ABSTRACT

DAS, PRIYAM. Bayesian Quantile Regression. (Under the direction of Subhashis Ghosal.)

Quantile regression (QR) is a fundamental approach to quantifying the relation a set of predictors and a response variable especially for skewed response data. In the first chapter, a literature overview has been provided about the existing quantile regression methods. We discuss the key methods from the literature in the field of quantile regression. A simulation study has been also provided to compare the performances of existing popular Bayesian and non-Bayesian methods for univariate case.

In the second chapter, we consider a Bayesian method for simultaneous quantile regression on a real variable. A representation of quantile function is given by a convex combination of two monotone increasing functions $\xi_1$ and $\xi_2$ not depending on the prediction variables. In a Bayesian approach, a prior is put on quantile functions by putting prior distributions on $\xi_1$ and $\xi_2$. The monotonicity constraint on the curves $\xi_1$ and $\xi_2$ are obtained through a spline basis expansion with coefficients increasing and lying in the unit interval. We put a Dirichlet prior distribution on the spacings of the coefficient vector. We compare our approach with a Bayesian method using Gaussian process prior through an extensive simulation study and some other Bayesian approaches proposed in the literature. Applications to a data on hurricane activities in the Atlantic region and population data of USA are given. This chapter is based on the material in Das and Ghosal (2016c).

In the third chapter, we extend our method of Bayesian quantile regression to the case where the predictor space is multidimensional and the quantile regression depends on the predictors through an unknown linear combination only. This chapter is based on the material in Das and Ghosal (2016c).

In the fourth chapter, we extend the proposed Bayesian method of quantile regression for spatio-temporal data. We develop a method for spatio-temporal simultaneous quantile regression. Unlike existing procedures, in the proposed method, smoothing across the sites is incorporated within modeling assumptions thus allowing borrowing of information across locations, an essential step when the number of samples in each location is low. The quantile function has been assumed to be linear in time and smooth over space. A B-spline basis expansion with increasing coefficients varying smoothly over the space
is used to put a prior and a Bayesian analysis is performed. We analyze the average daily 1-hour maximum and 8-hour maximum ozone concentration level data of US and California during 2006–2015 using the proposed method. This chapter is based on the material in Das and Ghosal (2016a).

In the fifth chapter, we consider Bayesian methods for non-parametric quantile regression with multiple continuous predictors ranging values in the unit interval. In the first method, the quantile function is assumed to be smooth in the explanatory variable and is expanded in tensor product of B-spline basis functions. While in the second method, the distribution function is assumed to be smooth in the explanatory variable and is expanded in tensor product of B-spline basis functions. Unlike other existing methods of non-parametric quantile regressions, the proposed methods estimate the whole quantile function instead of estimating on a grid of quantiles. The proposed methods have also been modified for quantile grid data where only the percentile range of each response observations are known. Simulations studies have been provided for both complete and quantile grid data. The proposed method has been used to estimate the quantiles of US household income data and North Atlantic hurricane intensity data. This chapter is based on the material in Das and Ghosal (2016b).
Bayesian Quantile Regression

by

Priyam Das

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2017

APPROVED BY:

______________________________  ________________________________
Alyson Wilson                  Brian Reich

______________________________
Yichao Wu

______________________________
Subhashis Ghoshal
Chair of Advisory Committee
DEDICATION

I would like to dedicate my whole work to my father Prasanta Kumar Das, my mother Pama Das, my brother Pramit Das, my aunt Rupali Mitra and my mentor Priyam Biswas, the people who always helped me and inspired me while going through the hardest time of my life.
BIOGRAPHY

Priyam Das is the younger son of Prasanta Kumar Das and Pama Das. He was born in Kolkata, India. He completed his 12 years of schooling in Hindu School during the period 1995–2007. After class 12, he got admission at Jadavpur University in the department of Electronics and Telecommunication engineering. After studying at Jadavpur for one year, he decided to change his line of study and got admitted to Indian Statistical Institute in 2008. He completed his bachelors and masters degrees in statistics (B.Stat and M.Stat) in 2011 and 2013 respectively. Completing the master’s degree, he joined Ph.D. program in the Department of Statistics at North Carolina State University in August, 2013. The main topics of his research includes Bayesian quantile regression and it’s spatial applications, non-convex constrained and unconstrained optimization techniques and Bayesian methodologies in modeling water data of US.

Other than being a statistician, Priyam is also a great Hip-hop dancer and teacher known under the nickname “La Krusade”. His dancing styles are namely popping, dub-step, liquid, robiting, krumping and tutting. He won several professional solo dance competitions and dance battles among which winning the solo dance competition ‘Sanskriti – 2009’ at Jadavpur University is considered as one of his best achievements of his dancing career. He also gave paid guest performances in several occasions. But unfortunately, following three surgeries related to his sports injuries, he is currently in rehabilitation with a hope to start dancing soon again.

Priyam is also a great sportsman who won ‘Best sportsman award’ at Indian Statistical Institute for two consecutive years in 2011 and 2012. Most of his medals were from the sports categories high jump, long jump, arm-wrestling, shot put and walking race. Priyam is a Taekwondo green belt certified by the world Taekwondo headquarter Kukkiwon. He has some training experience in Mixed martial arts as well. He also served as the skipper goalkeeper of the soccer team of Indian Statistical Institute.

Priyam is also a dedicated power-lifter belonging to 1100 lbs power-lifting club (squat 410 lbs, deadlift 435 lbs, bench-press 260 lbs). He is also a meme-artist and owner of three facebook meme pages namely ‘ISI Troll’ (5700+ followers), ‘Priyam aka La Krusade - La’s Troll’ (5400+ followers) and ‘The Fat Bro Rises’ (1500+ followers). Priyam composed several bengali songs among which “Bheshe Jay” remained quite popular among the audience. Priyam can juggle up to 4 balls or rings. Priyam likes leg days, food and handsome
amount of sleep. One of his motivating quotes which summarizes his life in a nut-shell is “INSPIRE or DIE”.
ACKNOWLEDGEMENTS

First, I would like to thank Dr. Subhashis Ghoshal, my Statistics Ph.D. advisor. He has been always with me during my journey starting from being a novice statistician to a dedicated researcher. I am also grateful to Dr. Alyson Wilson, Dr. Sankar Arumugan, Dr. Brian Reich and Dr. Yichao Wu for being in my thesis committee. I would like to thank Dr. Hua Zhou for enriching my computational skills which helped me a lot in improving my computational skills. The research work given in the Chapter 3 was partially supported by NSF grant number DMS 1510238.

