{1}{2}\sum_{k=1}^{m} \left(\boldsymbol{\psi}_{\tau_k} - \boldsymbol{\psi}^*_{\tau_k}\right)^{'} \Sigma^{-1}_{\tau_k} \left(\boldsymbol{\psi}_{\tau_k} - \boldsymbol{\psi}^*_{\tau_k}\right) \right\}.$$

Then, the posterior distribution is given by

$$p\left(\boldsymbol{\alpha}_\tau, B_\tau, \boldsymbol{\lambda}_\tau, \boldsymbol{\psi}_{0,\tau}, \Psi_\tau | \boldsymbol{y}\right) \propto l\left(\boldsymbol{\alpha}_\tau, B_\tau, \boldsymbol{\lambda}_\tau, \boldsymbol{\psi}_{0,\tau}, \Psi_\tau; \boldsymbol{y}\right) p(\boldsymbol{\alpha}_\tau) p(B_\tau) p(\boldsymbol{\lambda}_\tau) p(\boldsymbol{\psi}_{0,\tau}) p(\Psi_\tau).$$

To simulate samples from this posterior distribution, the Metropolis-Hastings algorithm can be applied. A simulation study to investigate the performance of this quantile regression method is implemented in Section 3.4.

## 3.3 Bayesian two-sided Box-Cox quantile regression

The standard Box-Cox regression can be extended to a two-sided Box-Cox regression which can be written as

$$\Lambda(y_i, \lambda) = \alpha + \Lambda\left(\boldsymbol{x}_i, \boldsymbol{\eta}\right) \boldsymbol{\beta} + \epsilon_i,$$

where $\Lambda\left(\boldsymbol{x}_i, \boldsymbol{\eta}\right) = \Big(\Lambda\left(x_{1i}, \eta_1\right), \Lambda\left(x_{2i}, \eta_2\right), ..., \Lambda\left(x_{pi}, \eta_p\right)\Big)$, and $\epsilon_i$ is independently distributed with a mean equal to zero and a constant variance. The quantile function is given by

$$Q(\tau_k | \boldsymbol{x}_i) = \Lambda^{-1}\Big(\alpha_{\tau_k} + \Lambda\left(\boldsymbol{x}_i, \boldsymbol{\eta}_{\tau_k}\right) \boldsymbol{\beta}_{\tau_k} | \lambda_{\tau_k}\Big).$$

Then, multiple quantile curves of the two-sided Box-Cox regression can be estimated by solving the optimisation problem:

$$\widehat{Q}\left(\tau | X\right) = \min_{\boldsymbol{\alpha}_\tau, B_\tau, \boldsymbol{\lambda}_\tau, H_\tau \in R} \sum_{k=1}^{m} \sum_{i=1}^{n} \rho_{\tau_k}\Big(\Lambda(y_i, \lambda_{\tau_k}) - \alpha_{\tau_k} - \Lambda\left(\boldsymbol{x}_i, \boldsymbol{\eta}_{\tau_k}\right) \boldsymbol{\beta}_{\tau_k}\Big),$$

which is equivalent to maximising the pseudo asymmetric Laplace likelihood function (PAL-TBC) given by

$$l\left(\boldsymbol{\alpha}_\tau, B_\tau, \boldsymbol{\sigma}_\tau, \boldsymbol{\lambda}_\tau, H_\tau; \boldsymbol{y}\right) \propto \left(\prod_{k=1}^{m} \sigma_{\tau_k}^{-n}\right) \exp\left\{-\sum_{k=1}^{m} \sum_{i=1}^{n} \rho_{\tau_k}\left(\frac{\Lambda(y_i, \lambda_{\tau_k}) - \alpha_{\tau_k} - \Lambda\left(\boldsymbol{x}_i, \boldsymbol{\eta}_{\tau_k}\right) \boldsymbol{\beta}_{\tau_k}}{\sigma_{\tau_k}}\right)\right\},$$

where $H_\tau$ is a $p \times m$ matrix whose $k^{\text{th}}$ column is given by $\boldsymbol{\eta}_{\tau_k} = (\eta_{1,\tau_k}, ..., \eta_{p,\tau_k})$. Also, the quantile functions can be estimated individually by maximising the

asymmetric Laplace likelihood function with the two-sided Box-Cox transformation (AL-TBC), which is obtained by setting $m = 1$ in the above likelihood (PAL-TBC). The prior distributions for coefficients of quantile regression and the vector of the Box-Cox transformation parameters $\boldsymbol{\lambda}_\tau$ are as specified in Section 2.4.2. The left-hand side transformation parameters are assumed to be a priori identical over all quantile levels and qual $\boldsymbol{\eta}$. Then,

$$p(\boldsymbol{\eta}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta}-\boldsymbol{\eta}^*\right)' \Sigma_{\boldsymbol{\eta}}^{-1}\left(\boldsymbol{\eta}-\boldsymbol{\eta}^*\right)\right\}$$

Hence, the posterior distribution is written as

$$p\left(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau, \lambda, \boldsymbol{\eta} | \boldsymbol{y}\right) \propto l\left(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau, \lambda, \boldsymbol{\eta}; \boldsymbol{y}\right) p(\boldsymbol{\alpha}_\tau) p(\boldsymbol{\beta}) p(\boldsymbol{\sigma}_\tau) p(\lambda) p(\boldsymbol{\eta}).$$

The Metropolis-Hastings algorithm can be applied to simulate samples from this posterior distribution. Alternatively, utilising the representation of the asymmetric Laplace distribution based on a mixture of normal distributions proposed by Kozumi & Kobayashi (2011) as described in Section 1.6.2, the Gibbs sampler can be used by employing the representation of the pseudo asymmetric Laplace likelihood function (PAL-TBC) given by

$$l\left(\boldsymbol{\alpha}_\tau, B_\tau, W_\tau, \boldsymbol{\sigma}_\tau, \boldsymbol{\lambda}_\tau, H_\tau; \boldsymbol{y}\right) \propto \left(\prod_{k=1}^{m}\prod_{i=1}^{n} \sigma_{\tau_k}^{-\frac{1}{2}} w_{\tau_k,i}^{-\frac{1}{2}}\right) \exp\left\{-\sum_{k=1}^{m}\sum_{i=1}^{n} \frac{\left(\Lambda(y_i, \lambda_{\tau_k}) - \alpha_{\tau_k} - \Lambda\left(\boldsymbol{x}_i, \boldsymbol{\eta}_{\tau_k}\right)\boldsymbol{\beta}_{\tau_k}\right)^2}{2\phi_{\tau_k}\sigma_{\tau_k} w_{\tau_k,i}}\right\}.$$

With the conditional distributions of all parameters derived, the Gibbs sampler including one step Metropolis-Hastings updating transformation parameters can be applied.

## 3.4 Simulations and applications

### 3.4.1 Simulated Data

We compare the performance of simultaneous Box-Cox quantile regression accommodating heteroscedastic error (PAL-HBC) with individual Box-Cox quantile regression with heteroscedastic error (AL-HBC). For the priors, we use the same specified parameters described in Section 2.5. We sample 300 observations

from the model:

$$y_i = (1 + x_i + 0.3x_i\epsilon_i)^2, \tag{3.3}$$

where $x_i$ is generated from the uniform distribution U(2,20) and $\epsilon_i$ is simulated from the normal distribution N(0,1). The simulated data are plotted in Figure 3.3. For these data, the PAL-BC is able to fix the violation of linearity, but the

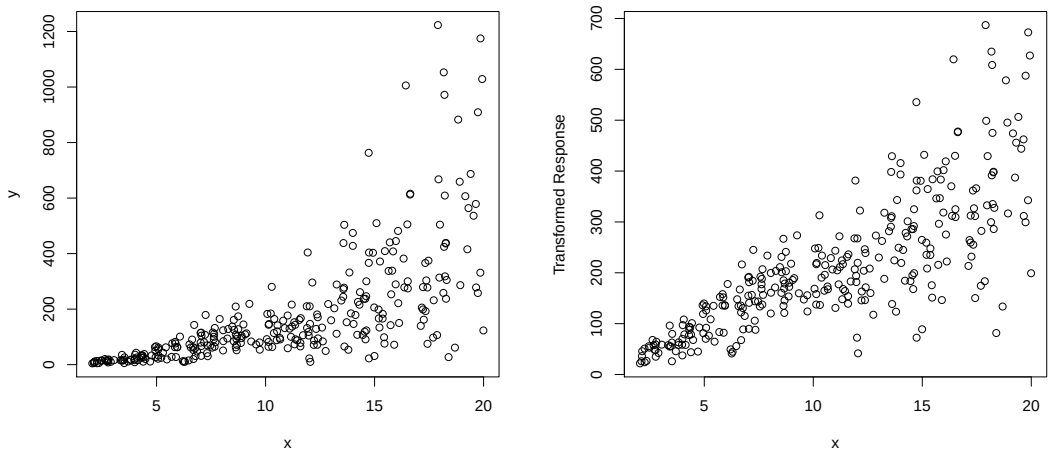



Figure 3.3: Scatter plot of the original response variable (left) and transformed response with $\lambda = 0.51$ estimated using PAL-HBC (right) against x.

errors are still heteroscedastic. Therefore, to achieve better Bayesian inference, it is required to take into account the heteroscedasticity by considering simultaneous Box-Cox quantile regression with heteroscedastic error (PAL-HBC). Since the PAL-HBC method depends on one assumption which is the linearity of relationship between the transformed response variable and the covariates, it could be considered more flexible and it may be useful to fit complex models. Thus, Figure 3.3 suggests that the PAL-HBC method is able to treat the violation of linearity. In contrast, Figure 3.4 shows that the AL-HBC method recommends different transformations for the quantile levels, since the AL-HBC method seeks to handle the violation of linearity at the desired quantile level. Since the true error distribution is normal, the median function that plays the role of central tendency measurement can provide much information about the suitable distribution for the data. For this reason, there is almost perfect agreement in the

estimation of the transformation parameters corresponding to $\tau = 0.5$ between the PAL-HBC and AL-HBC methods. For further illustration, we implement a simulation study using 300 samples of size 300 that just differ in the errors $\epsilon_i$, from the model given in (3.3).

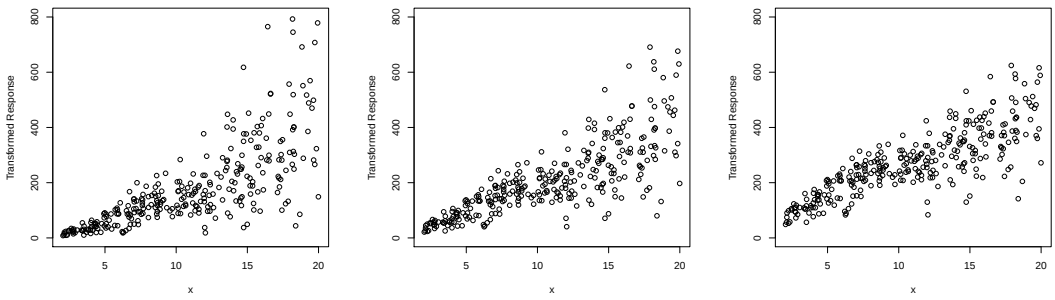



Figure 3.4: Scatter plot of the transformed response obtained using the AL-HBC method corresponding to $\tau = 0.25, 0.5, 0.75$ (left to right) against x, where $\lambda_{0.25} = 0.77$, $\lambda_{0.5} = 0.52$ and $\lambda_{0.75} = 0.30$.

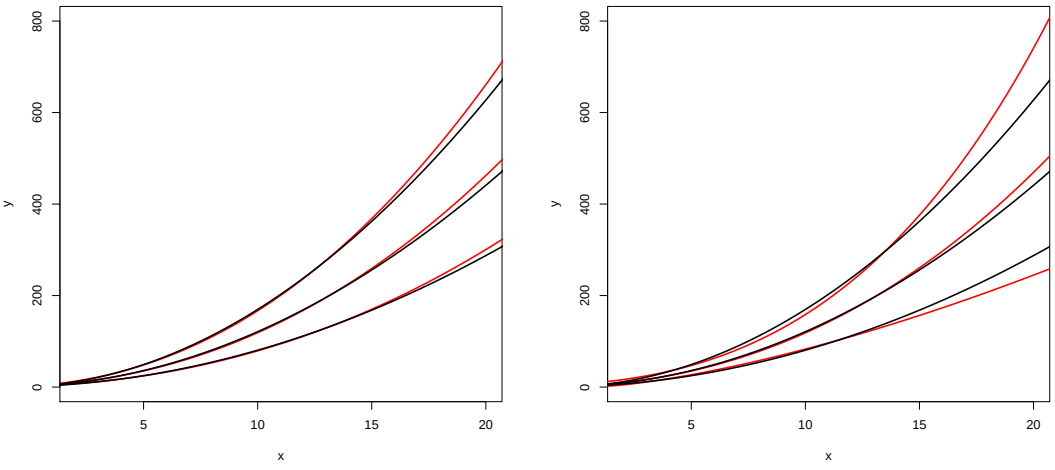



Figure 3.5: The average of the posterior mean of quantile curves (red) obtained by applying PAL-HBC (left) and AL-HBC (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75.

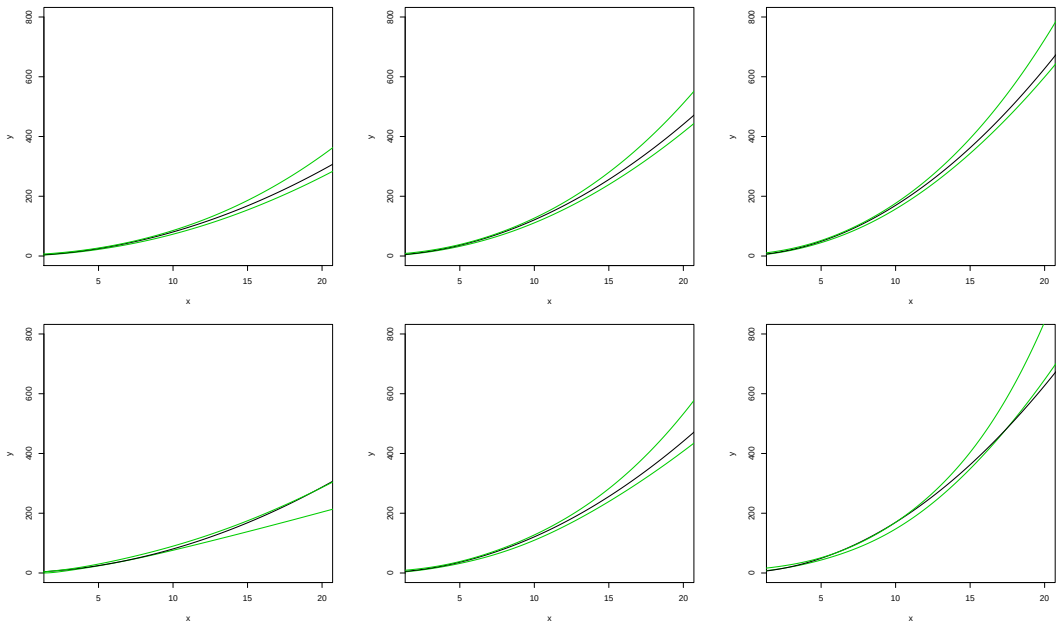


Figure 3.6: The average of the 95% highest posterior density intervals (green) estimated by applying PAL-HBC (top) and AL-HBC (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.25, 0.5 and 0.75 (left to right).

Figure 3.5 shows a comparison between the estimated quantile curves obtained using the two approaches and the true quantile functions. It can be seen that the PAL-HBC approach outperforms AL-HBC. Consistent with this conclusion, Table 3.1 suggests that the PAL-HBC method gives a greater reduction in $\text{MSE}_\tau$ and $\text{MAE}_\tau$ for all levels of quantile than the AL-HBC method.

| $\tau$ | $\text{MSE}_\tau$ | | $\text{MAE}_\tau$ | |
|---|---|---|---|---|
| | PAL-HBC | AL-HBC | PAL-HBC | AL-HBC |
| 0.25 | 114.22 ( 8.81) | 315.34 (17.41) | 6.30 (0.23) | 10.63 (0.28) |
| 0.5 | 190.84 (14.29) | 227.51 (16.26) | 8.17 (0.28) | 8.63 (0.28) |
| 0.75 | 359.25 (25.35) | 1475.39 (66.86) | 11.00 (0.36) | 21.20 (0.45) |

Table 3.1: The averages of mean squared error ($\text{MSE}_\tau$), mean absolute error ($\text{MAE}_\tau$) and their standard errors (s.e.) in the brackets, based on 300 samples.

Figure 3.6 suggests that the averages of the 95% highest posterior density intervals estimated using the two approaches for the median function include the true quantile regression curve. Also, it shows that the averages of the 95% highest posterior density intervals estimated using the PAL-HBC method for the 0.25 and 0.75 quantile functions include the true quantile curves while the averages of the 95% highest posterior density intervals estimated using the AL-HBC method does not. Thus, the PAL-HBC method is more confident about the location of the true 0.25 and 0.75 quantile functions.

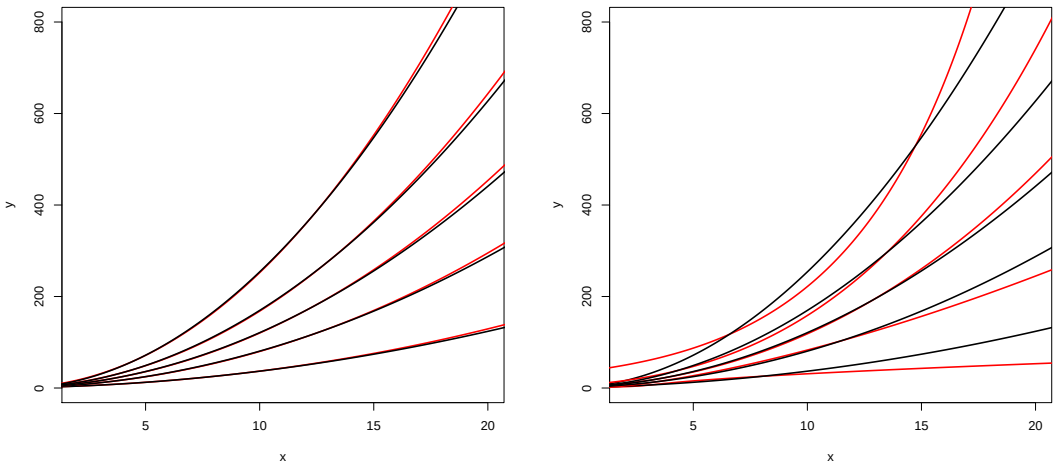



Figure 3.7: The average of the posterior mean of quantile curves (red) obtained by applying PAL-HBC (left) and AL-HBC (right) over 300 simulations, and true quantile curves (black) for quantiles: 0.05, 0.25, 0.5, 0.75 and 0.95.

To examine the effect of the number of quantiles included in simultaneous estimation on the performance of the PAL-HBC approach and to challenge the proposed methods to estimate more extreme quantiles, we estimate the 0.05 and 0.95 quantiles individually and simultaneously along with the 0.25, 0.5 and 0.75 quantiles. From Figure 3.7, the PAL-HBC method shows a great ability to estimate all quantiles of interest. On the other hand, the AL-HBC method gives poor estimates especially for the extreme quantiles. Also, Figures 3.5 & 3.7 suggest that the addition of 0.05 and 0.95 quantiles to the simultaneous estimation reduces the bias of estimates given by PAL-HBC for the three quartiles. Figure 3.8

shows how the PAL-HBC approach provides more accurate 95% highest posterior density intervals, including the true functions for the 0.05 and 0.95 quantiles, while the AL-HBC method fails to do that. Thus, the AL-HBC approach is unable to determine the correct transformation for the data using observed information available at individual quantile functions, especially the extreme ones. On the other hand, the PAL-HBC method, combining observed information over multiple quantiles, suggests an accurate estimation of the transformation.

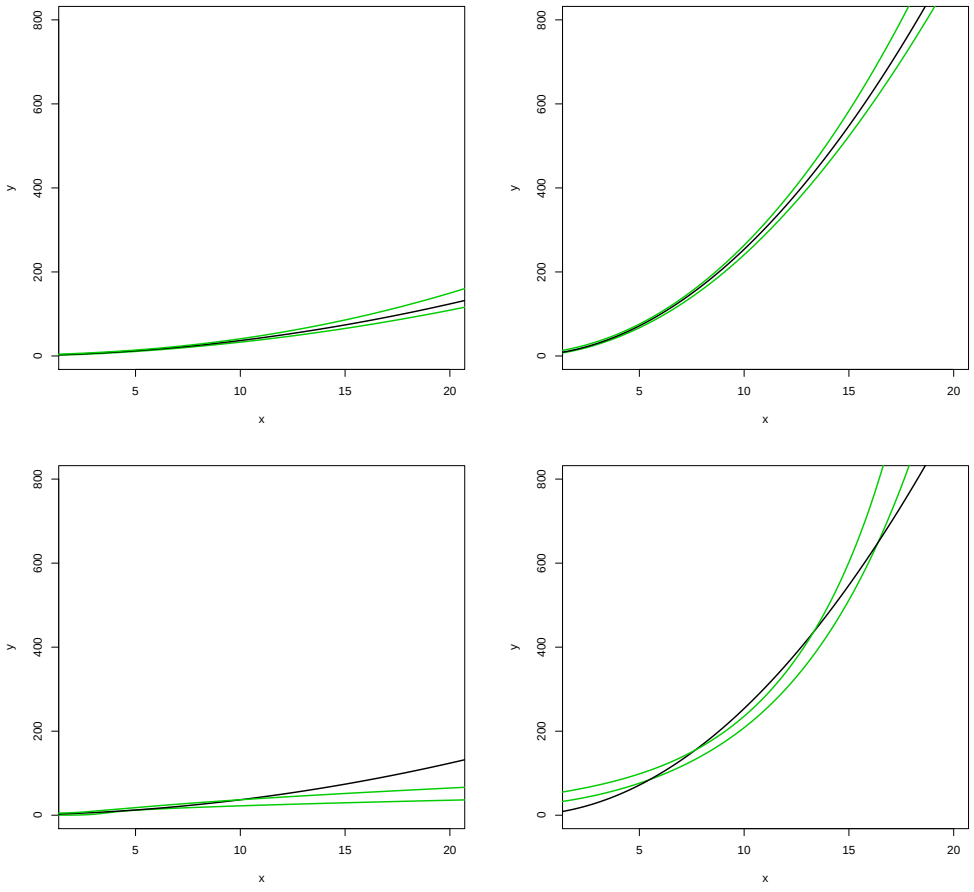



Figure 3.8: The average of the 95% highest posterior density intervals (green) estimated by applying PAL-HBC (top) and AL-HBC (bottom) over 300 simulations, and true quantile curves (black) for quantiles: 0.05 and 0.95 (left to right).

### 3.4.2 Housing values in Boston

We compare the performance of simultaneous two-sided Box-Cox quantile regression (PAL-TBC) with individual two-sided Box-Cox quantile regression (AL-TBC). We use the real data for the housing market in Boston. This dataset

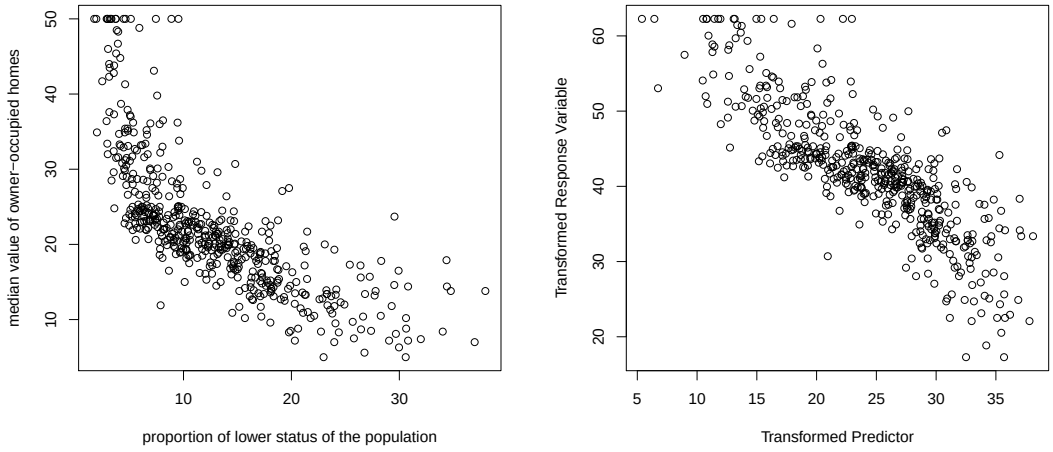



Figure 3.9: Scatter plot of the original response variable against the original predictor (left) and the transformed response variable against the transformed predictor obtained using PAL-TBC (right).

includes the measurement of median value of owner-occupied homes in $1000s, among other variables, crime rate by town, age (proportion of owner-occupied units built prior to 1940), the proportion of the lower status of the population and full value property tax rate per $10,000 (for more details, see Harrison & Rubinfeld, 1978). These data are now available publicly by Venables & Ripley (2002). Here we study the relationship between median value of owner-occupied homes and the proportion of the lower status of the population over different levels of quantiles. For the data given in Figure 3.9, the simultaneous Box-Cox quantile method (PAL-BC) can not suggest a suitable transformation, that is able to handle the violation of linearity and homoscedasticity, for the response variable. This complex model is considered beyond the ability of the PAL-BC method since a two-sided transformation is required. Then, to achieve the assumptions

of the linearity and homoscedasticity, it is required to consider a more powerful extension of the PAL-BC method which is simultaneous two-sided Box-Cox quantile regression (PAL-TBC). Figure 3.9 shows the capacity of the PAL-TBC method to handle the violation of linearity and constant variance by suggesting suitable transformations for the response variable and the covariate while Figure 3.10 shows the inability of the AL-TBC method to deal with these violations for the 0.25 and 0.75 quantile functions. Table 3.2 illustrates the differences in the estimation of transformation between the two proposed approaches and the effects of these differences on the estimates.

From Figure 3.11, which shows the estimation of quantile functions obtained by applying simultaneous two-sided Box-Cox quantile regression (PAL-TBC), individual two-sided Box-Cox quantile regression (AL-TBC) and local polynomial quantile regression (LPQR) described in Section 2.5.2, there is close agreement between the PAL-TBC and the AL-TBC methods in the estimation of median function. Also, compared to the LPQR method, it is obvious that the PAL-TBC and AL-TBC methods suggest reasonable transformations for the response variable and the predictor variable. Table 3.3 shows similar means of errors estimated using the leave-one-out cross-validation method, given in (2.11), for both the approaches.

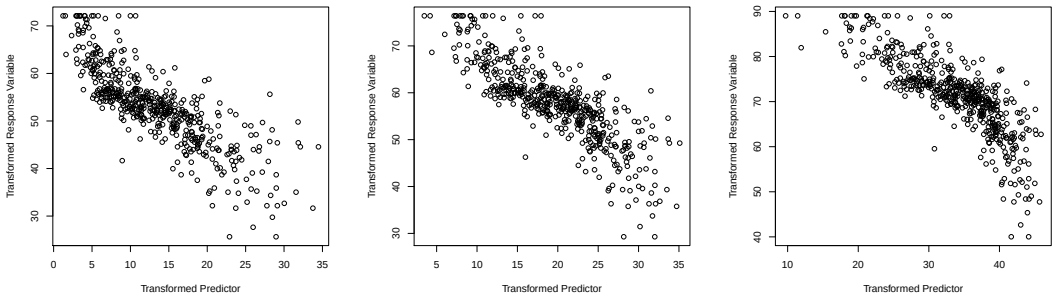



Figure 3.10: Scatter plot of the transformed response variable against the transformed predictor obtained using AL-TBC for quantiles: 0.25, 0.5 and 0.75 (left to right).

| PAL-TBC | | | AL-TBC | | |
|---|---|---|---|---|---|
| Coefficients | P.M. | 95% HPD | Coefficients | P.M. | 95% HPD |
| $\alpha_{0.25}$ | 64.89 | (60.55,69.55) | $\alpha_{0.25}$ | 45.41 | ( 40.35, 51.09) |
| $\alpha_{0.5}$ | 67.47 | (63.07,72.13) | $\alpha_{0.5}$ | 67.99 | ( 57.46, 79.42) |
| $\alpha_{0.75}$ | 70.42 | (66.03,75.10) | $\alpha_{0.75}$ | 123.47 | (103.64,144.35) |
| | | | $\beta_{0.25}$ | $-0.99$ | ( $-1.05$, $-0.93$) |
| $\beta$ | $-1.06$ | ($-1.10$,$-1.02$) | $\beta_{0.5}$ | $-1.04$ | ( $-1.11$, $-0.96$) |
| | | | $\beta_{0.75}$ | $-0.95$ | ( $-1.01$, $-0.89$) |
| | | | $\lambda_{0.25}$ | 0.75 | ( 0.56, 0.94) |
| $\lambda$ | 0.30 | ( 0.20, 0.38) | $\lambda_{0.5}$ | 0.28 | ( 0.05, 0.51) |
| | | | $\lambda_{0.75}$ | $-0.24$ | ( $-0.39$, $-0.09$) |
| | | | $\eta_{0.25}$ | 0.12 | ( $-0.11$, 0.35) |
| $\eta$ | 0.04 | ($-0.07$, 0.16) | $\eta_{0.5}$ | 0.06 | ( $-0.20$, 0.28) |
| | | | $\eta_{0.75}$ | $-0.08$ | ( $-0.25$, 0.10) |
| $\sigma_{0.25}$ | 1.38 | ( 1.26, 1.51) | $\sigma_{0.25}$ | 1.33 | ( 1.23, 1.44) |
| $\sigma_{0.5}$ | 1.79 | ( 1.65, 1.96) | $\sigma_{0.5}$ | 1.8 | ( 1.64, 1.95) |
| $\sigma_{0.75}$ | 1.49 | ( 1.36, 1.63) | $\sigma_{0.75}$ | 1.44 | ( 1.32, 1.59) |

Table 3.2: The posterior mean (P.M.) and the 95% highest posterior density intervals (95% HPD).

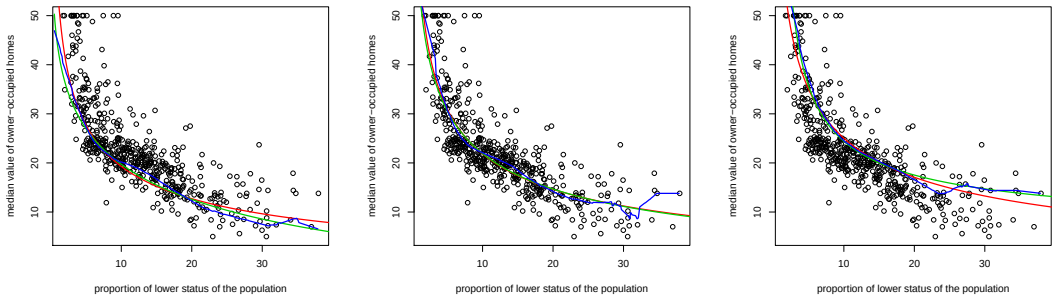



Figure 3.11: Quantile curve estimated by applying PAL-TBC, AL-TBC and local polynomial quantile regression (LPQR) are represented by the red, green and blue curves respectively for quantiles: 0.25, 0.5 and 0.75 (left to right).

| $\tau$ | PAL-TBC | AL-TBC |
|---|---|---|
| 0.25 | 1.37 | 1.36 |
| 0.5 | 1.90 | 1.91 |
| 0.75 | 1.69 | 1.65 |

Table 3.3: The mean of errors estimated using the leave-one-out cross-validation method, given in (2.11), for the different quantile functions.

### 3.4.3 Air Quality data

In this section, we illustrate the motivation of the two-sided Box-Cox regression to handle the violation of linearity and homoscedasticity for multiple regression. We use Air Quality data, described in Section 2.5.2, to investigate the relationship of *Ozone* against *Solar.R*, *Wind* and *Temp*. The original relationships between these variables are shown in Figure 3.12. It can be seen that there are different types of relationship between the response variable and the covariates. The relationship between *Ozone* and *Solar.R* is linear with heteroscedastic error, and the relationship between *Ozone* and *Wind* is inverse nonlinear with heteroscedastic error. Also, there is a nonlinear relationship with heteroscedastic error between *Ozone* and *Temp*. Due to these differences and complexity in the relationship between the response variable and the covariates, Box-Cox quantile regression discussed in Chapter 2 is not able to suggest a transformation, for the response variable, that is able to handle the violation of linearity and homoscedasticity. Therefore, these data can be considered as a good motivation to use two-sided Box-Cox quantile regression. We use this dataset to investigate the performance of simultaneous two-sided Box-Cox quantile regression (PAL-TBC) and individual two-sided Box-Cox quantile regression (AL-TBC). Table 3.4 show that the PAL-TBC approach suggests different transformations for the variables of interest. Also, Figure 3.13 shows that the suggested transformations by the PAL-TBC method succeed in accommodating nonlinear relationships between the variables. To examine the existence of heteroscedasticity, Koenker & Bassett (1982) suggested statistical test to examine the null hypothesis of the equality of the slope coefficients in $\boldsymbol{\tau}$. They assumed that

$$H_0 : RB_\tau = \boldsymbol{r},$$

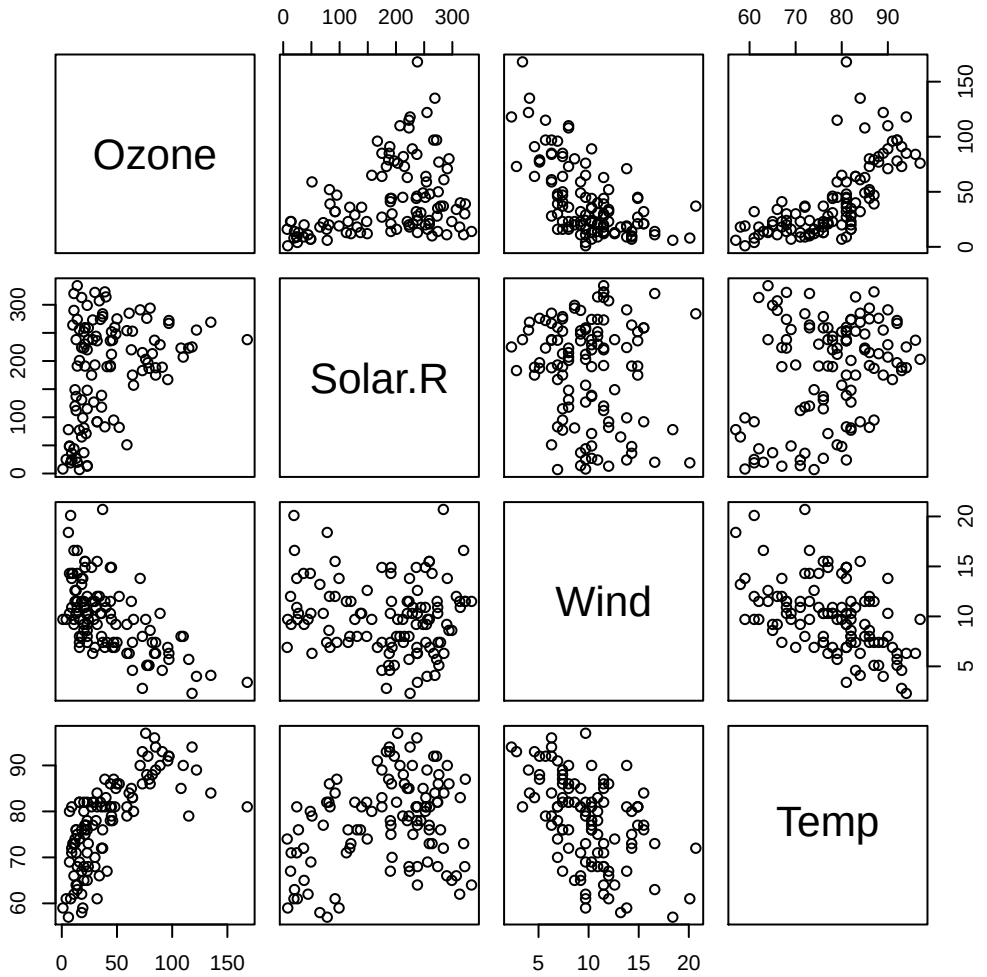



Figure 3.12: The scatter plot of original relationships between the variables.

where $B_\tau$ is a matrix of the slope coefficients corresponding to individual quantiles given by $B_\tau = [(\boldsymbol{\beta}_{\tau_1,1}, ..., \boldsymbol{\beta}_{\tau_1,p})', ..., (\boldsymbol{\beta}_{\tau_m,1}, \boldsymbol{\beta}_{\tau_m,p})']'$,. Under the null hypothesis $H_0$, the test statistic is given by

$$T_n = n\left(RB_\tau - \boldsymbol{r}\right)' \left[RV_n^{-1}R'\right]\left(RB_\tau - \boldsymbol{r}\right),$$

which is asymptotically chi-square with degrees of freedom equal to the rank of $R$, where $V_n$ is $mp \times mp$ with $(s,t)^{\text{th}}$ element given by

$$V_n(\tau_s, \tau_t) = \left[\min(\tau_s, \tau_t) - \tau_s\tau_t\right] H_n(\tau_s)^{-1} J_n H_n(\tau_t)^{-1},$$

where

$$H_n(\tau_k) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' f\big(Q(\tau_k)\big), \quad \text{and} \quad J_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i',$$

where $f$ is the conditional distribution of the response variable and $Q$ is the quantile function of the error distribution. To examine the equality of slope coefficients over different quantiles, we implement a general class of this test called the joint test of equality of slope coefficients of quantile regressions that is described by Koenker (2016). At the significance level of 0.05, the test suggests that slope coefficients of quantile regressions for the transformed data, shown in Figure 3.13, are identical under different numbers of quantile functions included in the joint test. This implies the absence of heteroscedasticity. On the other hand, the AL-TBC approach suggests different transformations for each quantile level (for more details see Table 3.4 and Figures 3.14 & 3.15 & 3.16). From Figure 3.16, nonlinearity may not well handled by this approach (e.g. an outlier with very low *Ozone* level is produced)

Table 3.4 shows that there is a similarity between the PAL-TBC and AL-TBC methods in estimating the parameters corresponding to the median function. Also, it suggests that there is no agreement between these approaches in estimating the parameters for the 0.25 and 0.75 quantile functions. By looking at Table 3.5, the AL-TBC method shows a slight reduction in the error, estimated using the leave-one-out cross-validation corresponding to the 0.25 quantile function, compared to the PAL-TBC method. However, the PAL-TBC method outperforms the AL-TBC method in terms of the errors corresponding to the 0.5 and 0.75 quantile functions.

## 3.5 Discussion

Simultaneous Box-Cox quantile regression with heteroscedastic error (PAL-HBC) shows an excellent ability to fit a nonlinear relationship with heteroscedasticity and shows superiority over individual Box-Cox quantile regression with heteroscedastic error (AL-HBC) by suggesting an optimal transformation for all quantiles rather than a different transformation for each quantile. Two-sided Box-Cox quantile regression can be considered as a good extension of Box-Cox

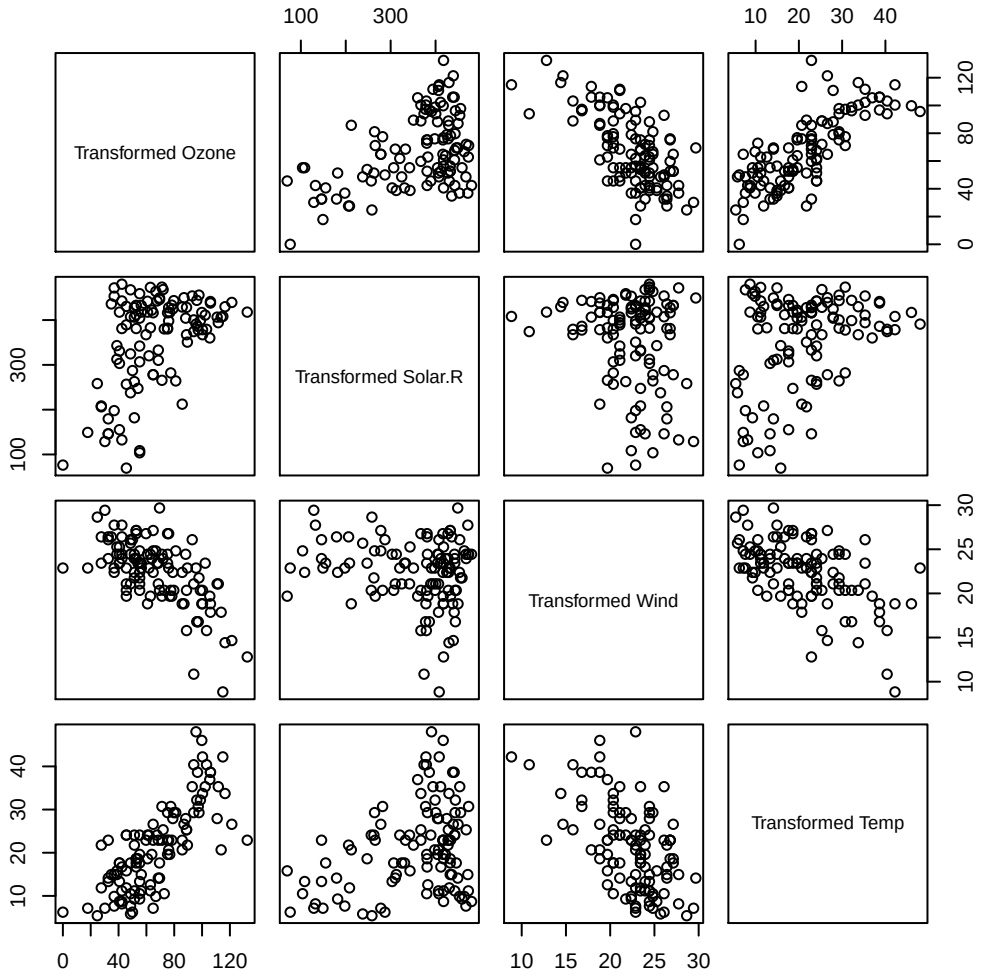



Figure 3.13: The scatter plot of relationships between the transformed variables obtained using the PAL-TBC method.

quantile regression that can deal with more complex models, which are considered beyond the simple Box-Cox transformation of the response variable. Simultaneous two-sided Box-Cox quantile regression (PAL-TBC) has a great ability to accommodate nonlinearity and heteroscedasticity for regression models with multiple covariates. Moreover, the simultaneous Box-Cox quantile methods can outperform the individual Box-Cox quantile methods, especially in estimating the extreme quantile functions.

The results of the analysis show that individual Box-Cox quantile methods handle the nonlinearity by suggesting a transformation estimated using observed

| PAL-TBC | | | AL-TBC | | |
|---|---|---|---|---|---|
| Coefficients | P.M. | 95% HPD | Coefficients | P.M. | 95% HPD |
| $\alpha_{0.25}$ | 53.90 | (26.85,78.44) | $\alpha_{0.25}$ | 31.94 | (–42.35, 87.00) |
| $\alpha_{0.5}$ | 62.21 | (34.76,85.89) | $\alpha_{0.5}$ | 56.80 | (–85.17,137.24) |
| $\alpha_{0.75}$ | 74.07 | (46.15,97.77) | $\alpha_{0.75}$ | 100.48 | ( 23.83,191.41) |
| | | | $\beta_{1,0.25}$ | 0.06 | ( 0.03, 0.10) |
| $\beta_1$ | 0.06 | ( 0.04, 0.08) | $\beta_{1,0.5}$ | 0.05 | ( 0.02, 0.09) |
| | | | $\beta_{1,0.75}$ | 0.04 | ( 0.01, 0.07) |
| | | | $\beta_{2,0.25}$ | −2.08 | ( −3.16, −0.99) |
| $\beta_2$ | −2.13 | (–2.54,–1.71) | $\beta_{2,0.5}$ | −2.09 | ( −2.95, −1.20) |
| | | | $\beta_{2,0.75}$ | −1.68 | ( −2.67, −0.64) |
| | | | $\beta_{3,0.25}$ | 1.44 | ( 1.08, 1.80) |
| $\beta_3$ | 1.39 | ( 1.20, 1.58) | $\beta_{3,0.5}$ | 1.39 | ( 1.06, 1.74) |
| | | | $\beta_{3,0.75}$ | 1.29 | ( 0.91, 1.66) |
| | | | $\lambda_{0.25}$ | 0.46 | ( 0.26, 0.67) |
| $\lambda$ | 0.32 | ( 0.21, 0.43) | $\lambda_{0.5}$ | 0.29 | ( 0.06, 0.52) |
| | | | $\lambda_{0.75}$ | 0.04 | ( −0.15, 0.24) |
| | | | $\eta_{1,0.25}$ | 0.50 | ( −0.54, 1.47) |
| $\eta_1$ | 0.36 | (–0.11, 0.88) | $\eta_{1,0.5}$ | 0.48 | ( −0.59, 1.81) |
| | | | $\eta_{1,0.75}$ | 0.61 | ( −0.81, 2.23) |
| | | | $\eta_{2,0.25}$ | 0.12 | ( −0.73, 1.11) |
| $\eta_2$ | −0.08 | (–0.55, 0.36) | $\eta_{2,0.5}$ | 0.00 | ( −1.11, 1.03) |
| | | | $\eta_{2,0.75}$ | 0.06 | ( −1.31, 1.52) |
| | | | $\eta_{3,0.25}$ | 4.60 | ( 2.17, 6.81) |
| $\eta_3$ | 4.11 | ( 2.49, 5.59) | $\eta_{3,0.5}$ | 3.96 | ( 0.78, 6.64) |
| | | | $\eta_{3,0.75}$ | 3.54 | ( 0.97, 6.29) |
| $\sigma_{0.25}$ | 4.22 | ( 3.44, 5.03) | $\sigma_{0.25}$ | 4.20 | ( 3.42, 5.02) |
| $\sigma_{0.5}$ | 5.58 | ( 4.53, 6.67) | $\sigma_{0.5}$ | 5.69 | ( 4.65, 6.80) |
| $\sigma_{0.75}$ | 4.64 | ( 3.83, 5.58) | $\sigma_{0.75}$ | 4.59 | ( 3.73, 5.46) |

Table 3.4: The posterior mean (P.M.) and the 95% highest posterior density intervals (95% HPD).

information, conveyed through a particular quantile function, that can not be sufficient to estimate the suitable transformation for the data. This leads to the suggestion of different transformations for the same dataset. Along with the fact that there is no theoretical justification for this variation in transformation

| $\tau$ | PAL-TBC | AL-TBC |
|---|---|---|
| 0.25 | 5.26 | 4.88 |
| 0.5 | 7.19 | 7.24 |
| 0.75 | 6.61 | 7.35 |

Table 3.5: The mean of errors estimated using the leave-one-out cross-validation, given in (2.11), for the different quantile functions.

parameters, it is complicated to interpret the quantile coefficients and transformation parameters. Thus, an optimal transformation parameter estimated using simultaneous Box-Cox quantile methods for all quantile levels has justification and practical motivations.

Although simultaneous Box-Cox quantile regression methods can be considered to improve the estimation and ensure the monotonic increase of the quantile function over different quantile levels, they could be sensitive to the number and locations of quantile functions included in the estimation. Therefore, to improve the performance of simultaneous Box-Cox quantile methods, this issue should be further investigated. Also, since the likelihood is a fundamental part of any Bayesian model, more research is needed on using the pseudo asymmetric Laplace likelihood to construct a Bayesian model to fit multiple quantile regressions simultaneously.

Despite the fact that Gibbs sampler algorithms - suggested to sample from the proposed posterior distributions in Chapters 2 and 3 - converge to the target distribution quicker than Metropolis-Hastings algorithms, these Gibbs samplers become less efficient as the number of observations or quantiles included in the estimation increases (that is, since $w_{\tau_k,i}$ in Algorithm 2 is generated separately for each $i$ and $k$). Therefore, to improve the performance of the proposed Box-Cox quantile methods, further research should be conducted on posterior computation methods, especially in the context of simultaneous quantile regression.

In terms of point estimation, the quantile methods based on the asymmetric Laplace likelihood and the pseudo asymmetric Laplace likelihood can show reliability. In the next chapters, we analyse and examine the use of the asymmetric Laplace likelihood and the pseudo asymmetric Laplace likelihood to make statistical inference on quantile coefficients from a Bayesian perspective by investigating the coverage probabilities.

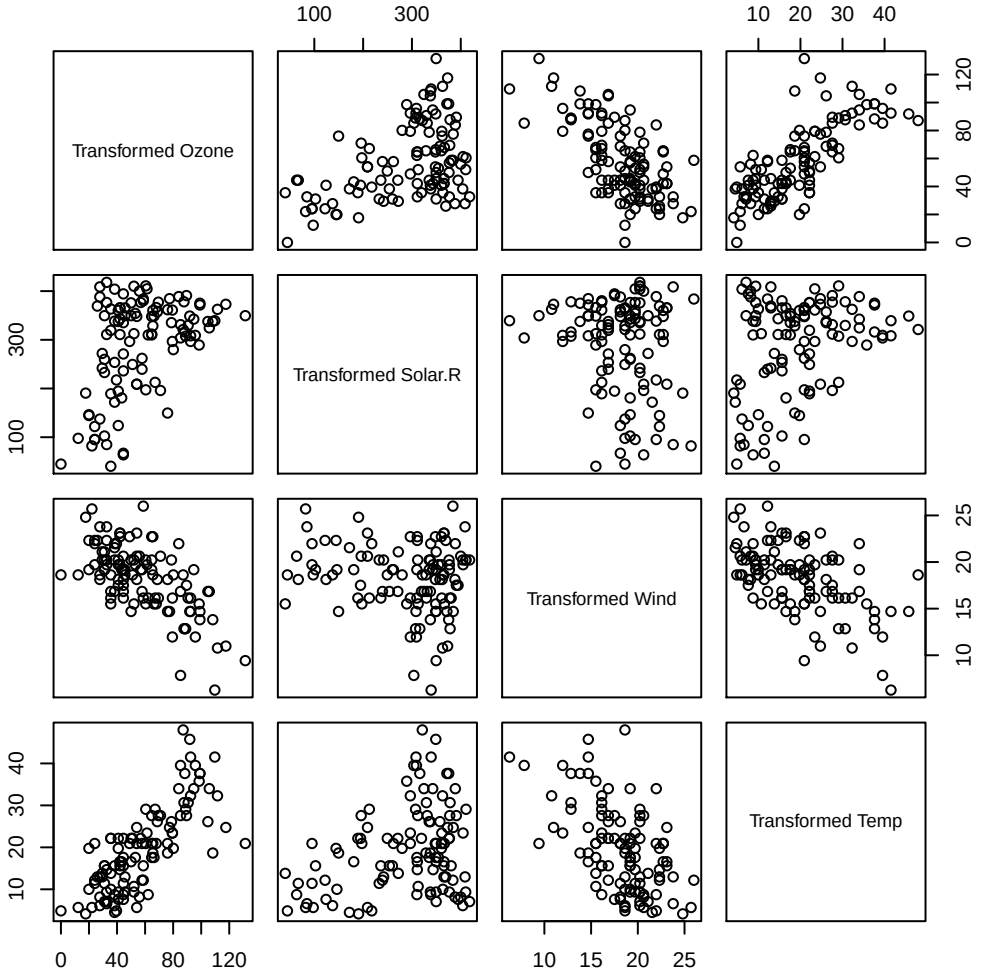



Figure 3.14: The scatter plot of relationships between the transformed variables obtained using the AL-TBC method corresponding to $\tau = 0.25$.

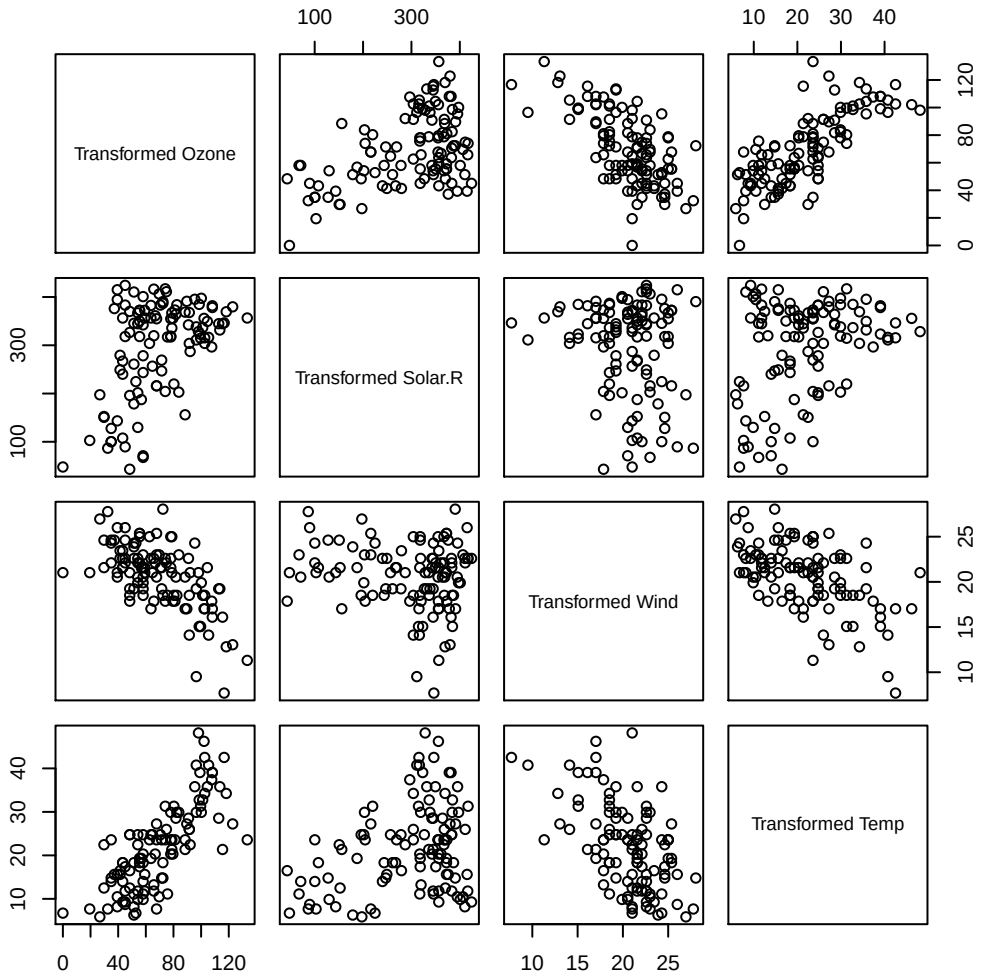



Figure 3.15: The scatter plot of relationships between the transformed variables obtained using the AL-TBC method corresponding to $\tau = 0.5$.

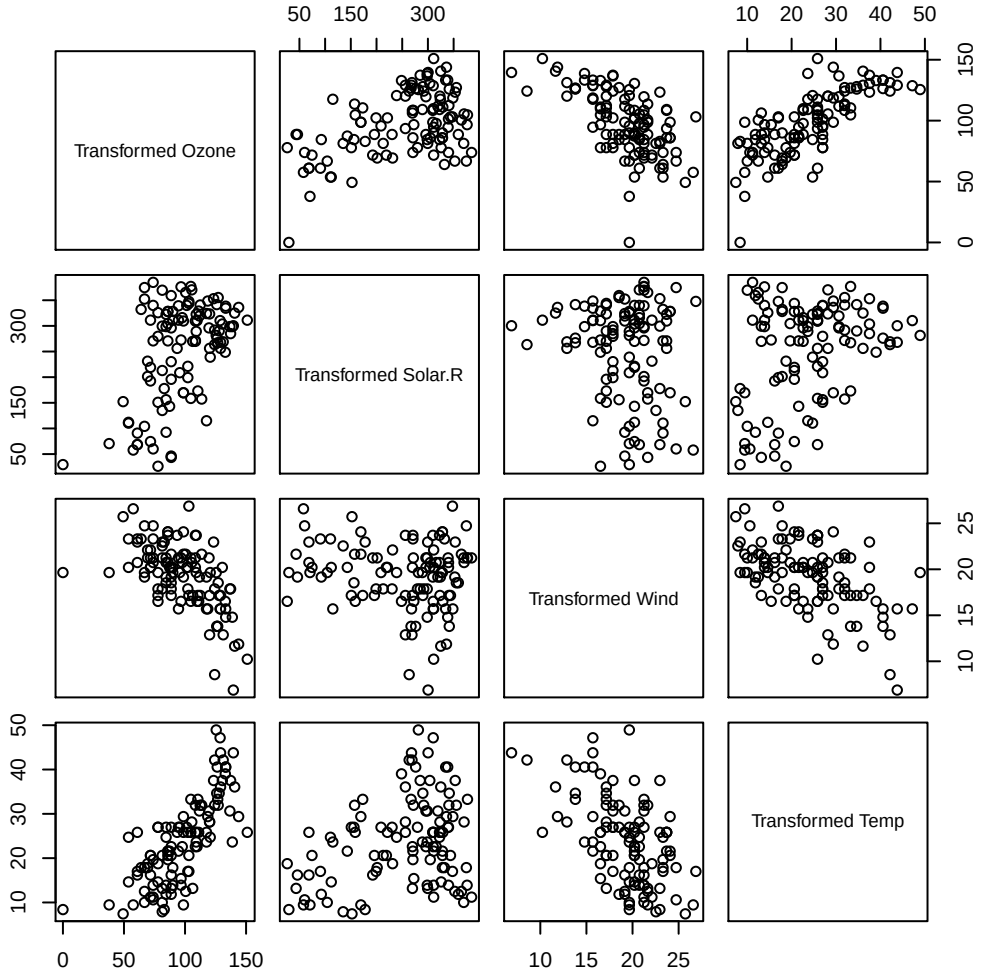



Figure 3.16: The scatter plot of relationships between the transformed variables obtained using the AL-TBC method corresponding to $\tau = 0.75$.

# Chapter 4

# On likelihood functions for Bayesian quantile estimation

## 4.1 Introduction

In frequentist statistics, there is no requirement to specify the joint distribution of the data and parameters corresponding to quantiles since the quantile regression is fitted by solving a linear programming problem. However, in the context of Bayesian statistics, the conditional distribution of the data given quantile parameters must be specified in order to determine the likelihood function that is a fundamental part of Bayesian models.

The literature on parametric quantile regression has focused on developing approximations for the underlying joint distribution of the data and parameters corresponding to quantile functions. The most widely used approximation is based on the asymmetric Laplace likelihood, considered by Yu & Moyeed (2001), whose maximisation is equivalent to the minimisation of a loss function considered by Koenker & Bassett (1978). Therefore, this likelihood function is considered a natural choice to construct a Bayesian model to fit quantile regression. There are a number of Bayesian quantile methods based on the asymmetric Laplace likelihood such as those proposed by Yu & Stander (2007), Geraci & Bottai (2007) and Rodrigues & Fan (2017); for more details see, Section 1.6.

The asymmetric Laplace distribution enjoys a number of attractive properties as an error distribution, such as the robustness against outliers and misspecification of the underlying distribution. The Bayesian quantile methods, based on

the asymmetric Laplace likelihood, can show a reliability in terms of point estimation. However, we need more investigation to understand this likelihood and to deal with its limitations. This aim can be achieved by checking other aspects of Bayesian inference such as coverage probabilities.

In this chapter, we discuss Bayesian quantile regression based on the asymmetric Laplace likelihood and we illustrate the limitations of this Bayesian approach. In addition, to deal with these limitations, we develop a new probability distribution that can accommodate a wide variety of underlying distributions whatever the quantile of interest.

## 4.2 Review of asymmetric Laplace likelihood

To develop a likelihood ratio test for quantile regression, Koenker & Machado (1999) used the asymmetric Laplace distribution which was employed later by Yu & Moyeed (2001) to construct a Bayesian method to estimate quantile curves separately. The asymmetric Laplace distribution can be represented by the following probability density function:

$$f_\tau(y|\mu_\tau,\sigma) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{y-\mu_\tau}{\sigma}\right)\right\}, \tag{4.1}$$

where $0 < \tau < 1$ is the skew parameter, $\sigma > 0$ is the scale parameter, $-\infty < \mu_\tau < \infty$ is the location parameter, and $\rho_\tau(.)$ is a loss function given by

$$\rho_\tau(u) = \frac{|u| + (2\tau-1)u}{2}. \tag{4.2}$$

The asymmetric Laplace distribution is skewed to the left when $\tau > 0.5$ and skewed to the right when $\tau < 0.5$. The reason behind choosing this likelihood to develop Bayesian models to fit quantile regressions is *not* an attempt to model the true error distribution. Rather, it is that the maximisation of the asymmetric Laplace likelihood function is exactly equivalent to the minimisation of the loss function given in (4.2). Yu & Moyeed (2001) concluded that the use of the asymmetric Laplace distribution to develop Bayesian quantile methods is satisfactory whatever the underlying distribution. Sriram *et al.* (2013) established posterior consistency to suggest a theoretical justification to use this likelihood for Bayesian

inference on quantiles under the misspecification of the true likelihood. They suggested an asymptotic justification for using the asymmetric Laplace distribution as an error distribution under several conditions on the covariates, the priors and the underlying distribution.

Although the asymmetric Laplace distribution has received growing attention in the context of Bayesian quantile regression, it has some limitations that can lead to concern about the reliability of Bayesian inference, obtained using this likelihood, on quantiles. The first limitation is that the asymmetric Laplace distribution has a fixed shape with sharp peak at the quantile of interest. Therefore, it will not approximate the underlying distribution well. For fixed $\tau$, the estimated mode of the asymmetric Laplace density given in (4.1) represented by $\mu_\tau$ does not correspond to the mode of the data. The second limitation is that the quantile of interest and skewness are controlled by the same parameter. For example, the median can be estimated by setting $\tau = 0.5$, and this implies that the symmetric Laplace likelihood must be used as an underlying distribution. Therefore, by using the Bayesian model based on the asymmetric Laplace distribution, we assume different distributions for the same dataset to be able to estimate a variety of quantile functions. Therefore, when this method leads to a good estimation for a particular quantile, this does not imply good estimation for other quantile functions. The quality of the estimation depends on the similarity between the asymmetric Laplace distribution corresponding to the desired quantile function and the underlying distribution. For example, if the true distribution is normal, the best coverage probability can be achieved for the median and the worst coverage probabilities are for extreme quantiles. Generally, we can only achieve good estimation with reasonable coverage probabilities for the quantile functions that are close to the mode of the data. These limitations make the asymmetric Laplace distribution powerless to accommodate the underlying distributions. Thus, the Bayesian inference on quantiles obtained using the asymmetric Laplace likelihood can be poor and unreliable. In an attempt to address this gap in the literature, we propose a new distribution that can accommodate the underlying distribution regardless of the quantile of interest.

## 4.3 Generalised distributions for quantile estimation

To handle the limitations of the asymmetric Laplace likelihood and increase its flexility to accommodate a variety of underlying distributions, we investigate a generalised asymmetric Laplace distribution, with a shape parameter $\gamma > 0$, given by the density function:

$$f(y|\mu_\tau, \sigma, \gamma) = \begin{cases} \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})} \exp\left\{-\left[(1-\tau)|\frac{y-\mu_\tau}{\sigma}|\right]^\gamma\right\}, & y \leq \mu_\tau, \\ \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})} \exp\left\{-\left[\tau|\frac{y-\mu_\tau}{\sigma}|\right]^\gamma\right\}, & y > \mu_\tau, \end{cases} \tag{4.3}$$

where $\mu_\tau$ is $\tau$ quantile and $\sigma$ is the scale parameter. This generalisation satisfies the properties:

$$\int_{-\infty}^{\infty} f(y|\mu_\tau, \sigma, \gamma)dy = 1 \text{ and } \int_{-\infty}^{\mu_\tau} f(y|\mu_\tau, \sigma, \gamma)dy = \tau.$$

For more details, see Appendix C.1.

It is obvious that the generalised asymmetric Laplace likelihood function no longer corresponds to the loss function given in (4.2), where it forms an approximation of the underlying distribution. This implies that the quality of the estimation depends on the ability of this generalisation to accommodate the true error distributions, and this may not be achieved for all quantile levels. The generalised form given in (4.3) includes a variety of distributions as special cases, such as the asymmetric Laplace distribution and a skewed normal distribution which correspond to $\gamma = 1$ and $\gamma = 2$, respectively. Also, as the shape parameter $\gamma$ approaches infinity, the density $f(y|\mu_\tau, \sigma, \gamma)$ becomes a uniform distribution over the interval $(\mu_\tau - \sigma, \mu_\tau + \sigma)$. By using this distribution, it is still assumed that the mode of the underlying distribution is represented by the quantile of interest. Therefore, the shape parameter can offer a flexibility that can lead to an improvement in the estimation of quantile functions which are very close to the true mode of the dataset. However, in the case of extreme quantiles, in attempt to accommodate the true mode of data, the shape parameter $\gamma$ approaches infinity; that is, this generalisation becomes an approximate uniform distribution. This can result in poor and biased estimation for quantiles located in the tails

of the true distribution. Moreover, since this approximate distribution no longer shares the estimation properties with the loss function given by (4.2), it is very sensitive to misspecification of the underlying distribution. Thus, this generalisation cannot be the promising solution for limitations of the asymmetric Laplace distribution.

To overcome these drawbacks, we assume that the joint distribution of the data is represented by a new generalisation of the Gumbel distribution described by Johnson *et al.* (1994, p.2). This generalised Gumbel distribution has five parameters that give it more flexibility to accommodate different types of underlying distributions. This generalisation is described by the following probability density:

$$f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma}) = \begin{cases} \frac{\gamma_1\tau(1-\tau)e^{\gamma_1}}{(e^{\gamma_1}-1)\sigma_1}\exp\left\{-(1-\tau)\left|\frac{y-\mu_\tau}{\sigma_1}\right| - \gamma_1\exp\left[-(1-\tau)\left|\frac{y-\mu_\tau}{\sigma_1}\right|\right]\right\}, & y \le \mu_\tau, \\ \frac{\gamma_2\tau(1-\tau)e^{\gamma_2}}{(e^{\gamma_2}-1)\sigma_2}\exp\left\{-\tau\left|\frac{y-\mu_\tau}{\sigma_2}\right| - \gamma_2\exp\left[-\tau\left|\frac{y-\mu_\tau}{\sigma_2}\right|\right]\right\}, & y > \mu_\tau, \end{cases}$$

where $\mu_\tau$ is $\tau$ quantile, $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ are the vectors of the scale and shape parameters. It is obvious that this density is discontinuous when $\mu_\tau = 0$ (that is, jumping). This is because that the density function $f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma})$ is restricted to satisfy the following properties:

$$\int_{-\infty}^{\infty} f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma})dy = 1 \text{ and } \int_{-\infty}^{\mu_\tau} f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma})dy = \tau,$$

in order to preserve that the estimator $\mu_\tau$ is interpreted as a quantile (for more details, see Appendix C.2). Also, since the shape and scale parameters are allowed to be different for each part of the distribution to ensure the flexibility of the density $f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma})$ to accommodate different distributions without assuming that the quantile of interest is the true mode of data. However, this discontinuity will be disappeared when $\gamma_1 \approx \gamma_2$ and $\sigma_1 \approx \sigma_2$ (for example, in the case of estimating the median for the normal distribution). In the context of approximation, it is complicated to figure out the effect of the discontinuity on the consistency and the bias of the estimation, since these properties depend on the overall approximation of the underlying distribution. Since the quality of the approximation obtained using this density can vary with respect to the underlying distribution and the quantile to be estimated, we use Kullback-Leibler

divergence to investigate numerically how this approximation diverges from the true distribution of the data. Along with this, the converge probabilities and the lack of bias are examined (for more details, see Sections 4.5 & 4.6). A number of Bayesian quantile methods were developed using this type of densities (for example, see Feng *et al.*, 2015; Hahn & Burgette, 2012). The cumulative distribution function of this distribution can be written as follows

$$F(y;\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma}) = \begin{cases} \frac{\tau e^{\gamma_1}}{e^{\gamma_1}-1} h\left(\gamma_1 \exp\left\{(1-\tau)\left[\frac{y-\mu_\tau}{\sigma_1}\right]\right\}\right), & y \leq \mu_\tau, \\ 1 - \frac{(1-\tau)e^{\gamma_2}}{e^{\gamma_2}-1} h\left(\gamma_2 \exp\left\{-\tau\left[\frac{y-\mu_\tau}{\sigma_2}\right]\right\}\right), & y > \mu_\tau, \end{cases} \tag{4.4}$$

where $h(x) = 1 - e^{-x}$. This cumulative distribution function is continuous and it has the value of $\tau$ at $y = \mu_\tau$.

## 4.4 An approximate quantile estimation for linear models

Consider the linear regression model

$$y_i = \alpha + \boldsymbol{x}_i^{'}\boldsymbol{\beta} + \epsilon_i, \quad \text{for } i = 1, ...n$$

where $\epsilon_i$ is independently distributed with a mean equal to zero and a constant variance; $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates for the $i^{\text{th}}$ observation; $\alpha$ is an intercept coefficient and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown slope coefficients. Then, the linear quantile function is given by

$$Q(\tau_k|\boldsymbol{x_i}) = \alpha_{\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}.$$

To develop a Bayesian quantile method to estimate this linear quantile function for $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$ individually, we can consider the generalised Gumbel likelihood given by

$$l(\alpha_{\tau_k}, \boldsymbol{\beta}_{\tau_k}, \boldsymbol{\sigma}, \boldsymbol{\gamma}; \boldsymbol{y}) \propto \prod_{i=1}^{n} \begin{cases} \frac{\gamma_1 e^{\gamma_1}}{(e^{\gamma_1}-1)\sigma_1} \exp\left\{-(1-\tau_k)\left|\frac{y_i-\mu_i}{\sigma_1}\right| - \gamma_1 \exp\left[-(1-\tau_k)\left|\frac{y_i-\mu_i}{\sigma_1}\right|\right]\right\}, & y_i \leq \mu_i, \\ \frac{\gamma_2 e^{\gamma_2}}{(e^{\gamma_2}-1)\sigma_2} \exp\left\{-\tau_k\left|\frac{y_i-\mu_i}{\sigma_2}\right| - \gamma_2 \exp\left[-\tau_k\left|\frac{y_i-\mu_i}{\sigma_2}\right|\right]\right\}, & y_i > \mu_i, \end{cases}$$

where $\mu_i = \alpha_{\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k}$. For all parameters included in the model assumed to be a priori independent, we consider relatively diffuse prior distributions that lead to a proper posterior distribution. For quantile regression coefficients, the prior distribution is given by

$$\alpha_{\tau_k} \sim \mathrm{N}(\alpha^*, \sigma_\alpha),$$
$$\boldsymbol{\beta}_{\tau_k} \sim \mathrm{N}_p(\boldsymbol{\beta}^*, \Sigma_\beta),$$

where each normal distribution has mean equal to zero and variance equal to $10^5$. Also, each multivariate normal distribution has a mean vector of zeros and a diagonal covariance matrix with entries equal to $10^5$. For the scale and shape parameters, we consider the inverse-gamma distribution (IG) that is given by

$$\sigma_t \sim \mathrm{IG}(a, b),$$
$$\gamma_t \sim \mathrm{IG}(a, b),$$

for $t = 1, 2$, where the inverse-gamma parameters are set to be $a = b = 0.01$. The posterior distribution is then given by

$$p(\alpha_{\tau_k}, \boldsymbol{\beta}_{\tau_k}, \boldsymbol{\sigma}, \boldsymbol{\gamma}|\boldsymbol{y}) \propto l(\alpha_{\tau_k}, \boldsymbol{\beta}_{\tau_k}, \boldsymbol{\sigma}, \boldsymbol{\gamma}; \boldsymbol{y}) p(\alpha_{\tau_k}) p(\boldsymbol{\beta}_{\tau_k}) p(\boldsymbol{\sigma}) p(\boldsymbol{\gamma}).$$

Then, a random walk Metropolis-Hastings algorithm is used to sample from this posterior distribution.

## 4.5 Univariate simulation study

In this section we implement a simulation study to investigate the performance of the asymmetric Laplace and the generalised Gumbel distributions, when used as error distributions, to develop Bayesian quantile methods. We use Bayesian methods based on the asymmetric Laplace likelihood and the generalised Gumbel likelihood functions to estimate quantiles corresponding to 0.05, 0.25, 0.5, 0.75 and 0.95 individually for 1000 samples of size 150 from normal distribution $\mathrm{N}(0, 1)$, Student's $t$ distribution $t(3)$ and gamma distribution $\mathrm{Gamma}(3, 3)$. Then, we plot the average of estimated density and cumulative functions computed at the maximum a posteriori estimators.

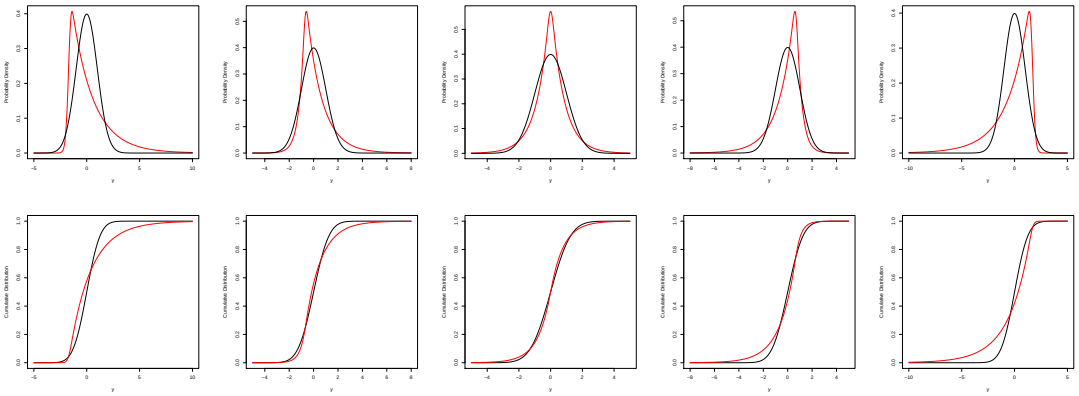


Figure 4.1: The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for normal distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the asymmetric Laplace distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves.

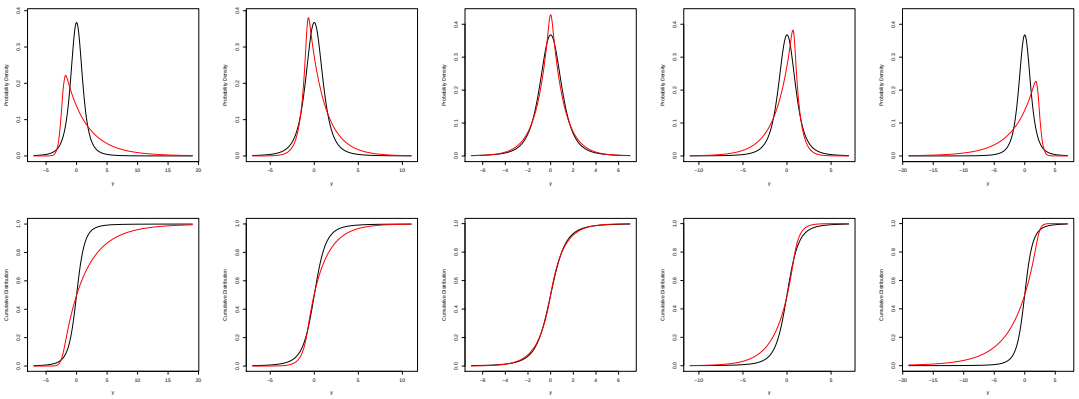


Figure 4.2: The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for Student's $t$ distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the asymmetric Laplace distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves.

Figures 4.1, 4.2 and 4.3 show how the asymmetric Laplace distribution is maximised at the desired quantile which may not match the true mode of the data. In addition, they show how this approach assumes different distributions for the

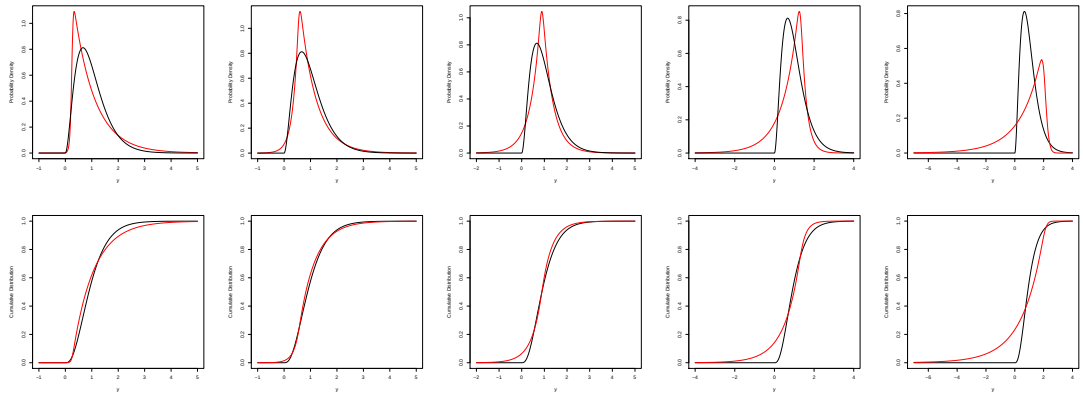


Figure 4.3: The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for gamma distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the asymmetric Laplace distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves.

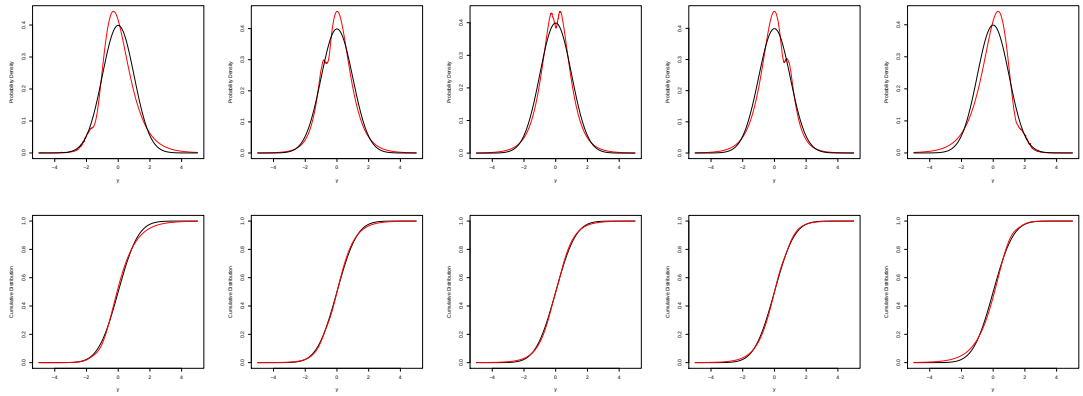


Figure 4.4: The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for normal distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the generalised Gumbel distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves.

same dataset and how this depends on the quantile to be estimated. This results in a good approximation for the underlying distribution when the desired quantile to be estimated is close to the true mode of the dataset, and poor approxima-

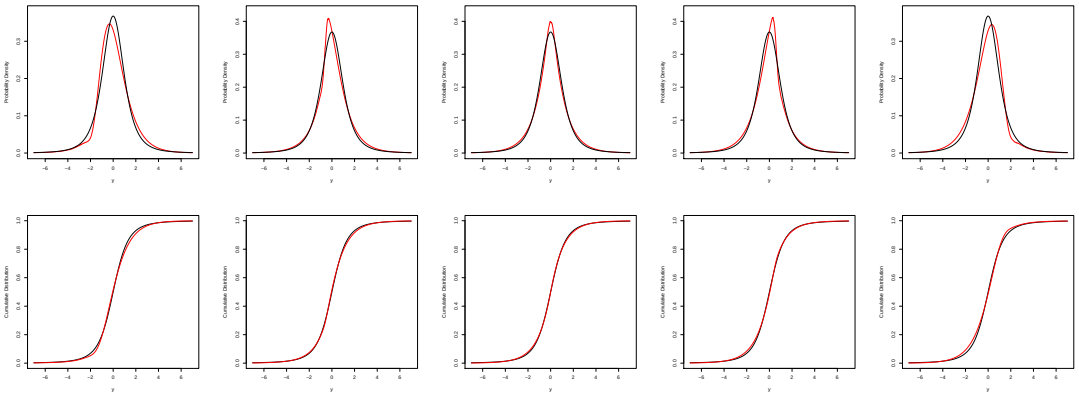


Figure 4.5: The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for Student's $t$ distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the generalised Gumbel distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves.

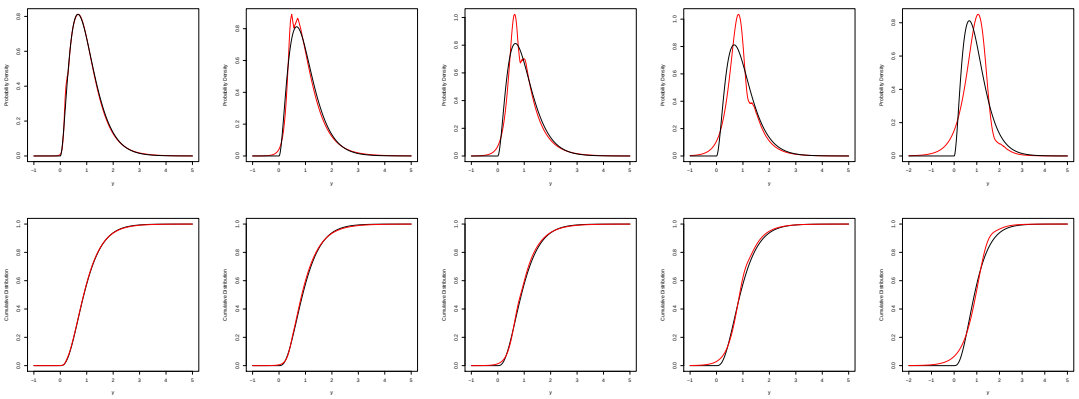


Figure 4.6: The average of the estimated probability density functions (top) and the estimated cumulative functions (bottom) for gamma distribution based on 1000 simulations. The true distribution is represented by black curves and its approximations obtained using the generalised Gumbel distribution, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (left to right), are represented by red curves.

tions in the case of other quantiles especially the extreme ones. Moreover, since the asymmetric Laplace distribution has a fixed shape with a sharp peak, it also cannot accommodate the underlying distribution when the quantile of interest

and the true mode of the dataset are identical (e.g. the estimation of the median for the normal distribution). On the other hand, Figures 4.4, 4.5 and 4.6 show that the generalised Gumbel distribution has great flexibility to provide good approximations for an underlying distribution regardless of the quantile of interest. Also, the generalised Gumbel distribution no longer deals with the desired quantile as the mode of the approximate distribution and attempts to estimate the true mode of the dataset whatever the required quantile. In addition, the generalised Gumbel distribution has a flexible shape that can accommodate different underlying distributions. For further investigation, we use the Kullback–Leibler divergence given by

$$KLD\left(f\|\widehat{f}\right) = \int_{-\infty}^{\infty} f(y)\log\left(\frac{f(y)}{\widehat{f}(y)}\right)dy,$$

where $f$ is the true distribution of the data and $\widehat{f}$ is the approximate distribution. Then, Table 4.1 shows that the Kullback–Leibler divergence of the asymmetric Laplace distribution with respect to the true distribution achieves its minimum when the quantile of interest is close to the true mode of the data and increases as the quantiles are further away. On the other hand, Table 4.2 shows how the generalised Gumbel likelihood tends to provide similarly accurate approximations for the underlying distribution whatever the quantile of interest. Also, it shows that by using the generalised Gumbel distribution, the Kullback–Leibler divergence tends to share the same minima for all quantiles.

| $\epsilon_i$ | 0.05 | 0.25 | 0.5 | 0.75 | 0.95 |
|---|---|---|---|---|---|
| N$(0,1)$ | 0.38 | 0.12 | 0.06 | 0.12 | 0.38 |
| Student's $t(3)$ | 0.68 | 0.14 | 0.03 | 0.14 | 0.68 |
| Gamma$(3,3)$ | 0.10 | 0.05 | 0.13 | 0.32 | 0.76 |

Table 4.1: The average of Kullback–Leibler divergence, from the approximation obtained using the asymmetric Laplace distribution to the true distribution, based on 1000 simulations.

## 4.6 Linear models

To investigate the quality of approximations achieved using the asymmetric Laplace

| $\epsilon_i$ | 0.05 | 0.25 | 0.5 | 0.75 | 0.95 |
|---|---|---|---|---|---|
| N$(0,1)$ | 0.07 | 0.04 | 0.05 | 0.04 | 0.06 |
| Student's $t(3)$ | 0.07 | 0.05 | 0.04 | 0.05 | 0.07 |
| Gamma$(3,3)$ | 0.04 | 0.05 | 0.06 | 0.10 | 0.19 |

Table 4.2: The average of Kullback–Leibler divergence, from the approximation obtained using the generalised Gumbel distribution to the true distribution, based on 1000 simulations.

distribution and the generalised Gumbel distribution, we examine the accuracy of the coverage probabilities of 95% HPD intervals estimated using Bayesian quantile methods. The asymmetric Laplace likelihood and the generalised Gumbel likelihood are used along with diffuse priors that could not deliver any significant information about the parameters of the model. We fit quantile regressions for 1000 datasets with sample size of 200 generated from the linear model given by

$$y_i = 1 + x_i + \epsilon_i,$$

where $x_i \sim \text{U}(20, 30)$ and $\epsilon_i$ is considered to follow different distributions. The simulated datasets only differ in the errors $\epsilon_i$.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\alpha_{0.95}$ | $\beta_{0.05}$ | $\beta_{0.25}$ | $\beta_{0.5}$ | $\beta_{0.75}$ | $\beta_{0.95}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| N$(0,1)$ | 0.68 | 0.87 | 0.89 | 0.87 | 0.66 | 0.69 | 0.87 | 0.89 | 0.86 | 0.66 |
| Student's $t(3)$ | 0.58 | 0.88 | 0.93 | 0.88 | 0.61 | 0.57 | 0.88 | 0.93 | 0.88 | 0.61 |
| Gamma$(3,3)$ | 0.79 | 0.88 | 0.89 | 0.84 | 0.61 | 0.78 | 0.88 | 0.89 | 0.85 | 0.60 |

Table 4.3: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated individually using Bayesian quantile method based on the asymmetric Laplace likelihood over 1000 simulations.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\alpha_{0.95}$ | $\beta_{0.05}$ | $\beta_{0.25}$ | $\beta_{0.5}$ | $\beta_{0.75}$ | $\beta_{0.95}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| N$(0,1)$ | 0.88 | 0.92 | 0.91 | 0.90 | 0.88 | 0.88 | 0.90 | 0.92 | 0.90 | 0.88 |
| Student's $t(3)$ | 0.90 | 0.91 | 0.93 | 0.91 | 0.91 | 0.90 | 0.92 | 0.93 | 0.91 | 0.91 |
| Gamma$(3,3)$ | 0.92 | 0.90 | 0.90 | 0.86 | 0.80 | 0.90 | 0.90 | 0.90 | 0.87 | 0.86 |

Table 4.4: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated individually using Bayesian quantile method based on the generalised Gumbel likelihood over 1000 simulations.