I am thankful to my best buddy Longshaokan Wang (Marshall) for helping me through tough times and for being there always to accompany me with good and evil thoughts. I am also really grateful to Debraj Das for helping me in difficult times. Besides, I am grateful to my friends and colleagues Srinjoy Chattopadhyay, Cassie Coleman-Smith, Arnab Bose, Arnab Hazra, Sohini Raha, Moumita Chakraborty, Indranil Sahoo, Arkaprava Roy, Arunava Nag, Sujatro Chakladar, Arkopal Choudhury, Sayan Patra, Poulamit Maitra, Mityl Biswas, Sudarshana Mukhopadhyay, Abhishek Banerjee, Ajay Kumar, Arnab Chakraborty, Kiran Nihlani, Debashis Chatterjee, Bharadwaja Srimat Tirumala Peddinti, Ankit Dubey, Suman Majumder, Rahul Ghosal, Indrabati Bhattacharya, Cai Li, So Young Park, Merve Yasemin Tekbudak, Marcela Alfar. Among my seniors, I am thankful to Rajarshi Das Bhowmik, Sayantan Banerjee, Prithwish Bhaumik, Jyotishka Datta, Shalini Choudhury, Souvik Chandra, Saswata Sahoo, Anuradha Biswas, Samarpan Majumder, Rinku Majumder. Among my distant friends, I am really glad to Sayan Roy, Deborshee Sen, Raka Mandal, Atin Ghosh, Rounak Dey for checking on me so regularly.

My heartiest thanks go to Dr. Shuva Gupta for motivating me what I am doing today, maintaining perfect balance between research work and physical fitness. Besides I am thankful to Abdullah Mughrabi, Mike Chang for motivating me into the world of fitness and bodybuilding. I would like to thank Eminem, Lil Wayne, Tupac, Nicki Minaj, Rupam Islam among the greatest music artists of my inspiration. Among the dancers, I would like to thank my biggest inspirations Robert Muraine, Les Twins, Chadd Smith, Tushar Kapoor, Harshvardhan Bhan, Anurag Ghosh, Siddharth Basu Roy, Rahul Roychowdhury, Mekhola Bose. I am also grateful to my Taekwondo instructor Master Pradipta Kumar Roy, Master Ruma Roy Chowdhury and my soccer coach Trijit Das. Among the other
helpful hands and inspirations I must thank Dr. Sourabh Bhattacharya and Dr. Arunava Goswami for helping me in the process of application to this university from my country, India. I am also really grateful to Dr. Alyson Wilson and Dr. Sankar Arumugam for their immense help throughout my application procedure for job search towards the end of my graduate program.
# Table of Contents

LIST OF TABLES ....................................................................................... ix

LIST OF FIGURES ...................................................................................... xi

Chapter 1 Introduction ............................................................................. 1
  1.1 Frequentist single Quantile Regression methods .......................... 1
  1.2 Bayesian single Quantile Regression ............................................. 2
  1.3 Drawbacks of single quantile regression ......................................... 4
  1.4 Non-crossing quantile regression .................................................... 6
  1.5 Simultaneous Quantile Regression .................................................. 10

Chapter 2 Univariate Regression Using B-spline Series Prior .............. 15
  2.1 Introduction ................................................................................. 15
  2.2 Model Assumptions ....................................................................... 16
  2.3 Regression with Spline ................................................................. 17
    2.3.1 Model Fitting .......................................................................... 19
    2.3.2 MCMC and Transition Step ................................................... 19
    2.3.3 Choosing value of $k$ and Model Averaging ........................... 22
  2.4 Large-sample Properties ............................................................... 25
  2.5 Simulation Study ........................................................................... 27
    2.5.1 First Study ............................................................................. 30
    2.5.2 Second Study .......................................................................... 31
  2.6 Application to Hurricane Intensity Data ....................................... 36
  2.7 Application to US Population Data ............................................... 40
  2.8 Conclusion .................................................................................... 44

Chapter 3 Single Index Model for Multivariate Predictors ................. 46
  3.1 Introduction ................................................................................. 46
    3.1.1 Likelihood evaluation, MCMC and Transition ....................... 47
    3.1.2 Choosing value of $k$ and Model Averaging ........................... 48
  3.2 Simulation Study ........................................................................... 48

Chapter 4 Bayesian Spatio-temporal Quantile Regression ................. 51
  4.1 Introduction ................................................................................. 51
  4.2 Ozone in troposphere ................................................................. 52
  4.3 Proposed Bayesian Method .......................................................... 53
    4.3.1 Prior ....................................................................................... 54
    4.3.2 Block Metropolis-Hastings MCMC algorithm ....................... 58
  4.4 Simulation study ........................................................................... 59
Chapter 5 Bayesian Non-parametric Simultaneous Quantile Regression for Complete and Grid Data

5.1 Introduction .......................................................... 76
5.2 Proposed Bayesian Method ......................................... 77
  5.2.1 Non-parametric Modeling of Quantile Function ............ 77
  5.2.2 Non-parametric Modeling of Distribution Function ........ 78
5.3 Likelihood Evaluation .............................................. 79
  5.3.1 Complete Data ................................................ 79
  5.3.2 Quantile Grid Data ........................................... 80
5.4 Block Metropolis-Hastings MCMC Algorithm .................... 81
  5.4.1 Warm Start .................................................. 83
  5.4.2 Automatic Controlling of Acceptance Probability .......... 84
5.5 Simulation study ................................................ 84
  5.5.1 Case of Complete Data ...................................... 85
  5.5.2 Case of Grid data ........................................... 90
5.6 Application to Hurricane Data ................................... 91
5.7 Application to US household income data ....................... 95
5.8 Conclusion ....................................................... 97

References ..................................................................... 100
LIST OF TABLES

Table 2.1 (First Simulation study) Comparison of RMISE of estimation of slope, intercept and quantile regression function at \( x = 0.3, 0.5, 0.7, \tau = 0.25, 0.50, 0.75 \) and \( n = 100 \) under different methods QSSQR(HB); QSSQR(EB); CSSQR(HB); CSSQR(EB); GSQR; BPSQR; BSquare 1 and BSquare 2 (denote qreg and qreg.spline function under BSquare package in \texttt{R} respectively). ............................................ 34

Table 2.2 (First Simulation study) Size and posterior coverage of inflated (in bold) and regular uniform 95\% posterior credible interval of estimated quantile regression function for \( x = 0.2, 0.5, 0.7 \) for \( \tau \in [0, 1] \) for three different sizes of sample \( n = 50, 100, 200 \) for QSSQR(HB), QSSQR(EB), CSSQR(HB) and CSSQR(EB). .................................................... 35

Table 2.3 (Second Simulation study) Comparison of RMISE of estimation of slope, intercept and quantile regression function at \( x = 0.2, 0.5, 0.7, \tau = 0.25, 0.50, 0.75 \) and \( n = 100 \) under different methods. QSSQR(HB); QSSQR(EB); CSSQR(HB); CSSQR(EB); GSQR; qreg; qreg.spline and BPSQR. .......................................................... 36