Consistent with the discussion in the previous section, Table 4.3 shows that the Bayesian quantile method based on the asymmetric Laplace likelihood gives different coverage probabilities for different quantiles. This approach gives coverage probabilities that approach the nominal coverage probabilities as the quantile to be estimated gets closer to the actual mode of the underlying distribution, and gives lower coverage probabilities for the extreme quantiles. Table 4.4 shows how using the generalised Gumbel likelihood to construct Bayesian quantile model gives improved coverage probabilities for different levels of quantiles where all coverage probabilities approach the nominal coverage probabilities with great improvement compared to those achieved using the Bayesian quantile method based on the asymmetric Laplace likelihood. Despite the relatively large number of parameters in the Bayesian model based on the generalised Gumbel likelihood, it still enjoys fast convergence to the target distribution and stability of the posterior computation using the Metropolis-Hastings algorithm. Thus, it is obvious that the Bayesian quantile method based on the generalised Gumbel likelihood outperforms the other method based on the asymmetric Laplace likelihood in terms of the coverage probability for all quantiles. By looking at the distributions of the maximum a posteriori estimators shown in Figures 4.7 and 4.8, neither approach shows any undesirable behaviour in the estimation of the maximum a posteriori. Regarding the lack of bias, the Bayesian quantile method based on the generalised Gumbel likelihood performs better when the generalised Gumbel distribution corresponding to the quantile of interest can approximate the underlying distribution well. Otherwise, the Bayesian quantile method based on the asymmetric Laplace can be better, although the maximum a posteriori estimators obtained using this method tend to be much more variable compared to others obtained using Bayesian quantile method based on the generalised Gumbel likelihood. This is because the generalised Gumbel distribution is no longer linked to the loss function given in (4.2) that is minimised when the number of observations under the quantile curve is almost equal to $\tau n$. Whereas, since the asymmetric Laplace likelihood has a direct link to this loss function, it is maximised when the number of observations under the quantile curve is almost equal to $\tau n$ and this gives the asymmetric Laplace distribution some flexibility to work under misspecification of the underlying distribution. Moreover, the Bayesian method based on the generalised Gumbel likelihood tends to show the same uncertainty

about the true quantile function for all quantile levels, while the other method based on the asymmetric Laplace likelihood shows a greater uncertainty about the true functions of the extreme quantiles.

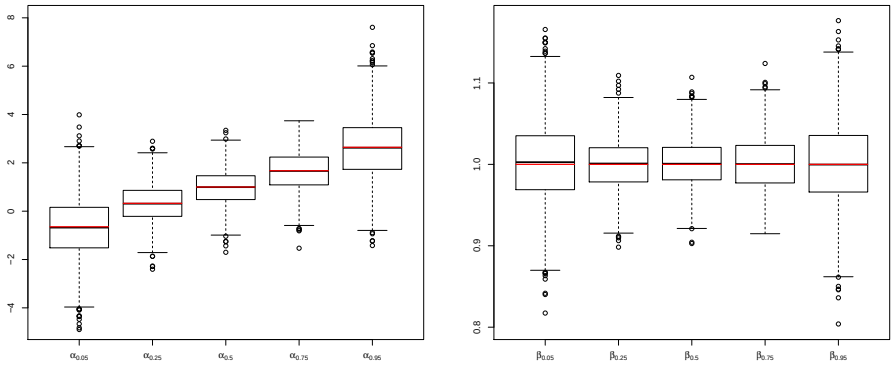
## 4.7 Discussion

The asymmetric Laplace likelihood has a direct link to the loss function used in classic statistics to estimate the quantile functions. Therefore, when the approximate distribution departs from the true distribution of the data, the Bayesian method based on the asymmetric Laplace likelihood may be better than other approaches, especially in terms of point estimation. However, the Bayesian method based on the asymmetric Laplace likelihood does not give a good approximation of the underlying distribution that can be used to draw reliable Bayesian inference on quantile curves, especially for non-complex models where its alternatives can provide more accurate and reliable inference.

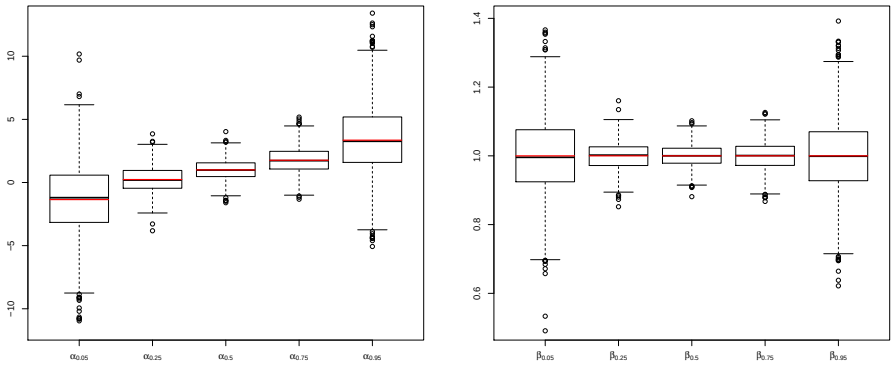
The generalised Gumbel likelihood can be considered a great alternative to the asymmetric Laplace likelihood to construct an efficient Bayesian quantile method. In contrast to the asymmetric Laplace distribution, the generalised Gumbel distribution attempts to approximate the underlying distribution whatever the quantile of interest. Although the quantile estimation based on the generalised Gumbel likelihood no longer corresponds to the traditional loss function, the Bayesian quantile method based on this likelihood shows flexibility to fit quantile regressions under different error distributions and can outperform the Bayesian quantile method based on the asymmetric Laplace likelihood.

Improving Bayesian estimation based on the asymmetric Laplace distribution for all quantiles is difficult. This is because, by using this approach to estimate several quantiles individually, we assume different distributions for the same dataset. Therefore, accurate Bayesian inference on a particular quantile does not imply accurate Bayesian inference for other quantiles. To overcome this problem, we may need to combine these different distributions in a suitable way to form the underlying distribution. In the next chapter, we discuss the pseudo asymmetric Laplace likelihood used in Chapter 2 and 3 from a Bayesian viewpoint. In addition, we develop a mixture of the asymmetric Laplace distributions corresponding to the individual quantiles to form the joint distribution

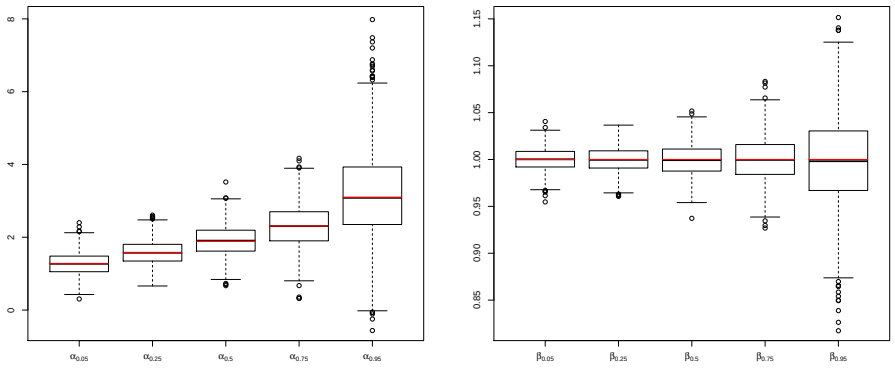
of the multiple quantiles and provide a good approximation for the underlying distribution of the data.



(a) $\epsilon_i \sim \mathrm{N}(0,1)$

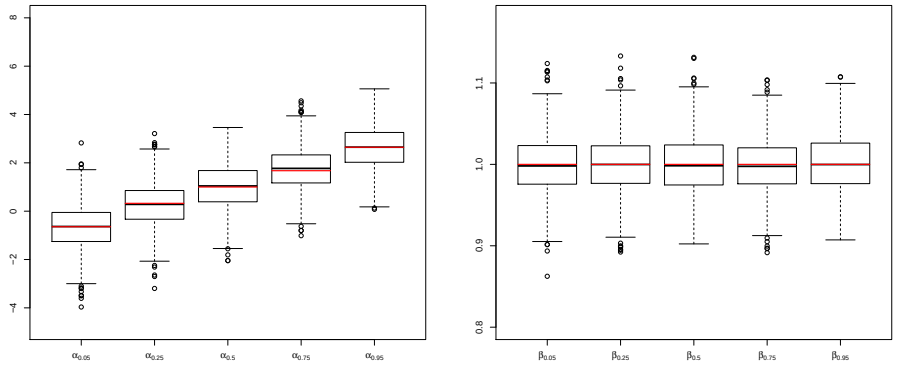


(b) $\epsilon_i \sim t(3)$

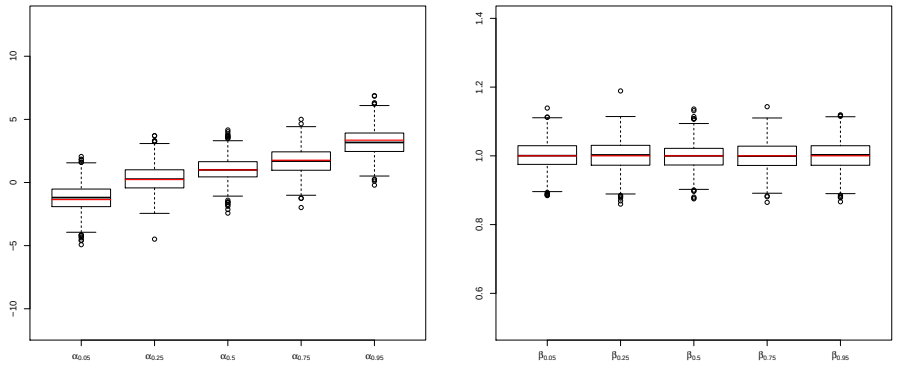


(c) $\epsilon_i \sim \mathrm{Gamma}(3,3)$

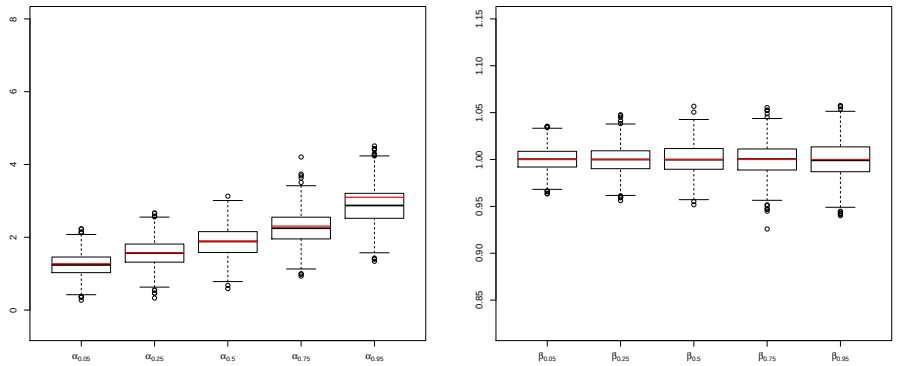
Figure 4.7: The box-plot of the maximum a posteriori estimators, obtained individually using Bayesian quantile method based on the asymmetric Laplace likelihood, over 1000 simulations. The red lines represent the true values.




(a) $\epsilon_i \sim \mathrm{N}(0, 1)$




(b) $\epsilon_i \sim t(3)$




(c) $\epsilon_i \sim \mathrm{Gamma}(3, 3)$

Figure 4.8: The box-plot of the maximum a posteriori estimators, obtained individually using Bayesian quantile method based on the generalised Gumbel likelihood, over 1000 simulations. The red lines represent the true values.

# Chapter 5

# On the approximation of the joint distribution of quantiles

## 5.1 Introduction

Although simultaneous estimation of multiple quantile functions is favoured over individual quantile estimation, in that simultaneous estimation can ensure monotonically increasing quantile functions in quantile level $\tau$ (that is, avoiding the issue of crossing quantile curves), investigation of the properties of simultaneous estimation from a Bayesian perspective is an outstanding research issue.

The asymmetric Laplace likelihood is the most widely used likelihood in Bayesian quantile regression. However, estimating quantile functions separately using a Bayesian method based on the asymmetric Laplace likelihood assumes a different error distribution for each quantile level. This implies that good estimation for a particular quantile does not mean good estimations for other quantile functions. The quality of estimation depends on the similarity between the asymmetric Laplace likelihood corresponding to the quantile of interest and the true error distribution. Also, the quantile functions can cross.

Simultaneous estimation of multiple quantile functions based on the pseudo asymmetric Laplace density discussed in Chapter 2 & 3 can be considered as a solution for these problems. However, it does not provide a good approximation for the underlying distribution and this may lead to poor Bayesian inference for the quantile regression coefficients.

Since valid Bayesian inference about quantile curves require a good approximation for the underlying distribution, we employ the asymmetric Laplace distributions corresponding to individual quantile functions to develop a novel approximation, based on a weighted pseudo asymmetric Laplace distribution, for the true error distribution.

In this chapter, we discuss the properties of the pseudo asymmetric Laplace likelihood. We introduce a weighted pseudo asymmetric Laplace distribution that can be used to approximate the underlying distribution very well. In addition, we discuss different specifications of weights. Then, we implement a simulation study to examine the ability of the proposed methods to approximate the true distribution of the data and to estimate the quantiles. Also, we consider a real-data application.

## 5.2 The pseudo asymmetric Laplace distribution

The pseudo asymmetric Laplace (PAL) distribution, formed by the product of asymmetric Laplace distributions corresponding to the quantile functions included in the estimation, can be represented by the following pseudo probability density $\widehat{f}$:

$$\widehat{f}\left(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau\right) \propto \left[\prod_{k=1}^{m} \frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{m} \rho_{\tau_k}\left(\frac{y-\mu_{\tau_k}}{\sigma_{\tau_k}}\right)\right\}.$$

Although the Bayesian method based on this pseudo distribution shows reliable estimation for the quantile regression coefficients in terms of point estimation, which is due to its link to the composite optimisation problem as shown in Chapter 2, it provides poor approximation for the underlying distribution. Accumulated errors over all quantile functions included in the simultaneous estimation result in low density at all observations. This results in some observations, which are from the tails of the underlying distribution, having an approximate density value of zero. In other words, the approximate distribution obtained using the pseudo asymmetric Laplace density has a smaller scale than the true distribution. In addition, the coverage probabilities decrease as the number of quantile func-

tions included in simultaneous estimation increases. Thus, this approximation becomes worse as the number of quantiles included in the simultaneous estimation increases. This leads to an actual concern about Bayesian inference based on this likelihood. Therefore, to approximate the underlying distribution effectively, we need to control the contribution played by each asymmetric Laplace density corresponding to individual quantiles in forming the joint pseudo asymmetric Laplace distribution of the quantiles. In the next section, we propose a weighted pseudo asymmetric Laplace distribution that can approximate the underlying distribution more accurately. This can then be used to develop efficient Bayesian multiple quantile regressions without crossing.

## 5.3 The weighted pseudo asymmetric Laplace distribution

For a finite set of probability density functions given by $f_1(y), ..., f_m(y)$, the mixture distribution formed by averaging these densities can be represented by the density function:

$$f(y) = \sum_{k=1}^{m} w_k f_k(y),$$

where $w_k$ is a weight such that $\sum_{k=1}^{m} w_k = 1$. Then, $f(y)$ is the weighted arithmetic mean of the given set of probability density functions. Similarly, the distribution formed by multiplying these densities can be represented by function:

$$g(y) = \prod_{k=1}^{m} \big[f_k(y)\big]^{w_k},$$

which is no longer a density function. Then, $g(y)$ is the weighted geometric mean of the given set of probability density functions. However, the properties of the estimator based on $g$ is related to the sum of the weighted logarithm of density functions:

$$h(y) = \sum_{k=1}^{m} w_k \log\Big(f_k(y)\Big).$$

For a random variable $y$, we assume that $y \sim F$, where $F$ is a continuous distribution. Then, the probability density $f$ of the distribution $F$ can be approximated

by the weighted pseudo asymmetric Laplace (WPAL) density given by

$$\begin{aligned}\widehat{f}(y|\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau) &\propto \begin{cases} 0 & ,\boldsymbol{\mu}_\tau \notin A, \\ \prod_{k=1}^{m}\left[\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\exp\left\{-\rho_{\tau_k}\left(\frac{y-\mu_{\tau_k}}{\sigma_{\tau_k}}\right)\right\}\right]^{w_{\tau_k}} & ,\boldsymbol{\mu}_\tau \in A, \end{cases} \\ &\propto \begin{cases} 0 & ,\boldsymbol{\mu}_\tau \notin A, \\ \left[\prod_{k=1}^{m}\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{-\sum_{k=1}^{m} w_{\tau_k}\rho_{\tau_k}\left(\frac{y-\mu_{\tau_k}}{\sigma_{\tau_k}}\right)\right\} & ,\boldsymbol{\mu}_\tau \in A, \end{cases} \end{aligned} \tag{5.1}$$

for $k = 1, 2, ..., m$, where $\mu_{\tau_k}$ is the $\tau_k$ quantile of the distribution $F$, $0 < \tau_1 < ... < \tau_m < 1$, $A = \{\boldsymbol{\mu}_\tau : \mu_{\tau_1} < ... < \mu_{\tau_k} < ... < \mu_{\tau_m}\}$ and $w_{\tau_k}$ is a weight for each component of the mixture to control the contribution of the asymmetric Laplace distribution corresponding to each individual quantile in forming the approximation of the underlying distribution.

## 5.4 Normalising term

The normalising term for the density given in (5.1) can be obtained by solving the following equation with respect to $\delta$:

$$\int_{-\infty}^{\infty} \frac{1}{\delta}\widehat{f}\,(y|\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau)\,dy = 1, \tag{5.2}$$

Assuming that $w_{\tau_k}$ does not depend on $y$ for any $k$, we can show that

$$\int_{-\infty}^{\mu_{\tau_1}} \widehat{f}\,(y|\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau)\,dy = \frac{\prod_{k=1}^{m}\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\exp\left\{-\sum_{k=1}^{m}\frac{w_{\tau_k}(1-\tau_k)}{\sigma_{\tau_k}}\mu_{\tau_k} + \mu_{\tau_1}\sum_{k=1}^{m} w_{\tau_k}\left(\frac{1-\tau_k}{\sigma_{\tau_k}}\right)\right\}}{\sum_{k=1}^{m} w_{\tau_k}\left(\frac{1-\tau_k}{\sigma_{\tau_k}}\right)},$$

and

$$\begin{aligned}\int_{\mu_{\tau_t}}^{\mu_{\tau_{t+1}}} \widehat{f}\,(y|\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau)\,dy &= \prod_{k=1}^{m}\left(\frac{\tau_k\,(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\exp\left\{w_{\tau_k}\left(\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\mu_{\tau_k} - \sum_{k=t+1}^{m}\frac{(1-\tau_k)}{\sigma_{\tau_k}}\mu_{\tau_k}\right)\right\} \\ &\quad\times\left[w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{(1-\tau_k)}{\sigma_{\tau_k}} - \sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)\right]^{-1}\left[\exp\left\{w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{(1-\tau_k)}{\sigma_{\tau_k}}\right.\right.\right. \\ &\quad\left.\left.\left. - \sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)\mu_{\tau_{t+1}}\right\} - \exp\left\{w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{(1-\tau_k)}{\sigma_{\tau_k}} - \sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)\mu_{\tau_t}\right\}\right].\end{aligned}$$

Also,

$$\int_{\mu_{\tau_m}}^{\infty} \widehat{f}\,(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\, dy \;=\; \frac{\prod_{k=1}^{m} \left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}} \exp\left\{\sum_{k=1}^{m} \frac{w_{\tau_k}\tau_k\mu_{\tau_k}}{\sigma_{\tau_k}} - \mu_{\tau_m}\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\right\}}{\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}}.$$

Then, the normalising term is given by

$$\delta = \int_{-\infty}^{\mu_{\tau_1}} \widehat{f}\,(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\, dy + \sum_{t=1}^{m-1} \int_{\mu_{\tau_t}}^{\mu_{\tau_{t+1}}} \widehat{f}\,(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\, dy + \int_{\mu_{\tau_m}}^{\infty} \widehat{f}\,(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\, dy. \quad (5.3)$$

For more details, see Appendix D.1. Since the approximation density given in (5.1) depends on the number and locations of quantile functions included in the simultaneous estimation, the normalising term is not a constant and it depends on location and scale parameters. Then, there is no guarantee that

$$\int_{-\infty}^{\mu_{\tau_k}} \frac{1}{\delta}\widehat{f}\,(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\, dy = \tau_k, \text{ for } k = 1, 2, .., m. \quad (5.4)$$

This implies that $\mu_{\tau_k}$ is no longer interpreted as $\tau_k$ quantile. Therefore, using this normalising term during the estimation process, it was seen that the approximate density given in (5.1) is not maximised at the quantile functions of interest. To avoid this issue, the normalising term $\delta$ should be obtained by solving the equation system given by Equations 5.2 & 5.4. Since, in the case of multiple quantiles this leads to an over-determined equation system, calculating the normalising term to satisfy these properties is intractable. However, although the normalising term $\delta$ cannot be involved in the estimation, it can be used after estimates are obtained as a normalising constant to normalise the density and cumulative functions for comparison purposes.

## 5.5 Maximising the weighted pseudo asymmetric Laplace likelihood

Assume that $y$ is a random variable from a continuous distribution $F$ with probability density $f$. To estimate quantile functions of $F$ corresponding to $\tau_k$ for all $k$, the likelihood function related to the WPAL distribution given in (5.1) can be

used. Then, maximising the WPAL likelihood can be achieved by

$$\max_{\boldsymbol{\mu}_\tau} \quad \prod_{i=1}^{n} \widehat{f}(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)$$
$$\max_{\boldsymbol{\mu}_\tau} \quad \sum_{i=1}^{n} \log\left(\widehat{f}(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\right)$$

The maximisation of the expected log-likelihood can be obtained as follow:

$$\max_{\boldsymbol{\mu}_\tau}\Big(E\big[L(y; \boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\big]\Big),$$

where $L(y; \boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau) = \log\left(\widehat{f}(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\right)$. Assuming that $w_{\tau_k}$ is a constant with respect to $y$, the first order conditions are given by

$$\frac{d}{d\boldsymbol{\mu}_\tau}\left\{\int_{-\infty}^{\infty} L(y; \boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau) f(y) dy\right\} = 0.$$

Then, after some straightforward calculations, we have that

$$F(\mu_{\tau_k}) = \tau_k, \text{ for } k = 1, 2, ..., m.$$

This implies that estimators obtained by maximising expected log-likelihood are the quantile functions of the distribution $F$. This property is shown for two quantile functions in Appendix D.2.

## 5.6 The specification of the weight function

To provide a good approximation for the underlying distribution of $y$, the information over multiple quantiles must be combined properly by specifying a suitable weight function. To achieve this aim, the weight function $w_{\tau_k}$ should be such that the WPAL density minimises the Kullback–Leibler divergence given by

$$KLD\left(f\|\widehat{f}\right) = \int_{-\infty}^{\infty} f(y) \log\left(\frac{f(y)}{\frac{1}{\delta}\widehat{f}(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)}\right) dy,$$

where $\delta$ is a constant normalising the WPAL density (i.e. $\delta$ can be computed using Equation 5.3). In the following sections, we discuss different types of weights that

can be used.

### 5.6.1 Fixed weights

Since the aim is to use the WPAL distribution to approximate an unknown distribution, it is complicated to find optimal weights that minimise the Kullback–Leibler divergence under any true distribution. However, it is rational to consider equal fixed weights, given by the function

$$w_{\tau_k} = \frac{1}{m}, \text{ for } k =, 1, 2, ..., m, \tag{5.5}$$

where $m$ is the number of quantiles included in the estimation. Under these weights, it is assumed that all quantiles included in the estimation contribute equally in approximating the underlying distribution. In addition, by using this weight function, the WPAL density is the geometric mean of asymmetric Laplace probability densities corresponding to quantile functions included in the simultaneous estimation.

### 5.6.2 Estimated weights

Instead of applying fixed weights, we can allow $w_{\tau_k}$ to be estimated along with other parameters included in the Bayesian model such that $w_{\tau_k} \in (0, 1)$ and $\sum_{k=1}^{m} w_{\tau_k} = 1$. However, using the weighted pseudo asymmetric Laplace density given in (5.1) gives the estimated weights approaching one for the asymmetric Laplace density that has a lower Kullback–Leibler divergence and zero for the others. This implies that the WPAL likelihood is maximised at the values of weights which are not related to the achievement of minimum Kullback–Leibler divergence. For example, if we use the WPAL density corresponding to the $0.25, 0.5$ and $0.75$ quantiles to estimate quantiles of two samples from $\mathrm{N}(0, 1)$ and $\mathrm{Gamma}(3, 3)$, then in the case of $\mathrm{N}(0, 1)$, the estimated weight corresponding to the $0.5$ quantile approaches one while the estimated weights corresponding to the $0.25$ and $0.75$ quantiles approach zero, and in the case of $\mathrm{Gamma}(3, 3)$, the estimated weight corresponding to the $0.25$ quantile approaches one while the estimated weights corresponding to the $0.5$ and $0.75$ quantiles approach zero.

A possible solution for this issue is to assume that $w_{\tau_k} \in (l_b, 1)$ where $l_b$ is a lower boundary of weights selected such that $0 < l_b < 1$, and $\sum_{k=1}^{m} w_{\tau_k} = 1$.

However, if $l_b = \frac{1}{m}$, the estimated weights become the equal fixed weights. Also, choosing a small value for $l_b$ may lead to biased and poor estimation for the quantile functions that receive lower weights. To handle this issue effectively, we propose a new version of the WPAL density as follows:

$$\widehat{f}(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau, \boldsymbol{w}_\tau) \propto \begin{cases} 0 & , \boldsymbol{\mu}_\tau \notin A, \\ \prod_{k=1}^{m} \left[ \frac{\tau_k(1-\tau_k)}{w_{\tau_k}\sigma_{\tau_k}} \exp\left\{ -\rho_{\tau_k}\left( \frac{y-\mu_{\tau_k}}{\sigma_{\tau_k}} \right) \right\} \right]^{w_{\tau_k}} & , \boldsymbol{\mu}_\tau \in A, \end{cases} \tag{5.6}$$

where $w_{\tau_k} > 0$ for $k = 1, 2, ..., m$. It is obvious that each asymmetric Laplace component included in the WPAL density is divided by $w_{\tau_k}$ as a penalty term to make the likelihood maximised at desired weights which can contribute to the achievement of minimum Kullback–Leibler divergence. Similarly to Section 5.5, it is straightforward to show that this likelihood function is maximised at the quantiles. Also, the normalising term can be derived similarly to that in Section 5.4. In the next section, we use the proposed approximate likelihood functions to develop Bayesian methods to estimate multiple linear quantile functions simultaneously without crossing.

## 5.7 Bayesian linear multiple quantile regressions

Consider the linear regression model

$$y_i = \alpha + \boldsymbol{x}_i^{'}\boldsymbol{\beta} + \epsilon_i, \quad \text{for } i = 1, ..., n,$$

where $\epsilon_i$ is independently distributed with a mean equal to zero and a constant variance; $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates for the $i^{\text{th}}$ observation; $\alpha$ is an intercept coefficient and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown slope coefficients. To develop a Bayesian quantile approach to estimate the linear quantile function given by

$$Q(\tau_k|\boldsymbol{x_i}) = \alpha_{\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\beta},$$

for $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$ simultaneously, we assume that the response variable $y_i$ follows a distribution which is approximated by the WPAL distribution.

### 5.7.1 Using equal fixed weights

To develop a Bayesian method to fit multiple quantile regressions, we can apply the following WPAL likelihood with equal fixed weights:

$$\widehat{l}(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau; \boldsymbol{y}) \propto \begin{cases} 0 & , \boldsymbol{\alpha}_\tau \notin A, \\ \left[\prod_{k=1}^{m} \left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{n w_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{m} w_{\tau_k} \sum_{i=1}^{n} \rho_{\tau_k}\left(\frac{y_i - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma_{\tau_k}}\right)\right\} & , \boldsymbol{\alpha}_\tau \in A, \end{cases}$$

$$\propto \begin{cases} 0 & , \boldsymbol{\alpha}_\tau \notin A, \\ \left[\prod_{k=1}^{m} \left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{\frac{n}{m}}\right] \exp\left\{-\frac{1}{m}\sum_{k=1}^{m} \sum_{i=1}^{n} \rho_{\tau_k}\left(\frac{y_i - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma_{\tau_k}}\right)\right\} & , \boldsymbol{\alpha}_\tau \in A, \end{cases}$$

where $A = \{\boldsymbol{\alpha}_\tau : \alpha_{\tau_1} < ... < \alpha_{\tau_k} < ... < \alpha_{\tau_m}\}$. For the priors, we consider diffuse prior distributions that can be used to obtain proper posterior distributions. For quantile regression coefficients, the prior distribution is given by

$$\alpha_{\tau_k} \sim \mathrm{N}(\alpha^*, \sigma_\alpha), \text{ for } k = 1, ..., m,$$
$$\boldsymbol{\beta} \sim \mathrm{N}_p(\boldsymbol{\beta}^*, \Sigma_\beta),$$

where each normal distribution has mean zero and variance equal to $10^5$. Also, each multivariate normal distribution has a mean vector of zeros and a diagonal covariance matrix with entries equal to $10^5$. For the scale parameters, we consider the inverse-gamma distribution (IG) given by

$$\sigma_k \sim \mathrm{IG}(a = 0.01, b = 0.01), \text{ for } k = 1, ..., m.$$

Then, the posterior distribution is given by

$$p(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau | \boldsymbol{y}) = \widehat{l}(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau; \boldsymbol{y}) p(\boldsymbol{\alpha}_\tau) p(\boldsymbol{\beta}) p(\boldsymbol{\sigma}_\tau).$$

### 5.7.2 Using estimated weights

To develop a Bayesian method to estimate multiple quantile functions, we can apply the following WPAL likelihood with estimated weights:

$$\widehat{l}(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau, \boldsymbol{w}_\tau; \boldsymbol{y}) \propto \begin{cases} 0 & , \boldsymbol{\alpha}_\tau \notin A, \\ \left[\prod_{k=1}^{m} \left(\frac{\tau_k(1-\tau_k)}{w_{\tau_k}\sigma_{\tau_k}}\right)^{n w_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{m} w_{\tau_k} \sum_{i=1}^{n} \rho_{\tau_k}\left(\frac{y_i - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma_{\tau_k}}\right)\right\} & , \boldsymbol{\alpha}_\tau \in A. \end{cases}$$

Along with prior distributions specified in the previous section, the prior distribution for the vector of weights $\boldsymbol{w}_\tau$ is given by

$$\boldsymbol{w}_\tau \sim \text{Dirichlet}(\boldsymbol{\pi}),$$

where concentration parameters $\boldsymbol{\pi}$ is the vector of ones. Then, the posterior distribution is given by

$$p(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau, \boldsymbol{w}_\tau | \boldsymbol{y}) = \widehat{l}(\boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, \boldsymbol{\sigma}_\tau, \boldsymbol{w}_\tau; \boldsymbol{y}) p(\boldsymbol{\alpha}_\tau) p(\boldsymbol{\beta}) p(\boldsymbol{\sigma}_\tau) p(\boldsymbol{w}_\tau).$$

### 5.7.3 Posterior computation

To obtain a sample from posterior distributions, we apply an automated random walk Metropolis-Hastings algorithm based on a multivariate normal distribution with an updated covariance matrix. We assume

$$\boldsymbol{\theta}^* \sim N(\boldsymbol{\theta}_{(r-1)}, \phi^2_{(r-1)} S_{(r-1)}).$$

where $r$ is the iteration number, the $\boldsymbol{\theta}^*$ is a vector of candidate values for parameters, $\boldsymbol{\theta}_{(r-1)}$ is a vector of the stored candidate values at the $(r-1)^{\text{th}}$ iteration, $S$ is a positive-definite matrix and $\phi^2$ is a positive value. For the parameters with positive support such as the scale, we use the transformation $\sigma = \exp(\theta^*)$. To automate the random walk Metropolis-Hastings algorithm and achieve the desired acceptance probability, Garthwaite *et al.* (2016) proposed an adaptive method based on the Robbins-Monro search process to estimate an optimal covariance matrix for the proposal distribution. They assume that

$$S_{(r)} = \Sigma_{(r)} + \frac{\phi^2_{(r)} I_v}{r},$$

where

$$\Sigma_{(r)} = \begin{cases} I_v & r \leq c, \\ \frac{1}{r-1} \sum_{r^*=1}^{r} \left(\boldsymbol{\theta}_{(r^*)} - \overline{\boldsymbol{\theta}}_{(r)}\right) \left(\boldsymbol{\theta}_{(r^*)} - \overline{\boldsymbol{\theta}}_{(r)}\right)' & r > c, \end{cases}$$

where $\overline{\boldsymbol{\theta}}_{(r)}$ is the sample mean vector of the stored candidate values until the $r^{\text{th}}$ iteration, $v$ is the length of $\boldsymbol{\theta}$, $I$ is the identity matrix and $c$ is a fixed number of

iterations. To reduce the computation, they use

$$\overline{\boldsymbol{\theta}}_{(r)} = \frac{1}{r}\left[(r-1)\overline{\boldsymbol{\theta}}_{(r-1)} + \boldsymbol{\theta}_{(r)}\right],$$

and if $r > c$, they use

$$\Sigma_{(r)} = \frac{r-2}{r-1}\Sigma_{(r-1)} + \overline{\boldsymbol{\theta}}_{(r-1)}\overline{\boldsymbol{\theta}}'_{(r-1)} - \frac{r}{r-1}\overline{\boldsymbol{\theta}}_{(r)}\overline{\boldsymbol{\theta}}'_{(r)} + \frac{1}{r-1}\boldsymbol{\theta}_{(r)}\boldsymbol{\theta}'_{(r)}.$$

Also, they assume that $\phi$ is updated using the formula:

$$\phi_{(r)} = \begin{cases} \phi_{(r-1)} + \frac{a(1-\widehat{p})}{\max\{2c,(r-1)/v\}} & \boldsymbol{\theta}^* \text{ is accepted,} \\ \phi_{(r-1)} - \frac{a\widehat{p}}{\max\{2c,(r-1)/v\}} & \boldsymbol{\theta}^* \text{ is rejected,} \end{cases}$$

where $\widehat{p} = 0.234$ (for more details, see Roberts *et al.*, 1997) and

$$a = \phi_{(r-1)}\left[\left(1-\frac{1}{v}\right)\frac{(2\pi)^{\frac{1}{2}}\, e^{\frac{\left[-\Phi^{-1}(\widehat{p}/2)\right]^2}{2}}}{2\left[-\Phi^{-1}\left(\widehat{p}/2\right)\right]} + \frac{1}{v\widehat{p}(1-\widehat{p})}\right],$$

where $\Phi$ is the cumulative function of the standard normal distribution. The value $\max\{2c, (r-1)/v)$ is used to ensure that the estimate of $\Sigma$ is reasonably stable before updating $\phi$. In the case of the WPAL likelihood with estimated weights, we update weight parameters at each iteration independently from other parameters in the model using a Dirichlet distribution, with suitable fixed variance, as a proposal distribution.

## 5.8 Simulation studies and real data analysis

### 5.8.1 Univariate models

In this section, we generate univariate samples, each of size 150, from normal distribution $\mathrm{N}(0,1)$, gamma distribution $\mathrm{Gamma}(3,3)$, Student's $t$ distribution $t(3)$ and mixture distribution $0.5\mathrm{N}(-2,1) + 0.5\mathrm{N}(2,1)$. Then, we investigate the performance of the pseudo asymmetric Laplace (PAL) and the weight pseudo asymmetric Laplace (WPAL) distributions corresponding to different sets of quantiles to approximate the true distributions by comparing the estimated density and

cumulative functions with true ones. The estimated density and cumulative functions are represented by the average of normalised function values computed at the maximum a posteriori estimators over 1000 simulations. A set of five quantiles corresponding to $\boldsymbol{\tau} = (0.05, 0.25, 0.5, 0.75, 0.95)$ are used in estimation. Also, we examine the effect of the number and locations of quantiles included in the estimation on the approximation by comparing the average of Kullback–Leibler divergence over 5 different sets of quantiles where the first set includes two quantiles corresponding to $\boldsymbol{\tau} = (0.45, 0.55)$. Then, two more quantiles corresponding to $\boldsymbol{\tau} = (0.45 - \frac{k}{10}, 0.55 + \frac{k}{10})$, for $k = 1, ..., 4$, are involved at each replication up to 10 quantiles (Scheme 1). In addition, we consider another 5 different sets of quantiles where the first set includes two quantiles corresponding to $\boldsymbol{\tau} = (0.05, 0.95)$. Then, two more quantiles corresponding to $\boldsymbol{\tau} = (0.05 + \frac{k}{10}, 0.95 - \frac{k}{10})$, for $k = 1, ..., 4$ are added at each replication up to 10 quantiles (Scheme 2).

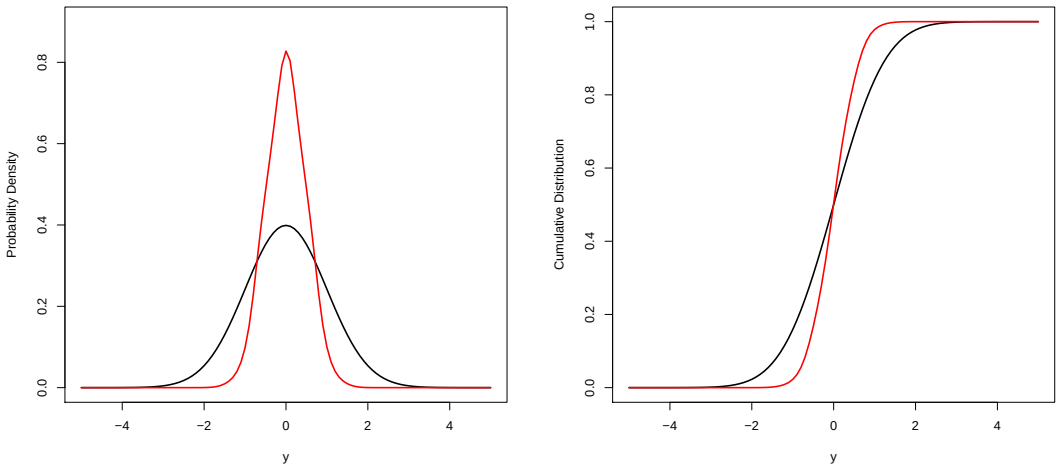



Figure 5.1: The probability density function (left) and the cumulative distribution function (right) for the true distribution (black) which is N(0,1) and their approximations obtained using the PAL distribution based on 5 quantiles (red).

Figure 5.1 shows poor approximation by the PAL distribution for the normal distribution. The approximate distribution has a smaller scale than the true distribution. This leads to unreliable Bayesian inference about the quantile coefficients (e.g. the highest probability density intervals). Also, it can be seen from

Figure 5.2 that, by increasing the number of quantiles included in estimation, the approximation becomes worse (i.e. the scale of approximate distribution approaches zero). Although the quantile method based on the PAL distribution performs well in terms of the point estimation, from a Bayesian perspective, the PAL distribution can be considered an asymptotic error distribution for quantile regression that may lead to invalid Bayesian inference.

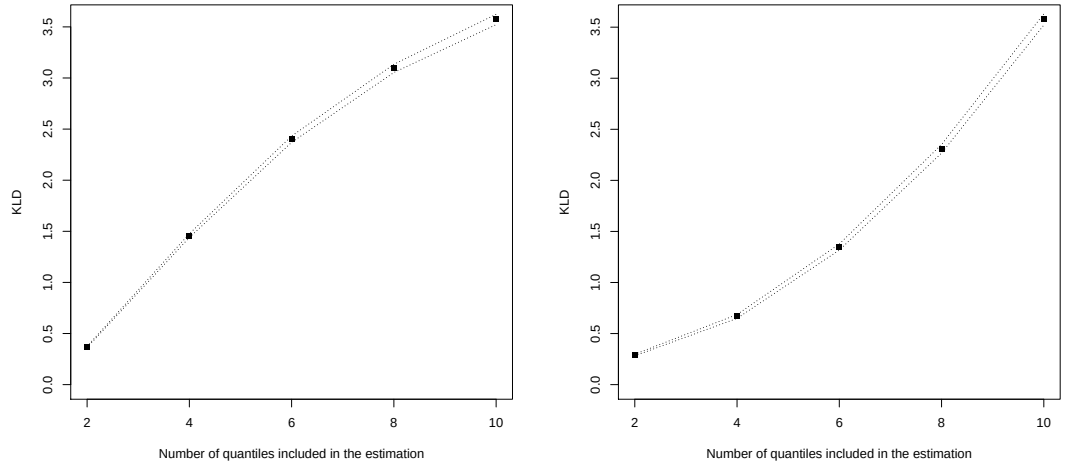



Figure 5.2: The average of Kullback–Leibler divergence, from the approximation obtained using the PAL distribution to N(0,1), over 500 simulations against the number of quantiles corresponding to Scheme 1 (left) and Scheme 2 (right). The dotted curves represent the 95% confidence intervals for the mean.

Figures 5.3 & 5.4 show the ability of the WPAL distribution to approximate the true distributions. It is obvious that the approximation performs better in the case of unimodal distributions. However, the WPAL distribution still provides a reasonable approximation for bimodal distributions. This can lead to reliable and accurate Bayesian inference for quantile coefficients.

From Table 5.1, by using the WPAL density with estimated weights, more weights assign to quantile functions which are close to the modes of the true distributions. In addition, the estimated weight shows a flexibility to deal with symmetric and asymmetric distributions and this could lead to an improvement in the approximation based on the WPAL distribution.

Figure 5.5 & 5.6 show the effects of the number and locations of quantiles included in the estimation on the quality of approximations obtained using the

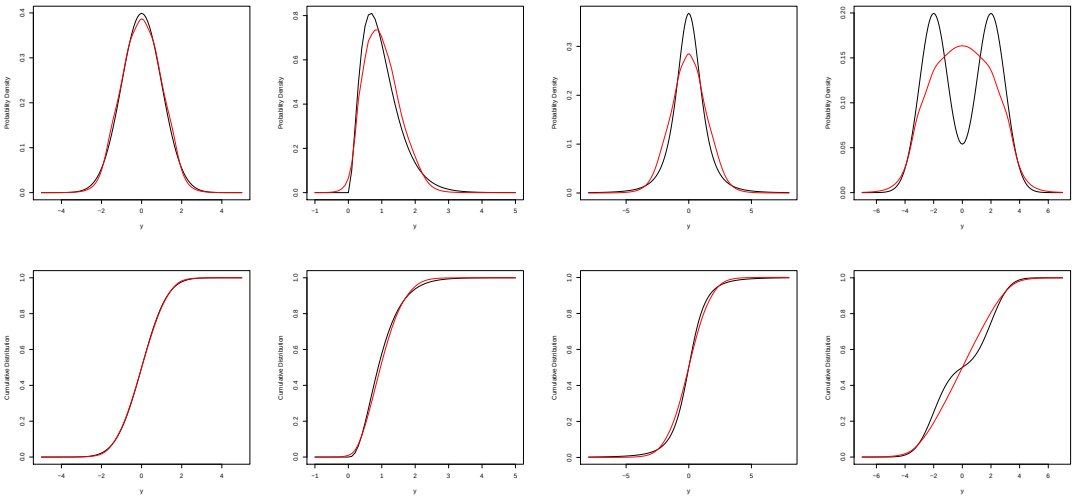


Figure 5.3: The probability density function (top) and the cumulative distribution function (bottom) for normal, gamma, Student's $t$ and mixture distributions (left to right). The true distributions are represented by black curves and their approximations obtained using the WPAL distribution based on 5 quantiles with equal fixed weights are represented by red curves.

| $\epsilon_i$ | 0.05 | 0.25 | 0.5 | 0.75 | 0.95 |
|---|---|---|---|---|---|
| $\mathrm{N}(0,1)$ | 0.17 | 0.22 | 0.23 | 0.22 | 0.17 |
| Student's $t(3)$ | 0.14 | 0.23 | 0.26 | 0.23 | 0.14 |
| $\mathrm{Gamma}(3,3)$ | 0.24 | 0.24 | 0.22 | 0.18 | 0.12 |
| $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$ | 0.20 | 0.21 | 0.19 | 0.21 | 0.20 |

Table 5.1: The average of the maximum a posteriori estimators for the weights, corresponding to the Bayesian method based on the WPAL distribution with estimated weights, over 1000 simulations.

WPAL distribution with equal fixed weights and estimated weights. Also, they show that the number and locations of quantiles included in the estimation affect the approximation differently according to characteristics of the true distribution. For all considered distributions, the Kullback–Leibler divergences decrease when the number of quantiles included in the estimation increases, except for Student's $t$-distribution in Figure 5.5 and the mixture distribution in Figure 5.6, where the Kullback–Leibler divergences increase slightly. This is because the asymmetric Laplace distributions corresponding to quantiles added to the estimation do not

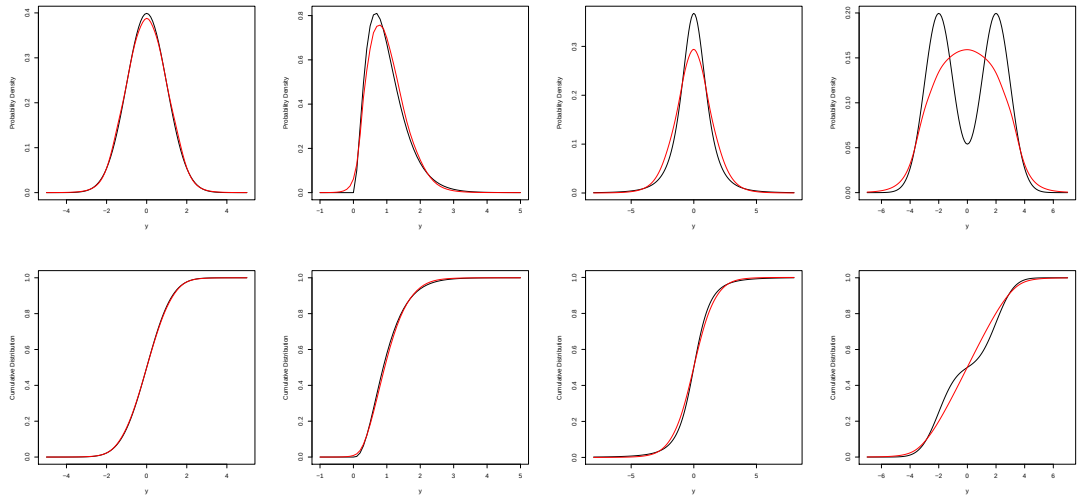


Figure 5.4: The probability density function (top) and the cumulative distribution function (bottom) for normal, gamma, Student's $t$ and mixture distributions (left to right). The true distributions are represented by black curves and their approximations obtained using the WPAL distribution based on 5 quantiles with estimated weights are represented by red curves.

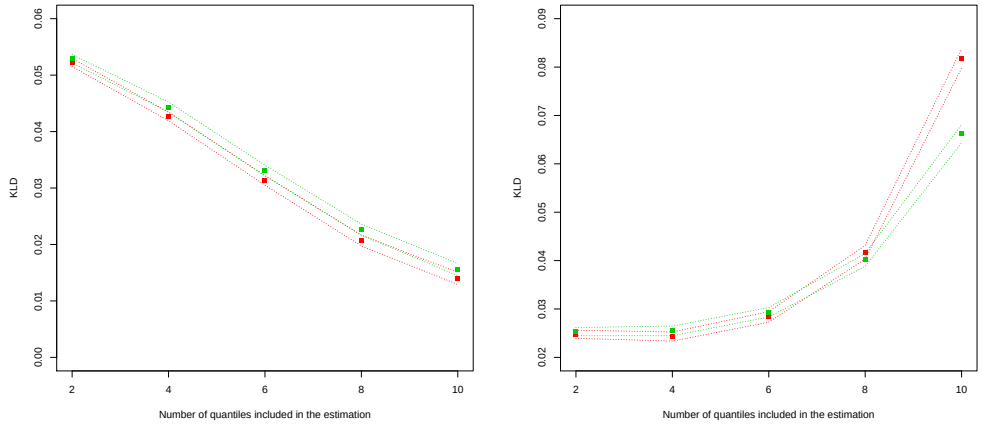
convey accurate information about these quantiles of the underlying distribution. However, despite this change in the quality of the approximation, the Kullback–Leibler divergences still lie in an acceptable range. In addition, Figure 5.5 & 5.6 show that the approximation based on the estimated weights may outperform the approximation based on the equal fixed weights.

### 5.8.2 Linear models

In this section, we evaluate the accuracy of the coverage probabilities of 95% HPD intervals estimated using Bayesian quantile methods based on the PAL likelihood and the WPAL likelihood functions. To examine the proposed methods beyond univariate models, we consider estimating quantiles curves for 200 observations from the linear model:

$$y_i = 1 + x_i + \epsilon_i$$

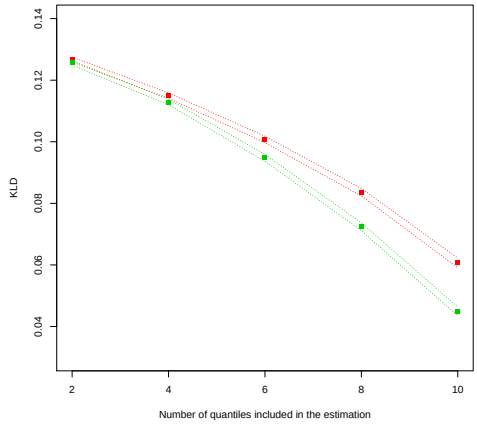
where $x_i \sim \mathrm{U}(20, 30)$ and $\epsilon_i$ is considered to follow different distributions. The simulated datasets, used in the simulation study, only differ in $\epsilon_i$. Tables 5.2 &



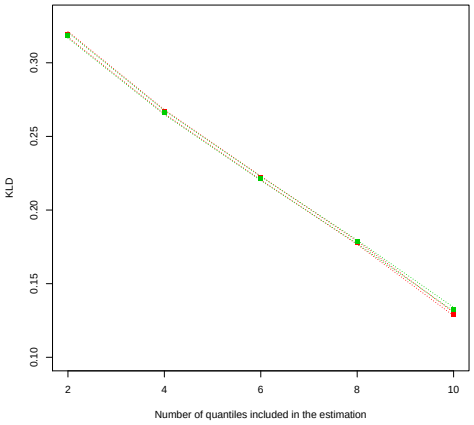

(a) N(0, 1)




(b) Student's $t(3)$




(c) Gamma(3, 3)

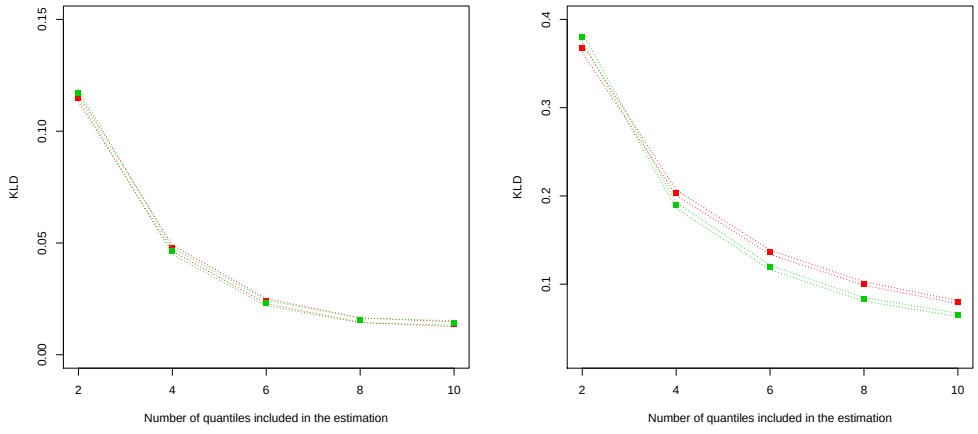



(d) $0.5\mathrm{N}(-2, 1) + 0.5\mathrm{N}(2, 1)$

Figure 5.5: The average of Kullback–Leibler divergence, from the approximation obtained using the WPAL distribution with equal fixed weights (red) and estimated weights (green) to the true distribution, over 500 simulations against the number of quantiles corresponding to Scheme 1. The dotted curves represent the 95% confidence intervals for the mean.

5.3 show how poor approximation obtained by the PAL distribution affects the estimation of the 95% highest posterior density intervals. Also, consistent with the discussion in Section 5.2, Tables 5.3 shows how that approximation becomes worse as the number of quantiles included in the estimation increases.

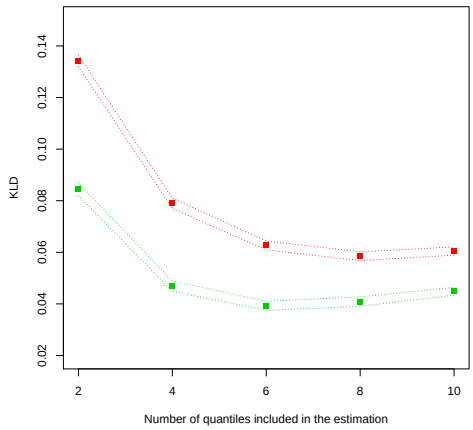
Table 5.4 shows that the Bayesian quantile method based on the WPAL likelihood with equal fixed weights offers a great improvement in coverage probabilities that approach desired levels of the nominal coverage probabilities for all quan-



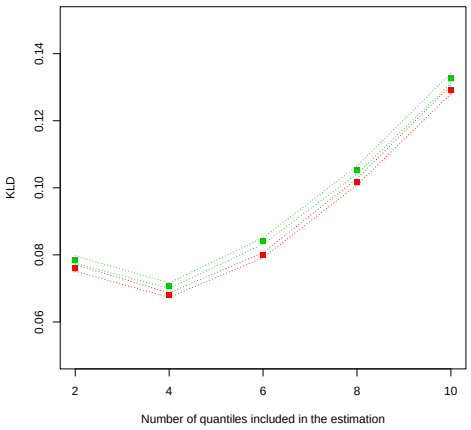

(a) N(0, 1)




(b) Student's $t(3)$




(c) Gamma(3, 3)




(d) $0.5\mathrm{N}(-2, 1) + 0.5\mathrm{N}(2, 1)$

Figure 5.6: The average of Kullback–Leibler divergence, from the approximation obtained using the WPAL distribution with equal fixed weights (red) and estimated weights (green) to the true distribution, over 500 simulations against the number of quantiles corresponding to Scheme 2. The dotted curves represent the 95% confidence intervals for the mean.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\alpha_{0.95}$ | $\beta$ |
|---|---|---|---|---|---|---|
| N(0,1) | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.56 |

Table 5.2: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated using the Bayesian quantile method based on the PAL likelihood over 1000 simulations.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.15}$ | $\alpha_{0.25}$ | $\alpha_{0.35}$ | $\alpha_{0.45}$ | $\alpha_{0.55}$ | $\alpha_{0.65}$ | $\alpha_{0.75}$ | $\alpha_{0.85}$ | $\alpha_{0.95}$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| N(0,1) | 0.48 | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | 0.48 | 0.48 | 0.47 | 0.44 |

Table 5.3: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated using the Bayesian quantile method based on the PAL likelihood over 1000 simulations.

tile functions under different true distributions. Table 5.5 shows that a similar result, with slight improvements for normal and Student's $t$-distributions, can be achieved by applying the Bayesian quantile method based on the WPAL likelihood with estimated weights. This method suggests different weights for each quantile as in Table 5.6 where more weights are assigned to quantiles close to the modes of the true distributions.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\alpha_{0.95}$ | $\beta$ |
|---|---|---|---|---|---|---|
| $\mathrm{N}(0,1)$ | 0.93 | 0.93 | 0.94 | 0.93 | 0.93 | 0.92 |
| Student's $t(3)$ | 0.96 | 0.97 | 0.97 | 0.97 | 0.95 | 0.96 |
| $\mathrm{Gamma}(3,3)$ | 0.97 | 0.97 | 0.98 | 0.98 | 0.97 | 0.97 |
| $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$ | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |

Table 5.4: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated using the Bayesian quantile method based on the WPAL likelihood with equal fixed weights over 1000 simulations.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\alpha_{0.95}$ | $\beta$ |
|---|---|---|---|---|---|---|
| $\mathrm{N}(0,1)$ | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.94 |
| Student's $t(3)$ | 0.95 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| $\mathrm{Gamma}(3,3)$ | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.97 |
| $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$ | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 | 0.97 |

Table 5.5: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated using the Bayesian quantile method based on the WPAL likelihood with estimated weights over 1000 simulations..

Tables 5.7 & 5.8 show that by increasing the number of quantile functions included in the estimation, the approximation based on the WPAL distribution may be improved or worsened slightly depending on the underlying distribution. It is seen that the coverage probabilities have little changes along with increasing

| $\epsilon_i$ | 0.05 | 0.25 | 0.5 | 0.75 | 0.95 |
|---|---|---|---|---|---|
| $\mathrm{N}(0,1)$ | 0.17 | 0.22 | 0.23 | 0.22 | 0.17 |
| Student's $t(3)$ | 0.14 | 0.23 | 0.26 | 0.23 | 0.14 |
| $\mathrm{Gamma}(3,3)$ | 0.23 | 0.24 | 0.22 | 0.18 | 0.12 |
| $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$ | 0.20 | 0.21 | 0.19 | 0.21 | 0.20 |

Table 5.6: The average of the maximum a posteriori estimators for the weights, corresponding to the Bayesian method based on the WPAL distribution with estimated weights, over 1000 simulations.

the number of quantile functions included in the estimation. However, Table 5.9 shows that the Bayesian quantile method based on the WPAL likelihood with estimated weights gives small values of weights for extreme quantiles and this can result in conservative intervals (e.g. 0.05 quantile for Student's $t$ distribution and 0.95 quantile for gamma distribution as shown in Table 5.8)

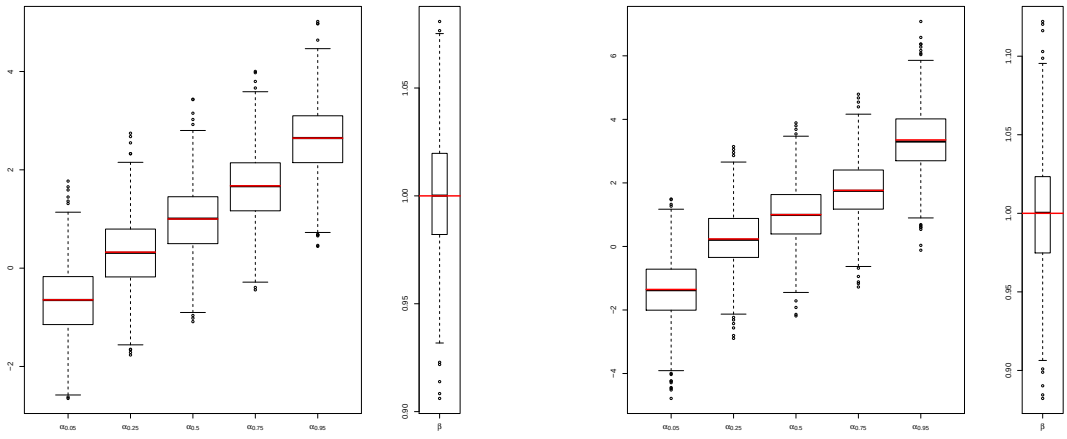
From Figures 5.7 – 5.10 that show the distributions of the maximum a posteriori estimators obtained using the Bayesian quantile method based on the WPAL likelihood over 1000 simulations, there is no strange behaviour of the maximum a posteriori estimators which are distributed around the true values. This is shown under different number of quantiles included in the simultaneous estimation.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.15}$ | $\alpha_{0.25}$ | $\alpha_{0.35}$ | $\alpha_{0.45}$ | $\alpha_{0.55}$ | $\alpha_{0.65}$ | $\alpha_{0.75}$ | $\alpha_{0.85}$ | $\alpha_{0.95}$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{N}(0,1)$ | 0.95 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.95 |
| Student's $t(3)$ | 0.97 | 0.97 | 0.96 | 0.96 | 0.97 | 0.96 | 0.96 | 0.97 | 0.97 | 0.97 | 0.96 |
| $\mathrm{Gamma}(3,3)$ | 0.96 | 0.96 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.98 | 0.97 | 0.96 |
| $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$ | 0.98 | 0.98 | 0.97 | 0.98 | 0.98 | 0.97 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 |

Table 5.7: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated using the Bayesian quantile method based on the WPAL likelihood with equal fixed weights over 1000 simulations.

| $\epsilon_i$ | $\alpha_{0.05}$ | $\alpha_{0.15}$ | $\alpha_{0.25}$ | $\alpha_{0.35}$ | $\alpha_{0.45}$ | $\alpha_{0.55}$ | $\alpha_{0.65}$ | $\alpha_{0.75}$ | $\alpha_{0.85}$ | $\alpha_{0.95}$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{N}(0,1)$ | 0.96 | 0.95 | 0.95 | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 | 0.95 | 0.96 | 0.95 |
| Student's $t(3)$ | 0.99 | 0.98 | 0.97 | 0.97 | 0.97 | 0.96 | 0.97 | 0.97 | 0.97 | 0.98 | 0.96 |
| $\mathrm{Gamma}(3,3)$ | 0.97 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.99 | 0.97 |
| $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$ | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.98 | 0.97 |

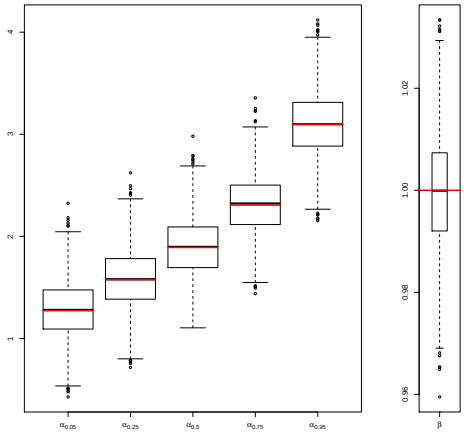
Table 5.8: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated using the Bayesian quantile method based on the WPAL likelihood with estimated weights over 1000 simulations.



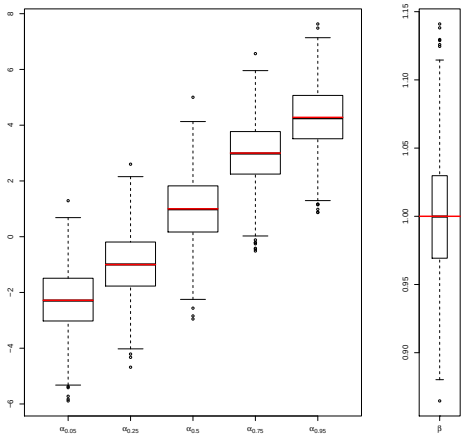
(a) N(0, 1)



(b) Student's $t(3)$



(c) Gamma(3, 3)

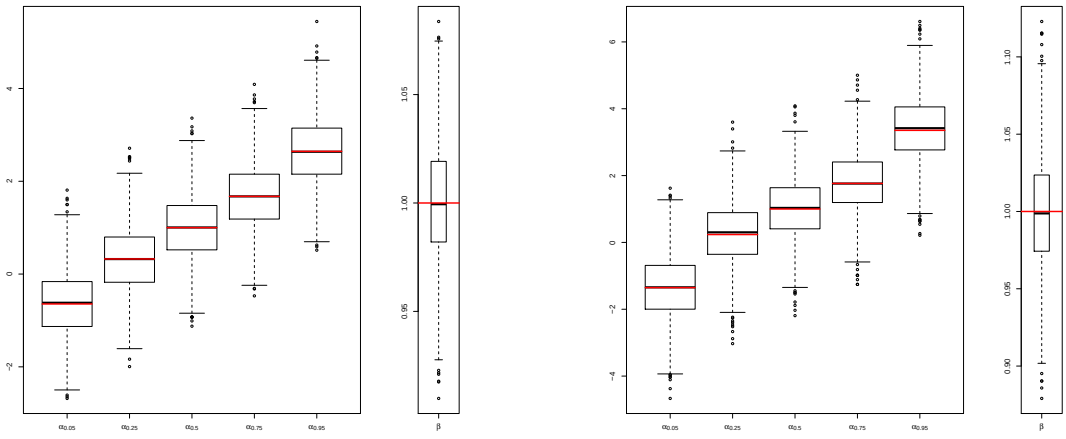


(d) $0.5\mathrm{N}(-2,1) + 0.5\mathrm{N}(2,1)$

Figure 5.7: The box-plot of the maximum a posteriori estimators, obtained using the Bayesian quantile method based on the WPAL likelihood with equal fixed weights, over 1000 simulations. The red lines represent the true values.

| $\epsilon_i$ | 0.05 | 0.15 | 0.25 | 0.35 | 0.45 | 0.55 | 0.65 | 0.75 | 0.85 | 0.95 |
|---|---|---|---|---|---|---|---|---|---|---|
| N(0, 1) | 0.08 | 0.10 | 0.10 | 0.11 | 0.11 | 0.11 | 0.11 | 0.10 | 0.10 | 0.08 |
| Student's $t(3)$ | 0.07 | 0.09 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.11 | 0.09 | 0.07 |
| Gamma(3, 3) | 0.11 | 0.12 | 0.12 | 0.11 | 0.11 | 0.10 | 0.10 | 0.09 | 0.08 | 0.06 |
| $0.5\mathrm{N}(-2,1) + 0.5\mathrm{N}(2,1)$ | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |

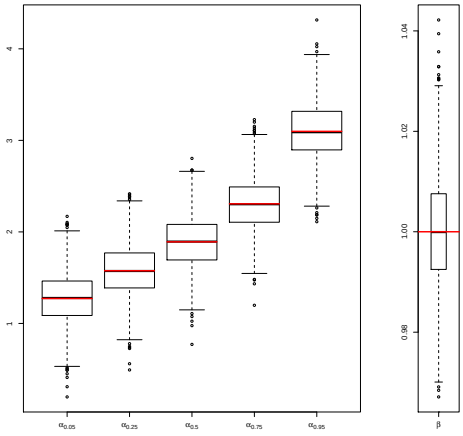
Table 5.9: The average of the maximum a posteriori estimators for the weights, corresponding to the Bayesian quantile method based on the WPAL distribution with estimated weights, over 1000 simulations.



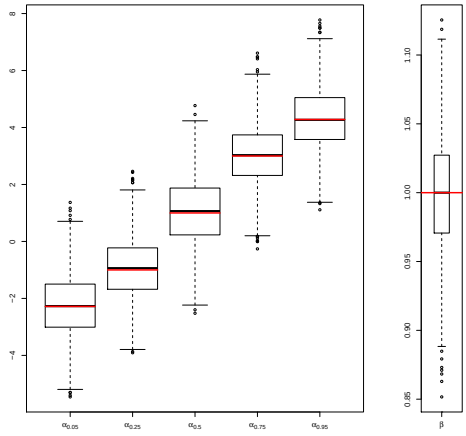
(a) N(0, 1)



(b) Student's $t(3)$



(c) Gamma(3, 3)

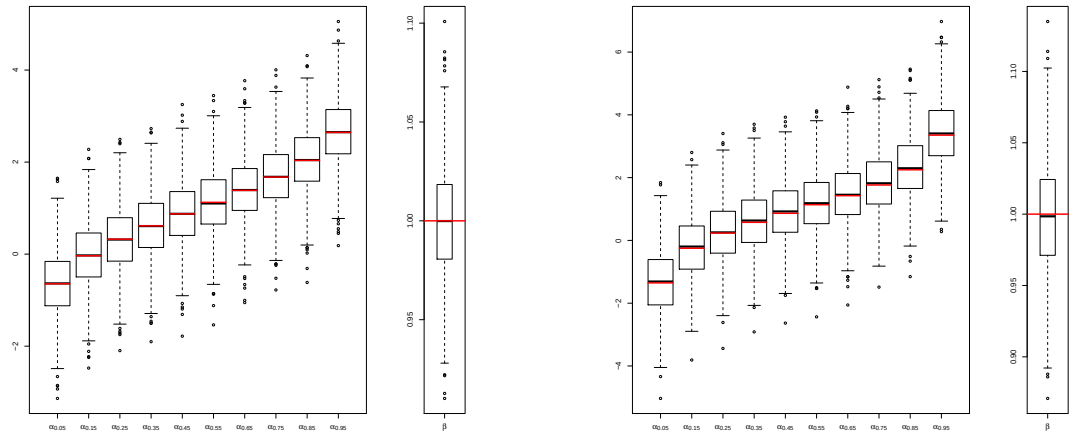


(d) $0.5\mathrm{N}(-2, 1) + 0.5\mathrm{N}(2, 1)$

Figure 5.8: The box-plot of the maximum a posteriori estimators, obtained using the Bayesian quantile method based on the WPAL likelihood with estimated weights, over 1000 simulations. The red lines represent the true values.

### 5.8.3 The Chatterjee-Price Attitude data

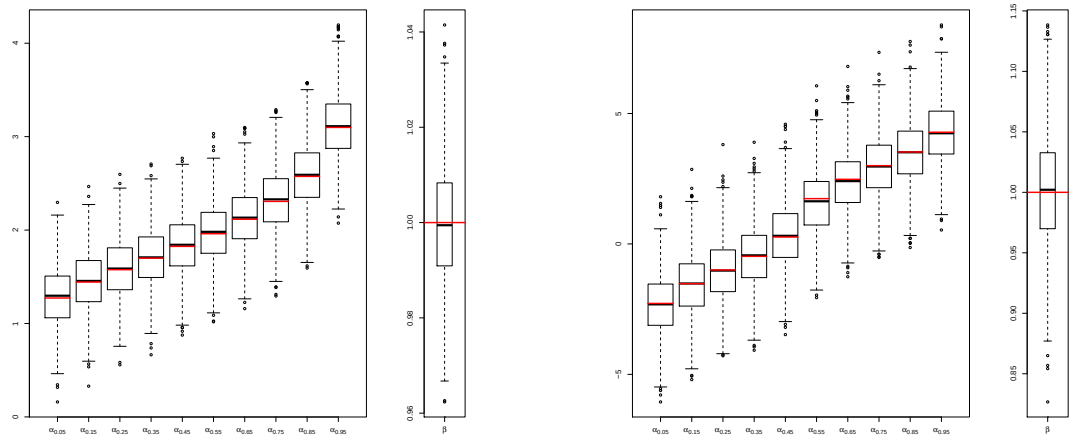
These data are from a survey collected from nearly 35 employees corresponding to 30 randomly selected departments to present the percentage of favourable responses, in each department, to seven questions related to the overall rating, the handling of employee complaints, not allowing special privileges, the opportunity to learn, raises based on performance, too critical and advancement. This dataset is available publicly by R Core Team (2016). In this section, we use our proposed



(a) N(0, 1)



(b) Student's $t(3)$



(c) Gamma(3, 3)

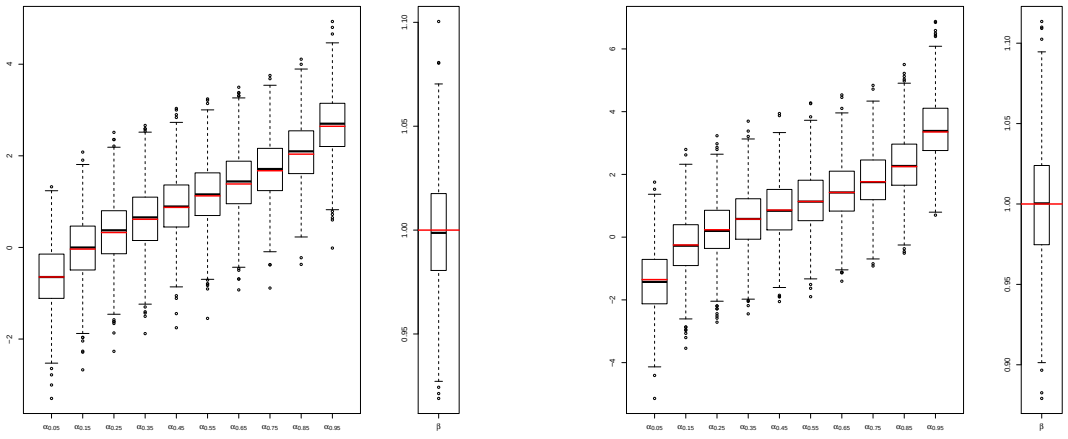


(d) $0.5\mathrm{N}(-2, 1) + 0.5\mathrm{N}(2, 1)$

Figure 5.9: The box-plot of the maximum a posteriori estimators, obtained using the Bayesian quantile method based on the WPAL likelihood with equal fixed weights, over 1000 simulations. The red lines represent the true values.

methods to study the relationship between the overall rating and the handling of employee complaints given in Figure 5.11.

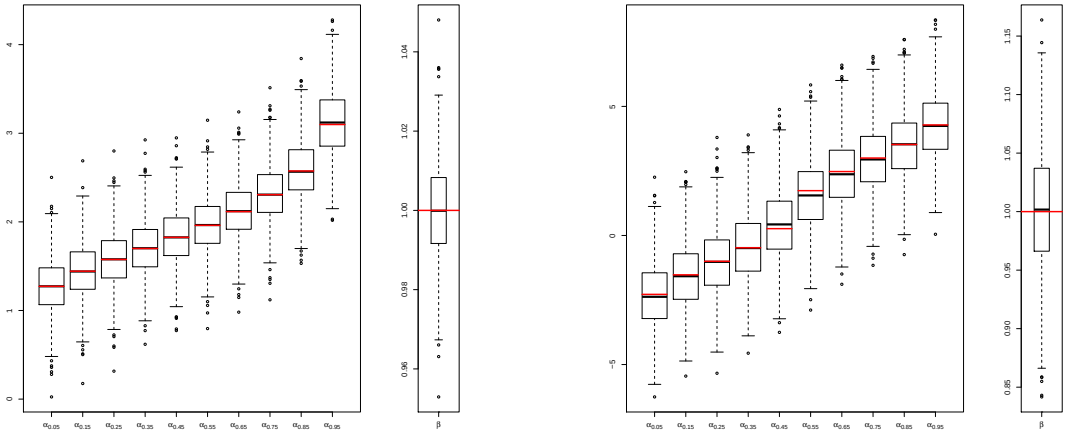
Figure 5.11 shows how the individual estimations of quantile curves based on the asymmetric Laplace likelihood fails in describing the conditional distribution of the overall rating. Figure 5.12 shows the ability of simultaneous estimation to improve the estimation of quantile curves. It shows that nearly all simultaneous methods suggest that the conditional distribution of the overall rating is not symmetric with a short upper tail and long lower tail. From Table 5.10, it can be seen



(a) N(0, 1)



(b) Student's $t(3)$



(c) Gamma(3, 3)



(d) $0.5\mathrm{N}(-2,1) + 0.5\mathrm{N}(2,1)$

Figure 5.10: The box-plot of the maximum a posteriori estimators, obtained using Bayesian quantile method based the WPAL likelihood with estimated weights, over 1000 simulations. The red lines represent the true values.

that there is a clear difference in the 95% highest posterior density intervals obtained using the different approaches. Also, there is some agreement between the methods in terms of the maximum a posteriori estimators especially the methods based on the WPAL likelihood with equal fixed weights and the PAL likelihood. Using the Bayesian quantile method based on the WPAL likelihood with estimated weights, the maximum a posteriori estimators for weights corresponding to $0.05, 0.25, 0.5, 0.75$ and $0.95$ quantiles respectively are $0.17, 0.18, 0.16, 0.23$ and $0.26$.

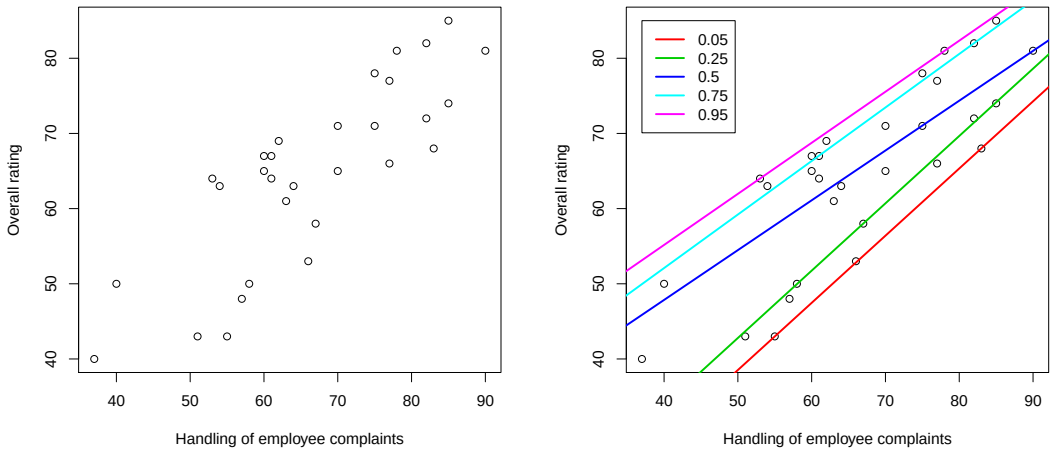



Figure 5.11: The scatter plot of the overall rating against the handling of employee complaints (left). The estimated quantile functions, based on the maximum a posteriori estimators obtained individually using the asymmetric Laplace likelihood, for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ (right).

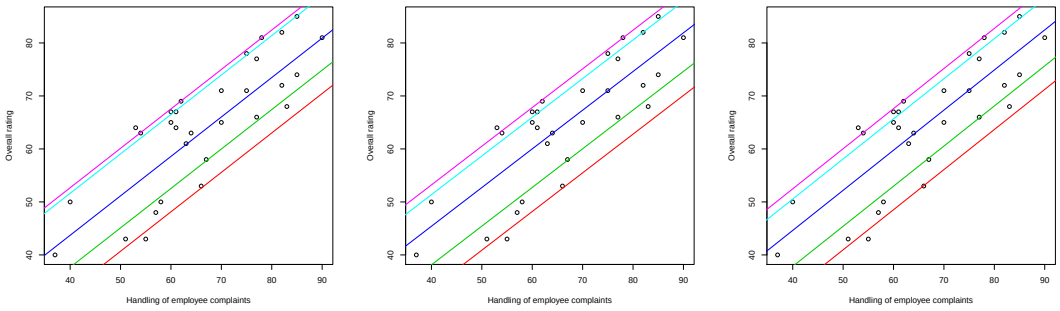


Figure 5.12: The estimated quantile functions, based on the maximum a posteriori estimators obtained simultaneously using the WPAL likelihood with equal fixed weights and estimated weights, and the PAL likelihood (left to right), for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$.

To investigate the performance of the proposed methods, we apply leave-one-out cross-validation method for quantile regression which is given by

$$\text{QLCV}_{\tau_k} = \frac{1}{n}\sum_{i=1}^{n} \rho_{\tau_k}\left(y_i - \widehat{Q}^{-i}(\tau_k|\boldsymbol{x}_i)\right)$$

| | WPAL$^{(1)}$ | | WPAL$^{(2)}$ | | PAL | |
|---|---|---|---|---|---|---|
| Coefficients | MAP | 95% HPD | MAP | 95% HPD | MAP | 95% HPD |
| $\alpha_{0.05}$ | 3.49 | (−12.63, 14.86) | 4.39 | (−203.89, 24.36) | 3.12 | ( −0.66, 9.06) |
| $\alpha_{0.25}$ | 7.90 | ( −5.88, 22.76) | 8.93 | ( −5.67, 23.07) | 7.56 | ( 3.51, 15.70) |
| $\alpha_{0.5}$ | 13.94 | ( 0.71, 29.57) | 16.21 | ( 1.80, 30.45) | 14.35 | ( 10.27, 22.50) |
| $\alpha_{0.75}$ | 21.86 | ( 7.01, 33.96) | 22.17 | ( 7.98, 34.87) | 20.27 | ( 16.62, 27.10) |
| $\alpha_{0.95}$ | 22.91 | ( 12.09, 38.85) | 24.01 | ( 12.46, 40.20) | 22.20 | ( 19.24, 28.87) |
| $\beta$ | 0.74 | ( 0.56, 0.94) | 0.73 | ( 0.55 , 0.94 ) | 0.76 | ( 0.66 , 0.81 ) |

Table 5.10: The maximum a posteriori (MAP) estimators and the 95% highest posterior density intervals (95% HPD) obtained using the weighted pseudo asymmetric Laplace likelihood with equal fixed weights (WPAL$^{(1)}$) and estimated weights (WPAL$^{(2)}$), and the pseudo asymmetric Laplace likelihood (PAL).

where $\widehat{Q}^{-i}$ is the estimated quantile function based on the maximum a posteriori estimators that is estimated when the $i^{\text{th}}$ observation is left out. Also, we use the average of leave-one-out cross-validation estimation of the error over all quantile functions, included in the simultaneous estimation, given by

$$\text{AQLCV} = \frac{1}{m}\sum_{k=1}^{m} \text{QLCV}_{\tau_k}.$$

where $m$ is the number of quantiles. Table 5.11 shows similar leave-one-out cross-

| $\tau$ | WPAL$^{(1)}$ | WPAL$^{(2)}$ | PAL |
|---|---|---|---|
| 0.05 | 0.71 | 0.70 | 0.69 |
| 0.25 | 2.63 | 2.62 | 2.75 |
| 0.5 | 3.25 | 3.30 | 3.27 |
| 0.75 | 2.09 | 2.06 | 2.10 |
| 0.95 | 0.53 | 0.50 | 0.49 |
| AQLCV | 1.85 | 1.84 | 1.86 |

Table 5.11: The Leave-one-out cross-validation estimation of the error for quantile functions.

validation estimation of the error for the individual quantiles and its average over all quantiles included in the simultaneous estimation. This implies that there is some agreement between the different simultaneous quantile methods in terms of the maximum a posteriori estimators. To investigate the estimated conditional distribution of the response variable, we compare the estimated conditional den-

sity of the response variable obtained using simultaneous quantile regression based on the WPAL likelihood with those obtained using other methods which can be used to fit this dataset, such as Bayesian linear regression based on normal distribution and Bayesian linear regression based on skew normal distribution (for more details, see Wuertz *et al.*, 2016). Figure 5.13, which shows the estimated conditional density of the response variable (the overall rating) given the mean of the covariate (the handling of employee complaints) using the different methods, illustrates the flexible approximation of the conditional distribution obtained by using the WPAL distribution compared to other methods.

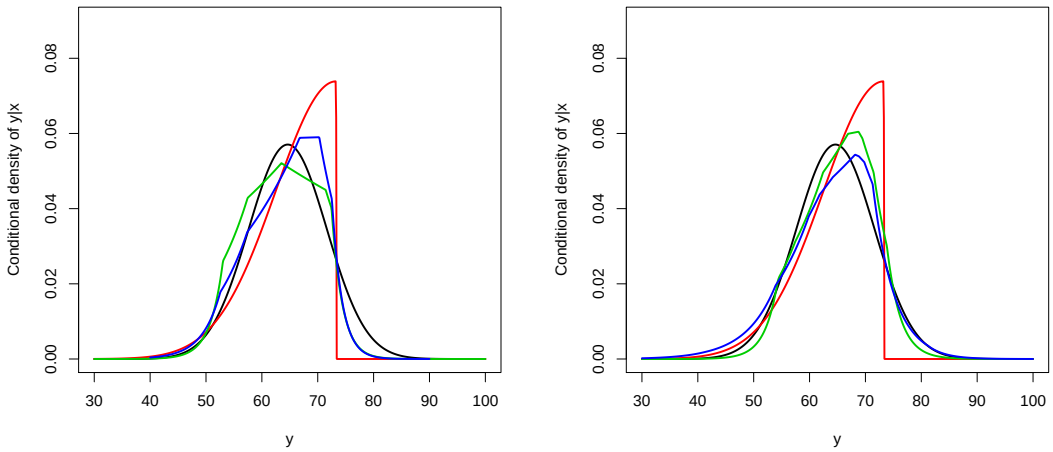



Figure 5.13: Comparison of estimated conditional densities of $y|x = E(x)$ obtained using Bayesian quantile regression based on the WPAL distribution corresponding to 5 quantiles (left) and 10 quantiles (right) with equal fixed weights (green) and estimated weights (blue), and Bayesian regression based on normal distribution (black) and skew normal distribution (red). Overall rating is represented by $y$ and handling of employee complaints is represented by $x$.

## 5.9 Conclusion

The proposed quantile methods based on the weighted pseudo asymmetric Laplace (WPAL) likelihood with equal fixed weights and estimated weights show a flexibility to approximate a variety of error distributions including symmetric, asym-

metric and bimodal distributions. This leads to great improvement in the reliability of Bayesian inference on quantile coefficients (e.g. the coverage probabilities approach the nominal coverage probabilities). Both equal fixed weights and estimated weights have their own motivation. The equal fixed weights reduce the uncertainty in the Bayesian model and estimated weights can show more flexibility to improve the approximation of underlying distributions. However, the weighted pseudo asymmetric Laplace likelihood with estimated weights might be more conservative in terms of the highest posterior density intervals for the extreme quantiles of distributions with long tails, especially in the case of a large number of quantiles included in the simultaneous estimation. This is because of the small weights assigned for these quantiles.

Bayesian inference on quantile coefficients should not be based on the asymmetric Laplace likelihood blindly and alternatives that are parameterised by the quantiles and can provide a good approximation to the underlying distributions should be considered. In the next chapter, we propose a family of approximate likelihood that can be used to draw reliable and valid Bayesian inference on conditional quantile functions.

# Chapter 6

# A family of approximate likelihood functions for quantile regression

## 6.1 Introduction

Bayesian quantile methods based on the asymmetric Laplace likelihood have been used in the literature more than other Bayesian quantile approaches. However, the asymmetric Laplace distribution has several limitations which were discussed in Chapter 4 (for example, the quantile of interest is assumed to be the mode of the underlying distribution of the data and the quantile of interest and skewness are determined by the same parameter). Also, it does not have a direct application to accommodate a vector of quantiles. Therefore, to overcome these shortcomings and to improve the quality and reliability of Bayesian inference on quantile functions, Dunson & Taylor (2005), Reich *et al.* (2010), Lancaster & Jae Jun (2010), Hahn & Burgette (2012) and Feng *et al.* (2015) have proposed a number of alternatives to Bayesian quantile methods based on the asymmetric Laplace likelihood (for more details see, Section 1.6).

To achieve more accurate and reliable Bayesian inference on quantiles, we propose a family of approximate distributions that are parameterised by quantiles and can accommodate different types of underlying distributions. This family of approximate distributions can achieve their maximum at the quantile of interest without assuming that this quantile represents the mode of the true distribution

of the data. These approximate likelihood functions can be used to construct Bayesian methods to estimate quantile curves individually and simultaneously. Then, we extend these methods to fit linear quantile regression with homoscedastic and heteroscedastic errors. To investigate the performance of proposed methods, we propose a bootstrap test to check the lack of fit for quantile regressions. In addition, we implement simulation studies and use real data to investigate the properties of estimators obtained using the proposed quantile methods.