Table 2.4 (Second Simulation study) Size and posterior coverage of inflated (in bold) and regular uniform 95\% posterior credible interval of estimated quantile regression function for \( x = 0.2, 0.5, 0.7 \) for \( \tau \in [0, 1] \) for three different sizes of sample \( n = 50, 100, 200 \) for QSSQR(HB), QSSQR(EB), CSSQR(HB) and CSSQR(EB). .................................................... 38

Table 2.5 Computation time (in seconds) of QSSQR, CSSQR and GSQR for simulation study 1 and 2 with sample size \( n = 100 \). .......................................................... 40

Table 3.1 (Simulation study : Multivariate) Comparison of RMISE of estimation of quantile regression functions (QRF) at the Points 1-8 which are \((X_1, X_2, X_3) = (-0.5, -0.5, -0.5), (-0.5, -0.5, 0.5), (-0.5, 0.5, -0.5), (-0.5, 0.5, 0.5), (0.5, -0.5, -0.5), (0.5, -0.5, 0.5), (0.5, 0.5, -0.5), (0.5, 0.5, 0.5)\) respectively under different methods QSSQR(HB); QSSQR(EB); CSSQR(HB); CSSQR(EB); GSQR; qreg; qreg.spline and BPSQR. .......................................................... 49

Table 4.1 Average MSE (based on \( S = 50 \) repetitions of simulation study) of estimated slope, intercept and estimated quantile function value at three time-points \( X = 0.2, 0.5, 0.8 \) using SSTQR (both MLE based and Bayesian approach) and GBSTQR over all \( L = 50 \) locations and \( T = 19 \) quantile levels \( \tau = 0.05, 0.10, \ldots, 0.95 \). .......................................................... 62
Table 5.1  (First Simulation Study (Complete Data)) Prediction Mean Squared Errors using NPSQR, NPDFSQR, LLQR and LSQR based on simulation studies for sample sizes $n = 50, 100, 200$.  

Table 5.2  (Second Simulation Study (Complete Data)) Prediction Mean Squared Errors using NPSQR, NPDFSQR, LLQR and LSQR based on simulation studies for sample sizes $n = 50, 100, 200$.  

Table 5.3  (First Simulation Study (Grid Data)) Prediction Mean Squared Errors using NPSQR and NPDFSQR based on simulation studies with 5, 10, 20 distant percentile grids for sample sizes $n = 50, 100, 200$.  

Table 5.4  (Second Simulation Study (Grid Data)) Prediction Mean Squared Errors using NPSQR and NPDFSQR based on simulation studies with 5, 10, 20 distant percentile grids for sample sizes $n = 50, 100, 200$.  

Table 5.5  A snapshot of US household income data table showing the income distribution of all population during 2010 – 2015. 
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>P-values after fitting single quantile regression (KB method) curve for $\tau \in {0.01, 0.02, \ldots, 0.99}$ in north Atlantic cyclone intensity data to test the significance of the slope term.</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Estimated $Q(\tau</td>
<td>x)$ at $\tau = 0.1$ and $\tau = 0.2$ over the period 1981–2006 by fitting single quantile regression curve in north Atlantic cyclone intensity data with KB method.</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Estimated quantile curves of wind speed of cyclones over the period 1981 – 2006 at North Atlantic for $\tau = {0.05, 0.1, 0.2, \ldots, 0.8, 0.9, 0.95}$ using (a) frequentist approach (KB method) and (b) Bayesian approach (YM method).</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Estimated quantile curves of wind speed of cyclones over the period 1981 – 2006 at North Atlantic for $\tau = {0.05, 0.1, 0.2, \ldots, 0.8, 0.9, 0.95}$ using (a) RFD method and (b) Re method for non-crossing QR.</td>
<td>10</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>(First Simulation study) Comparison of true and estimated intercept, slope and estimated quantile regression functions (QRF) at $x = 0.3, 0.5, 0.7, \tau = 0.25, 0.5, 0.75$ and $n = 100$ for different methods QSSQR(HB), GSQR, qreg, qreg_spline and BPSQR.</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>(First Simulation study) MCMC auto-correlation and trace plots of $\xi_1(\tau)$ at $\tau = 0.25, 0.5, 0.75$ and $\xi_2(\tau)$ at $\tau = 0.25, 0.5, 0.75$ respectively in row 1–6 for QSSQR ($k = 6$), CSSQR ($k = 9$) and GSQR.</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>(Second Simulation study) Comparison of true and estimated intercept, slope and estimated quantile regression functions (QRF) at $x = 0.2, 0.5, 0.7$, $\tau = 0.25, 0.5, 0.75$ and $n = 100$ for different methods QSSQR(HB), GSQR, qreg, qreg_spline and BPSQR.</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>(a) Comparison of $\text{wmaxST}$ during first and last 10 years: The dotted superimposed line is assumed power-Pareto density used for transforming velocities; the solid superimposed lines are approximate density of $\text{wmaxST}$ for corresponding decades. (b) Simultaneous Quantiles: Estimated quantile curves of velocities over the period 1981–2006 has been shown for the quantiles $\tau \in {0.05, 0.10, 0.20, \ldots, 0.80, 0.90, 0.95}$. (c) $\text{wmaxST}$ as a function of quantiles for the years 1981, 1986, 1991, 1996, 2001 and 2006. (d) Posterior probability of slope being negative for $\tau \in [0.5, 1]$.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Estimated quantile curves of county wise population of Northeast, Midwest, South and West regions of USA for the quantiles $\tau \in {0.05, 0.10, 0.20, \ldots, 0.80, 0.90, 0.95}$ over the years 1985–2010.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Estimated population for $\tau \in [0, 1]$ of Northeast, Midwest, South and West regions of USA for the years 1985, 1990, 1995, 2000, 2005 and 2010.</td>
<td>42</td>
</tr>
</tbody>
</table>
Figure 2.7  Estimated probability of negative slope for $\tau \in [0,1]$ of Northeast, Midwest, South and West regions of USA for the years 1985, 1990, 1995, 2000, 2005 and 2010 .................................................. 43

Figure 3.1  (Simulation study : Multivariate) Comparison of true and estimated quantile regression functions (QRF) at $(X_1, X_2, X_3) = (\pm 0.5, \pm 0.5, \pm 0.5)$ for $n = 100$ using QSSQR(HB), GSQR, qreg, qreg_spline and BPSQR. 50

Figure 4.1  Climate region-wise division of the US. ........................................... 63

Figure 4.2  Daily 1-hour maximum average ozone concentration (in ppb) of the US in 2006, 2010 and 2015 at $\tau = 0.5$. The dots denote weather stations where data have been collected. ................................. 65