## 6.2 Approximate likelihood functions for individual quantile estimation

For a random variable $y$, we assume that $y \sim F$, where $F$ is a continuous distribution with $\tau_k$ quantile given by $q_{\tau_k}$. Then, for any known density $g(y|\boldsymbol{\theta})$, the true probability density $f(y|q_{\tau_k})$ can be approximated by a density function

$$\widehat{f}(y|\boldsymbol{\theta}, q_{\tau_k}, \boldsymbol{w}) = \begin{cases} w_1 g(y|\boldsymbol{\theta}), & y \leq q_{\tau_k}, \\ w_2 g(y|\boldsymbol{\theta}), & y > q_{\tau_k}, \end{cases} \tag{6.1}$$

where $g(y|\boldsymbol{\theta})$ is a known density. Then, to specify $w_1$ and $w_2$ such that $q_{\tau_k}$ is the $\tau_k$ quantile of the underlying distribution, the following two conditions must be satisfied:

$$\int_{-\infty}^{q_{\tau_k}} w_1 g(y|\boldsymbol{\theta}) dy = \tau_k, \tag{6.2}$$

$$\int_{q_{\tau_k}}^{\infty} w_2 g(y|\boldsymbol{\theta}) dy = 1 - \tau_k. \tag{6.3}$$

This implies that

$$\begin{aligned} w_1 &= \frac{\tau_k}{\int_{-\infty}^{q_{\tau_k}} g(y|\boldsymbol{\theta}) dy} \\ &= \frac{\tau_k}{G(q_{\tau_k}; \boldsymbol{\theta})}, \\ w_2 &= \frac{1 - \tau_k}{\int_{q_{\tau_k}}^{\infty} g(y|\boldsymbol{\theta}) dy} \\ &= \frac{1 - \tau_k}{1 - G(q_{\tau_k}; \boldsymbol{\theta})}, \end{aligned}$$

where $G$ is the cumulative distribution function corresponding to the density $g$. It can be seen that the restriction of the density function $\widehat{f}(y|\boldsymbol{\theta}, q_{\tau_k}, \boldsymbol{w})$ to meet the condition given in Equations 6.2 & 6.3 could results in discontinuity (that is, jumping) at the quantile $q_{\tau_k}$. This discontinuity would disappeared as $g \approx f$, since $w_1$ and $w_2$ would approach one. The cumulative distribution function $\widehat{F}$ corresponding to the density $\widehat{f}$ is continuous with $\widehat{F}(q_{\tau_k}) = \tau_k$. The properties of the estimation based on this approximation are assessed numerically as in Section 6.7. Consider the Kullback–Leibler divergence given by

$$KLD(f\|\widehat{f}) = \int_{-\infty}^{\infty} f(y) \log\left(\frac{f(y)}{\widehat{f}(y|\boldsymbol{\theta}, q_{\tau_k}, \boldsymbol{w})}\right) dy.$$

It is obvious that if $g = f$, then $\widehat{f} = f$ and $KLD(f\|\widehat{f}) = 0$. It is obvious that the choice of function $g$ plays a fundamental role in determining the quality of approximation. Therefore, we investigate the performance of this approximation under three different choices of the function $g$ as discussed in the next sections. We consider the normal distribution, the generalised normal distribution that includes a variety of symmetric distributions as special cases (for example, symmetric Laplace and normal distribution), and a mixture of skewed Laplace and skew normal distributions.

### 6.2.1 The weighted normal distribution

To begin with a simple case, assume $g$ is the density of normal distribution, then the approximate density is given by

$$\widehat{f}(y|\boldsymbol{\theta}, q_{\tau_k}) = \begin{cases} \frac{w_1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}, & y \leq q_{\tau_k}, \\ \frac{w_2}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}, & y > q_{\tau_k}, \end{cases}$$

where $\mu$, $\sigma^2$ and $q_{\tau_k}$ are location, variance and $\tau_k$ quantile parameters respectively. Also, the weights are given by

$$\begin{aligned} w_1 &= \tau_k \left[\Phi\left(\frac{q_{\tau_k} - \mu}{\sigma}\right)\right]^{-1}, \\ w_2 &= (1 - \tau_k) \left[1 - \Phi\left(\frac{q_{\tau_k} - \mu}{\sigma}\right)\right]^{-1}. \end{aligned}$$

where $\Phi$ is the cumulative function of the standard normal distribution.

### 6.2.2 The weighted generalised normal distribution

To add more flexibility to the approximate distribution to be able to accommodate a variety of symmetric distributions and approximate other distributions, we assume $g$ to be the generalised normal distribution. The associated approximate density can be written as

$$\widehat{f}(y|\mu,\sigma,q_{\tau_k},\gamma) = \begin{cases} \frac{w_1}{2\gamma\sigma\Gamma(\gamma)}\exp\left\{-\left(\frac{|y-\mu|}{\sigma}\right)^{\frac{1}{\gamma}}\right\}, & y \leq q_{\tau_k}, \\ \frac{w_2}{2\gamma\sigma\Gamma(\gamma)}\exp\left\{-\left(\frac{|y-\mu|}{\sigma}\right)^{\frac{1}{\gamma}}\right\}, & y > q_{\tau_k}, \end{cases}$$

where $\mu$, $\sigma$, $\gamma$ and $q_{\tau_k}$ are location, scale, shape and $\tau_k$ quantile parameters. Also, the weights are given by

$$\begin{aligned} w_1 &= 2\tau_k\left(1+\text{sign}(q_\tau-\mu)\frac{1}{\Gamma(\gamma)}\overline{\Gamma}\left[\gamma,\left(\frac{|q_\tau-\mu|}{\sigma}\right)^{\frac{1}{\gamma}}\right]\right)^{-1}, \\ w_2 &= 2(1-\tau_k)\left(1-\text{sign}(q_\tau-\mu)\frac{1}{\Gamma(\gamma)}\overline{\Gamma}\left[\gamma,\left(\frac{|q_\tau-\mu|}{\sigma}\right)^{\frac{1}{\gamma}}\right]\right)^{-1}. \end{aligned}$$

where $\overline{\Gamma}(.,.)$ is the lower incomplete gamma function: $\overline{\Gamma}(a,x) = \int_0^x t^{a-1}e^{-t}dt$.

### 6.2.3 The weighted mixture distribution

To improve the performance of the approximation distribution in terms of symmetric and asymmetric distribution, we consider $g$ to be a mixture of skewed Laplace and skewed normal distributions. Then, the approximate density is given by

$$\widehat{f}(y|\mu,q_{\tau_k},\sigma,\kappa,\nu) = \begin{cases} w_1\sum_{t=1}^{2}\frac{(1-\nu)^{t-1}}{\nu^{t-2}}g_t(y|\mu,\sigma,\kappa), & y \leq q_{\tau_k}, \\ w_2\sum_{t=1}^{2}\frac{(1-\nu)^{t-1}}{\nu^{t-2}}g_t(y|\mu,\sigma,\kappa), & y > q_{\tau_k}, \end{cases}$$

where

$$g_t(y|\mu,\sigma,\kappa) = \begin{cases} \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})}\exp\left\{-\left(\frac{1}{\kappa\sigma}|y-\mu|\right)^t\right\}, & y \leq \mu, \\ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})}\exp\left\{-\left(\frac{\kappa}{\sigma}|y-\mu|\right)^t\right\}, & y > \mu. \end{cases}$$

which is a density function for any $t > 0$ (for more details, see Appendix E.1). Also,

$$\begin{aligned} w_1 &= \frac{\tau_k}{G(q_{\tau_k};\mu,\sigma,\kappa,\nu)}, \\ w_2 &= \frac{1-\tau_k}{1-G(q_{\tau_k};\mu,\sigma,\kappa,\nu)}, \end{aligned}$$

where

$$G(q_{\tau_k};\mu,\sigma,\kappa) = \begin{cases} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left(\frac{1}{t}, \left(\frac{|q_{\tau_k}-\mu|}{\kappa\sigma}\right)^t\right)\right], & q_{\tau_k} \le \mu, \\ \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ 1 - \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left(\frac{1}{t}, \left(\frac{\kappa|q_{\tau_k}-\mu|}{\sigma}\right)^t\right)\right], & q_{\tau_k} > \mu, \end{cases}$$

where $\widetilde{\Gamma}(.,.)$ is the upper incomplete gamma function: $\widetilde{\Gamma}(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$ and $\nu$ is a mixture weight such that $\nu \in (0,1)$; for more details, see Appendix E.2. This mixture distribution includes a number of distributions as special cases, such as normal and Laplace distribution families.

## 6.3 An approximate likelihood function for simultaneous estimation of quantiles

To prevent the crossing of estimated quantile functions, the family of approximate likelihood functions that are discussed in the previous section can be easily generalised to accommodate multiple quantiles corresponding to $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$. To achieve this goal, we reformulate the density function given in (6.1) to be written as

$$\widehat{f}(y|\boldsymbol{\theta},\boldsymbol{q}_\tau) = \begin{cases} w_1 g(y|\boldsymbol{\theta}), & y \le q_{\tau_1}, \\ w_2 g(y|\boldsymbol{\theta}), & q_{\tau_1} < y \le q_{\tau_2}, \\ \quad\vdots & \quad\vdots \\ w_{m+1} g(y|\boldsymbol{\theta}), & y > q_{\tau_m}, \end{cases}$$

where

$$w_1 = \frac{\tau_1}{G(q_{\tau_1};\boldsymbol{\theta})},$$

$$w_2 = \frac{\tau_2 - \tau_1}{G(q_{\tau_2};\boldsymbol{\theta}) - G(q_{\tau_1};\boldsymbol{\theta})},$$
$$\vdots$$
$$w_{m+1} = \frac{1-\tau_m}{1 - G(q_{\tau_m};\boldsymbol{\theta})},$$

where $G$ is the cumulative distribution function corresponding to the density $g$. It is obvious that $KLD(f\|\widehat{f}) = 0$ if $g = f$, since if $g = f$, $\widehat{f} = f$.

As this approximation is affected by the choices of $g$, it can also be affected by the number and locations of quantiles included in simultaneous estimation. In the following section, we employ this family of approximate likelihood functions to construct Bayesian methods to fit multiple quantile linear regressions simultaneously.

## 6.4 Simultaneous estimation of linear quantile functions

Consider the linear regression model

$$y_i = \beta_o + \boldsymbol{x}_i^{'}\boldsymbol{\beta} + \epsilon_i, \text{ for } i = 1, 2, ..., n, \tag{6.4}$$

where $\epsilon_i$ is independently distributed with a mean equal to zero and a constant variance; $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates for the $i^{\text{th}}$ observation; $\beta_o$ is an intercept coefficient and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown slope coefficients. To develop a Bayesian quantile method to estimate the linear quantile function given by $Q(\tau_k|\boldsymbol{x_i}) = \alpha_{\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}$, for $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$ simultaneously using the family of approximate likelihood functions, we can consider the weighted normal likelihood written as

$$l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma^2; \boldsymbol{y}) \propto \prod_{i=1}^{n} \begin{cases} \frac{w_1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left(y_i - \beta_o - \boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)^2}{2\sigma^2}\right\}, & y_i \leq \alpha_{\tau_1} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}, \\ \frac{w_2}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left(y_i - \beta_o - \boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)^2}{2\sigma^2}\right\}, & \alpha_{\tau_1} + \boldsymbol{x}_i^{'}\boldsymbol{\beta} < y_i \leq \alpha_{\tau_2} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}, \\ \vdots & \vdots \\ \frac{w_{m+1}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left(y_i - \beta_o - \boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)^2}{2\sigma^2}\right\}, & y_i > \alpha_{\tau_m} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}, \end{cases}$$

where the weights are given by

$$
\begin{aligned}
w_1 &= \tau_k \left[ \Phi\left( \frac{\alpha_{\tau_k} - \beta_o}{\sigma} \right) \right]^{-1}, \\
w_k &= (\tau_k - \tau_{k-1}) \left[ \Phi\left( \frac{\alpha_{\tau_k} - \beta_o}{\sigma} \right) - \Phi\left( \frac{\alpha_{\tau_{k-1}} - \beta_o}{\sigma} \right) \right]^{-1} \text{ for } k = 2, ...m, \\
w_{m+1} &= (1 - \tau_m) \left[ 1 - \Phi\left( \frac{\alpha_{\tau_m} - \beta_o}{\sigma} \right) \right]^{-1}.
\end{aligned}
$$

To illustrate the performance of the proposed approximate likelihood function, we use diffuse proper prior distributions, which yield proper posterior distributions, for all model parameters. For linear quantile regression coefficients, the prior distribution is given by

$$
\begin{aligned}
\alpha_{\tau_k} &\sim \mathrm{N}(\alpha^*, \sigma_\alpha), \text{ for } k = 1, ..., m, \\
\boldsymbol{\beta} &\sim \mathrm{N}_p(\boldsymbol{\beta}^*, \Sigma_{\boldsymbol{\beta}}),
\end{aligned}
$$

where each normal distribution has mean equal to zero and variance equal to $10^5$ and each multivariate normal distribution has a mean vector of zero and a diagonal covariance matrix with entries equal to $10^5$. For the variance parameter, we consider the inverse-gamma distribution (IG) that is given by

$$
\sigma^2 \sim \mathrm{IG}(a = 0.01, b = 0.01).
$$

Then, the posterior distribution is given by

$$
p(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma^2 | \boldsymbol{y}) \propto l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma^2; \boldsymbol{y}) p(\beta_o) p(\boldsymbol{\beta}) p(\boldsymbol{\alpha}_\tau) p(\sigma^2).
$$

Also, we can consider the weighted generalised normal likelihood given by

$$
l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma, \gamma; \boldsymbol{y}) \propto \prod_{i=1}^{n} \begin{cases}
\frac{w_1}{2\gamma\sigma\Gamma(\gamma)} \exp\left\{ - \left( \frac{\left| y_i - \beta_o - \boldsymbol{x}_i' \boldsymbol{\beta} \right|}{\sigma} \right)^{\frac{1}{\gamma}} \right\}, & y_i \leq \alpha_{\tau_1} + \boldsymbol{x}_i' \boldsymbol{\beta}, \\
\frac{w_2}{2\gamma\sigma\Gamma(\gamma)} \exp\left\{ - \left( \frac{\left| y_i - \beta_o - \boldsymbol{x}_i' \boldsymbol{\beta} \right|}{\sigma} \right)^{\frac{1}{\gamma}} \right\}, & \alpha_{\tau_1} + \boldsymbol{x}_i' \boldsymbol{\beta} < y_i \leq \alpha_{\tau_2} + \boldsymbol{x}_i' \boldsymbol{\beta}, \\
\vdots & \vdots \\
\frac{w_{m+1}}{2\gamma\sigma\Gamma(\gamma)} \exp\left\{ - \left( \frac{\left| y_i - \beta_o - \boldsymbol{x}_i' \boldsymbol{\beta} \right|}{\sigma} \right)^{\frac{1}{\gamma}} \right\}, & y_i > \alpha_{\tau_m} + \boldsymbol{x}_i' \boldsymbol{\beta},
\end{cases}
$$

where the weights are given by

$$
\begin{aligned}
w_1 &= 2\tau_k \left(1 + \text{sign}(\alpha_{\tau_1} - \beta_o)\frac{1}{\Gamma(\gamma)}\overline{\Gamma}\left[\gamma, \left(\frac{|\alpha_{\tau_1} - \beta_o|}{\sigma}\right)^{\frac{1}{\gamma}}\right]\right)^{-1}, \\
w_k &= 2(\tau_k - \tau_{k-1})\left(\text{sign}(\alpha_{\tau_k} - \beta_o)\frac{1}{\Gamma(\gamma)}\overline{\Gamma}\left[\gamma, \left(\frac{|\alpha_{\tau_k} - \beta_o|}{\sigma}\right)^{\frac{1}{\gamma}}\right]\right. \\
&\quad \left. -\text{sign}(\alpha_{\tau_{k-1}} - \beta_o)\frac{1}{\Gamma(\gamma)}\overline{\Gamma}\left[\gamma, \left(\frac{|\alpha_{\tau_{k-1}} - \beta_o|}{\sigma}\right)^{\frac{1}{\gamma}}\right]\right)^{-1} \quad \text{for } k = 2, ...m, \\
w_{m+1} &= 2(1 - \tau_m)\left(1 - \text{sign}(\alpha_{\tau_m} - \beta_o)\frac{1}{\Gamma(\gamma)}\overline{\Gamma}\left[\gamma, \left(\frac{|\alpha_{\tau_m} - \beta_o|}{\sigma}\right)^{\frac{1}{\gamma}}\right]\right)^{-1}.
\end{aligned}
$$

For the scale parameter, we consider the inverse-gamma distributions (IG) given by

$$\sigma \sim \text{IG}(a = 0.01, b = 0.01).$$

Also, the shape parameter has a prior distribution:

$$\gamma \sim \text{IG}(a = 0.01, b = 0.5),$$

which is a diffuse distribution with a mode that is approximately equal to 0.5 that results to the shape of normal distribution. Then, the posterior distribution is given by

$$p(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma, \gamma | \boldsymbol{y}) \propto l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma, \gamma; \boldsymbol{y}) p(\beta_o) p(\boldsymbol{\beta}) p(\boldsymbol{\alpha}_\tau) p(\sigma) p(\gamma).$$

Moreover, we can consider the weighted mixture of skewed Laplace and skewed normal likelihood functions given by

$$
l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma, \kappa, \nu; \boldsymbol{y}) \propto \prod_{i=1}^{n} \begin{cases}
w_1 \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} g_t(y_i | \beta_o, \boldsymbol{\beta}, \sigma, \kappa), & y_i \leq \alpha_{\tau_1} + \boldsymbol{x}_i' \boldsymbol{\beta}, \\
w_2 \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} g_t(y_i | \beta_o, \boldsymbol{\beta}, \sigma, \kappa), & \alpha_{\tau_1} + \boldsymbol{x}_i' \boldsymbol{\beta} < y_i \leq \alpha_{\tau_2} + \boldsymbol{x}_i' \boldsymbol{\beta}, \\
\vdots & \vdots \\
w_{m+1} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} g_t(y_i | \beta_o, \boldsymbol{\beta}, \sigma, \kappa), & y_i > \alpha_{\tau_m} + \boldsymbol{x}_i' \boldsymbol{\beta},
\end{cases}
$$

where, for $t = 1, 2$,

$$g_t(y_i|\beta_o, \boldsymbol{\beta}, \sigma, \kappa) = \begin{cases} \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left(\frac{1}{\kappa\sigma}\left|y_i - \beta_o - \boldsymbol{x}_i'\boldsymbol{\beta}\right|\right)^t\right\}, & y_i \leq \beta_o + \boldsymbol{x}_i'\boldsymbol{\beta}, \\ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left(\frac{\kappa}{\sigma}\left|y_i - \beta_o - \boldsymbol{x}_i'\boldsymbol{\beta}\right|\right)^t\right\}, & y_i > \beta_o + \boldsymbol{x}_i'\boldsymbol{\beta}. \end{cases}$$

Also, the weights are given by

$$\begin{aligned} w_1 &= \frac{\tau_1}{G(\alpha_{\tau_1}; \beta_o, \sigma, \kappa, \nu)}, \\ w_2 &= \frac{\tau_2 - \tau_1}{G(\alpha_{\tau_2}; \beta_o, \sigma, \kappa, \nu) - G(\alpha_{\tau_1}; \beta_o, \sigma, \kappa, \nu)}, \\ &\vdots \\ w_{m+1} &= \frac{1 - \tau_m}{1 - G(\alpha_{\tau_m}; \beta_o, \sigma, \kappa, \nu)}, \end{aligned}$$

where

$$G(\alpha_{\tau_k}; \beta_o, \sigma, \kappa, \nu) = \begin{cases} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[\frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left(\frac{1}{t}, \left(\frac{|\alpha_{\tau_k} - \beta_o|}{\kappa\sigma}\right)^t\right)\right], & \alpha_{\tau_k} \leq \beta_o, \\ \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[1 - \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left(\frac{1}{t}, \left(\frac{\kappa|\alpha_{\tau_k} - \beta_o|}{\sigma}\right)^t\right)\right], & \alpha_{\tau_k} > \beta_o. \end{cases}$$

For the mixture weight parameter, we consider the Beta distribution (Beta) given by

$$\nu \sim \text{Beta}(1, 1),$$

which is the uniform distribution over the interval (0,1). For the skew parameter, we consider the inverse-gamma distribution (IG) given by

$$\kappa \sim \text{IG}(a = 0.01, b = 0.01).$$

Then, the posterior distribution is given by

$$p(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma, \kappa, \nu|\boldsymbol{y}) \propto l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, \sigma, \kappa, \nu; \boldsymbol{y}) p(\beta_o) p(\boldsymbol{\beta}) p(\boldsymbol{\alpha}_\tau) p(\sigma) p(\kappa) p(\nu).$$

## 6.5 Simultaneous estimation of quantiles for heteroscedastic linear models

Consider the heteroscedastic linear regression model

$$y_i = \beta_o + \boldsymbol{x}_i^{'}\boldsymbol{\beta} + \sigma(\boldsymbol{x_i})\epsilon_i,$$

where $\epsilon_i$ is independently distributed with a mean equal to zero and a constant variance and $\sigma(\boldsymbol{x}_i)$, is standard deviation of the error term, assumed to be a function of the covariates. To accommodate the heteroscedastic scale in the Bayesian quantile model, we assume that $\sigma(\boldsymbol{x}_i) = \exp\left(\psi_o + \boldsymbol{x}_i^{'}\boldsymbol{\psi}\right)$. Then, we propose a Bayesian quantile method to estimate linear quantile function:

$$Q(\tau_k|\boldsymbol{x_i}) = \alpha_{\tau_k} + \boldsymbol{x}_i^{'}\boldsymbol{\beta}_{\tau_k},$$

for $0 < \tau_1 < \tau_2 < ... < \tau_m < 1$ simultaneously using the family of approximate likelihood functions by considering the weighted mixture of skewed Laplace and skewed normal likelihood functions given by

$$l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, B_\tau, \psi_o, \boldsymbol{\psi}, \kappa, \nu; \boldsymbol{y}) \propto \prod_{i=1}^{n} \begin{cases} w_1 \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} g_t(y_i|\beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa), & y_i \le Q(\tau_1|\boldsymbol{x_i}), \\ w_2 \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} g_t(y_i|\beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa), & Q(\tau_1|\boldsymbol{x_i}) < y_i \le Q(\tau_2|\boldsymbol{x_i}), \\ \vdots & \vdots \\ w_{m+1} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} g_t(y_i|\beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa), & y_i > Q(\tau_m|\boldsymbol{x_i}), \end{cases}$$

where

$$g_t(y_i|\beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa) = \begin{cases} \frac{t\kappa}{\sigma(\boldsymbol{x}_i)(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left(\frac{1}{\kappa\sigma(\boldsymbol{x}_i)}\left|y_i - \beta_o - \boldsymbol{x}_i^{'}\boldsymbol{\beta}\right|\right)^t\right\}, & y_i \le \beta_o + \boldsymbol{x}_i^{'}\boldsymbol{\beta}, \\ \frac{t\kappa}{\sigma(\boldsymbol{x}_i)(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left(\frac{\kappa}{\sigma(\boldsymbol{x}_i)}\left|y_i - \beta_o - \boldsymbol{x}_i^{'}\boldsymbol{\beta}\right|\right)^t\right\}, & y_i > \beta_o + \boldsymbol{x}_i^{'}\boldsymbol{\beta}. \end{cases}$$

Also, the weights are given by

$$\begin{aligned} w_1 &= \frac{\tau_1}{G(\alpha_{\tau_1}, \boldsymbol{\beta}_{\tau_1}; \beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa, \nu)}, \\ w_2 &= \frac{\tau_2 - \tau_1}{G(\alpha_{\tau_2}, \boldsymbol{\beta}_{\tau_2}; \beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa, \nu) - G(\alpha_{\tau_1}, \boldsymbol{\beta}_{\tau_1}; \beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa, \nu)}, \\ &\vdots \end{aligned}$$

$$w_{m+1} = \frac{1-\tau_m}{1-G(\alpha_{\tau_m}, \boldsymbol{\beta}_{\tau_m}; \beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa, \nu)},$$

where

$$G(\alpha_{\tau_k}\boldsymbol{\beta}_{\tau_k}; \beta_o, \boldsymbol{\beta}, \psi_o, \boldsymbol{\psi}, \kappa, \nu) = \begin{cases} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[\frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})}\widetilde{\Gamma}\left(\frac{1}{t}, \xi_1\right)\right], & Q(\tau_k|\boldsymbol{x_i}) \leq \beta_o + \boldsymbol{x}_i'\boldsymbol{\beta}, \\ \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[1 - \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})}\widetilde{\Gamma}\left(\frac{1}{t}, \xi_2\right)\right], & Q(\tau_k|\boldsymbol{x_i}) > \beta_o + \boldsymbol{x}_i'\boldsymbol{\beta}, \end{cases}$$

where

$$\xi_1 = \left(\frac{\left|\alpha_{\tau_k} - \beta_o + \boldsymbol{x}_i'(\boldsymbol{\beta}_{\tau_k} - \boldsymbol{\beta})\right|}{\kappa \exp\left(\psi_o + \boldsymbol{x}_i'\boldsymbol{\psi}\right)}\right)^t, \quad \xi_2 = \left(\frac{\kappa\left|\alpha_{\tau_k} - \beta_o + \boldsymbol{x}_i'(\boldsymbol{\beta}_{\tau_k} - \boldsymbol{\beta})\right|}{\exp\left(\psi_o + \boldsymbol{x}_i'\boldsymbol{\psi}\right)}\right)^t.$$

The prior distributions for the quantile slope coefficients and scale components are given by

$$\begin{aligned} \boldsymbol{\beta}_{\tau_k} &\sim \mathrm{N}_p(\boldsymbol{\beta}^*, \Sigma_{\boldsymbol{\beta}}) \text{ for } k = 1, ..., m, \\ \psi_o &\sim \mathrm{N}(\psi_o^*, \sigma_{\psi_o}), \\ \boldsymbol{\psi} &\sim \mathrm{N}_p(\boldsymbol{\psi}^*, \Sigma_{\boldsymbol{\psi}}), \end{aligned}$$

with the same specified parameters used, in Section 6.4, for prior distributions of quantile coefficients. Then, the posterior distribution is given by

$$\begin{aligned} p(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, B_\tau, \psi_o, \boldsymbol{\psi}, \kappa, \nu|\boldsymbol{y}) \propto\ & l(\beta_o, \boldsymbol{\beta}, \boldsymbol{\alpha}_\tau, B_\tau, \psi_o, \boldsymbol{\psi}, \kappa, \nu; \boldsymbol{y}) p(\beta_o) p(\boldsymbol{\beta}) p(\boldsymbol{\alpha}_\tau) p(B_\tau) \\ & \times p(\psi_o) p(\boldsymbol{\psi}) p(\kappa) p(\nu). \end{aligned}$$

## 6.6 Checking the goodness of fit

To check the goodness of fit for quantile regression, a number of hypothesis test methods have been developed (for example, see Wang, 2008; Zheng, 1998). In this section we proposed a new bootstrap test to examine the goodness of fit for linear quantile regression given in (6.4). We consider

$$\begin{aligned} H_o &: P\big(\epsilon(\tau_k) \leq 0\big) = \tau_k, \\ H_1 &: P\big(\epsilon(\tau_k) \leq 0\big) \neq \tau_k, \end{aligned}$$

where $P\big(\epsilon(\tau_k) \leq 0\big)$ is the cumulative distribution function of the residuals, corresponding to $\tau_k$ quantile function, which given by $\epsilon_i(\tau_k) = y_i - (\alpha_{\tau_k} + \boldsymbol{x}_i'\boldsymbol{\beta})$. Then, the test statistic is given by

$$D = \widehat{P}\big(\epsilon(\tau_k) \leq 0\big) - \tau_k,$$

where $\widehat{P}\big(\epsilon(\tau_k) \leq 0\big)$ is a non-parametric estimation of unknown cumulative function $P\big(\epsilon(\tau_k) \leq 0\big)$. Then, the p-value of the two-sided test is given by

$$\begin{aligned} \text{p-value} &= 2\min\left\{P(D \leq 0|H_o), P(D > 0|H_o)\right\}, \\ &= 2\min\left\{\frac{\sum_{b=1}^{B} 1_{(D_b \leq 0)}}{B}, \frac{\sum_{b=1}^{B} 1_{(D_b > 0)}}{B}\right\}, \end{aligned} \tag{6.5}$$

where $B$ is the number of bootstrap simulations obtained by resampling from the observed data with replacement, and $1_{(A)}$ is an indicator function which equals one if $A$ holds and zero otherwise. To estimate the cumulative function $P\big(\epsilon(\tau_k) \leq 0\big)$, we could use the empirical cumulative function given by

$$\widehat{P}\big(\epsilon(\tau_k) \leq 0\big) = \frac{1}{n}\sum_{i=1}^{n} 1_{\big(\epsilon_i(\tau_k) \leq 0\big)}.$$

However, the hypothesis test would then favor quantile curves satisfying the sample definition of quantiles. To improve the performance of the hypothesis test, we estimate $P\big(\epsilon(\tau_k) \leq 0\big)$ using the cumulative function estimation considered by Nadaraya (1964) and that can be defined for identically and independently distributed random variables $Z_1, ...Z_n$ as follows

$$\widehat{P}(Z \leq z) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{z - Z_i}{h}\right),$$

where $K$ is the cumulative distribution function of a positive kernel (here we use Epanechnikov kernel function). For bandwidth selection, we use a plug-in estimator of optimal bandwidth proposed by Altman & Leger (1995). To illustrate the performance of this method to approximate cumulative distributions, Figure 6.1 shows the lack of bias in the estimation of the cumulative distribution functions corresponding to normal and gamma distributions (for more details about implementation of these methods, see Wang, 2012).

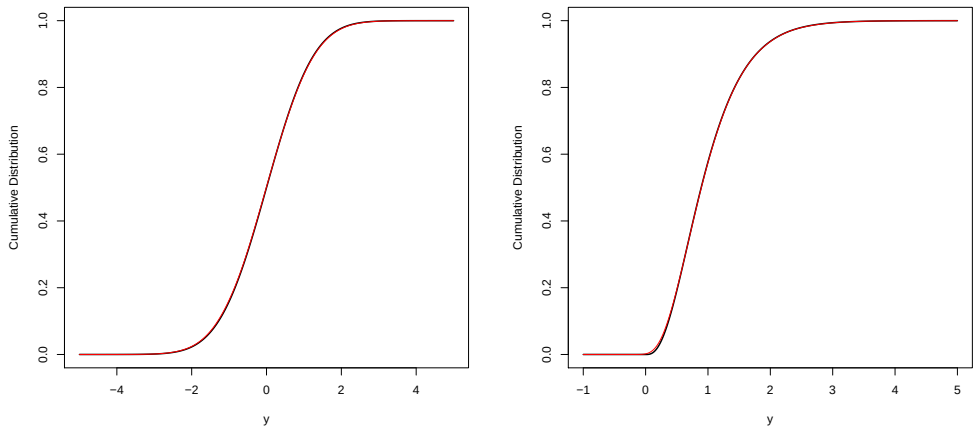



Figure 6.1: The average of the estimated cumulative functions, obtained using the method developed by Nadaraya (1964) with a plug-in estimator of optimal bandwidth proposed by Altman & Leger (1995) for $\mathrm{N}(0,1)$ and $\mathrm{Gamma}(3,3)$, over 1000 simulations of size 150.

## 6.7 Simulation studies and real data analysis

### 6.7.1 Univariate models

In this section, we generate univariate samples, each of size 150, from normal distribution N(0,1), gamma distribution Gamma(3,3) and Student's $t$ distribution $t(3)$. Then, we investigate the performance of the proposed family of approximate distributions, discussed in Section 6.4, to construct Bayesian methods to draw Bayesian inference on the first quartile, the median, and the third quartile by comparing the estimated density and cumulative functions with true ones. The estimated density and cumulative functions are represented by the average of function values computed at the maximum a posteriori estimators over 1000 simulations.

Figures 6.2, 6.3 & 6.4 illustrate the good quality of approximation obtained using the proposed family of approximate distributions under the different choices of the function $g$. They show how the choice of the function $g$ can affect the approximation. Also, they show that the approximation based on the weighted mixture of skewed Laplace and skewed normal distributions outperform other approximations based on the weighted normal and the weighted generalised normal

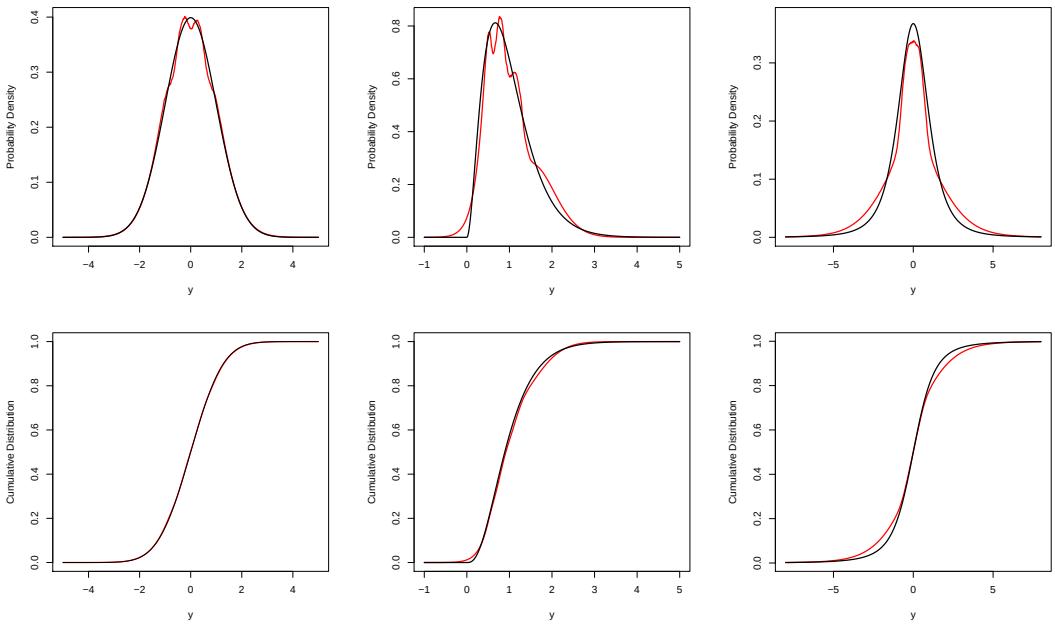



Figure 6.2: The probability density function (top) and the cumulative distribution function (bottom) for the normal, gamma and Student's $t$ distributions (left to right). The true distributions are represented by black curves and their approximations, obtained using the weighted normal distribution, are represented by red curves.

distributions. In addition, these figures suggest that the quantiles included in the simultaneous estimation can play a significant role in improving the approximations of the underlying distributions, in the case that the $g$ density function does not accommodate these underlying distributions. Although the proposed family of approximate distributions does not assume that any quantile of interest is the mode of the true distribution, Figures 6.2, 6.3 & 6.4 show that the proposed family of approximate distributions jumps at estimated quantiles. This can affect the quality of the approximation especially in the case of small size of datasets. Thus, in the case that the function $g$ can accommodate the underlying distribution, the proposed approximation may perform better for fewer quantiles.

To examine the accuracy of the proposed bootstrap test, we implement a

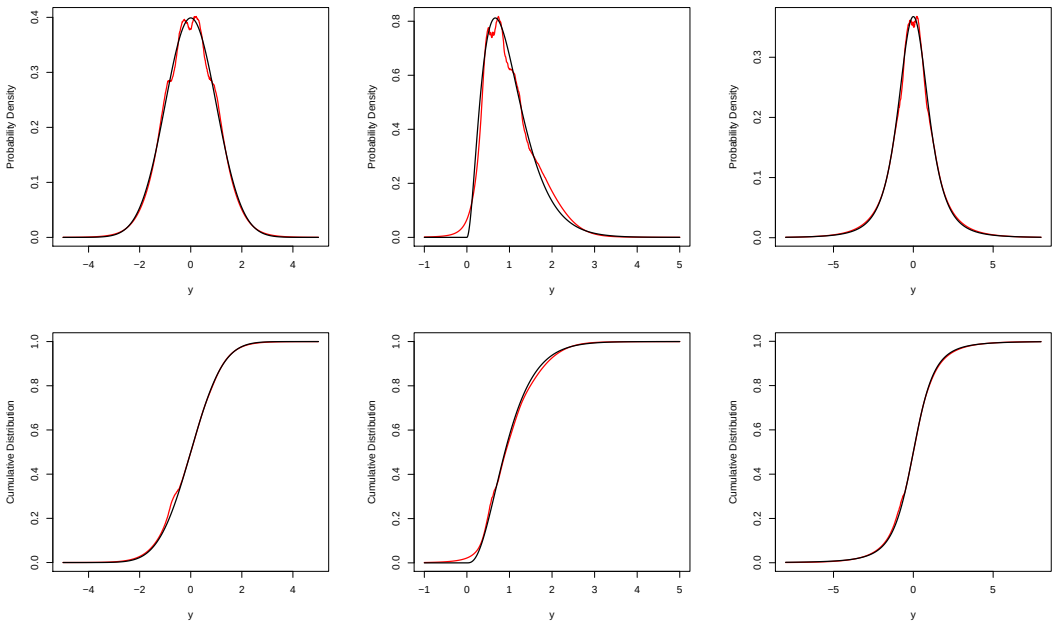



Figure 6.3: The probability density function (top) and the cumulative distribution function (bottom) for the normal, gamma and Student's $t$ distributions (left to right). The true distributions are represented by black curves and their approximations, obtained using the weighted generalised normal distribution, are represented by red curves.

simulation study to calculate the p-value for the hypothesis test:

$$H_o : P(y \leq \mu_{\tau_k}) = \tau_k,$$
$$H_1 : P(y \leq \mu_{\tau_k}) \neq \tau_k,$$

where $\mu_{\tau_k}$ is set to be the true quantile. Then, we compute $P(\text{reject } H_o | H_o \text{ is true})$ at the significance level 0.05. From Table 6.1, the estimated significance level is very close to the nominal level for all distributions. This implies that the proposed test can be useful in measuring the goodness of fit for quantile regression and comparing different quantile methods.

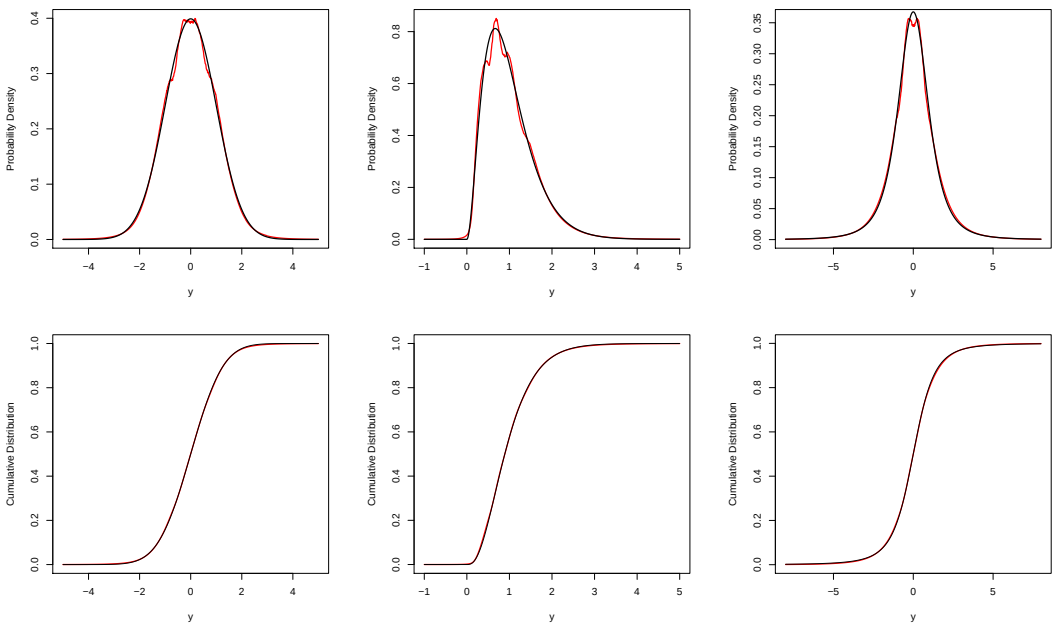



Figure 6.4: The probability density function (top) and the cumulative distribution function (bottom) for the normal, gamma and Student's $t$ distributions (left to right). The true distributions are represented by black curves and their approximations, obtained using the weighted mixture of skewed Laplace and skewed normal distributions, are represented by red curves.

| $\epsilon_i$ | 0.05 | 0.25 | 0.5 | 0.75 | 0.95 |
|---|---|---|---|---|---|
| N$(0,1)$ | 0.07 | 0.06 | 0.04 | 0.05 | 0.06 |
| Student's $t(3)$ | 0.08 | 0.06 | 0.05 | 0.05 | 0.06 |
| Gamma$(3,3)$ | 0.07 | 0.05 | 0.05 | 0.05 | 0.07 |

Table 6.1: The estimated $P(\text{reject } H_o | H_o \text{ is true})$ over 1000 simulations.

### 6.7.2 Linear models

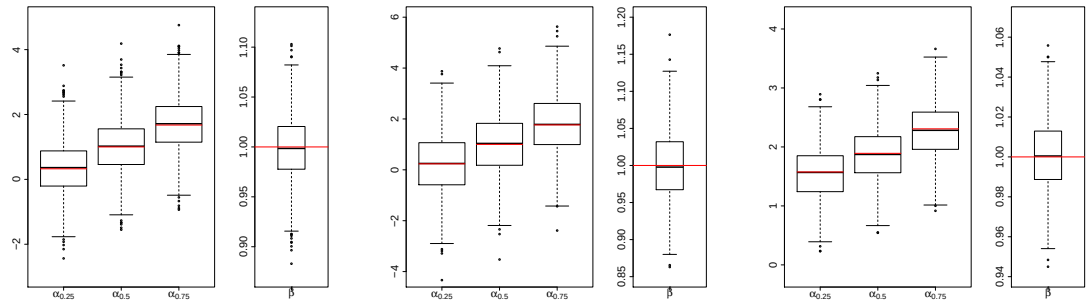
In this section, we evaluate the accuracy of the coverage probabilities of 95% HPD intervals estimated using Bayesian quantile methods based on the family of approximate likelihood functions. We consider estimating quantile curves for 200 observations from the linear model

$$y_i = 1 + x_i + \epsilon_i,$$

where $x_i \sim \mathrm{U}(20, 30)$ and $\epsilon_i$ is considered to follow different distributions. In the simulation study, we use 1000 simulated datasets which only differ in the errors $\epsilon_i$.

| $\epsilon_i$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\beta$ |
|---|---|---|---|---|
| $\mathrm{N}(0, 1)$ | 0.94 | 0.94 | 0.94 | 0.94 |
| Student's $t(3)$ | 0.86 | 0.85 | 0.86 | 0.85 |
| Gamma$(3, 3)$ | 0.81 | 0.81 | 0.80 | 0.81 |

Table 6.2: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated simultaneously using the Bayesian quantile method based on the weighted normal likelihood over 1000 simulations.



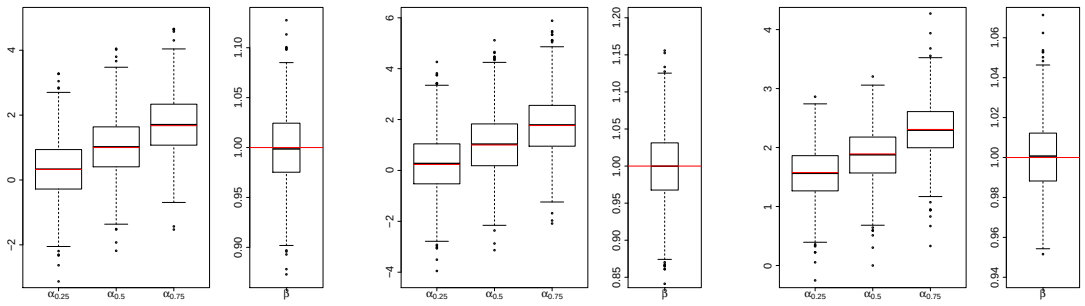
(a) $\mathrm{N}(0, 1)$ (b) Student's $t(3)$ (c) Gamma$(3, 3)$

Figure 6.5: The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted normal likelihood, over 1000 simulations. The red lines represent the true values.

| $\epsilon_i$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\beta$ |
|---|---|---|---|---|
| $\mathrm{N}(0, 1)$ | 0.91 | 0.90 | 0.90 | 0.90 |
| Student's $t(3)$ | 0.86 | 0.85 | 0.85 | 0.85 |
| Gamma$(3, 3)$ | 0.84 | 0.83 | 0.84 | 0.84 |

Table 6.3: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated simultaneously using the Bayesian quantile method based on the weighted generalised normal likelihood over 1000 simulations.

Tables 6.2, 6.3 & 6.4 show that the coverage probabilities, obtained using simultaneous quantile methods based on the proposed family of approximate dis-



(a) N(0, 1) (b) Student's $t(3)$ (c) Gamma(3, 3)

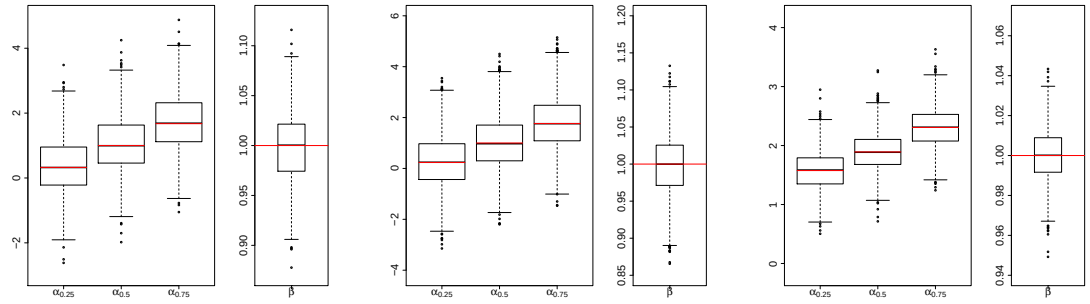
Figure 6.6: The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted generalised normal likelihood, over 1000 simulations. The red lines represent the true values.

| $\epsilon_i$ | $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\beta$ |
|---|---|---|---|---|
| N(0, 1) | 0.91 | 0.90 | 0.91 | 0.90 |
| Student's $t(3)$ | 0.90 | 0.90 | 0.90 | 0.90 |
| Gamma(3, 3) | 0.91 | 0.91 | 0.91 | 0.91 |

Table 6.4: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated simultaneously using the Bayesian quantile method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions over 1000 simulations.

tributions, are always below the nominal 95% and approach the nominal coverage probabilities more closely as the function $g$ can represent the conditional distribution of the response variable perfectly. Therefore, the approximation based on the weighted mixture of skewed Laplace and skewed normal likelihood functions gives the best result, and this is consistent with the discussion in the previous section. In addition, these tables suggest that, under the misspecified choice for the function $g$, the proposed family of approximations still performs well.

Figures 6.5, 6.6 & 6.7 show the distribution of the maximum a posteriori estimators, obtained using the Bayesian methods based on the proposed family of approximate likelihood functions, for linear quantile coefficients over 1000 simulations. It can be seen that the distribution of the maximum a posteriori estimates given by the Bayesian quantile methods based on the proposed family of approximations, over replicated datasets, are symmetric and centered at the true values



(a) N(0, 1) (b) Student's $t(3)$ (c) Gamma(3, 3)

Figure 6.7: The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions, over 1000 simulations. The red lines represent the true values.

under different choices of the density function $g$.

### 6.7.3 Heteroscedastic linear models

In this section, we examine the performance of the Bayesian method, based on the weighted mixture of skewed Laplace and skewed normal likelihood functions described in Section 6.5, to estimate the quantile functions for linear models with heteroscedastic errors. We consider 1000 datasets, differing in $\epsilon_i$, of size 200 generated from the heteroscedastic linear model

$$y_i = 1 + x_i + (1 + \psi x_i)\epsilon_i$$

where $x_i \sim \mathrm{U}(20, 30)$, $\psi = 0.3$ and $\epsilon_i$ is considered to have a chi-square distribution with 3 degrees of freedom. From Table 6.5, the coverage probabilities

| $\alpha_{0.25}$ | $\alpha_{0.5}$ | $\alpha_{0.75}$ | $\beta_{0.25}$ | $\beta_{0.5}$ | $\beta_{0.75}$ |
|---|---|---|---|---|---|
| 0.92 | 0.91 | 0.93 | 0.93 | 0.91 | 0.91 |

Table 6.5: The coverage probabilities of the 95% HPD intervals for the quantile regression coefficients estimated simultaneously using the Bayesian quantile method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions with heteroscedastic scale over 1000 simulations.

corresponding to heteroscedastic linear quantile regression coefficients indicate that the Bayesian method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions succeeds in effectively handling this complex model that has a long tail with heteroscedasticity. Also, Figure 6.8 illustrates a good behaviour of the distributions of the maximum a posteriori estimators.

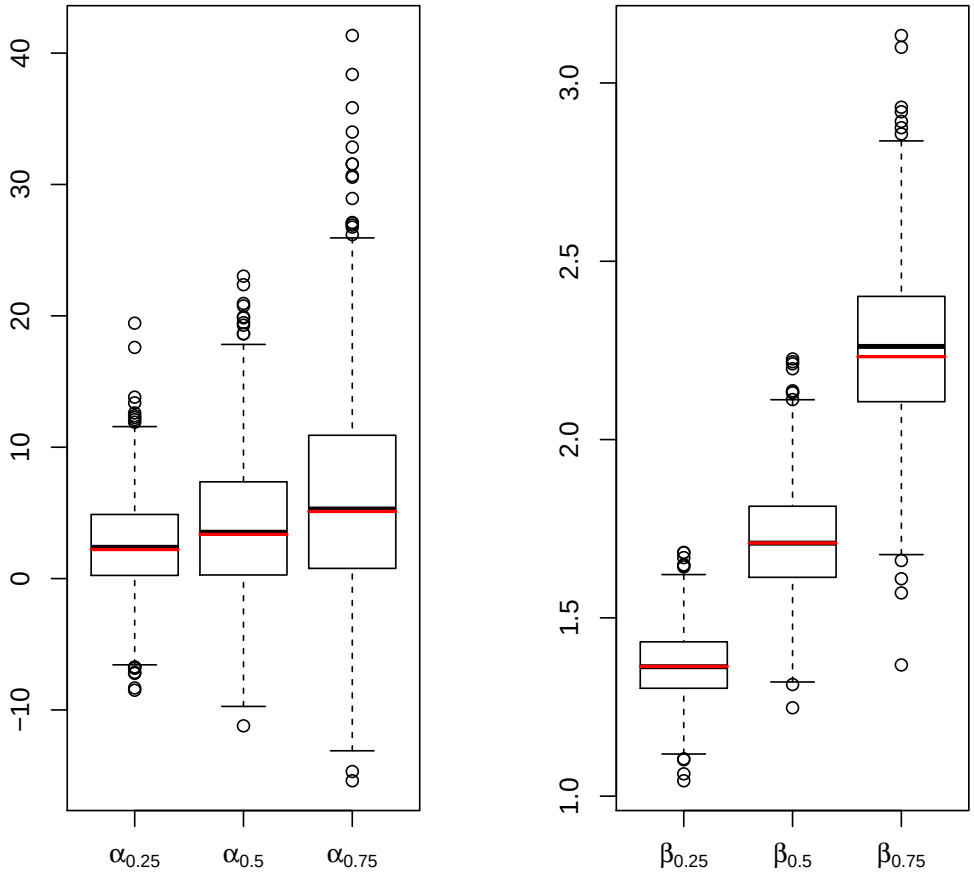



Figure 6.8: The box-plot of the maximum a posteriori estimators, obtained simultaneously using the Bayesian quantile method based on the weighted mixture of skewed Laplace and skewed normal likelihood functions with heteroscedastic scale, over 1000 simulations. The red lines represent the true values.

### 6.7.4 Cats data

In contrast to the asymmetric Laplace distribution estimating the quantile that minimises the traditional loss function, the proposed family of approximations estimates the quantile satisfying the definition of the quantile with respect to the approximate underlying distribution. This implies that the quality of estimation depends on the quality of the approximation for the distribution of the data. Therefore, for fair investigation, we use a new dataset rather than using the Chatterjee-Price Attitude data used in Section 5.8.3, which is relatively small to provide enough information for approximating the underlying distribution. We use Cats data, that includes the measurements on the heart weights (in grams) and body weights ( in kilograms) of 144 male and female cats, to investigate performance of Bayesian methods based on the proposed family of approximations to estimate the quantile curves corresponding to the first quartile, the median, and the third quartile simultaneously. This dataset is available publicly by R Core Team (2016) and Venables & Ripley (2002).

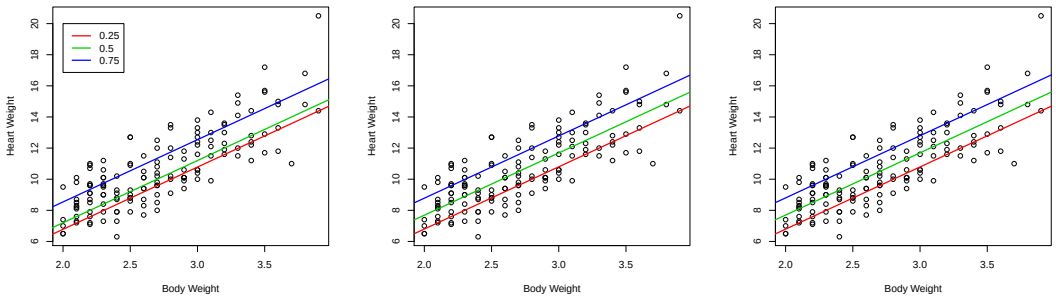



Figure 6.9: The estimated quantile functions based on the maximum a posteriori estimators obtained simultaneously using the weighted normal likelihood, the weighted generalised normal likelihood and the weighted mixture of skewed Laplace and skewed normal likelihoods (left to right), for $\tau = 0.25, 0.5, 0.75$.

From Figure 6.9 and Table 6.6, there is a similarity in the maximum a posteriori estimators obtained using the Bayesian quantile methods based on the weighted generalised normal likelihood, and the weighted mixture of skewed Laplace and skewed normal likelihoods. In addition, the Bayesian quantile method based on the weighted normal likelihood do not agree with other methods in terms

| | QR-WN | | QR-WGN | | QR-WM | |
|---|---|---|---|---|---|---|
| Coefficients | MAP | 95% HPD | MAP | 95% HPD | MAP | 95% HPD |
| $\alpha_{0.25}$ | $-1.22$ | $(-2.45, -0.17)$ | $-1.20$ | $(-2.36, -0.31)$ | $-1.21$ | $(-1.88, 0.17)$ |
| $\alpha_{0.5}$ | $-0.80$ | $(-1.69, 0.61)$ | $-0.31$ | $(-1.40, 0.41)$ | $-0.30$ | $(-1.15, 0.92)$ |
| $\alpha_{0.75}$ | 0.54 | $(-0.41, 1.89)$ | 0.78 | $(-0.26, 1.59)$ | 0.80 | $(0.08, 2.21)$ |
| $\beta$ | 4 | $(3.56, 4.38)$ | 4 | $(3.57, 4.33)$ | 4 | $(3.42, 4.16)$ |

Table 6.6: The maximum a posteriori (MAP) estimators and the 95% highest posterior density intervals (95% HPD) obtained simultaneously using the weighted normal likelihood (QR-WN), the weighted generalised normal likelihood (QR-WGN) and the weighted mixture of skewed Laplace and skewed normal likelihood functions (QR-WM).

of some maximum a posteriori estimators especially for the intercept coefficients corresponding to the 0.5 and 0.75 quantiles. This implies that the approximate distribution based on the weighted normal density could not accommodate the true distribution of the data well. However, despite some agreement between the proposed methods in terms of the maximum a posteriori estimators, the proposed methods suggest different 95% highest posterior density intervals for quantile regression coefficients.

According to Figure 6.10 and Table 6.7, the bootstrap test suggests that the Bayesian quantile methods based on the weighted generalised normal likelihood and the weighted mixture of skewed Laplace and skewed normal likelihood functions succeed in estimating all quantiles of interest, while the method based on the weighted normal likelihood fails to estimate the median. This suggests that the flexibility of the function $g$ to accommodate the conditional distribution of the response variable can lead to a more accurate and reliable estimation of the quantile functions.

| $\tau$ | QR-WN | QR-WGN | QR-WM |
|---|---|---|---|
| 0.25 | 0.30 | 0.21 | 0.25 |
| 0.5 | 0.02 | 0.57 | 0.54 |
| 0.75 | 0.24 | 0.72 | 0.63 |

Table 6.7: The p-values corresponding to the proposed bootstrap test method for quantile functions estimated simultaneously using the weighted normal likelihood (QR-WN), the weighted generalised normal likelihood (QR-WGN) and the weighted mixture of skewed Laplace and skewed normal likelihoods (QR-WM).

## 6.8 Discussion

The proposed family of approximate likelihood functions can provide flexible and reliable alternatives to the asymmetric Laplace likelihood for developing Bayesian quantile methods to estimate quantile functions individually and simultaneously. Also, it can outperform the asymmetric Laplace likelihood in terms of Bayesian inference on quantile coefficients. It still performs well under the assumption of heteroscedasticity. Moreover, the proposed methods are easy to be implemented and interpreted from a Bayesian perspective.

The choice of the function $g$ can affect the approximation fundamentally. Therefore, the proposed family of approximate likelihood functions can be extended by considering a suitable form of the function $g$ that can handle complex models and draw reliable Bayesian inference on quantiles. In addition, the proposed method can be extended by considering the function $g$ to be a nonparametric kernel density.

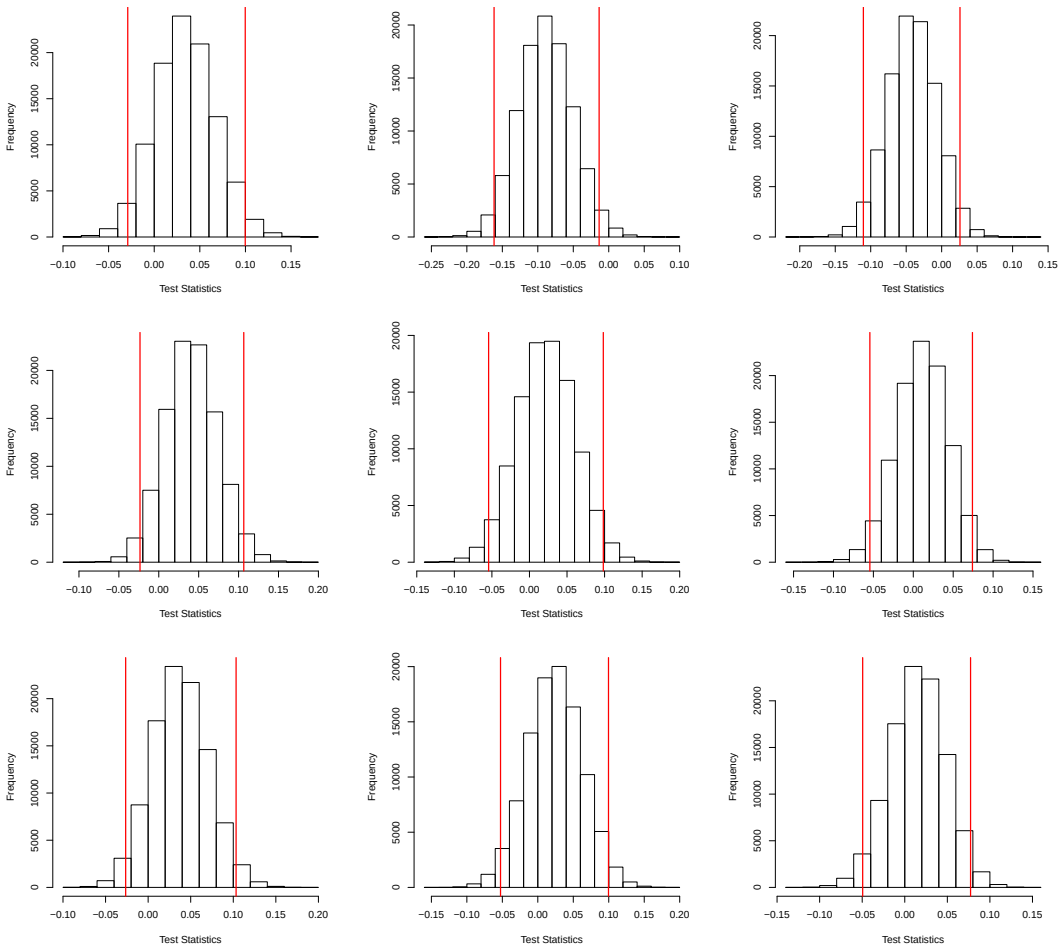



Figure 6.10: The histogram of test statistics, corresponding to the 0.25, 0.5 and 0.75 quantile curves (left to right) estimated simultaneously using the weighted normal likelihood, the weighted generalised normal likelihood and the weighted mixture of skewed Laplace and skewed normal likelihoods (top to bottom), over 100,000 bootstrap samples. The red lines represent 95% percentile confidence interval.

# Chapter 7

# Comparison and Conclusion

## 7.1 Summary

The aim of this research was to provide efficient methods to draw Bayesian inference for multiple quantile curves. This goal could be achieved by addressing three challenges. Firstly, the full specification of the residual distribution was required and this was the main interest. Secondly, it was necessary to specify suitable priors for all quantile regression coefficients. Finally, since it was impossible to draw samples from complex posterior distributions directly, efficient numerical sampling methods should be used to simulate from the posterior distributions. In addition, the issue of the crossing of quantile curves was taken into account while developing the Bayesian quantile models.

The Box-Cox transformation, which is a nonlinear monotonic transformation function including several known transformations as special cases, could be used to extend linear regression to deal with more complex nonlinear models. We employed the Box-Cox transformation in the context of Bayesian linear quantile regression based on the asymmetric Laplace likelihood function to deal with violations of linearity and homoscedasticity. Then, we proposed simultaneous Box-Cox quantile regression to estimate multiple quantile curves without crossing. To improve the efficiency of posterior computation, we proposed a Gibbs sampler with a Metropolis-Hastings step. Also, to overcome the limitations of this Box-Cox quantile regression, we considered the extensions to Box-Cox quantile regression with heteroscedastic errors and two-sided Box-Cox quantile regression.

Although the asymmetric Laplace likelihood has been used extensively in the literature to develop Bayesian quantile methods due to its direct link to the traditional loss function used to estimate quantiles, it has some limitations that can lead to a real concern about Bayesian inference drawn using this likelihood on quantiles' coefficients. Namely, these limitations are that skewness is controlled by the same parameter used to estimate the quantile of interest. Also, by using this likelihood, it is assumed that the mode of the underlying distribution is represented by the quantile of interest. To overcome this lack of flexibility to accommodate the distribution of the data, we proposed a new distribution called the generalised Gumbel distribution that showed a great ability to approximate the underlying distribution. Then, we employed this distribution to draw a Bayesian inference on quantile functions of linear models individually.

Simultaneous estimation of multiple quantile functions has been favoured over individual estimation given that it can be employed to overcome the issue of crossing quantile curves. In this project, we employed simultaneous estimation of quantiles to approximate the true distribution of the data. We formed the joint distribution of multiple quantile functions using a mixture of asymmetric Laplace distributions with suitable weights that can be fixed or estimated along with other model parameters. The proposed approach offered a great improvement in the accuracy and reliability of Bayesian inference on multiple linear quantile functions without crossing. This method has shown promising results and can be extended easily to handle more complex models.

To improve the quality of Bayesian inference on quantile regression, we proposed a family of approximate distributions that are parameterised by quantiles and maximized at the quantile of interest. In contrast to the asymmetric Laplace distribution, the proposed family of approximations does not assume that the mode of the data is represented by the quantile of interest. Also, it can be used to develop Bayesian methods to estimate the quantiles individually and simultaneously. This family of approximate likelihood functions was used to estimate multiple quantile functions without crossing for linear models with homoscedastic and heteroscedastic error. In addition, we proposed a new bootstrap test to check the lack of fit for quantile regression.

The proposed methods enjoy flexible assumptions and provide the Bayesian

framework for the analysis of quantile regression. Moreover, they can be implemented easily and can accommodate different types of regression models.

## 7.2 Comparison

As mentioned, a number of methods for handling quantile regression within a Bayesian context were explored in this thesis. In this section, we compare the different proposed Bayesian quantile methods based on the generalised Gumbel likelihood (GGUM), the pseudo asymmetric Laplace likelihood (PAL), the weighted pseudo asymmetric Laplace likelihood with fixed weights (WPAL$^{(1)}$) and with estimated weights (WPAL$^{(2)}$), and the weighted mixture of skewed Laplace and skewed normal distributions (QR-WM). We also compare them with the traditional Bayesian quantile regression based on the asymmetric Laplace (AL) likelihood. Tables 7.1 & 7.2 show the key differences between using the different approximate distributions as an error distribution for quantile regression. Moreover, we compare Bayesian inference, obtained using these methods, on quantile regression coefficients for Cats data described in Section 6.7.4. In addition, we use the bootstrap hypothesis test discussed in Section 6.6 to check the goodness of fit for these different Bayesian models.

Figure 7.1 shows how the different Bayesian quantile methods describe the conditional distribution of the response variable. It can be seen that there are differences in the estimates between the individual estimation and the simultaneous estimation of quantile functions. The individual estimation recommends that the characteristics of the conditional distribution change over the covariate. Although this behaviour can have a reasonable interpretation such as in the case of heteroscedasticity, it can be a result of the lack of information at some quantile levels, especially if the behaviour is not consistent over multiple quantiles. The simultaneous methods give similar estimation where they suggest that the conditional distribution is asymmetric with a longer upper tail. Table 7.3 illustrate the difference in the maximum a posteriori estimators obtained using the different methods. There are obvious differences especially between the individual and simultaneous approaches. To study and understand this variation, we apply the bootstrap test to examine the null hypothesis that the estimated functions can be defined as quantile functions for the true distribution of the data. Table 7.4

| AL | GGUM |
|---|---|
| • Its maximisation is equivalent to the minimisation of the loss function given in (1.2). | • Its maximum is equivalent to the approximate maximum of the underlying distribution. |
| • It provides poor approximation of the underlying distribution. | • It can give good approximation of the underlying distribution. |
| • The quality of the estimation depends on the quantile function of interest. | • The quality of the estimation tends to be the same for all quantiles of interest. |
| • The estimated quantile functions can cross. | • The estimated quantile functions can cross. |
| • The coverage probability could be very low for some quantiles of interest. | • The coverage probability approaches the nominal coverage probability for all quantiles. |
| • It has no flexible shape to accommodate different underlying distributions. | • It has flexible shape to accommodate different underlying distributions. |
| • It is assumed that the quantile of interest represents the mode of the data. | • It is not assumed that the quantile of interest represents the mode of the data. |

Table 7.1: Comparison between the use of the asymmetric Laplace (AL) and the generalised Gumbel distributions (GGUM) for estimating the quantiles individually.

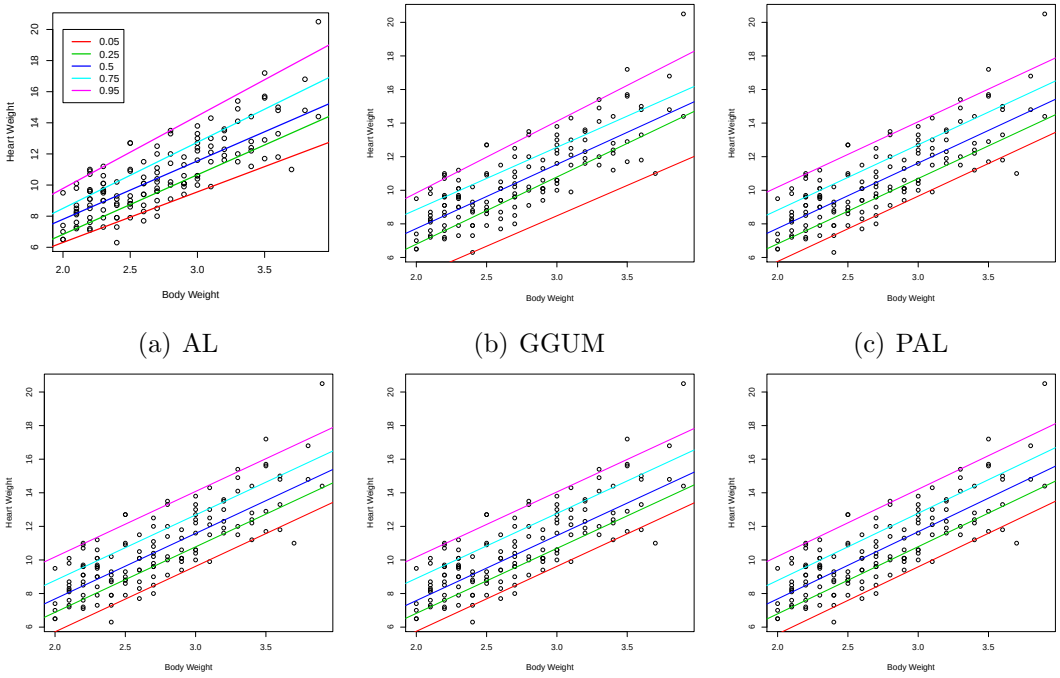
giving the p-values corresponding to quantile functions estimated using the different quantile approaches shows that the bootstrap test suggests that the quantile method based on the generalised Gumbel likelihood fails to estimate the true 0.05 quantile function where the null hypothesis is rejected with p-value, which obtained by formula (6.5), equal to zero. Also, Figure 7.1(b) shows that there are no observations under the 0.05 quantile line, which is unlikely.

Figures 7.2 & 7.3, which show the 95% highest posterior density (HPD) intervals for quantile coefficients, reveal the similarity in the estimated 95% HPD

| PAL | WPAL$^{(1)}$ & WPAL$^{(2)}$ | QR-WM |
|---|---|---|
| • Its maximisation is equivalent to the minimisation of the composite loss function given in (1.5). | • Its maximisation is approximately equivalent to the maximisation of the underlying distribution. | • Its maximisation is approximately equivalent to the maximisation of the underlying distribution. |
| • It gives poor approximation of the underlying distribution and the coverage probability could be very low. | • It gives good approximation of the underlying distribution and the coverage probability approaches the nominal level. | • It gives good approximation of the underlying distribution and the coverage probability approaches the nominal level. |
| • It can accommodate a solution for crossing of quantile functions. | • It can accommodate a solution for crossing of quantile functions. | • It can accommodate a solution for crossing of quantile functions. |
| • The quality of approximation is affected by the number and locations of quantiles included in the estimation. | • The quality of approximation is affected by the number and locations of quantiles included in the estimation. | • The quality of approximation can be robust against the number and locations of quantiles included in the estimation. |
| • The approximation become worse as the number of quantiles included in the estimation increases. | • The approximation improve as the number of quantiles included in the estimation increases. | • The approximation can be stable as the number of quantiles included in the estimation increases. |

Table 7.2: Comparison between the use of the pseudo asymmetric Laplace (PAL) distribution, the weighted pseudo asymmetric Laplace distribution with equal fixed weights (WPAL$^{(1)}$) and with estimated weights (WPAL$^{(2)}$) and the weighted mixture of skewed Laplace and skewed normal distributions (QR-WM) for estimating the quantiles simultaneously.

intervals between the approach based on the weighted pseudo asymmetric Laplace likelihood with fixed weights and estimated weights. For all quantiles, these figures show that the method based on the pseudo asymmetric Laplace likelihood




(a) AL (b) GGUM (c) PAL

(d) WPAL$^{(1)}$ (e) WPAL$^{(2)}$ (f) QR-WM

Figure 7.1: The colored lines represent the estimated quantile function, based on the maximum a posteriori estimators, corresponding to $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$.

| Individual estimation | | | Simultaneous estimation | | | | |
|---|---|---|---|---|---|---|---|
| Coefficients | AL | GGUM | Coefficients | PAL | WPAL$^{(1)}$ | WPAL$^{(2)}$ | QR-WM |
| $\alpha_{0.05}$ | -0.19 | -2.38 | $\alpha_{0.05}$ | -2.02 | -2.08 | -1.99 | -2.41 |
| $\alpha_{0.25}$ | -0.83 | -1.21 | $\alpha_{0.05}$ | -0.97 | -0.92 | -0.92 | -1.21 |
| $\alpha_{0.5}$ | 0.30 | 0.08 | $\alpha_{0.05}$ | -0.05 | -0.12 | -0.16 | -0.32 |
| $\alpha_{0.75}$ | 0.00 | 1.43 | $\alpha_{0.05}$ | 1.04 | 0.98 | 1.16 | 0.78 |
| $\alpha_{0.95}$ | 0.46 | 1.30 | $\alpha_{0.05}$ | 2.41 | 2.38 | 2.44 | 2.21 |
| $\beta_{0.05}$ | 3.25 | 3.62 | | | | | |
| $\beta_{0.25}$ | 3.83 | 4.00 | | | | | |
| $\beta_{0.5}$ | 3.75 | 3.82 | $\beta$ | 3.89 | 3.90 | 3.87 | 4 |
| $\beta_{0.75}$ | 4.25 | 3.71 | | | | | |
| $\beta_{0.95}$ | 4.66 | 4.27 | | | | | |

Table 7.3: The maximum a posteriori estimators obtained using different quantile methods

| $\tau$ | AL | GGUM | PAL | WPAL$^{(1)}$ | WPAL$^{(2)}$ | QR-WM |
|---|---|---|---|---|---|---|
| 0.05 | 0.32 | 0.00 | 0.64 | 0.79 | 0.78 | 0.98 |
| 0.25 | 0.77 | 0.26 | 0.60 | 0.20 | 0.63 | 0.26 |
| 0.5 | 0.72 | 0.87 | 0.66 | 0.84 | 0.66 | 0.61 |
| 0.75 | 0.74 | 0.57 | 0.93 | 0.91 | 0.62 | 0.72 |
| 0.95 | 0.83 | 0.39 | 0.73 | 0.72 | 0.61 | 0.83 |

Table 7.4: The p-value corresponding to the bootstrap hypothesis test for each level of quantiles.

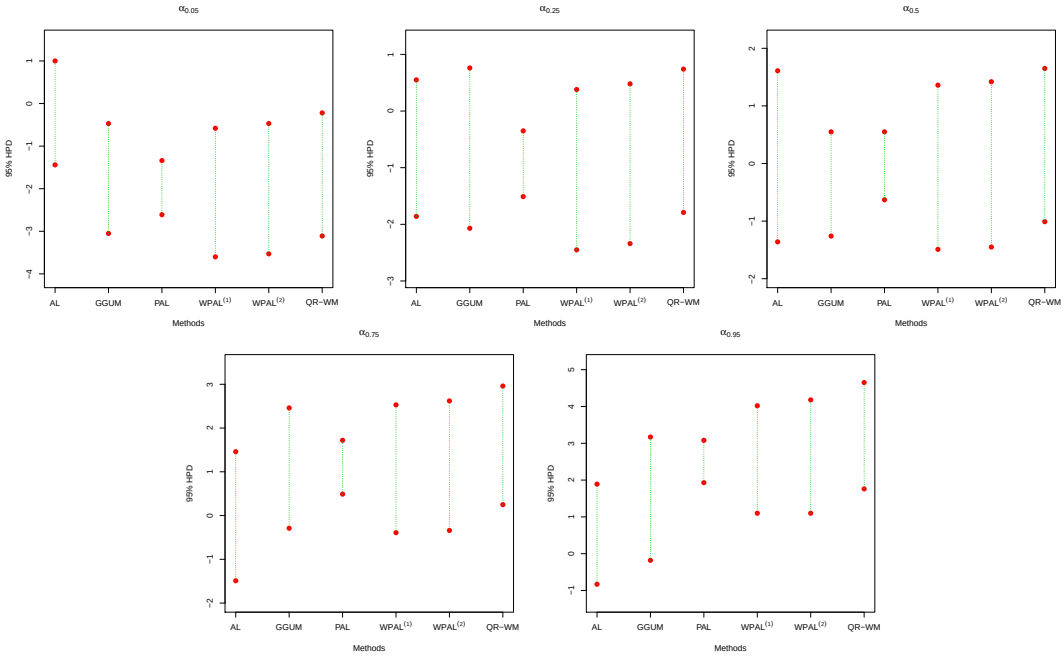



Figure 7.2: The 95% highest posterior density intervals for intercept coefficients are represented by the colored lines.

provides shorter 95% HPD intervals which are nested in other 95% HPD intervals, except those obtained by the approach based on the asymmetric Laplace likelihood. In addition, they show great intersections between the 95% HPD intervals suggested by the different approaches. However, these intersections can be explained better by Figure 7.4 which shows joint 95% highest posterior density regions of intercept and slope coefficients estimated using the method described by Bolker (2016) and Bolker (2008). It can be seen that 95% HPD regions obtained

using the quantile method based on the pseudo asymmetric Laplace likelihood represent the smallest credible regions shared by all simultaneous approaches and included all maximum a posteriori estimators suggested by these methods. This implies that the differences in the maximum a posteriori estimators, obtained using the simultaneous methods and shown in Table 7.3, are not quite significant. For the 0.25 and 0.5 quantiles, the individual estimation methods tend to agree with simultaneous estimation methods regarding the maximum a posteriori estimators and sharing the HPD region including all of these estimators. Moreover, all methods share the smallest 95% HPD region, including all maximum a posteriori estimators, obtained using the pseudo asymmetric Laplace likelihood for the 0.25 quantile coefficients. The reason behind this type of agreement is that the 0.25 and 0.5 quantiles are close to the mode of the data where the individual estimation methods can provide better approximation for the underlying distribution as discussed in Chapter 4. To illustrate the mode of the conditional distribution of the response variable, the conditional kernel density, estimated using the local polynomial estimation described by Hyndman (2013) and Hyndman *et al.* (1996), is shown in Figure 7.5.

## 7.3 Future work

In this research, much of the effort was on constructing likelihood functions that can be used to obtain accurate and reliable Bayesian inference on quantile curves. However, there are some research areas in Bayesian quantile analysis that need more attention. In the following section, we highlight some potential research gaps in the context of quantile regression that can be addressed in future.

### 7.3.1 The asymptotic properties of Bayesian estimators

A number of new Bayesian estimators were proposed in this thesis. Therefore, in order to extend this work to deal with more complex models, we will investigate some asymptotic prosperities of these estimators under regular and irregular conditions. In particular, the effect of the discontinuity on the consistency and the bias of the Bayesian estimators based on the likelihood functions considered in Chapters 4 & 6 will be examined. Also, we will derive these asymptotic properties under different priors.

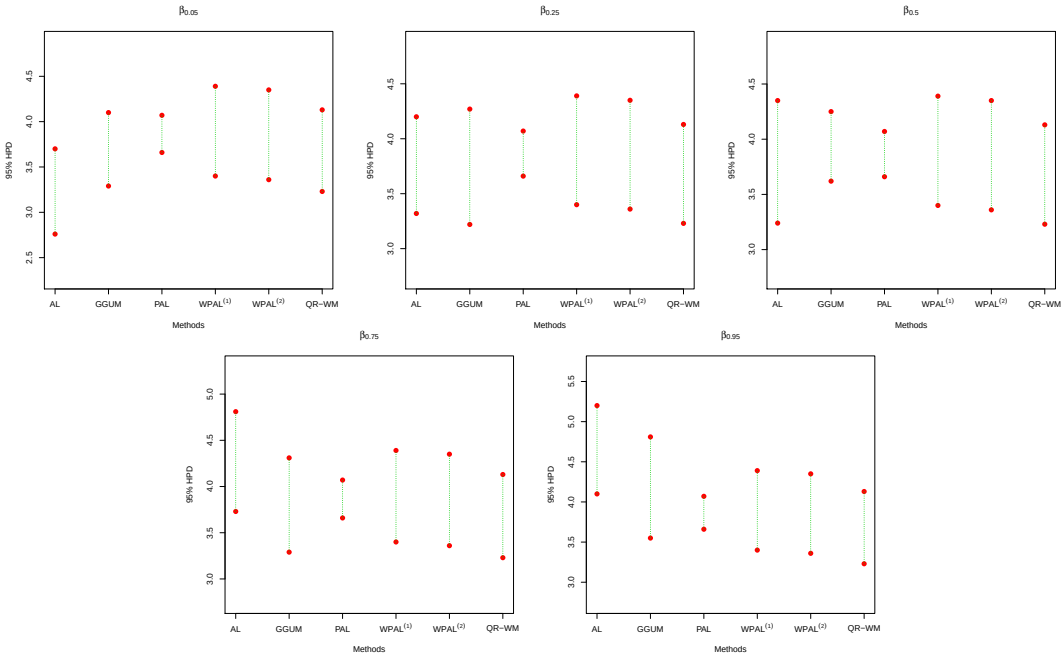



Figure 7.3: The 95% highest posterior density intervals for slope coefficients are represented by the colored lines.

### 7.3.2 Simultaneous non-linear quantile regression

The Bayesian linear quantile method based on weighted pseudo asymmetric Laplace likelihood proposed in Chapter 5 can be easily extended to polynomial and spline quantile regression. In addition, the simultaneous quantile methods based on the weighted pseudo asymmetric Laplace likelihood can be employed to handle Bayesian model selection rather than using an individual quantile, which can be subjective to the chosen quantile.

### 7.3.3 Goodness of fit

The work on goodness of fit measures for quantile regression in literature is limited in the sense they depend on asymptotic theories considered in classical statistics, or they are based on assuming the asymmetric Laplace distribution as an error distribution. Therefore, these methods require improvement in order to achieve an accurate investigation and reliable examination of the obtained inferences on quantile functions. To achieve this aim, we may need to adapt the

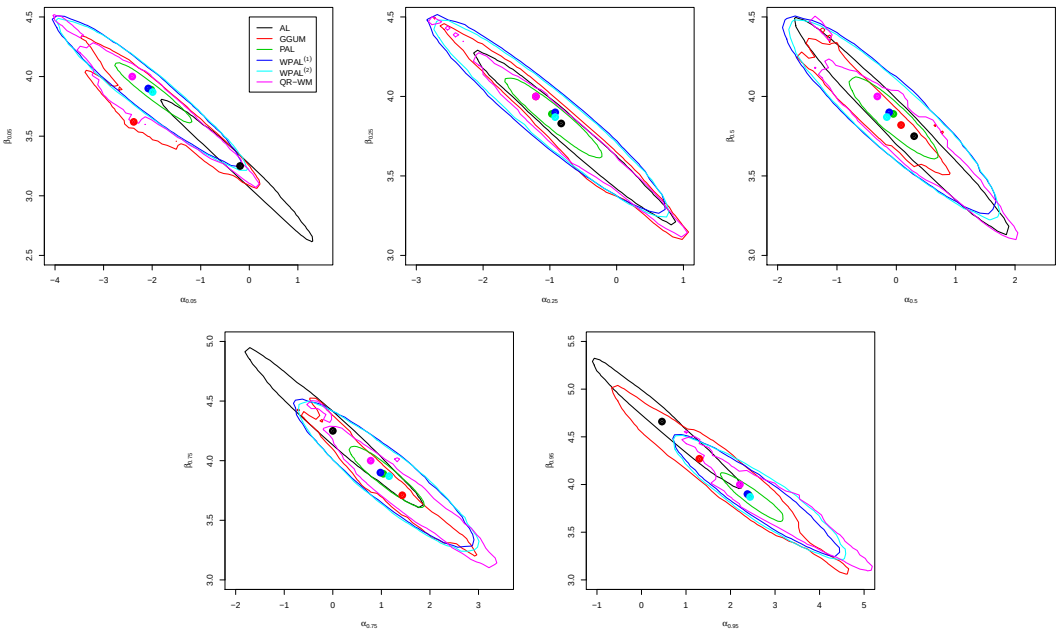



Figure 7.4: The joint 95% highest posterior density regions of the intercept and slope coefficients of quantile regression.

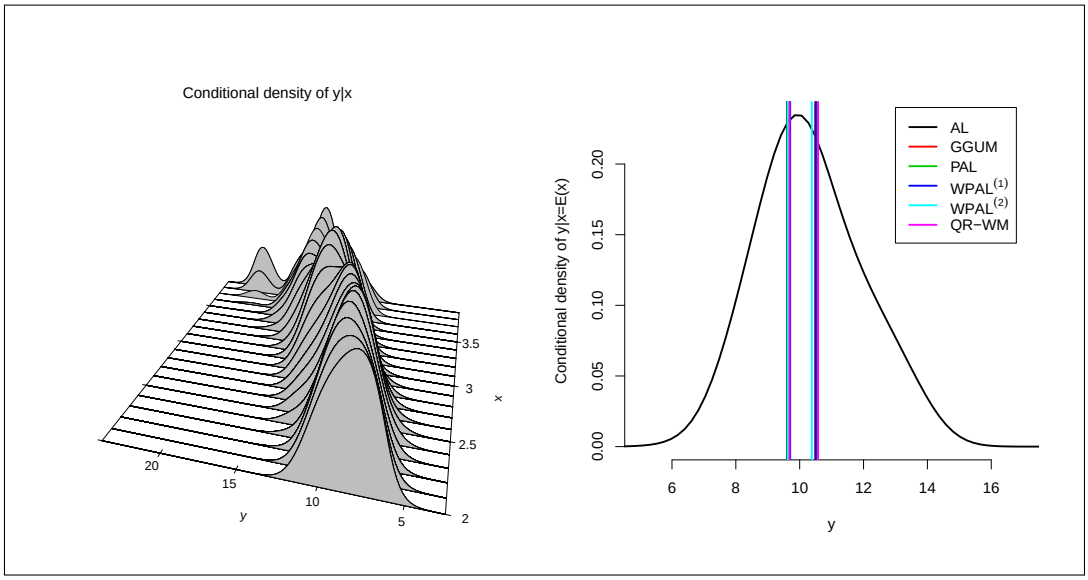



Figure 7.5: The conditional distribution of $y$=Heart Weight given $x$=Body Weight. The colored lines represent the estimated 0.25 and 0.5 quantiles.

goodness of fit test based on posterior predictive p-values – that are discussed, for example, by Meng (1994) – in the context of quantile regression. Moreover, although the Bayesian model selection based on Bayes factor is considered straightforward, its practical implementation may need carefully specified priors and efficient posterior computation. Therefore, more research on developing the methods of Bayesian model selection for quantile regression is required.

### 7.3.4 Prior distribution

Informative priors play a significant role in Bayesian analysis, especially in real applied research where using the informative priors available about the investigated problem can be useful to obtain more reliable Bayesian inference. Also, informative priors can be required to implement model selection using Bayes factors. Applying informative priors can improve the stability of the posterior estimation and convergence. However, to be able to employ informative prior beliefs about the model parameters effectively, we need to consider suitable theoretical formulas that can accommodate these beliefs easily and simplify the posterior form. To achieve this in the context of the simultaneous quantile estimation, we may need to consider the joint prior distribution for all quantile coefficients.

### 7.3.5 Prediction

Bayesian prediction has the advantage that it takes into account the uncertainty caused by the parameter being unknown. However, predictive inference obtained using the Bayesian quantile model based on the asymmetric Laplace likelihood may be misleading, because the asymmetric Laplace likelihood is unrelated to the distribution generating data. The proposed likelihood functions in this project can offer solutions for this issue, since they correspond to the underlying distribution of data. Moreover, as the proposed methods show a great ability to accommodate a variety of error distributions, this should improve the prediction of the distribution of new unobserved data points. Then, we can carry out predictive inference and decisions about future observables.