Figure 4.3  Yearly rate of change of daily 1-hour maximum average ozone concentration (in ppb/year) of the US at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected. ......................... 66

Figure 4.4  Daily 8-hour maximum average ozone concentration (in ppb) of the US in 2006, 2010 and 2015 at $\tau = 0.5$. The dots denote weather stations where data have been collected. 67

Figure 4.5  Yearly rate of change of daily 8-hour maximum average ozone concentration (in ppb/year) of the US at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected. 68

Figure 4.6  Quantiles at which 4th highest daily 8-hour maximum ozone concentration (in ppb) of the US is 70 ppb in 2006, 2010 and 2015. .......................... 69

Figure 4.7  Region-wise division of California. .................................................. 70

Figure 4.8  (a-c) Daily 1-hour maximum average ozone concentration (in ppb) of California in 2006, 2010 and 2015 at $\tau = 0.5$. (d-f) Yearly rate of change of daily 1-hour maximum average ozone concentration (in ppb/year) of California at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected. ................................. 72

Figure 4.9  (a-c) Daily 8-hour maximum average ozone concentration (in ppb) of California in 2006, 2010 and 2015 at $\tau = 0.5$. (d-f) Yearly rate of change of daily 8-hour maximum average ozone concentration (in ppb/year) of California at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected. ................................. 73

Figure 4.10 (a-c) Quantiles at which 4th highest daily 8-hour maximum ozone concentration (in ppb) of California is 70 ppb in 2006, 2010 and 2015. (d-f) Quantiles at which 4th highest daily 1-hour maximum ozone concentration (in ppb) of California is 90 ppb in 2006, 2010 and 2015. 74

Figure 5.1  (First Simulation Study (Complete Data)) True and estimated quantiles at $\tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR, NPDFSQR, LLQR and LSQR with the data points. .......... 87
Figure 5.2 (Second Simulation Study (Complete Data)) True and estimated quantiles at $\tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR, NPDFSQR, LLQR and LSQR with the data points. .................. 88

Figure 5.3 (First Simulation Study (Grid Data)) True and estimated quantiles at $\tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR and NPDFSQR with the data points. .................. 92

Figure 5.4 (Second Simulation Study (Grid Data)) True and estimated quantiles at $\tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR and NPDFSQR with the data points. .................. 93

Figure 5.5 Simultaneous quantiles of hurricane velocities in North Atlantic region during 1981-2006. ................................. 95

Figure 5.6 Estimated simultaneous quantiles of household income of (a) All population (1967 – 2015) (b) Asian (2002 – 2015) (c) Black (1967 – 2015) (d) Hispanic (1972 – 2015) (e) White (1967 – 2015) and (f) White Non-Hispanic (1972 – 2015) population at $\tau = \{0.05, 0.10, 0.20, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR. ................................. 98
Chapter 1

Introduction

In a regression model, if the distribution of the response variable is highly skewed, traditional mean regression model may fail to describe interesting aspect of the relationship between the prediction and response variables. For instance, if we deal with the income distribution data of a certain state or country, the traditional mean regression will get affected by the outliers, i.e., the income of top 1–2 % people of that region, and hence is of limited use for prediction of income of general people. For example, as mentioned in this link\(^1\), more than 35% of the total wealth of the US belongs to top 1% people. So it can be understood that this distribution being heavy-tailed, regular mean regression might not be a good representative of the average per-capita wealth of the US. This situation arises often in business, economics, environmental and many other fields. As an alternative to traditional linear regression, quantile regression (QR) is one of the most popular and useful regression technique.

Frequentist single Quantile Regression methods

Quantile regression (QR) was first proposed in Koenkar and Bassett (1978). Suppose our univariate observations are given by \( \{Y_t : t = 1, \ldots, T\} \). So the \( \tau \)-th quantile \((0 < \tau < 1)\) of the observations can be found solving the optimization problem given by

\[
\min_{b \in \mathbb{R}} \left[ \sum_{t \in \{t : Y_t \geq b\}} \tau(Y_t - b) + \sum_{t \in \{t : Y_t < b\}} (1 - \tau)(b - Y_t) \right].
\]

It is noted that putting \( \tau = \frac{1}{2} \) in the above equation yields the median. But, this more general form to find the quantiles is helpful specially when we are interested in the

\(^1\)Source https://en.wikipedia.org/wiki/Recent_Distribution_of_Wealth_in_the_World
behavior of the tail distributions. In case we have the explanatory variables \( \{X_t \in \mathbb{R}^k : t = 1, \ldots, T, k \in \mathbb{N}\} \), the \( \tau \)-th \((0 < \tau < 1)\) linear quantile regression coefficient \( \beta \in \mathbb{R}^k \) is given by

\[
\min_{b \in \mathbb{R}} \left[ \sum_{t \in \{t : Y_t \geq X_t \beta\}} \tau |Y_t - X_t \beta| + \sum_{t \in \{t : Y_t < X_t \beta\}} (1 - \tau) |Y_t - X_t \beta| \right].
\]

Thus the method proposed by Koenkar and Bassett (1978) (KB method) can be used to find the linear regression coefficients for any specified quantile.

Due to computational convenience and other theoretical properties, the method proposed by Koenkar and Bassett (1978) remained popular for long time. Later on, several works emerged studying the asymptotic properties and procedures of estimating its standard error along with some extensions and improvements of single quantile regression from the frequentist’s point of view; see Gutenbrunner and Jureckova (1992), Gutenbrunner et al. (1993), Koenker and Xiao (2002), Zhou and Portnoy (1996), Koenker and Machado (1999). Geraci and Bottai (2007) proposed the single level quantile regression model with subject specific random intercept term which accounts for within group correlation.

**Bayesian single Quantile Regression**

One of the first attempts for Bayesian quantile regression was proposed in Yu and Moyeed (2001) (YM method). Consider the standard form of the linear model

\[
Y_t = \mu(X_t) + \epsilon_t
\]

where \( \mu(X_t) \) is the conditional mean of \( Y_t \) given the regressors \( X_t \) and error \( \epsilon_t \) with mean zero and constant variance. The distribution of \( \epsilon_t \) is not needed to be specified. Typically, for incorporation of linear relationship of the regressors with the explained variable, assume \( \mu(X_t) = X_t'\beta \). Then the \( \tau \)-th \((0 < \tau < 1)\) conditional quantile of \( Y_t \) given \( X_t \) is given by

\[
q_{\tau}(Y_t|X_t) = X_t'\beta(\tau)
\]

where \( \beta(\tau) \) is the coefficient vector which depends on the value of \( \tau \) \((0 < \tau < 1)\). As mentioned in Koenkar and Bassett (1978), for any value of \( \tau \) in \((0, 1)\), \( \beta(\tau) \) can be found.
by solving the minimization problem

$$\min_{\beta} \sum_t \rho_{\tau}(Y_t - X'_t\beta(\tau)),$$

Yu and Moyeed (2001) showed that minimization of the loss function $\rho_{\tau}(u)$ is equivalent to the maximization of a likelihood function formed with combination of asymmetric Laplace densities which are independently distributed. The probability density function of random variable $U$ coming from asymmetric Laplace distribution is given by

$$f_{\tau}(u) = \tau (1 - \tau) \exp\{-u(\tau - I(u < 0))\} = \tau (1 - \tau) \exp\{\rho_{\tau}(u)\}$$

(1.1)

where $0 < \tau < 1$. It is noted that for all values of $\tau$ except $\frac{1}{2}$, the density given by (1.1) is asymmetric with

$$E(U) = \frac{1 - 2\tau}{\tau(1 - \tau)}, \quad V(U) = \frac{1 - 2\tau + 2\tau^2}{\tau^2(1 - \tau^2)}.$$

Note that the variance term increases very rapidly whenever $\tau$ approaches to 0 or 1. Further after incorporating location and scale parameters in Equation (1.1), we get

$$f_{\tau}(u; \mu, \sigma) = \frac{\tau(1 - \tau)}{\sigma} \exp\{-\rho_{\tau}(\frac{u - \mu}{\sigma})\}.$$  

(1.2)

Suppose we have data $Y = \{Y_t : t = 1, \ldots, n\}$. To estimate $\beta(\tau)$, i.e., the coefficients of the $\tau$-th quantile curve, we need to maximize the likelihood function $L(\beta|Y)$ given by

$$L(\beta|Y) = \tau^n(1 - \tau)^n \exp\left\{-\sum_{i=1}^{n} \rho_{\tau}(Y_i - X'_i\beta)\right\}$$

using Equation (1.1) with a location parameter $\mu_i = X'_i\beta$. Yu and Moyeed (2001) adopted a Bayesian approach to estimate the values of $\hat{\beta}(\tau)$ for any $0 < \tau < 1$. Let $\pi(\beta)$ denote the prior distribution of the coefficient vector $\beta$. Then the posterior distribution is given by

$$\pi(\beta|Y) \propto L(\beta|Y)\pi(\beta)$$

If no prior information is available, improper uniform prior distribution could be used on each component of $\beta$. Yu and Moyeed (2001) also showed that choosing improper uniform prior on $\beta$ yields a proper joint posterior distribution. A few other generalization and
extension of quantile regression were proposed in Kottas and Gelfand (2001), Gelfand and Kottas (2003) and Kottas and Krnjajic (2009)). For continuous dependent variable, this method was further extended with incorporation of adaptive lasso variable selection in Alhamzawi et al. (2012). Benoit and Van den Poel (2012) extended it for binary dependent variable and later Benoit et al. (2013) incorporated adaptive lasso for binary quantile regression.

**Drawbacks of single quantile regression**

Although single quantile regression remained popular for a long period of time, there are some drawbacks which comes into the scenario while using it. If our motivation is to find a single quantile curve, these above-mentioned methods seem to work fine. But, to compare the estimated quantile curves for multiple quantiles, the drawbacks of single quantile curve estimation becomes more visible. If we want to draw inference on the quantile regression coefficients simultaneously for a range of quantile values, we might face some practical as well as philosophical issues. For example we analyze the North Atlantic hurricane intensity data² over the period 1981–2006 with the quantile regression method proposed in Koenkar and Bassett (1978). Elsner et al. (2008) argued that that the strongest hurricanes in the North Atlantic basin have gotten stronger over the last couple of decades. Here our explanatory variable is year and the response variable is the wind speed. Suppose the quantile regression curve is given by

\[ Q(\tau|x) = \beta_0(\tau) + x'\beta(\tau) \] for \( 0 \leq \tau \leq 1 \).

Similar to Tokdar and Kadane (2012), to perform the hypothesis test

\[ H_0 : \beta(\tau) = 0 \text{ vs } H_1 : \beta(\tau) \neq 0 \]

we fit separate quantile curves for each \( \tau \in \{0.01, 0.02, \ldots, 0.99\} \) in North Atlantic cyclone intensity data and calculate the p-values for each of them. In Figure 1.1, it is noted that the p-values fluctuate a lot across the quantiles. For example, in this case, we note that p-value drops from 1 to \( 3.56 \times 10^{-5} \) as we move from \( \tau = 0.14 \) to \( \tau = 0.15 \). That indicates a poor borrowing of information across the different quantile estimation. So it might be totally inconclusive if we want to draw inference on \( \beta(\tau) \) using this pooling method of single quantile regression for even small intervals.

Figure 1.1: P-values after fitting single quantile regression (KB method) curve for $\tau \in \{0.01, 0.02, \ldots, 0.99\}$ in north Atlantic cyclone intensity data to test the significance of the slope term.

Another philosophical issue we face while using the single quantile regression techniques is that the monotonicity of the quantile curves are not ensured while estimating them separately using these methods. Which means that there is a possibility that for a given value of the explanatory variable, the predicted value of the response value at $\tau = \tau_1$ might be smaller than it’s predicted value at some lower quantile level $\tau = \tau_2$ ($\tau_2 < \tau_1$). In other words, $\hat{Q}(\tau_1|x)$ might be smaller than $\hat{Q}(\tau_2|x)$ even if $\tau_2 < \tau_1$.

In Figure (1.2), the obtained value of $\hat{Q}(\tau|x)$ after fitting single quantile regression (Koenkar and Bassett (1978)) to the north Atlantic cyclone intensity data, has been plotted for $\tau = 0.1$ and $\tau = 0.2$. It is noted that the estimated quantile curves cross each other violating the monotonicity property of the quantiles.

In Figure (1.3) the simultaneous quantiles of the wind velocities at north Atlantic have been plotted for during the period 1981–2006 using both frequentist (Koenkar and Bassett (1978)) and Bayesian (Yu and Moyeed (2001)) single quantile regression techniques. We use quantreg (Koenker et al. (2016)) and bayesQR (Benoit et al. (2015)) packages in R for implementing the frequentist and the Bayesian approach respectively. For the Bayesian method, in MCMC, we perform 20000 iterations and disregard first 5000 as the burn-in. It is to be noted that the lower quantile curves of the frequentist method have crossed each other violating the monotonicity property of the quantile regression.
Figure 1.2: Estimated $Q(\tau|x)$ at $\tau = 0.1$ and $\tau = 0.2$ over the period 1981–2006 by fitting single quantile regression curve in north Atlantic cyclone intensity data with KB method.