# Appendix A

## A.1 The link between the pseudo AL likelihood and the weighted optimisation problem.

The weighted optimisation problem to estimate the quantile functions can be written as

$$\widehat{Q}\left(\boldsymbol{\tau}|X\right) = \min_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} \sum_{k=1}^{m} \sum_{i=1}^{n} w_{\tau_k} \rho_{\tau_k} \left( \Lambda\left(y_i; \lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta}_{\tau_k} \right)$$

where $\sum_{k=1}^{m} w_{\tau_k} = 1$. By assuming that $\sigma_{\tau_k}$ is the scale parameter corresponding to the $\tau_k$ quantile function, we can write

$$\widehat{Q}\left(\boldsymbol{\tau}|X\right) = \min_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{\sigma_{\tau_k}} \rho_{\tau_k} \left( \Lambda\left(y_i; \lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta}_{\tau_k} \right),$$

since, for a constant $c > 0$, $\frac{1}{\sigma_{\tau_k}} = c w_{\tau_k}$.

$$\begin{aligned}
\widehat{Q}\left(\boldsymbol{\tau}|X\right) &= \max_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} - \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{\sigma_{\tau_k}} \rho_{\tau_k} \left( \Lambda\left(y_i; \lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta}_{\tau_k} \right) \\
&= \max_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} \exp \left\{ - \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{\sigma_{\tau_k}} \rho_{\tau_k} \left( \Lambda\left(y_i; \lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta}_{\tau_k} \right) \right\} \\
&= \max_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} \left( \prod_{k=1}^{m} \sigma_{\tau_k}^{-n} \right) \exp \left\{ - \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{\sigma_{\tau_k}} \rho_{\tau_k} \left( \Lambda\left(y_i; \lambda_{\tau_k}\right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta}_{\tau_k} \right) \right\},
\end{aligned}$$

since $\prod_{k=1}^{m} \sigma_{\tau_k}^{-n}$ is a constant with respect to $\boldsymbol{\alpha_\tau}$, $B_{\boldsymbol{\tau}}$ and $\boldsymbol{\lambda_\tau}$.

$$= \max_{\boldsymbol{\alpha_\tau}, B_{\boldsymbol{\tau}}, \boldsymbol{\lambda_\tau} \in R} \left( \prod_{k=1}^{m} \sigma_{\tau_k}^{-n} \right) \exp \left\{ - \sum_{k=1}^{m} \sum_{i=1}^{n} \rho_{\tau_k} \left( \frac{\Lambda \left( y_i; \lambda_{\tau_k} \right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta}_{\tau_k}}{\sigma_{\tau_k}} \right) \right\}$$

# A.2 Conditional distributions for parameters of Bayesian Box-Cox quantile model

In following sections, we derive the conditional distributions for parameters of simultaneous Bayesian Box-Cox quantile model (PAL-BC). The pseudo posterior distribution is given by

$$\begin{aligned} p \left( \boldsymbol{\alpha}_\tau, \boldsymbol{\beta}, W_\tau, \boldsymbol{\sigma}_\tau, \lambda | \boldsymbol{y} \right) \propto & \prod_{i=1}^{m} \prod_{i=1}^{n} \sigma_{\tau_k}^{-\frac{1}{2}} w_{\tau_k, i}^{-\frac{1}{2}} \exp \left\{ - \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{\left( \Lambda \left( y_i; \lambda \right) - \alpha_{\tau_k} - \boldsymbol{x}_i^{'} \boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k, i} \right)^2}{2 \sigma_{\tau_k} \phi_{\tau_k} w_{\tau_k, i}} \right\} \\ & \times \prod_{k=1}^{m} \sigma_{\tau_k}^{-\left( a_{\tau_k} + 1 \right)} \exp \left\{ - \sum_{k=1}^{m} \frac{b_{\tau_k}}{\sigma_{\tau_k}} \right\} \exp \left\{ - \frac{1}{2} \left( \boldsymbol{\beta} - \boldsymbol{\beta}^* \right)^{'} S^{-1} \left( \boldsymbol{\beta} - \boldsymbol{\beta}^* \right) \right\} \\ & \times \exp \left\{ - \frac{\left( \lambda - \lambda^* \right)^2}{2 v^2} \right\} \exp \left\{ - \sum_{k=1}^{m} \frac{\left( \alpha_{\tau_k} - \alpha_{\tau_k}^* \right)^2}{2 c_{\tau_k}^2} \right\} 1_{\left( \alpha_{\tau_1} < \alpha_{\tau_2} < \ldots < \alpha_{\tau_m} \right)}. \end{aligned}$$

## A.2.1 Conditional Distribution of $\alpha_{\tau_k}$

We assume that

$$\widetilde{\alpha}_{\tau_k} = \psi_{\tau_k}^2 \left( \frac{\alpha_{\tau_k}^*}{c_{\tau_k}^2} + \sum_{i=1}^{n} \frac{\left( \Lambda(y_i; \lambda) - \boldsymbol{x}_i^{'} \boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k, i} \right)}{\sigma_{\tau_k} \phi_{\tau_k} w_{\tau_k, i}} \right), \quad \psi_{\tau_k}^2 = \left( \frac{1}{c_{\tau_k}^2} + \sum_{i=1}^{n} \frac{1}{\sigma_{\tau_k} \phi_{\tau_k} w_{\tau_k, i}} \right)^{-1}$$

Hence, for $k = 1, 2, ..., m-1$, the conditional distribution of $\alpha_{\tau_k}$ given by

$$\begin{aligned} p \left( \alpha_{\tau_k} | \boldsymbol{\beta}, \boldsymbol{w}_{\tau_k}, \boldsymbol{\sigma}_\tau, \lambda \right) \propto & \exp \left\{ - \frac{1}{2} \left[ \frac{\alpha_{\tau_k}^2}{\psi_{\tau_k}^2} - 2 \frac{\alpha_{\tau_k} \widetilde{\alpha}_{\tau_k}}{\psi_{\tau_k}^2} + \sum_{i=1}^{n} \frac{\left( \Lambda \left( y_i; \lambda \right) - \boldsymbol{x}_i^{'} \boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k, i} \right)^2}{\sigma_{\tau_k} \phi_{\tau_k} w_{\tau_k, i}} \right. \right. \\ & \left. \left. + \frac{\alpha_{\tau_k}^{*2}}{c_{\tau_k}^2} \right] \right\} 1_{\left( \alpha_{\tau_k} < \alpha_{\tau_{k+1}} \right)} \\ \propto & \exp \left\{ - \frac{1}{2} \left[ \frac{\alpha_{\tau_k}^2}{\psi_{\tau_k}^2} - 2 \frac{\alpha_{\tau_k} \widetilde{\alpha}_{\tau_k}}{\psi_{\tau_k}^2} + \frac{\widetilde{\alpha}_{\tau_k}^2}{\psi_{\tau_k}^2} - \frac{\widetilde{\alpha}_{\tau_k}^2}{\psi_{\tau_k}^2} \right. \right. \\ & \left. \left. + \sum_{i=1}^{n} \frac{\left( \Lambda \left( y_i; \lambda \right) - \boldsymbol{x}_i^{'} \boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k, i} \right)^2}{\sigma_{\tau_k} \phi_{\tau_k} w_{\tau_k, i}} + \frac{\alpha_{\tau_k}^{*2}}{c_{\tau_k}^2} \right] \right\} 1_{\left( \alpha_{\tau_k} < \alpha_{\tau_{k+1}} \right)} \end{aligned}$$

$$
\begin{aligned}
p\left(\alpha_{\tau_k}|\boldsymbol{\beta}_\tau, \boldsymbol{w}_{\tau_k}, \sigma_{\tau_k}, \lambda\right) \quad \propto \quad & \left[\exp\left\{-\frac{\left(\alpha_{\tau_k}-\widetilde{\alpha}_{\tau_k}\right)^2}{2\psi_{\tau_k}^2}\right\} + \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{\left(\Lambda\left(y_i;\lambda\right)-\boldsymbol{x}_i^{'}\boldsymbol{\beta}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\right.\right. \\
& \left.\left. - \frac{\alpha_{\tau_k}^{*2}}{2c_{\tau_k}^2} + \frac{\widetilde{\alpha}_{\tau_k}^2}{2\psi_{\tau_k}^2}\right\}\right] 1_{\left(\alpha_{\tau_k}<\alpha_{\tau_{k+1}}\right)} \\
\propto \quad & \exp\left\{-\frac{\left(\alpha_{\tau_k}-\widetilde{\alpha}_{\tau_k}\right)^2}{2\psi_{\tau_k}^2}\right\} 1_{\left(\alpha_{\tau_k}<\alpha_{\tau_{k+1}}\right)}.
\end{aligned}
$$

Then, this can be written

$$
\alpha_{\tau_k}|\boldsymbol{\beta}_{\tau_k}, \boldsymbol{w}_{\tau_k}, \sigma_{\tau_k}, \lambda \sim \mathrm{N}\left(\widetilde{\alpha}_{\tau_k}, \psi_{\tau_k}^2\right) 1_{\left(\alpha_{\tau_k}<\alpha_{\tau_{k+1}}\right)}.
$$

### A.2.2 Conditional Distribution of $\beta$

The conditional distribution of $\boldsymbol{\beta}$ can be written as

$$
\begin{aligned}
p\left(\boldsymbol{\beta}|\alpha_\tau, W_\tau, \boldsymbol{\sigma}_\tau, \lambda\right) \quad \propto \quad & \exp\left\{-\sum_{i=1}^{n}\sum_{k=1}^{m}\left[\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i^{'}\boldsymbol{\beta}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{2\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\right]\right. \\
& \left. - \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)^{'} S^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)\right\} \\
\propto \quad & \exp\left\{-\sum_{i=1}^{n}\sum_{k=1}^{m}\left[\frac{1}{2\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\left(\boldsymbol{\beta}^{'}\boldsymbol{x}_i\boldsymbol{x}_i^{'}\boldsymbol{\beta} - 2\boldsymbol{x}_i^{'}\boldsymbol{\beta}\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)\right.\right.\right. \\
& \left.\left.\left. + \left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)^2\right)\right] - \frac{1}{2}\left[\boldsymbol{\beta}^{'}S^{-1}\boldsymbol{\beta} - 2\beta^{'}S^{-1}\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*'}S^{-1}\boldsymbol{\beta}^*\right]\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\sum_{k=1}^{m}\frac{1}{\sigma_{\tau_k}\phi w_{\tau_k,i}}\left(\boldsymbol{\beta}^{'}\boldsymbol{x}_i\boldsymbol{x}_i^{'}\boldsymbol{\beta} - 2\boldsymbol{x}_i^{'}\boldsymbol{\beta}\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)\right.\right.\right. \\
& \left.\left.\left. + \left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)^2\right) + \boldsymbol{\beta}^{'}S^{-1}\boldsymbol{\beta} - 2\beta^{'}S^{-1}\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*'}S^{-1}\boldsymbol{\beta}^*\right]\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^{'}\left(S^{-1}+\sum_{i=1}^{n}\sum_{k=1}^{m}\frac{\boldsymbol{x}_i\boldsymbol{x}_i^{'}}{\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\right)\boldsymbol{\beta} - \boldsymbol{\beta}^{'}\left(S^{-1}\boldsymbol{\beta}^*\right.\right.\right. \\
& \left. + \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{\boldsymbol{x}_i\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)}{\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\right) - \left(\boldsymbol{\beta}^{*'}S^{-1}\right. \\
& \left. + \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)\boldsymbol{x}_i^{'}}{\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\right)\boldsymbol{\beta} \\
& \left.\left. + \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}} + \boldsymbol{\beta}^{*'}S^{-1}\boldsymbol{\beta}^*\right]\right\}
\end{aligned}
$$

Now, we assume that

$$\widetilde{\boldsymbol{\beta}} = \Sigma \left( S^{-1}\boldsymbol{\beta}^* + \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\boldsymbol{x}_i\left(\Lambda(y_i;\lambda)-\alpha_{\tau_k}-\theta_\tau \boldsymbol{w}_{\tau_k,i}\right)}{\sigma_{\tau_k}\phi_{\tau_k}\boldsymbol{w}_{\tau_k,i}} \right), \quad \Sigma = \left( S^{-1} + \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\boldsymbol{x}_i\boldsymbol{x}_i'}{\sigma_{\tau_k}\phi_{\tau_k}\boldsymbol{w}_{\tau_k,i}} \right)^{-1}$$

Hence, the conditional distribution of $\boldsymbol{\beta}$

$$\begin{aligned} p\left(\boldsymbol{\beta}|\boldsymbol{\alpha}_\tau, W_\tau, \boldsymbol{\sigma}_\tau, \lambda\right) \quad \propto \quad & \exp\left\{ -\frac{1}{2}\left[ \boldsymbol{\beta}'\Sigma^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}'\Sigma^{-1}\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}'\Sigma^{-1}\boldsymbol{\beta} \right.\right. \\ & \left.\left. + \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} + \boldsymbol{\beta}^{*'} S^{-1}\boldsymbol{\beta}^* \right] \right\} \end{aligned}$$

$$\begin{aligned} p\left(\boldsymbol{\beta}|\boldsymbol{\alpha}_\tau, W_\tau, \boldsymbol{\sigma}_\tau, \lambda\right) \quad \propto \quad & \exp\left\{ -\frac{1}{2}\left[ \boldsymbol{\beta}'\Sigma^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}'\Sigma^{-1}\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}'\Sigma^{-1}\boldsymbol{\beta} + \widetilde{\boldsymbol{\beta}}'\Sigma^{-1}\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}'\Sigma^{-1}\widetilde{\boldsymbol{\beta}} \right.\right. \\ & \left.\left. + \sum_{i=1}^{n}\sum_{k=1}^{m} \frac{1}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \theta_{\tau_k} w_{\tau_k,i}\right)^2 + \boldsymbol{\beta}^{*'} S^{-1}\boldsymbol{\beta}^* \right] \right\} \\ \propto \quad & \exp\left\{ -\frac{1}{2}\left(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right)' \Sigma^{-1}\left(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right) \right\} \\ & \times \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{m} \frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} + \boldsymbol{\beta}^{*'} S^{-1}\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}}'\Sigma^{-1}\widetilde{\boldsymbol{\beta}} \right\} \\ \propto \quad & \exp\left\{ -\frac{1}{2}\left(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right)' \Sigma^{-1}\left(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\right) \right\} \end{aligned}$$

Then, this can be written

$$\boldsymbol{\beta}|\boldsymbol{\alpha}_\tau, W_\tau, \boldsymbol{\sigma}_\tau, \lambda \sim \mathrm{N}\left(\widetilde{\boldsymbol{\beta}}, \Sigma\right)$$

### A.2.3 Conditional Distribution of $w_{\tau_k,i}$

The conditional distribution of $w_{\tau_k,i}$ is given by

$$\begin{aligned} p\left(w_{\tau_k,i}|\alpha_{\tau_k}, \boldsymbol{\beta}, \sigma_{\tau_k}, \lambda\right) \quad \propto \quad & w_{\tau_k,i}^{-\frac{1}{2}} \exp\left\{ -\frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta_{\tau_k} w_{\tau_k,i}\right)^2}{2\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right\} \times \exp(-\frac{1}{\sigma} w_{\tau_k,i}) \\ \propto \quad & w_{\tau_k,i}^{-\frac{1}{2}} \exp\left\{ -\left[ \frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2 - 2\theta_{\tau_k} w_{\tau_k,i}\left(\Lambda\left(y_i\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}\right) + \theta_{\tau_k}^2 w_{\tau_k,i}^2}{2\sigma_{\tau_k}\phi_{\tau_k} w_{\tau_k,i}} \right] \right. \\ & \left. -\frac{1}{\sigma} w_{\tau_k,i} \right\} \\ \propto \quad & w_{\tau_k,i}^{-\frac{1}{2}} \exp\left\{ -\left[ \frac{\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2 w_{\tau_k,i}^{-1} - 2\theta_{\tau_k}\left(\Lambda\left(y_i;\lambda\right) - \alpha_{\tau_k} - \boldsymbol{x}_i'\boldsymbol{\beta}\right) + \theta_{\tau_k}^2 w_{\tau_k,i}}{2\sigma_{\tau_k}\phi_{\tau_k}} \right] \right. \end{aligned}$$

$$
\begin{aligned}
& -\frac{2w_{\tau_k,i}}{2\sigma}\Bigg\} \\
\propto\ & w_{\tau_k,i}^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[w_{\tau_k,i}^{-1}\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k}}-2\frac{\theta_{\tau_k}\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)}{\sigma_{\tau_k}\phi_{\tau_k}}\right]\right. \\
& \left.+w_{\tau_k,i}\left(\frac{\theta_{\tau_k}^2}{\sigma_{\tau_k}\phi_{\tau_k}}\right)\right]-\frac{2w_{\tau_k,i}}{2\sigma_{\tau_k}}\Bigg\} \\
\propto\ & w_{\tau_k,i}^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[w_{\tau_k,i}^{-1}\left(\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k}}\right)+\left(\sum_{k=1}^{m}\frac{\theta_{\tau_k}^2}{\sigma_{\tau_k}\phi_{\tau_k}}+\frac{2}{\sigma_{\tau_k}}\right)w_{\tau_k,i}\right]\right\}
\end{aligned}
$$

which the density of generalized inverse Gaussian distribution (GIG), then, it is can be written

$$
w_{\tau_k,i}|\alpha_{\tau_k},\boldsymbol{\beta},\lambda,\sigma\sim\text{GIG}\left(\frac{1}{2},\sqrt{\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{\sigma_{\tau_k}\phi_{\tau_k}}},\sqrt{\frac{\theta_{\tau_k}^2}{\sigma_{\tau_k}\phi_{\tau_k}}+\frac{2}{\sigma_{\tau_k}}}\right)
$$

where the density of generalized inverse Gaussian distribution (GIG) is given by

$$
f\left(x|t,a,b\right)=\frac{\left(b/a\right)^t}{2K_t\left(ab\right)}x^{t-1}\exp\left\{-\frac{1}{2}\left(a^2x^{-1}+b^2x\right)\right\}
$$

### A.2.4 Conditional Distribution of $\sigma_{\tau_k}$

The conditional distribution of $\sigma_{\tau_k}$ is given by

$$
\begin{aligned}
p\left(\sigma_{\tau_k}|\alpha_{\tau_k},\boldsymbol{\beta},\boldsymbol{w}_{\tau_k},\lambda\right) \propto\ & \sigma_{\tau_k}^{-\left(\frac{n}{2}+a_{\tau_k}+1\right)}\exp\left\{-\sum_{i=1}^{n}\left[\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{2\sigma_{\tau_k}\phi_{\tau_k}w_{\tau_k,i}}\right]-\frac{b_{\tau_k}}{\sigma_{\tau_k}}\right\} \\
& \times\left(\frac{1}{\sigma_{\tau_k}}\right)^n\exp\left(-\frac{1}{\sigma_{\tau_k}}\sum_{i=1}^{n}w_{\tau_k,i}\right) \\
\propto\ & \sigma^{-\left(\frac{3n}{2}+a_{\tau_k}+1\right)}\exp\left\{-\frac{1}{\sigma_{\tau_k}}\left[b_{\tau_k}+\sum_{i=1}^{n}\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{2\phi_{\tau_k}w_{\tau_k,i}}\right]\right. \\
& \left.-\frac{1}{\sigma_{\tau_k}}\sum_{i=1}^{n}w_{\tau_k,i}\right\} \\
\propto\ & \sigma_{\tau_k}^{-\left(\frac{3n}{2}+a_{\tau_k}+1\right)}\exp\left\{-\frac{1}{\sigma_{\tau_k}}\left[b_{\tau_k}+\sum_{i=1}^{n}\frac{\left(\Lambda\left(y_i;\lambda\right)-\alpha_{\tau_k}-\boldsymbol{x}_i'\boldsymbol{\beta}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{2\phi_{\tau_k}w_{\tau_k,i}}\right.\right.
\end{aligned}
$$

$$+\sum_{i=1}^{n} w_{\tau_k,i}\Bigg]\Bigg\}$$

Thus

$$\sigma_{\tau_k}|\alpha_{\tau_k},\boldsymbol{\beta},\boldsymbol{w}_{\tau_k},\lambda \sim \text{IG}\left(\frac{3n}{2}+a_{\tau_k},b_{\tau_k}+\sum_{i=1}^{n}\frac{\left(\Lambda\left(y_i;\lambda\right)-\boldsymbol{x}_i'\boldsymbol{\beta}-\theta_{\tau_k}w_{\tau_k,i}\right)^2}{2\phi_{\tau_k}w_{\tau_k,i}}+\sum_{i=1}^{n}w_{\tau_k,i}\right)$$

where IG is the inverse gamma distribution.

# Appendix B

## B.1 Markov chain Monte Carlo (MCMC)

All MCMC methods used in this thesis to simulate from the posterior distributions are diagnosed carefully to achieve the best results. In this section, as an example, we diagnose and compare Metropolis-Hastings algorithm and Gibbs sampler used in Chapter 2 to simulate from the posterior distributions. To do this, we use the simulated data from the model given in (2.12). To make sure that MCMC converges to the target distribution properly, we implement Metropolis-Hastings algorithm with 100000 iterations and a burn-in of 20000 iterations. We seek for an acceptance probability of approximately 0.234 (for more details, see Roberts *et al.*, 1997). Also, we run the Gibbs samplers with 10000 iterations and a burn-in of 1000 iterations. To see how MCMC move around the model parameter space and the chain mixing, we can look at the trace plots and in this case, we need to check each parameter. Figures B.1, B.2, B.3 and B.4 show the good mixing of all chains for both the sampling methods. We apply Geweke's convergence test (for more details, see Plummer *et al.*, 2006), to compare the mean of the first 30% and the last 30% of each chain. Geweke's statistic is given by the difference between the means of the two parts of the Markov chain divided by its estimated standard error and asymptotically follows the standard normal distribution. At the significance level of 0.05, Table B.1 suggests that all chains converge to the stationary distributions. Also, we use Gelman and Rubins convergence diagnostic to monitor convergence of five parallel chains run with overdispersed starting values (for more details, see Plummer *et al.*, 2006). Table B.2, which show the estimates of the potential scale reduction factor, indicates that all Markov chains are converged. By looking at Figures B.5, B.6, B.7 and B.8 that shows the auto-

correlation of the simulated samples from the posterior distribution. It is obvious that the samples generated using Metropolis-Hastings algorithm have higher autocorrelation than those obtained using the Gibbs sampler. Therefore, we use the thinning method to handle this issue. We thin the samples by keeping every $50^{\text{th}}$ simulated value obtain using Metropolis-Hastings algorithm and $6^{\text{th}}$ simulated value obtained using the Gibbs sampler. The the results are given in Figures B.9, B.10, B.11 and B.12. However, there are no considerable changes in the summary Bayesian statistics such as the posterior mean, the posterior standard deviation and highest posterior density (HPD) intervals. From Table B.3, the both sampling approaches approximately share the same posterior means and 95% HPD intervals for all model parameters. These two methods do not require a computational power and much time to be implemented. However, Metropolis-Hastings algorithm is faster than the Gibbs sampler. All R codes are available to be ordered from the author and will be available publicly along with publications related to this research.

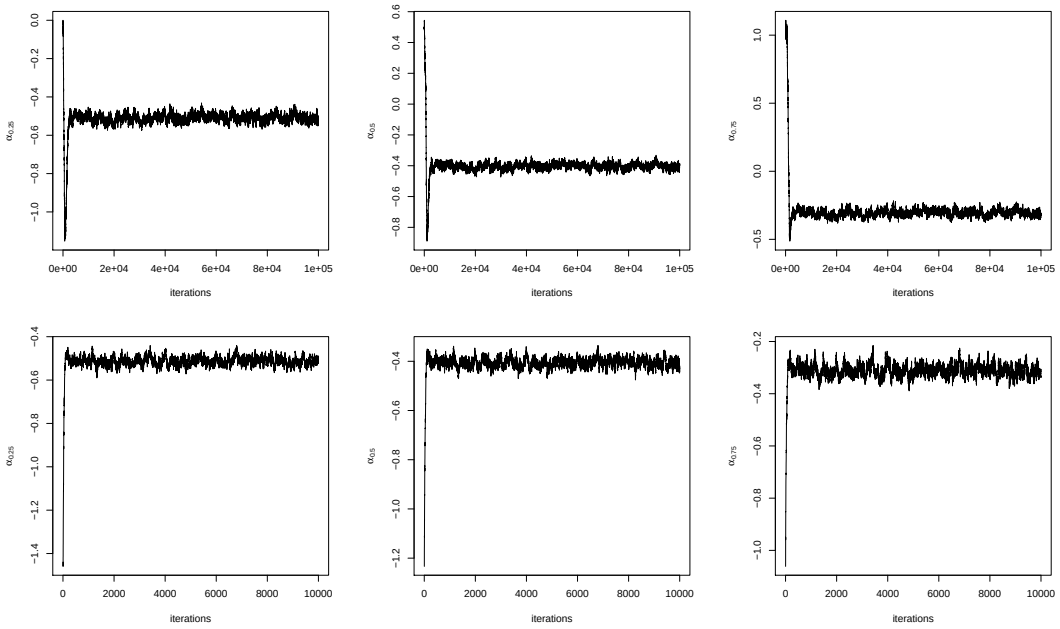



Figure B.1: The trace plots for the samples of the intercept coefficients obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom).

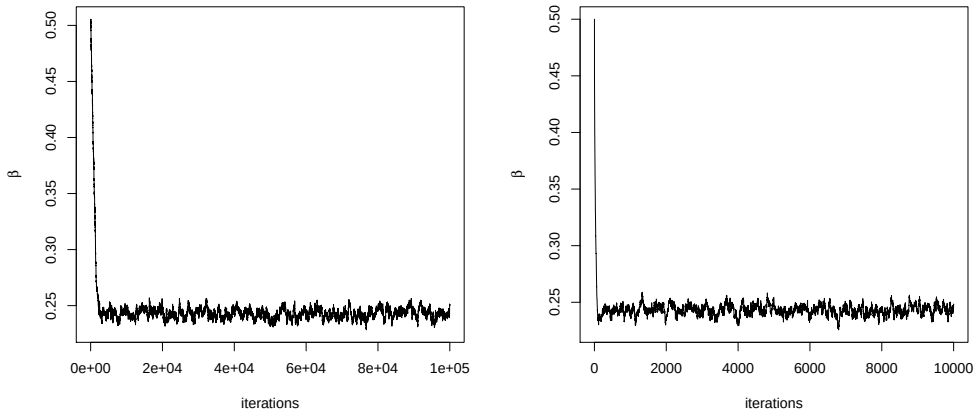



Figure B.2: The trace plots for the samples of the slope coefficient obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right).

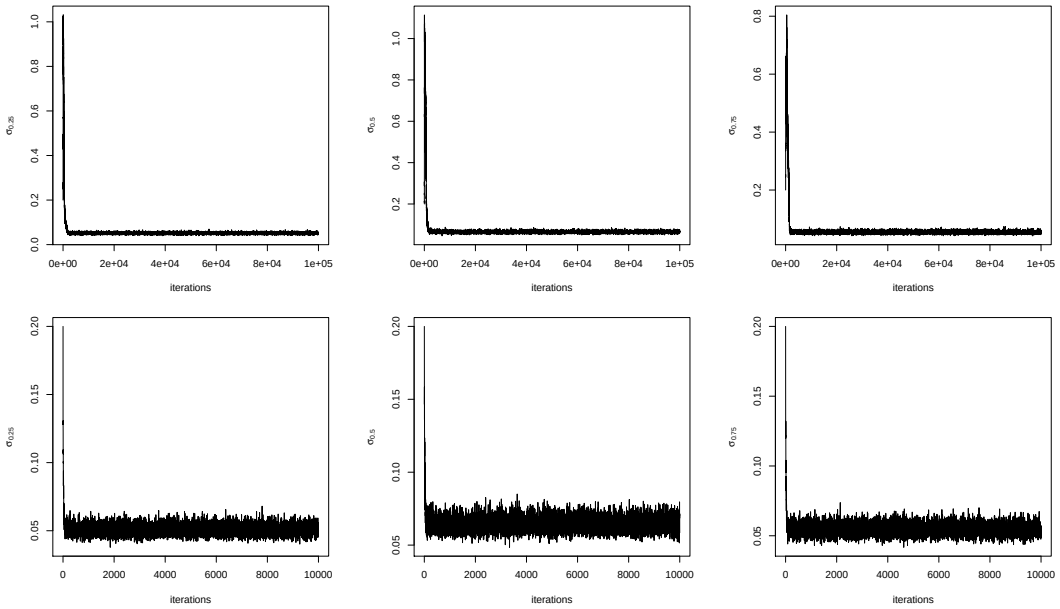



Figure B.3: The trace plots for the samples of the scale parameters obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom).

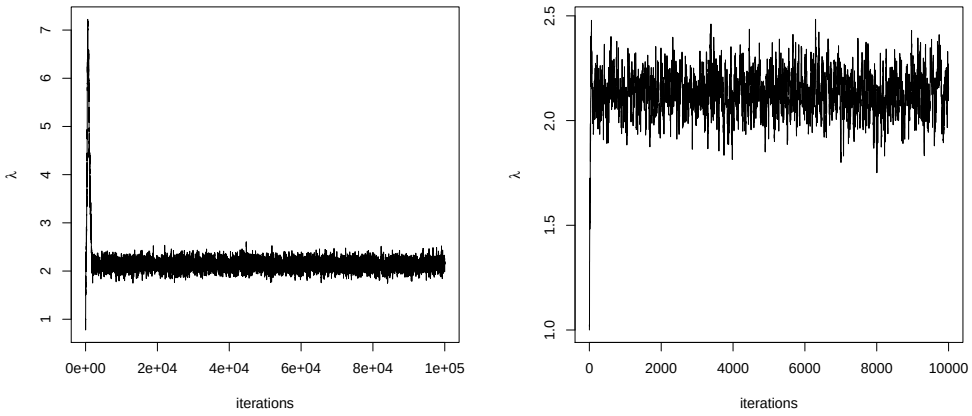



Figure B.4: The trace plots for the samples of the transformation parameter obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right).

| Coefficients | Metropolis-Hastings | Gibbs |
|---|---|---|
| $\alpha_{0.25}$ | 0.91 | 0.46 |
| $\alpha_{0.5}$ | 0.89 | 0.86 |
| $\alpha_{0.75}$ | 0.89 | 0.95 |
| $\beta$ | 0.88 | 0.69 |
| $\lambda$ | 0.24 | 0.54 |
| $\sigma_{0.25}$ | 0.42 | 0.63 |
| $\sigma_{0.5}$ | 0.36 | 0.92 |
| $\sigma_{0.75}$ | 0.13 | 0.31 |

Table B.1: The *p-value* corresponding to Geweke's convergence test.

| | Metropolis-Hastings | | Gibbs | |
|---|---|---|---|---|
| Coefficients | P.E. | U.C.L. | P.E. | U.C.L. |
| $\alpha_{0.25}$ | 1.01 | 1.02 | 1.01 | 1.02 |
| $\alpha_{0.5}$ | 1.01 | 1.02 | 1.01 | 1.02 |
| $\alpha_{0.75}$ | 1.01 | 1.02 | 1.01 | 1.02 |
| $\beta$ | 1.01 | 1.03 | 1.01 | 1.03 |
| $\lambda$ | 1.00 | 1.00 | 1.01 | 1.02 |
| $\sigma_{0.25}$ | 1.00 | 1.00 | 1.00 | 1.00 |
| $\sigma_{0.5}$ | 1.00 | 1.00 | 1.00 | 1.00 |
| $\sigma_{0.75}$ | 1.00 | 1.00 | 1.00 | 1.00 |
| M.S.R. | 1.01 | | 1.01 | |

Table B.2: The point estimates (P.E.), the upper confidence limit (U.C.L.) and the multivariate potential scale reduction factor (M.S.R.) obtained using Gelman and Rubin's convergence diagnostic for chains corresponding to the two proposed sampling methods.

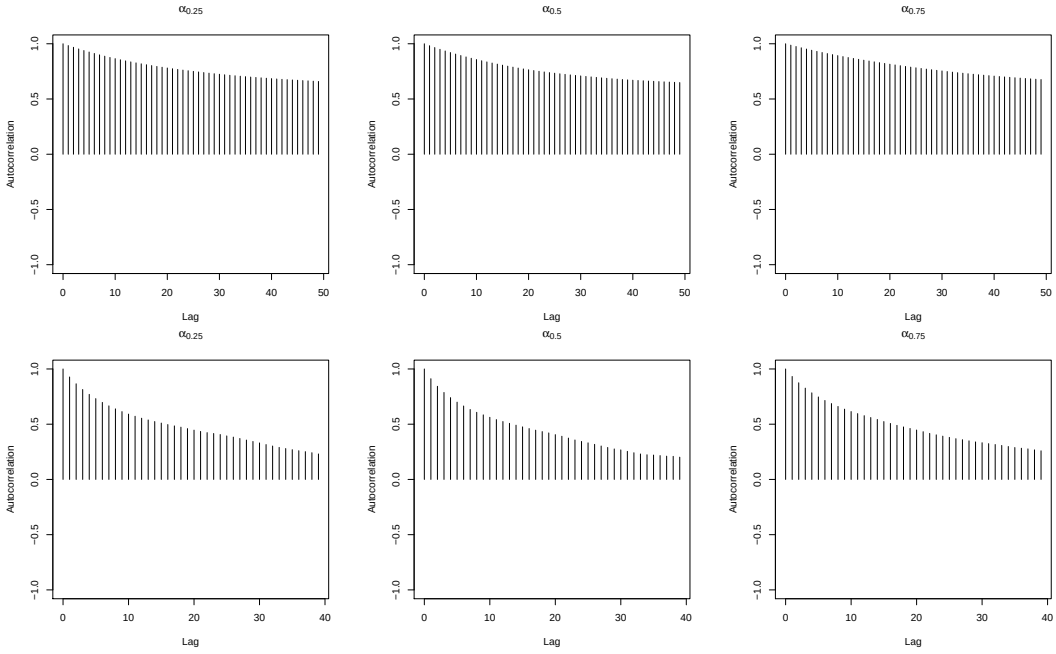



Figure B.5: The autocorrelation plot for the samples of the intercept coefficients obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom).

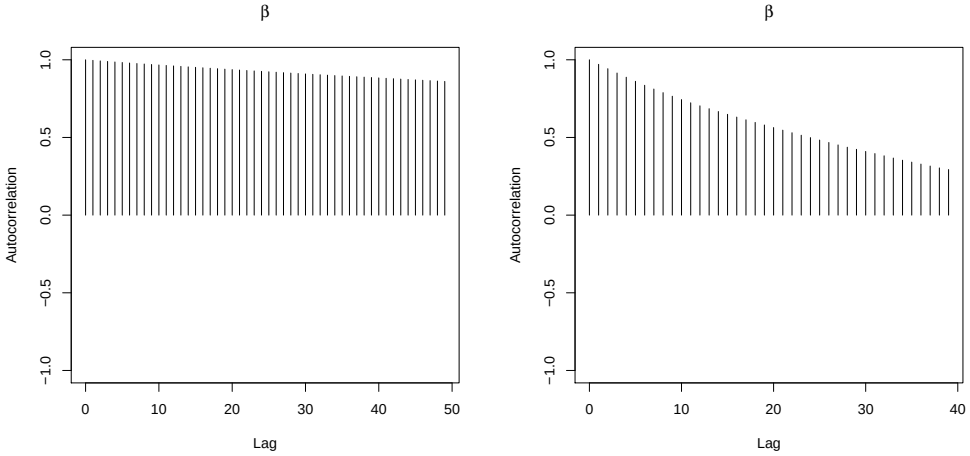



Figure B.6: The autocorrelation plot for the samples of the slope coefficient obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right).

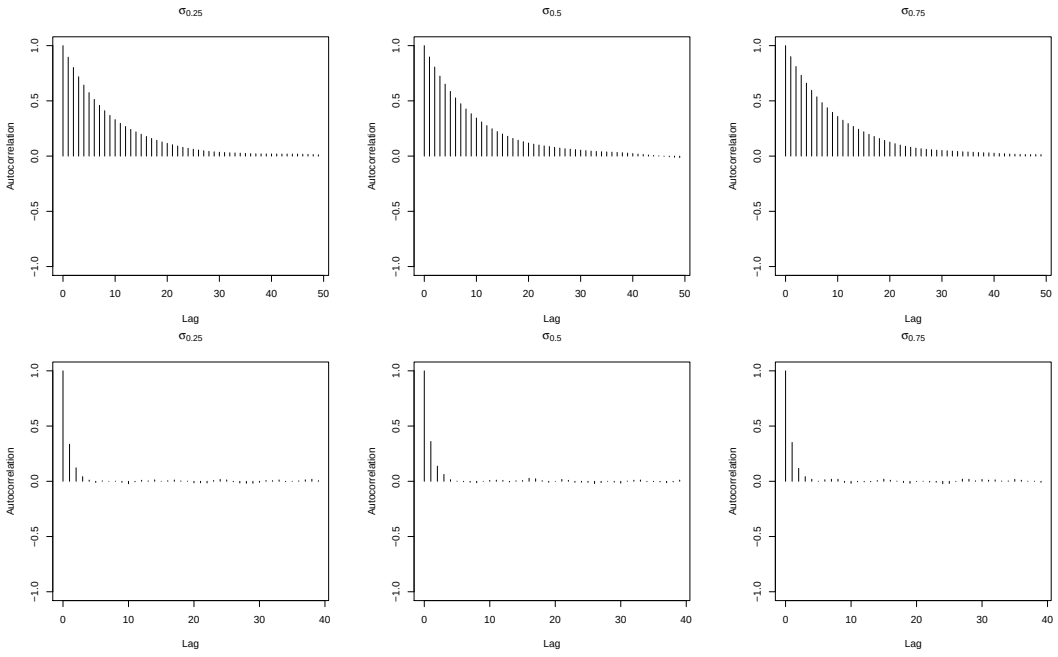



Figure B.7: The autocorrelation plot for the samples of the scale parameters obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom).

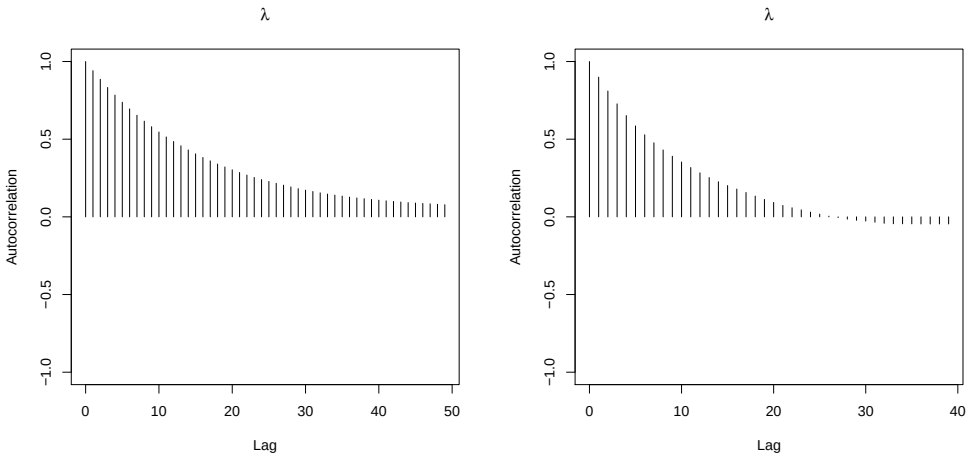



Figure B.8: The autocorrelation plot for the samples of the transformation parameter obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right).

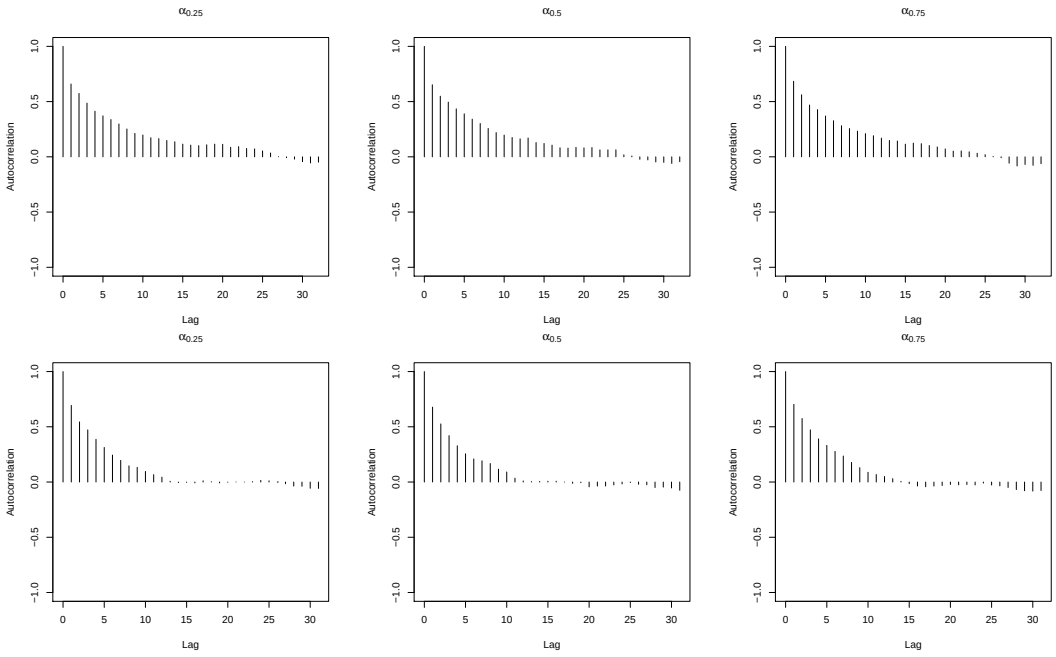



Figure B.9: The autocorrelation plot (after thinning) for the samples of the intercept coefficients obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom).

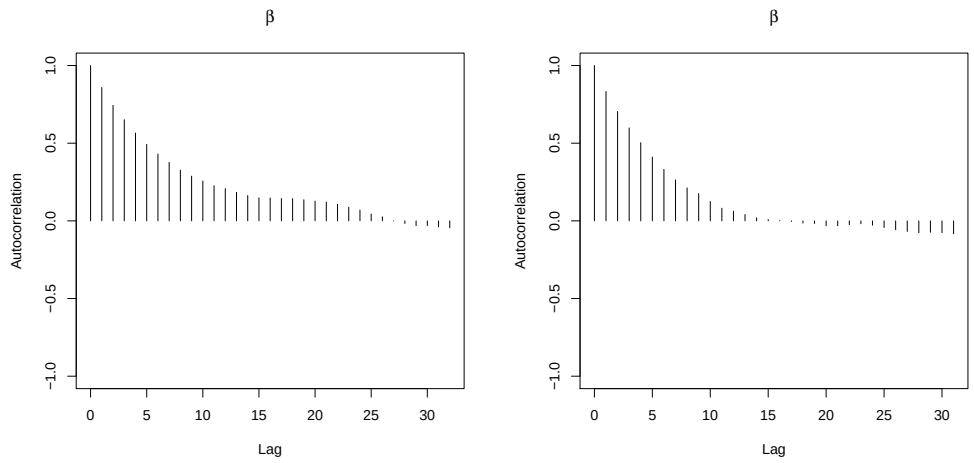



Figure B.10: The autocorrelation plot (after thinning) for the samples of the slope coefficient obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right).

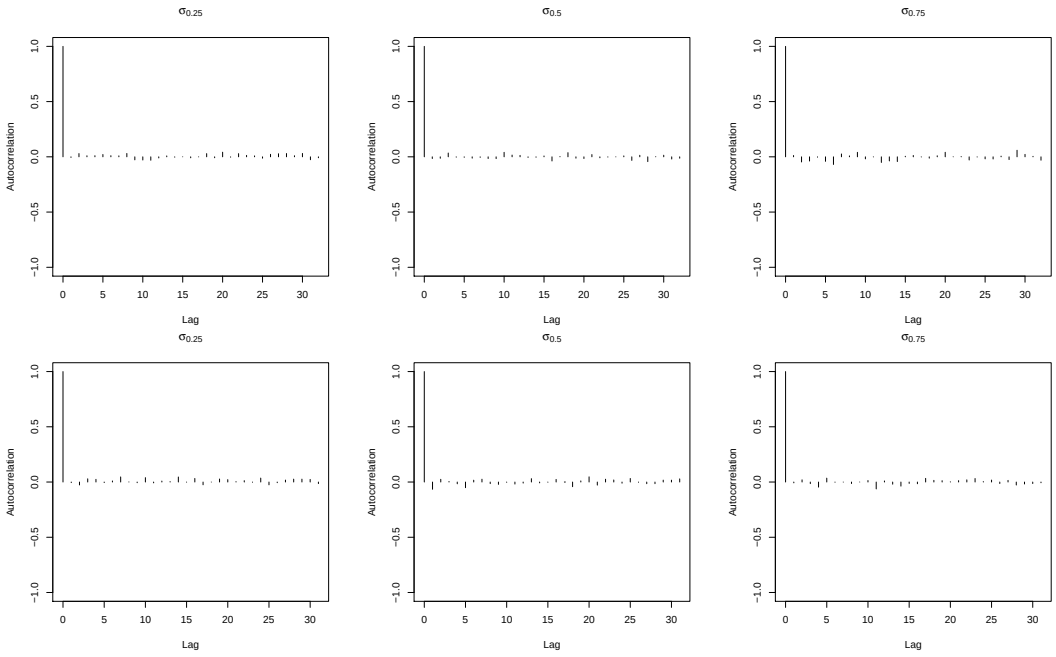



Figure B.11: The autocorrelation plot (after thinning) for the samples of the scale parameters obtained using Metropolis-Hastings algorithm (top) and Gibbs sampler (bottom).