Non-crossing quantile regression

Due to the drawbacks of the single quantile regression approach, several methods emerged addressing the crossing issue while estimating multiple quantiles simultaneously. He (1997) proposed a method of estimating non-crossing quantile curves assuming a heteroskedastic model for response variable. Under this assumption, predictors might affect the response variable via a location and scale change. Neocleous and Portnoy (2008) proposed a method of estimating the quantile curve by linear interpolation on the estimated quantile curves on a grid of quantiles. Takeuchi (2004) and Takeuchi et al. (2006) used support vector machines (SVM, Vapnik (1995)) for non-crossing quantile regression. But, as mentioned in Shim and Lee (2010), a disadvantage of using SVM for non-crossing quantile regression is that when multiple quantiles are needed, every adjacent pair of conditional quantile functions should be computed. Shim et al. (2009) proposed non-crossing quantile regression using doubly penalized kernel machine (DPKM). For this method also, we need to compute quantile curves separately for each quantile. The method proposed in Wu and Liu (2009) sequentially updates the quantile curves under the constraint that the upper quantile curves stay above the lower ones. As mentioned in Bondell et al. (2010), a drawback of this method is that the predicted quantile curves are dependent on the grid of quantiles for which we want to calculate the quantile curves. Quantile regression for
Figure 1.3: Estimated quantile curves of wind speed of cyclones over the period 1981 – 2006 at North Atlantic for $\tau = \{0.05, 0.1, 0.2, \ldots, 0.8, 0.9, 0.95\}$ using (a) frequentist approach (KB method) and (b) Bayesian approach (YM method).

A fixed number of levels of quantiles with monotonicity constraint was proposed in Dunson and Taylor (2005), Liu and Wu (2011) and Bondell et al. (2010).

Reich et al. (2011) (RFD method) proposed spatio-temporal quantile regression which is regarded as one of the first articles addressing the non-crossing issue of quantile regression for analyzing spatio-temporal data. They proposed a two-stage model where at the first stage the quantile regression coefficients are found using KB method for any desired grid of quantiles. Then the monotonicity of the quantile curves are preserved re-estimating the estimated coefficients found by KB method with Bernstein polynomial basis function with some constraints on the coefficients of the basis function. At first the explanatory variables are transformed to the unit interval. So, as long as the coefficients of each components of the explanatory variables are non-decreasing function of $\tau$, the quantile function also remains non-decreasing since the values of the explanatory variables are non-negative. Consider the following basis expansion of any coefficient term

$$\beta(\tau) = \sum_{i=1}^{M} B_m(\tau) \alpha_m$$

where $M$ is the number of basis functions, $\alpha_m$ are unknown coefficients that controls the shape of the quantile function and $B_m(\tau)$ is a Bernstein polynomial basis function of $\tau$ which is given by

$$B_m(\tau) = \binom{M}{m} \tau^m (1 - \tau)^{M-m}.$$
By the property of the Bernstein polynomial basis functions, if we ensure \( \alpha_m \geq \alpha_{m-1} \) for \( m = 2, \ldots, M \), then \( \beta(\tau) \) is non-decreasing function of \( \tau \). First, consider the intercept only model where the quantile function \( q(\tau) \) is same as \( \beta(\tau) \). Define \( \delta_m = \alpha_m - \alpha_{m-1} \) for \( m = 2, \ldots, M \). So the basis function coefficients can be written as \( \alpha_m = \sum_{l=1}^{m} \delta_l \) for \( m = 2, \ldots, M \) and \( \alpha_1 = \delta_1 \). Then a latent unconstrained variable \( \delta^*_m \) is introduced taking \( \delta_1 = \delta^*_1 \) and

\[
\delta_m = \delta^*_m I[\delta^*_m \geq 0].
\]

Independent normal priors \( \delta^*_m \sim N(\bar{\delta}_m(\Theta), \sigma^2) \) are put for sampling where \( \Theta \) is some set of hyper-parameters may depend on various factors like spatial location etc., and \( \bar{\delta}_m(\Theta) \) is picked for centering the quantile process based on parametric distribution \( f_0(y|\Theta) \). Suppose, \( q_0(\tau|\Theta) \) be the quantile function of \( f_0(y|\Theta) \), then \( \bar{\delta}_m(\Theta) \) are chosen in such a way that

\[
q_0(\tau|\Theta) \approx \sum_{m=1}^{M} B_m(\tau)\bar{\alpha}_m(\Theta).
\]

where \( \bar{\alpha}_m(\Theta) = \sum_{l=1}^{m} \bar{\delta}_l(\Theta) \). The value of \( \bar{\delta}_m(\Theta) \) are obtained corresponding to the ridge regression estimator

\[
(\bar{\delta}_1(\Theta), \ldots, \bar{\delta}_M(\Theta)) = \arg \min_{d} \sum_{k=1}^{K} \left( q_0(\tau_k|\Theta) - \sum_{m=1}^{M} B_m(\tau_k)\left[ \sum_{l=1}^{m} d_l \right] \right) + \lambda \sum_{m=1}^{M} d_m^2,
\]

where \( d_m \geq 0 \) for \( m = 2, \ldots, M \), and \( \{\tau_1, \ldots, \tau_K\} \) is a dense grid on \((0, 1)\). They add the ridge penalty term \( \lambda \sum_{m=1}^{M} d_m^2 \) with \( \lambda = 1 \) for numerical stability.

Now consider the quantile process with covariates. For the \( i \)-th covariate, the conditional quantile function is given by

\[
q(\tau|X_i) = X'_i \beta(\tau) = \sum_{j=1}^{p} X_{ij} \beta_j(\tau)
\]

where \( \beta_j(\tau) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm} \) for unknown coefficients \( \alpha_{jm} \). Here again, \( \beta_j(\tau) \) is constructed in a way so that \( q(\tau|X_i) \) remains non-decreasing in \( \tau \) for all \( X_i \). The quantile
function can be written in the following form

\[ q(\tau | X_i) = X'_i \beta(\tau) = \sum_{m=1}^{M} B_m(\tau) \theta_m(X_i), \]

where \( \theta_m(X_i) = \sum_{j=1}^{p} X_{ij} \alpha_{jm} \). Thus ensuring the condition \( \theta_m(X_i) \geq \theta_{m-1}(X_i) \) for all \( m = 2, \ldots, M \), the monotonicity of \( q(\tau | X_i) \) can be ensured. As mentioned earlier, \( X_{ij} \in [0, 1] \) for \( j = 2, \ldots, p \) and \( X_{i1} = 1 \) (intercept term). Similar to the previously mentioned re-parameterization, they take \( \delta_{j1} = \alpha_{j1} \) and \( \delta_{jm} = \alpha_{jm} - \alpha_{j(m-1)} \) for \( m = 2, \ldots, M \). The monotonicity constraint is ensured by taking unconstrained variable \( \delta_{jm}^{*} \sim N(\bar{\delta}_{jm}(\Theta), \sigma_j^2) \) and

\[ \delta_{jm} = \delta_{jm}^{*} I[\delta_{1m}^{*} + \sum_{j=2}^{p} I(\delta_{jm}^{*} < 0)\delta_{jm}^{*} \geq 0] \]

for \( j = 1, \ldots, p \) and \( m = 1, \ldots, M \). Since \( X_{i1} = 1 \) and \( X_{ij} \in [0, 1] \), it can be shown that \( \theta(X_i) - \theta_{m-1}(X_i) \geq 0 \) for all \( X_i \).

To estimate a quantile regression with possibly multiple covariates, first they obtain the quantile function estimate for \( \tau_k \)-th quantile using the KB method by solving

\[ (\hat{\beta}_1(\tau_k), \ldots, \hat{\beta}_p(\tau_k)) = \arg \min_{\beta} \sum_{Y_i > X'_i \beta} \tau_k |Y_i - X'_i \beta| + \sum_{Y_i < X'_i \beta} (\tau_k - 1)|Y_i - X'_i \beta|. \]

To incorporate the monotonicity property, they estimate these first-stage estimated quantile function with the above-mentioned method using Bernstein polynomial basis functions. They also extend it to analyze spatial data where they assume different quantile process for each spatial location and assuming some well-known spatial co-variance structure.

Reich (2012) (Re method) proposed a full Bayesian non-crossing multiple quantile regression using piece-wise Gaussian basis functions to interpolate the quantile function at different quantile levels. Reich and Smith (2013) suggested the use of different symmetric and asymmetric quantile functions for modeling based on the data. In the following figure, the simultaneous quantiles have been estimated using RFD method and Re method. For RFD method, 10000 iterations have been performed with 1000 burn-in. For Re method, we perform 50000 iterations with 10000 burn-in. We use \texttt{qreg} function from the \texttt{BSquare} R package (Smith and Reich (2013)) for implementing Re method.
Simultaneous Quantile Regression

Models based on a fixed number of quantiles do not give the estimates of all quantiles and may also be sensitive to the number and the location of the grid points of the quantile levels, which is not desirable. Instead of fitting quantile curves for a fixed set of quantiles, a more informative picture emerges by estimating the entire quantile curve. Suppose $Q(\tau|x) = \inf\{q : P(Y \leq q | X = x) \geq \tau\}$ denote the $\tau$-th conditional quantile ($0 \leq \tau \leq 1$) of a response $Y$ at $X = x$, $X$ being the predictor. A linear simultaneous quantile regression model for $Q(\tau|x)$ at a given $\tau$ is given by

$$Q(\tau|x) = \beta_0(\tau) + x\beta(\tau)$$

where $\beta_0(\tau)$ is the intercept and $\beta(\tau)$ is the slope smoothly varying as function of $\tau$. Thus estimating the quantile function involves estimation of the function $\beta_0(\tau)$ and $\beta(\tau)$. The main challenge in fitting this kind of model remains in complying with the monotonicity restriction of the predicted quantile lines $\beta_0(\tau) + x\beta_1(\tau)$ as a function of all values of the predictor $X$.

Tokdar and Kadane (2012) (TK method) first proposed the simultaneous quantile regression for univariate explanatory variable. They assumed the domain of the univariate response variable $\chi$ to be bounded and convex. So by monotonic transformation, domain can be transformed to $[-1, 1]$. Tokdar and Kadane (2012) showed that

**Theorem 1.1.** A linear specification $Q(\tau|x) = \beta_0(\tau) + x\beta(\tau)$, $\tau \in [0, 1]$ is monotonically
increasing in $\tau$ for every $x \in \chi = [-1, 1]$ if and only if

$$Q(\tau|x) = \mu + \gamma x + \frac{1 - x}{2} \eta_1(\tau) + \frac{1 + x}{2} \eta_2(\tau)$$  \hspace{1cm} (1.3)

where $\eta_1(\tau)$ and $\eta_2(\tau)$ are monotonically increasing in $\tau \in [0, 1]$.

For constructing the monotonically increasing functions $\eta_1(\tau)$ and $\eta_2(\tau)$, they defined $\xi_1(\tau)$ and $\xi_2(\tau)$ which are monotonically increasing maps from $[0, 1]$ to $[0, 1]$. Then they relate the these functions with $\xi_1(\tau)$ and $\xi_2(\tau)$ using the following relation

$$\eta_1(\tau) = \sigma_1 \bar{Q}(\xi_1(\tau)), \ \eta_2(\tau) = \sigma_1 \bar{Q}(\xi_1(\tau)),$$

where $\sigma_1, \sigma_2 > 0$ and $\bar{Q}$ is chosen in such a way that $\bar{Q}(0) = \underline{y}$ and $\bar{Q}(1) = \overline{y}$, $\underline{y}$ and $\overline{y}$ being the lower and the upper bound values of the response variable $Y$. They assumed $\sigma_1 = \sigma_2 = 1$ using which it can be showed that under that assumption $\mu = \gamma = 0$. $\bar{Q}$ can be chosen depending on the distribution of the response variable.

To estimate $\xi_1(\tau)$ and $\xi_2(\tau)$, they used zero-mean Gaussian process. Suppose $\omega(i, \tau)$ denotes a Gaussian process with mean zero for $i = 1, 2$ and $\tau \in [0, 1]$ with co-variance $\text{Cov}(\omega(i, \tau), \omega(i', \tau')) = \kappa^2 c_{ii'} \exp(-\lambda^2 (\tau - \tau')^2)$. They took $c_{11} = c_{22} = 1$ and $c_{12} = c_{21} = \rho \in [0, 1]$. $\xi_i(\tau)$ is defined as

$$\xi_i(\tau) = \frac{\int_0^\tau \exp(\omega(i, t)) dt}{\int_0^1 \exp(\omega(i, t)) dt}, \ i = 1, 2.$$

The priors on the parameters are given by

$$(\rho, \lambda^2, \kappa^{-2}) \sim U(0, 1) \times \text{Ga}(5, \frac{1}{10}) \times \text{Ga}(3, \frac{1}{3}).$$

The conditional density for the response variable $Y$ is given by

$$f_Y(y|x) = \frac{1}{\partial_{\tau} Q(\tau|x)} \bigg|_{\tau = \tau_x(y)}$$

where $\tau_x(y)$ solves $y = Q(\tau|x)$. So the log-likelihood for given dataset $\{(X_i, Y_i) : i =$
\[
\sum_{i=1}^{n} \log f_Y(Y_i|X_i) = -\sum_{i=1}^{n} \log \frac{\partial}{\partial \tau} Q(X_i(Y_i)|\tau) \\
= -\sum_{i=1}^{n} \log \left( \frac{1 - X_i}{2} \frac{\partial}{\partial \tau} \eta_1(X_i(Y_i)) + \frac{1 + X_i}{2} \frac{\partial}{\partial \tau} \eta_2(X_i(Y_i)) \right).
\]

To solve the equation \( Y_i = Q(\tau|X_i) \), they used Newton’s method.