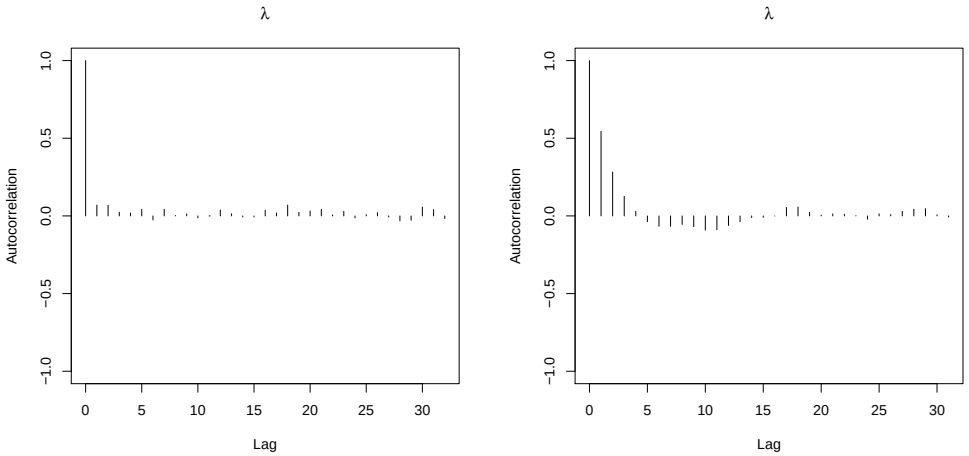



Figure B.12: The autocorrelation plot (after thinning) for the samples of the transformation parameter obtained using Metropolis-Hastings algorithm (left) and Gibbs sampler (right).

| | Metropolis-Hastings | | Gibbs | |
|---|---|---|---|---|
| Coefficients | P.M. | 95% HPD | P.M. | 95% HPD |
| $\alpha_{0.25}$ | $-0.51$ | $(-0.55, -0.47)$ | $-0.51$ | $(-0.55, -0.47)$ |
| $\alpha_{0.5}$ | $-0.40$ | $(-0.44, -0.37)$ | $-0.41$ | $(-0.44, -0.37)$ |
| $\alpha_{0.75}$ | $-0.31$ | $(-0.36, -0.27)$ | $-0.31$ | $(-0.36, -0.27)$ |
| $\beta$ | 0.24 | $(0.23, 0.25)$ | 0.24 | $(0.23, 0.25)$ |
| $\lambda$ | 2.13 | $(1.93, 2.33)$ | 2.13 | $(1.94, 2.33)$ |
| $\sigma_{0.25}$ | 0.05 | $(0.04, 0.05)$ | 0.05 | $(0.04, 0.06)$ |
| $\sigma_{0.5}$ | 0.06 | $(0.06, 0.07)$ | 0.06 | $(0.06, 0.07)$ |
| $\sigma_{0.75}$ | 0.05 | $(0.05, 0.06)$ | 0.05 | $(0.05, 0.06)$ |

Table B.3: The posterior mean (P.M.) and the 95% highest posterior density intervals (95% HPD) obtained the two proposed sampling methods.

# Appendix C

## C.1 Properties of the generalised asymmetric Laplace distribution

For the generalised asymmetric Laplace distribution given in (4.3), we can show that

$$\int_{-\infty}^{\infty} f(y|\mu_\tau,\sigma,\gamma) = 1 \text{ and } \int_{-\infty}^{\mu_\tau} f(y|\mu_\tau,\sigma,\gamma) = \tau,$$

as follows

$$\begin{aligned}
\int_{-\infty}^{\mu_\tau} f(y|\mu_\tau,\sigma,\gamma)dy &= \int_{-\infty}^{\mu_\tau} \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})} \exp\left\{-\left[(1-\tau)\left|\frac{y-\mu_\tau}{\sigma}\right|\right]^\gamma\right\} dy \\
&= \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})} \int_{-\infty}^{\mu_\tau} \exp\left\{-\left[(1-\tau)\left|\frac{y-\mu_\tau}{\sigma}\right|\right]^\gamma\right\} dy \\
&= \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})} \int_{-\infty}^{\mu_\tau} \exp\left\{-\left[-\frac{(1-\tau)(y-\mu_\tau)}{\sigma}\right]^\gamma\right\} dy.
\end{aligned}$$

Substituting

$$\begin{aligned}
u &= \left[-\frac{(1-\tau)(y-\mu_\tau)}{\sigma}\right]^\gamma \\
du &= -\frac{(1-\tau)\gamma}{\sigma}\left[-\frac{(1-\tau)(y-\mu_\tau)}{\sigma}\right]^{\gamma-1} dy \\
&= -\frac{(1-\tau)\gamma}{\sigma}\left[u^{\frac{1}{\gamma}}\right]^{\gamma-1} dy \\
&= -\frac{(1-\tau)\gamma}{\sigma} u^{1-\frac{1}{\gamma}} dy,
\end{aligned}$$

gives

$$\begin{aligned}\int_{-\infty}^{\mu_\tau} f(y|\mu_\tau,\sigma,\boldsymbol{\gamma})dy &= \frac{\tau}{\Gamma(\frac{1}{\gamma})}\int_{\infty}^{0} -u^{-\left(1-\frac{1}{\gamma}\right)}\exp\{-u\}\,du \\ &= \frac{\tau}{\Gamma(\frac{1}{\gamma})}\int_{0}^{\infty} u^{\frac{1}{\gamma}-1}\exp\{-u\}\,du \\ &= \frac{\tau}{\Gamma(\frac{1}{\gamma})}\Gamma(\frac{1}{\gamma}) = \tau.\end{aligned}$$

Also,

$$\begin{aligned}\int_{\mu_\tau}^{\infty} f(y|\mu_\tau,\sigma,\boldsymbol{\gamma})dy &= \int_{\mu_\tau}^{\infty}\frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})}\exp\left\{-\left[\tau\left|\frac{y-\mu_\tau}{\sigma}\right|\right]^{\gamma}\right\}dy \\ &= \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})}\int_{\mu_\tau}^{\infty}\exp\left\{-\left[\tau\left|\frac{y-\mu_\tau}{\sigma}\right|\right]^{\gamma}\right\}dy \\ &= \frac{\gamma\tau(1-\tau)}{\sigma\Gamma(\frac{1}{\gamma})}\int_{\mu_\tau}^{\infty}\exp\left\{-\left[\frac{\tau\,(y-\mu_\tau)}{\sigma}\right]^{\gamma}\right\}dy.\end{aligned}$$

Substituting

$$\begin{aligned}u &= \left[\frac{\tau\,(y-\mu_\tau)}{\sigma}\right]^{\gamma} \\ du &= \frac{\tau\gamma}{\sigma}\left[\frac{\tau\,(y-\mu_\tau)}{\sigma}\right]^{\gamma-1}dy \\ &= \frac{\tau\gamma}{\sigma}\left[u^{\frac{1}{\gamma}}\right]^{\gamma-1}dy \\ &= \frac{\tau\gamma}{\sigma}u^{1-\frac{1}{\gamma}}dy,\end{aligned}$$

gives

$$\begin{aligned}\int_{\mu_\tau}^{\infty} f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma})dy &= \frac{(1-\tau)}{\Gamma(\frac{1}{\gamma})}\int_{0}^{\infty} u^{-\left(1-\frac{1}{\gamma}\right)}\exp\{-u\}\,du \\ &= \frac{(1-\tau)}{\Gamma(\frac{1}{\gamma})}\int_{0}^{\infty} u^{\frac{1}{\gamma}-1}\exp\{-u\}\,du \\ &= \frac{(1-\tau)}{\Gamma(\frac{1}{\gamma})}\Gamma(\frac{1}{\gamma}) = 1-\tau.\end{aligned}$$

## C.2 Properties of the generalised Gumbel distribution

For the generalised Gumbel distribution, we can show that

$$\int_{-\infty}^{\infty} f(y|\mu_\tau, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = 1 \text{ and } \int_{-\infty}^{\mu_\tau} f(y|\mu_\tau, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \tau,$$

as follows

$$\begin{aligned}
\int_{-\infty}^{\mu_\tau} f(y|\mu_{\tau_k}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) dy &= \int_{-\infty}^{\mu_{\tau_k}} \frac{\gamma_1 \tau(1-\tau)e^{\gamma_1}}{\sigma_1(e^{\gamma_1}-1)} \exp\left\{-(1-\tau)\left|\frac{y-\mu_\tau}{\sigma_1}\right| - \gamma_1 \exp\left[-(1-\tau)\left|\frac{y-\mu_\tau}{\sigma_1}\right|\right]\right\} dy \\
&= \int_{-\infty}^{\mu_\tau} \frac{\gamma_1 \tau(1-\tau)e^{\gamma_1}}{\sigma_1(e^{\gamma_1}-1)} \exp\left\{(1-\tau)\left(\frac{y-\mu_\tau}{\sigma_1}\right) - \gamma_1 \exp\left[(1-\tau)\left(\frac{y-\mu_\tau}{\sigma_1}\right)\right]\right\} dy.
\end{aligned}$$

Substituting

$$\begin{aligned}
u &= \gamma_1 \exp\left\{(1-\tau)\left(\frac{y-\mu_\tau}{\sigma_1}\right)\right\} \\
du &= \frac{\gamma_1(1-\tau)}{\sigma_1} \exp\left\{(1-\tau)\left(\frac{y-\mu_\tau}{\sigma_1}\right)\right\} dy \\
&= \frac{(1-\tau)u}{\sigma_1} dy,
\end{aligned}$$

gives

$$\begin{aligned}
\int_{-\infty}^{\mu_\tau} f(y|\mu_\tau, \boldsymbol{\sigma}, \boldsymbol{\gamma}) dy &= \frac{\gamma_1 \tau(1-\tau)e^{\gamma_1}}{\sigma_1(e^{\gamma_1}-1)} \int_0^{\gamma_1} \exp\left\{\log\left(\frac{u}{\gamma_1}\right) - u\right\} \frac{\sigma_1}{(1-\tau)u} du \\
&= \frac{\tau e^{\gamma_1}}{e^{\gamma_1}-1} \int_0^{\gamma_1} \exp\{-u\} du \\
&= \tau.
\end{aligned}$$

Also,

$$\begin{aligned}
\int_{\mu_\tau}^{\infty} f(y|\mu_\tau, \boldsymbol{\sigma}, \boldsymbol{\gamma}) dy &= \int_{\mu_\tau}^{\infty} \frac{\gamma_2 \tau(1-\tau)e^{\gamma_2}}{\sigma_2(e^{\gamma_2}-1)} \exp\left\{-\tau\left|\frac{y-\mu_\tau}{\sigma_2}\right| - \gamma_2 \exp\left[-\tau\left|\frac{y-\mu_\tau}{\sigma_2}\right|\right]\right\} dy \\
&= \int_{\mu_\tau}^{\infty} \frac{\gamma_2 \tau(1-\tau)e^{\gamma_2}}{\sigma_2(e^{\gamma_2}-1)} \exp\left\{-\tau\left(\frac{y-\mu_\tau}{\sigma_2}\right) - \gamma_2 \exp\left[-\tau\left(\frac{y-\mu_\tau}{\sigma_2}\right)\right]\right\} dy.
\end{aligned}$$

Substituting

$$u = \gamma_2 \exp\left\{-\tau\left(\frac{y-\mu_\tau}{\sigma_2}\right)\right\}$$

$$\begin{aligned} du &= -\frac{\gamma_2\tau}{\sigma_2}\exp\left\{-\tau\left(\frac{y-\mu_\tau}{\sigma_2}\right)\right\}dy \\ &= -\frac{\tau u}{\sigma_2}dy, \end{aligned}$$

gives

$$\begin{aligned} \int_{\mu_\tau}^{\infty} f(y|\mu_\tau,\boldsymbol{\sigma},\boldsymbol{\gamma})dy &= \frac{\gamma_2\tau(1-\tau)e^{\gamma_2}}{\sigma_2(e^{\gamma_2}-1)}\int_{\gamma_2}^{0} -\exp\left\{\log(\frac{u}{\gamma_2})-u\right\}\frac{\sigma_2}{\tau u}du \\ &= \frac{(1-\tau)e^{\gamma_2}}{e^{\gamma_2}-1}\int_0^{\gamma_2}\exp\{-u\}\,du \\ &= 1-\tau. \end{aligned}$$

# Appendix D

## D.1 Normalising term

To find the normalising term for the density given (5.1), we can calculate

$$\delta = \int_{-\infty}^{\infty} \widehat{f}\left(y_i|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau\right) dy.$$

Assumed that $w_{\tau_k}$ is a constant with respect to $y$ for all $k$, we can show that

$$\begin{aligned}
\int_{-\infty}^{\mu_{\tau_1}} \widehat{f}(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau)\, dy &= \int_{-\infty}^{\mu_{\tau_1}} \left[\prod_{k=1}^{m} \left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} \left|y-\mu_{\tau_k}\right|\right\} dy \\
&= \int_{-\infty}^{\mu_{\tau_1}} \left[\prod_{k=1}^{m} \left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right] \exp\left\{\sum_{k=1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} \left(y-\mu_{\tau_k}\right)\right\} dy \\
&= \left[\prod_{k=1}^{m} \left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right] \int_{-\infty}^{\mu_{\tau_1}} \exp\left\{\sum_{k=1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} y - \sum_{k=1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} \mu_{\tau_k}\right\} dy \\
&= \left[\prod_{k=1}^{m} \left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} \mu_{\tau_k}\right\} \\
&\times \int_{-\infty}^{\mu_{\tau_1}} \exp\left\{y \sum_{k=1}^{m} w_{\tau_k} \left(\frac{1-\tau_k}{\sigma_{\tau_k}}\right)\right\} dy \\
&= \frac{\left[\prod_{k=1}^{m} \left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} \mu_{\tau_k}\right\} \exp\left\{\mu_{\tau_1} \sum_{k=1}^{m} w_{\tau_k} \left(\frac{1-\tau_k}{\sigma_{\tau_k}}\right)\right\}}{\sum_{k=1}^{m} w_{\tau_k} \left(\frac{1-\tau_k}{\sigma_{\tau_k}}\right)}
\end{aligned}$$

and

$$\begin{aligned}
\int_{\mu_{\tau_t}}^{\mu_{\tau_{t+1}}} \widehat{f}\left(y|\boldsymbol{\mu}_\tau, \boldsymbol{\sigma}_\tau\right) dy_i &= \int_{\mu_{\tau_t}}^{\mu_{\tau_{t+1}}} \left[\prod_{k=1}^{m} \left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right] \exp\left\{-\sum_{k=1}^{t} \frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}} \left(y-\mu_{\tau_k}\right)\right. \\
&\quad \left. + \sum_{k=t+1}^{m} \frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}} \left(y-\mu_{\tau_k}\right)\right\} dy
\end{aligned}$$

$$
\begin{aligned}
&= \left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\int_{\mu_{\tau_t}}^{\mu_{\tau_{t+1}}}\exp\left\{-\sum_{k=1}^{t}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}y+\sum_{k=1}^{t}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\mu_{\tau_k}\right.\\
&\quad\left.+\sum_{k=t+1}^{m}\frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}}y-\sum_{k=t+1}^{m}\frac{w_{\tau_k}\left(1-\tau_k\right)}{\sigma_{\tau_k}}\mu_{\tau_k}\right\}dy\\
&= \left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{w_{\tau_k}\left(\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\mu_{\tau_k}-\sum_{k=t+1}^{m}\frac{\left(1-\tau_k\right)}{\sigma_{\tau_k}}\mu_{\tau_k}\right)\right\}\\
&\quad\times\int_{\mu_{\tau_t}}^{\mu_{\tau_{t+1}}}\exp\left\{w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{\left(1-\tau_k\right)}{\sigma_{\tau_k}}-\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)y\right\}dy\\
&= \left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{w_{\tau_k}\left(\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\mu_{\tau_k}-\sum_{k=t+1}^{m}\frac{\left(1-\tau_k\right)}{\sigma_{\tau_k}}\mu_{\tau_k}\right)\right\}\\
&\quad\times\left[w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{\left(1-\tau_k\right)}{\sigma_{\tau_k}}-\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)\right]^{-1}\left[\exp\left\{w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right.\right.\right.\\
&\quad\left.\left.\left.-\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)\mu_{\tau_{t+1}}\right\}-\exp\left\{w_{\tau_k}\left(\sum_{k=t+1}^{m}\frac{\left(1-\tau_k\right)}{\sigma_{\tau_k}}-\sum_{k=1}^{t}\frac{\tau_k}{\sigma_{\tau_k}}\right)\mu_{\tau_t}\right\}\right]
\end{aligned}
$$

Also,

$$
\begin{aligned}
\int_{\mu_{\tau_m}}^{\infty}\widehat{f}\left(y|\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau\right)dy &= \int_{\mu_{\tau_m}}^{\infty}\left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{-\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\left(y-\mu_{\tau_k}\right)\right\}dy\\
&= \left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\int_{\mu_{\tau_m}}^{\infty}\exp\left\{-\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\left(y-\mu_{\tau_k}\right)\right\}dy\\
&= \left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\int_{\mu_{\tau_m}}^{\infty}\exp\left\{-\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}y+\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\mu_{\tau_k}\right\}dy\\
&= \left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k\mu_{\tau_k}}{\sigma_{\tau_k}}\right\}\int_{\mu_{\tau_m}}^{\infty}\exp\left\{-y\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\right\}dy\\
&= \frac{\left[\prod_{k=1}^{m}\left(\frac{\tau_k\left(1-\tau_k\right)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k\mu_{\tau_k}}{\sigma_{\tau_k}}\right\}\exp\left\{-\mu_{\tau_m}\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}\right\}}{\sum_{k=1}^{m}\frac{w_{\tau_k}\tau_k}{\sigma_{\tau_k}}}
\end{aligned}
$$

## D.2 Maximizing the weighted pseudo asymmetric Laplace likelihood

Assume that $y$ a random variable from a continuous distribution $F$ with probability density $f$ . Then, to estimate two quantile functions of $F$ corresponding to $\tau_1$ and $\tau_2$, we can use the weighted pseudo-asymmetric Laplace distribution given

by

$$\widehat{f}(y|\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau)=\begin{cases}\left[\prod_{k=1}^{2}\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{-\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}|\right\} & y\le\mu_{\tau_1}\\ \left[\prod_{k=1}^{2}\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}|\right\} & \mu_{\tau_1}<y\le\mu_{\tau_2}\\ \left[\prod_{k=1}^{2}\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)^{w_{\tau_k}}\right]\exp\left\{-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}|y-\mu_{\tau_2}|\right\} & y>\mu_{\tau_2}\end{cases}.$$

Then, the logarithm of $\widehat{f}$ is given by

$$\begin{aligned}L(y;\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau) &= \begin{cases}\left[\sum_{k=1}^{2}w_{\tau_k}\log\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)\right]-\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}| & y\le\mu_{\tau_1}\\ \left[\sum_{k=1}^{2}w_{\tau_k}\log\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)\right]-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}| & \mu_{\tau_1}<y\le\mu_{\tau_2}\\ \left[\sum_{k=1}^{2}w_{\tau_k}\log\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)\right]-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}|y-\mu_{\tau_2}| & y>\mu_{\tau_2}\end{cases}\\ &= \begin{cases}\phi-\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}| & y\le\mu_{\tau_1}\\ \phi-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}| & \mu_{\tau_1}<y\le\mu_{\tau_2}\\ \phi-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}|y-\mu_{\tau_2}| & y>\mu_{\tau_2}\end{cases}\end{aligned},$$

where $\phi=\sum_{k=1}^{2}w_{\tau_k}\log\left(\frac{\tau_k(1-\tau_k)}{\sigma_{\tau_k}}\right)$. To show that the location parameters $\mu_{\tau_1}$ and $\mu_{\tau_2}$ is $\tau_1$ and $\tau_2$ quantiles of distribution $F$ respectively, we need to find

$$\max_{\boldsymbol{\mu}_\tau}\Big(E\big[L(y;\boldsymbol{\mu}_\tau,\boldsymbol{\sigma}_\tau)\big]\Big).$$

Assuming that $w_{\tau_k}$ is a constant with respect to $y$, First order conditions are given by

$$\begin{aligned}0 &= \frac{d}{d\mu_\tau}\left\{\int_{-\infty}^{\mu_{\tau_1}}\left(\phi-\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}|\right)f(y)\,dy\right.\\ &\quad+\int_{\mu_{\tau_1}}^{\mu_{\tau_2}}\left(\phi-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}|y-\mu_{\tau_2}|\right)f(y)\,dy\\ &\quad\left.+\int_{\mu_{\tau_2}}^{\infty}\left(\phi-\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}|y-\mu_{\tau_1}|-\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}|y-\mu_{\tau_2}|\right)f(y)\,dy\right\}\\ 0 &= \frac{d}{d\mu_\tau}\left\{\int_{-\infty}^{\mu_{\tau_1}}\phi f(y)\,dy+\int_{-\infty}^{\mu_{\tau_1}}\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}(y-\mu_{\tau_1})f(y)\,dy+\int_{-\infty}^{\mu_{\tau_1}}\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}(y-\mu_{\tau_2})f(y)\,dy\right.\\ &\quad+\int_{\mu_{\tau_1}}^{\mu_{\tau_2}}\phi f(y)\,dy-\int_{\mu_{\tau_1}}^{\mu_{\tau_2}}\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}(y-\mu_{\tau_1})f(y)\,dy+\int_{\mu_{\tau_1}}^{\mu_{\tau_2}}\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}(y-\mu_{\tau_2})f(y)\,dy\\ &\quad\left.+\int_{\mu_{\tau_2}}^{\infty}\phi f(y)\,dy-\int_{\mu_{\tau_2}}^{\infty}\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}(y-\mu_{\tau_1})f(y)\,dy-\int_{\mu_{\tau_2}}^{\infty}\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}(y-\mu_{\tau_2})f(y)\,dy\right\}\\ 0 &= \frac{d}{d\mu_\tau}\left\{\int_{-\infty}^{\mu_{\tau_1}}\phi f(y)\,dy+\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}\left[\int_{-\infty}^{\mu_{\tau_1}}yf(y)\,dy-\int_{-\infty}^{\mu_{\tau_1}}\mu_{\tau_1}f(y)\,dy\right]\right.\\ &\quad\left.+\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[\int_{-\infty}^{\mu_{\tau_1}}yf(y)\,dy-\int_{-\infty}^{\mu_{\tau_1}}\mu_{\tau_2}f(y)\,dy\right]+\int_{\mu_{\tau_1}}^{\mu_{\tau_2}}\phi f(y)\,dy\right.\end{aligned}$$

$$
\begin{aligned}
& -\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[\int_{\mu_{\tau_1}}^{\mu_{\tau_2}} yf(y)\,dy - \int_{\mu_{\tau_1}}^{\mu_{\tau_2}} \mu_{\tau_1} f(y)\,dy\right] + \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[\int_{\mu_{\tau_1}}^{\mu_{\tau_2}} yf(y)\,dy - \int_{\mu_{\tau_1}}^{\mu_{\tau_2}} \mu_{\tau_2} f(y)\,dy\right] \\
& + \int_{\mu_{\tau_2}}^{\infty} \phi f(y)\,dy - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[\int_{\mu_{\tau_2}}^{\infty} yf(y)\,dy - \int_{\mu_{\tau_2}}^{\infty} \mu_{\tau_1} f(y)\,dy\right] - \frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}\left[\int_{\mu_{\tau_2}}^{\infty} yf(y)\,dy - \int_{\mu_{\tau_2}}^{\infty} \mu_{\tau_2} f(y)\,dy\right]\Bigg\} \\
0 &= \frac{d}{d\mu_\tau}\Bigg\{\phi F(\mu_{\tau_1}) + \frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}\left[\int_{-\infty}^{\mu_{\tau_1}} yf(y)\,dy - \mu_{\tau_1} F(\mu_{\tau_1})\right] \\
& + \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[\left(\int_{-\infty}^{\mu_{\tau_1}} yf(y)\,dy\right) - \mu_{\tau_2} F(\mu_{\tau_1})\right] + \phi F(\mu_{\tau_2}) - \phi F(\mu_{\tau_1}) \\
& - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[\left(\int_{\mu_{\tau_1}}^{\mu_{\tau_2}} yf(y)\,dy\right) - \mu_{\tau_1} F(\mu_{\tau_2}) + \mu_{\tau_1} F(\mu_{\tau_1})\right] \\
& + \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[\left(\int_{\mu_{\tau_1}}^{\mu_{\tau_2}} yf(y)\,dy\right) - \mu_{\tau_2} F(\mu_{\tau_2}) + \mu_{\tau_2} F(\mu_{\tau_1})\right] + \phi - \phi F(\mu_{\tau_2}) \\
& - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[\left(\int_{\mu_{\tau_2}}^{\infty} yf(y)\,dy\right) - \mu_{\tau_1} + \mu_{\tau_1} F(\mu_{\tau_2})\right] - \frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}\left[\left(\int_{\mu_{\tau_2}}^{\infty} yf(y)\,dy\right) - \mu_{\tau_2} + \mu_{\tau_2} F(\mu_{\tau_2})\right]\Bigg\}
\end{aligned}
$$

By Differentiating with respect to $\mu_{\tau_1}$, we can find that

$$
\begin{aligned}
0 &= \phi f(\mu_{\tau_1}) + \frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}}\left[\mu_{\tau_1} f(\mu_{\tau_1}) - F(\mu_{\tau_1}) - \mu_{\tau_1} f(\mu_{\tau_1})\right] \\
&\quad + \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[\mu_{\tau_1} f(\mu_{\tau_1}) - \mu_{\tau_2} f(\mu_{\tau_1})\right] - \phi f(\mu_{\tau_1}) \\
&\quad - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[-\mu_{\tau_1} f(\mu_{\tau_1}) - F(\mu_{\tau_2}) + F(\mu_{\tau_1}) + \mu_{\tau_1} f(\mu_{\tau_1})\right] \\
&\quad + \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[-\mu_{\tau_1} f(\mu_{\tau_1}) + \mu_{\tau_2} f(\mu_{\tau_1})\right] - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[-1 + F(\mu_{\tau_2})\right] \\
0 &= -\frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}} F(\mu_{\tau_1}) - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}} F(\mu_{\tau_1}) + \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}} \\
\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}} &= \frac{w_{\tau_1}(1-\tau_1)}{\sigma_{\tau_1}} F(\mu_{\tau_1}) + \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}} F(\mu_{\tau_1}) \\
\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}} &= \frac{w_{\tau_1} F(\mu_{\tau_1})}{\sigma_{\tau_1}}\left[(1-\tau_1) + \tau_1\right] \\
\frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}} &= \frac{w_{\tau_1} F(\mu_{\tau_1})}{\sigma_{\tau_1}} \\
\tau_1 &= F(\mu_{\tau_1})
\end{aligned}
$$

By Differentiating with respect to $\mu_{\tau_2}$, we can find that

$$
\begin{aligned}
0 &= \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[-F(\mu_{\tau_1})\right] + \phi f(\mu_{\tau_2}) - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[\mu_{\tau_2} f(\mu_{\tau_2}) - \mu_{\tau_1} f(\mu_{\tau_2})\right] \\
&\quad + \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}}\left[\mu_{\tau_2} f(\mu_{\tau_2}) - F(\mu_{\tau_2}) - \mu_{\tau_2} f(\mu_{\tau_2}) + F(\mu_{\tau_1})\right] - \phi f(\mu_{\tau_2}) \\
&\quad - \frac{w_{\tau_1}\tau_1}{\sigma_{\tau_1}}\left[-\mu_{\tau_2} f(\mu_{\tau_2}) + \mu_{\tau_1} f(\mu_{\tau_2})\right] - \frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}}\left[-\mu_{\tau_2} f(\mu_{\tau_2}) - 1 + F(\mu_{\tau_2}) + \mu_{\tau_2} f(\mu_{\tau_2})\right]
\end{aligned}
$$

$$
\begin{aligned}
0 &= -\frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}} F\left(\mu_{\tau_2}\right) + \frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}} - \frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}} F\left(\mu_{\tau_2}\right) \\
\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}} &= \frac{w_{\tau_2}(1-\tau_2)}{\sigma_{\tau_2}} F\left(\mu_{\tau_2}\right) + \frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}} F\left(\mu_{\tau_2}\right) \\
\frac{w_{\tau_2}\tau_2}{\sigma_{\tau_2}} &= \frac{w_{\tau_2} F\left(\mu_{\tau_2}\right)}{\sigma_{\tau_2}} \left[(1-\tau_2) + \tau_2\right] \\
\tau_2 &= F\left(\mu_{\tau_2}\right)
\end{aligned}
$$

# Appendix E

## E.1 The weighted mixture distribution

In this section, we show that the function given by

$$g_t(y|\mu,\sigma,\kappa) = \begin{cases} \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left(\frac{1}{\kappa\sigma}\left|y-\mu\right|\right)^t\right\}, & y \leq \mu \\ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left(\frac{\kappa}{\sigma}\left|y-\mu\right|\right)^t\right\}, & y > \mu \end{cases}.$$

is a density function for any $t > 0$ as follows

$$\begin{aligned} \int_{-\infty}^{\mu} g_t(y|\mu,\sigma,\kappa)dy &= \int_{-\infty}^{\mu} \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{-\left|\frac{y-\mu}{\kappa\sigma}\right|^t\right\} dy \\ &= \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{-\infty}^{\mu} \exp\left\{-\left(-\frac{y-\mu}{\kappa\sigma}\right)^t\right\} dy \end{aligned}$$

Substituting

$$\begin{aligned} u &= \left(-\frac{y-\mu}{\kappa\sigma}\right)^t \\ du &= -\frac{t}{\kappa\sigma}\left(-\frac{y-\mu}{\kappa\sigma}\right)^{t-1} dy \\ &= -\frac{t}{\kappa\sigma} u^{1-\frac{1}{t}} dy, \end{aligned}$$

gives

$$\int_{-\infty}^{\mu} g_t(y|\mu,\sigma,\kappa)dy = \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{\infty}^{0} -u^{-\left(1-\frac{1}{t}\right)} \exp\{-u\}\, du$$

$$\begin{aligned} &= \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})}\int_0^\infty u^{\frac{1}{t}-1}\exp\{-u\}\,du \\ &= \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})}\Gamma(\frac{1}{t}) \\ &= \frac{\kappa^2}{1+\kappa^2} \end{aligned}$$

Also

$$\begin{aligned} \int_\mu^\infty g(y|\mu,\sigma,\kappa)dy &= \int_\mu^\infty \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})}\exp\left\{-\left|\frac{\kappa(y-\mu)}{\sigma}\right|^t\right\}dy \\ &= \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})}\int_\mu^\infty \exp\left\{-\left(\frac{\kappa(y-\mu)}{\sigma}\right)^t\right\}dy \end{aligned}$$

Substituting

$$\begin{aligned} u &= \left(\frac{\kappa(y-\mu_\tau)}{\sigma}\right)^t \\ du &= \frac{t\kappa}{\sigma}\left(\frac{\kappa\,(y-\mu_\tau)}{\sigma}\right)^{t-1}dy \\ &= \frac{t\kappa}{\sigma}u^{1-\frac{1}{t}}dy, \end{aligned}$$

gives

$$\begin{aligned} \int_\mu^\infty g_t(y|\mu,\sigma,\kappa)dy &= \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})}\int_0^\infty u^{-\left(1-\frac{1}{t}\right)}\exp\{-u\}\,du \\ &= \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})}\int_0^\infty u^{\frac{1}{t}-1}\exp\{-u\}\,du \\ &= \frac{1}{1+\kappa^2} \end{aligned}$$

Then,

$$\begin{aligned} \int_{-\infty}^\mu g_t(y|\mu,\sigma,\kappa)dy + \int_\mu^\infty g_t(y;\mu,\sigma,\kappa)dy &= \frac{\kappa^2}{1+\kappa^2}+\frac{1}{1+\kappa^2} \\ &= 1 \end{aligned}$$

## E.2 The cumulative distribution function for the weighted mixture distribution

In this section, we show how we can derive the cumulative distribution function in the case of the mixture of skewed Laplace and skewed normal distributions. We assume that

$$g(y|\mu,\sigma,\kappa) = \begin{cases} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{ -\left(\frac{1}{\kappa\sigma}\left|y-\mu\right|\right)^t \right\} \right], & y \leq \mu \\ \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \exp\left\{ -\left(\frac{\kappa}{\sigma}\left|y-\mu\right|\right)^t \right\} \right], & y > \mu \end{cases}.$$

Then, the cumulative function is given by

$$\begin{aligned} G(q_{\tau_k};\mu,\sigma,\kappa) &= \begin{cases} \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{-\infty}^{q_{\tau_k}} \exp\left\{ -\left(\frac{1}{\kappa\sigma}\left|y-\mu\right|\right)^t \right\} dy \right] & q_{\tau_k} \leq \mu \\ \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ 1 - \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{q_{\tau_k}}^{\infty} \exp\left\{ -\left(\frac{\kappa}{\sigma}\left|y-\mu\right|\right)^t \right\} dy \right] & q_{\tau_k} > \mu \end{cases} \\ &= \begin{cases} \underbrace{\sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{-\infty}^{q_{\tau_k}} \exp\left\{ -\left(-\frac{1}{\kappa\sigma}\left[y-\mu\right]\right)^t \right\} dy \right]}_{A}, & q_{\tau_k} \leq \mu \\ \underbrace{\sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ 1 - \frac{t\kappa}{\sigma(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{q_{\tau_k}}^{\infty} \exp\left\{ -\left(\frac{\kappa}{\sigma}\left[y-\mu\right]\right)^t \right\} dy \right]}_{B}, & q_{\tau_k} > \mu \end{cases}. \end{aligned}$$

First, we assume that

$$u = \left(-\frac{1}{\kappa\sigma}\left[y-\mu\right]\right)^t.$$

This implies that

$$\begin{aligned} du &= -\frac{t}{\kappa\sigma}\left(-\frac{1}{\kappa\sigma}\left[y-\mu\right]\right)^{t-1} dy \\ &= -\frac{t}{\kappa\sigma}\left(u^{\frac{1}{t}}\right)^{t-1} dy \\ &= -\frac{t}{\kappa\sigma} u^{1-\frac{1}{t}} dy. \end{aligned}$$

Then,

$$\begin{aligned}
A &= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ -\frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{\infty}^{\left(-\frac{1}{\kappa\sigma}[q_{\tau_k}-\mu]\right)^t} u^{\frac{1}{t}-1} \exp\{-u\}\, du \right] \\
&= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{\left(-\frac{1}{\kappa\sigma}[q_{\tau_k}-\mu]\right)^t}^{\infty} u^{\frac{1}{t}-1} \exp\{-u\}\, du \right] \\
&= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left( \frac{1}{t}, \left( -\frac{1}{\kappa\sigma}\left[q_{\tau_k}-\mu\right] \right)^t \right) \right] \\
&= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ \frac{\kappa^2}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left( \frac{1}{t}, \left( \frac{1}{\kappa\sigma}\left|q_{\tau_k}-\mu\right| \right)^t \right) \right].
\end{aligned}$$

$\widetilde{\Gamma}$ is upper incomplete gamma function. Second, we assume that

$$u = \left( \frac{\kappa}{\sigma}\left[y-\mu\right] \right)^t.$$

This implies that

$$\begin{aligned}
du &= \frac{t\kappa}{\sigma} \left( \frac{\kappa}{\sigma}\left[y-\mu\right] \right)^{t-1} dy \\
&= \frac{t\kappa}{\sigma} \left( u^{\frac{1}{t}} \right)^{t-1} dy \\
&= \frac{t\kappa}{\sigma} u^{1-\frac{1}{t}} dy.
\end{aligned}$$

Then,

$$\begin{aligned}
B &= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ 1 - \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})} \int_{\left(\frac{\kappa}{\sigma}[q_{\tau_k}-\mu]\right)^t}^{\infty} u^{\frac{1}{t}-1} \exp\{-u\}\, du \right] \\
&= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ 1 - \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left( \frac{1}{t}, \left( \frac{\kappa}{\sigma}\left[q_{\tau_k}-\mu\right] \right)^t \right) \right] \\
&= \sum_{t=1}^{2} \frac{(1-\nu)^{t-1}}{\nu^{t-2}} \left[ 1 - \frac{1}{(1+\kappa^2)\Gamma(\frac{1}{t})} \widetilde{\Gamma}\left( \frac{1}{t}, \left( \frac{\kappa}{\sigma}\left|q_{\tau_k}-\mu\right| \right)^t \right) \right].
\end{aligned}$$

# References


Alhamzawi, R. & Yu, K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics & Data Analysis*, **64**, 209–219. 12

Altman, N. & Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46**, 195–214. xvii, 126, 127

Bel, G., Bolancé, C., Guillén, M. & Rosell, J. (2015). The environmental effects of changing speed limits: A quantile regression approach. *Transportation Research Part D: Transport and Environment*, **36**, 76–85. 1

Bolker, B. (2016). *emdbook: Ecological Models and Data in R*. Package version 1.3.9. 145

Bolker, B.M. (2008). *Ecological models and data in R*. Princeton University Press. 145

Bondell, H.D., Reich, B.J. & Wang, H. (2010). Non-crossing quantile regression curve estimation. *Biometrika*, **97**, 825–838. 2, 10, 12

Box, G.E. & Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211–252. 25, 26, 27

Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the US wage structure, 1963–1987. *Journal of Econometrics*, **65**, 109–154. 26, 28

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In *Advances in Econometrics: Sixth World Congress*, vol. 2, 171–209. 26

Dunson, D.B. & Taylor, J.A. (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics*, **17**, 385–400. 2, 11, 12, 13, 18, 115

Feng, Y., Chen, Y., He, X. *et al.* (2015). Bayesian quantile regression with approximate likelihood. *Bernoulli*, **21**, 832–850. 2, 20, 74, 115

Fin, D.E.A.F., Gerrans, P., Singh, A.K. & Powell, R. (2009). Quantile regression: Its application in investment analysis. *The Finsia Journal of Applied Finance*, **4**, 7. 1

Fitzenberger, B., Wilke, R.A. & Zhang, X. (2009). Implementing Box–Cox quantile regression. *Econometric Reviews*, **29**, 158–181. 26, 29

Garthwaite, P.H., Fan, Y. & Sisson, S.A. (2016). Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. *Communications in Statistics-Theory and Methods*, **45**, 5098–5111. 96

Geraci, M. & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140–154. 2, 14, 69

Hahn, P.R. & Burgette, L.F. (2012). The mesa distribution: an approximation likelihood for simultaneous nonlinear quantile regression. *University of Chicago, Tech. Rep*. 2, 19, 74, 115

Harrison, D. & Rubinfeld, D.L. (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management*, **5**, 81–102. 56

He, X. & Zhu, L.X. (2003). A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, **98**, 1013–1022. 2

Hendricks, W. & Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American statistical Association*, **87**, 58–68. 1

Hyndman, R.J. (2013). *hdrcde: Highest density regions and conditional density estimation*. R package version 3.1. 146

Hyndman, R.J., Bashtannyk, D.M. & Grunwald, G.K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, **5**, 315–336. 146

Hyndman, R.J. & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, **50**, 361–365. 3

Jeffreys, H. (1961). *Theory of Probability*. Oxford, 3rd edn. 18, 19

Jiang, X., Jiang, J. & Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, **22**, 1479–1506. 2

John, J. & Draper, N. (1980). An alternative family of transformations. *Applied Statistics*, **29**, 190–197. 29

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, vol. 2. Wiley, 2nd edn. 73

Koenker, R. (2005). *Quantile regression*. Cambridge university press. 2, 3, 40

Koenker, R. (2016). *quantreg: Quantile Regression*. R package version 5.26. 6, 47, 61

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33–50. 1, 3, 69

Koenker, R. & Bassett, G. (1982). Robust tests of heteroscedasticity based on regression quantiles. *Econometrica*, **50**, 43–61. 59

Koenker, R. & Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, **15**, 143–156. 2

Koenker, R. & Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of American Statistical Association*, **94**, 1296–1310. 70

Kottas, A. & Gelfand, A.E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **96**, 1458–1468. 2

Kozumi, H. & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of statistical computation and simulation*, **81**, 1565–1578. 2, 12, 13, 16, 33, 50

Lancaster, T. & Jae Jun, S. (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics*, **25**, 287–307. 2, 11, 19, 115

Lavine, M. (1995). On an approximate likelihood for quantiles. *Biometrika*, **82**, 220–222. 18

Lum, K. & Gelfand, A.E. (2012). Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Analysis*, **7**, 235–258. 17

Machado, J.A. & Mata, J. (2000). Box–Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics*, **15**, 253–274. 26

Manly, B. (1976). Exponential data transformations. *The Statistician*, **25**, 37–42. 29

Meng, X.L. (1994). Posterior predictive p-values. *The Annals of Statistics*, **22**, 1142–1160. 149

Monahan, J.F. & Boos, D.D. (1992). Proper likelihoods for Bayesian analysis. *Biometrika*, **79**, 271–278. 18

Nadaraya, E.A. (1964). Some new estimates for distribution functions. *Theory of Probability & Its Applications*, **9**, 497–500. xvii, 126, 127

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, **6**, 7–11. 157

Powell, J.L. (1991). Estimation of monotonic regression models under quantile restrictions. In W.A. Barnett, J. Powell & G.E. Tauchen, eds., *Nonparametric and semiparametric methods in Econometrics*, chap. 14, 357–384, Cambridge University Press, New York. 25, 28

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 40, 107, 135

Reich, B.J., Bondell, H.D. & Wang, H.J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, **11**, 337–352. 2, 17, 115

Roberts, G.O., Gelman, A., Gilks, W.R. *et al.* (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability*, **7**, 110–120. 32, 97, 157

Rodrigues, T. & Fan, Y. (2017). Regression adjustment for noncrossing Bayesian quantile regression. *Journal of Computational and Graphical Statistics*, **26**, 275–284. 12, 14, 69

Sakia, R. (1992). The Box-Cox transformation technique: a review. *The statistician*, **41**, 169–178. 27

Serfling, R. (2011). Asymptotic relative efficiency in estimation. In *International encyclopedia of statistical science*, 68–72, Springer. 8

Sherwood, B., Wang, L. & Zhou, X.H. (2013). Weighted quantile regression for analyzing health care cost data with missing covariates. *Statistics in medicine*, **32**, 4967–4979. 1

Sriram, K., Ramamoorthi, R. & Ghosh, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian analysis*, **8**, 479–504. 70

Tsionas, E.G. (2003). Bayesian quantile inference. *Journal of statistical computation and simulation*, **73**, 659–674. 2, 13, 15, 17

Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*. Springer, New York, 4th edn. 56, 135

Walker, S. & Mallick, B.K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, **55**, 477–483. 2

Wang, L. (2008). Nonparametric test for checking lack of fit of the quantile regression model under random censoring. *Canadian Journal of Statistics*, **36**, 321–336. 125

Wang, X.F. (2012). *sROC: Nonparametric Smooth ROC Curves for Continuous Data*. R package version 0.1-2. 126

Wu, Y. & Liu, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and Its Interface*, **2**, 299–310. 2

Wuertz, D., Chalabi, Y., with contribution from Miklovic, M., Boudt, C., Chausse, P. & others (2016). *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. R package version 3010.82.1. 112

Xiong, W. & Tian, M. (2015). Simultaneous variable selection and parametric estimation for quantile regression. *Journal of the Korean Statistical Society*, **44**, 134–149. 2

Yu, K. & Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, **137**, 260–276. 2, 14, 69

Yu, K. & Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in StatisticsTheory and Methods*, **34**, 1867–1879. 5

Yu, Y. & Moyeed, R. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, **54**, 437–447. 2, 4, 7, 11, 12, 13, 14, 21, 23, 25, 69, 70

Zheng, J.X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, **14**, 123–138. 125

Zou, H. & Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, **36**, 1108–1126. 2, 10