Tokdar and Kadane (2012) also proposed a single-index generalization of the univariate model when we have multiple explanatory variables. Here, the dimension of the explanatory variable \( X \) is reduced to 1 by taking a linear projection of the covariate vector. Suppose \( \chi_{\alpha,a,b} = \{ x \in \mathbb{R}^p : x' \alpha \in (a - b, a + b) \} \) for \(-\infty < \alpha < \infty \) and \( b > 0 \) and \( p_{\alpha,a,b}(x) = \frac{(x' - a)}{b} \). Then similar to the Equation (1.3), the quantile function is given by

\[
Q(\tau|x) = \mu + x' \gamma + \frac{1 - p_{\alpha,a,b}}{2} \eta_1(\tau) + \frac{1 + p_{\alpha,a,b}}{2} \eta_2(\tau), \; x \in \chi_{\alpha,a,b}.
\]

where \( \eta_1(\tau) = \sigma_1 \hat{Q}(\xi_1(\tau)) \) and \( \eta_2(\tau) = \sigma_2 \hat{Q}(\xi_2(\tau)) \) and the interpretation of all the parameters are same as mentioned previously in the univariate case.

Later Yang and Tokdar (2016) extended the simultaneous quantile regression for multivariate case. Suppose we have \( p \) covariates and \( \beta_0 : (0,1) \mapsto \mathbb{R} \beta : (0,1) \mapsto \mathbb{R}^p \) are the intercept term and the slope coefficient vector as a function of the quantile \( \tau \in (0,1) \). The main challenge while estimating the functions \( \beta_0(\tau) \) and \( \beta(\tau) \) is that the following inequality must hold.

\[
\beta_0(\tau_1) + x' \beta(\tau_1) \geq \beta_0(\tau_2) + x^T \beta(\tau_2) \text{ for every } 0 < \tau_2 < \tau_1 < 1 \text{ and for all } x \in \chi \tag{1.4}
\]

Equation (1.4) can be also written as

\[
\dot{\beta}_0(\tau_1) + x^T \dot{\beta}(\tau_1) \geq 0 \text{ for every } \tau \in (0,1) \text{ and } \forall x \in \chi. \tag{1.5}
\]

Consider the map \( b \mapsto a(b, \chi) \) on \( \mathbb{R}^p \cup \{ \infty \} \) such that

\[
a(b, \chi) = \begin{cases} 
\sup_{x \in \chi} \{ -x'b / ||b|| \} & b \neq 0 \\
\infty & b = 0.
\end{cases}
\]

Regarding the structure of the coefficient functions \( \beta_0(\tau) \) and \( \beta(\tau) \), Yang and Tokdar (2016) proved that if \( \chi \) is a bounded convex set in \( \mathbb{R}^p \) with zero as an interior point and
let $\beta_0(\tau)$ and $\beta(\tau) = (\beta_1(\tau), \ldots, \beta_p(\tau))'$ be real, differentiable functions in $\tau \in (0, 1)$. Then $\dot{\beta}_0(\tau_1) + x' \dot{\beta}(\tau_1) > 0$ for all $\tau \in (0, 1)$ at every $x \in \chi$ if and only if

$$\dot{\beta}_0(\tau) > 0, \ \dot{\beta}(\tau) = \frac{v(\tau)}{a(v(\tau), \chi)\sqrt{1 + \|v(\tau)\|^2}}, \ \tau \in (0, 1)$$

for some $p$-variate real function $v(\tau) = (v_1(\tau), \ldots, v_p(\tau))'$ in $\tau \in (0, 1)$. Using this, the set of monotonicity constraints of the quantile hyper-planes reduces to the monotonicity constraint of a single function $\beta_0(\tau)$. Note that $(\beta_0(0), \beta_0(1))$ denotes the support of the conditional density of $Y$ given $X = 0$. A model for $\beta_0(\tau)$ is looked upon based on a specified or default prior density $f_0(y)$ for that conditional density. In general, $f_0$ should be chosen in such a way that it has support $(-\infty, \infty)$, like normal or $t$-distribution in case of heavy-tailed response values. Suppose $f_0$ has support $(-\infty, \infty)$. Then the quantile density is given by $q_0(\tau) = Q_0(\tau)$ where $Q_0(\tau) = F_0^{-1}(\tau)$. Let $\tau_0 = F_0(0)$. Then $\beta_0$ and $\beta$ are modeled in a way such that

$$\beta_0(\tau_0) = \gamma_0, \ \beta(\tau_0) = \gamma$$  \hspace{0.5cm} (1.6)

$$\beta_0(\tau) - \beta_0(\tau_0) = \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} q_0(u)du, \ \tau \in (0, 1)$$  \hspace{0.5cm} (1.7)

$$\beta(\tau) - \beta(\tau_0) = \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{w(u)}{a(w(u), \chi)\sqrt{1 + \|w(u)\|^2}}q_0(u)du, \ \tau \in (0, 1)$$  \hspace{0.5cm} (1.8)

where $\gamma_0 \in \mathbb{R}$, $\gamma \in \mathbb{R}^p$, $\sigma > 0$, $w : (0, 1) \mapsto \mathbb{R}^p$, an unconstrained $p$-variate function on $(0, 1)$ and $\zeta : [0, 1] \mapsto [0, 1]$ is a diffeomorphism (i.e., differentiable, monotonically increasing bijection). Using the aforementioned theorem by Yang and Tokdar (2016), it can be shown that $\beta_0(\tau)$ and $\beta(\tau)$ constructed using Equations (1.6), (1.7) and (1.8) satisfies condition given by (1.5).

For likelihood evaluation, they took the same approach as mentioned in Tokdar and Kadane (2012). For the given data-set $\{(X_i, Y_i) : i = 1, \ldots, n\}$ the log-likelihood is given by

$$\sum_{i=1}^{n} \log f_Y(Y_i, X_i) = -\sum_{i=1}^{n} \log \left\{ \dot{\beta}_0(\tau_{X_i}(Y_i)) + X_i' \dot{\beta}(\tau_{X_i}(Y_i)) \right\}$$

$$= -\sum_{i=1}^{n} \log \sigma q_0(\zeta(\tau)) \zeta(\tau) \left\{ 1 + \frac{w(\zeta(\tau))}{a(w(\zeta(\tau), \chi)\sqrt{1 + \|w(\zeta(\tau))\|^2}} \right\}$$

where $\tau_{X_i}(Y_i)$ is obtained by solving $Y_i = \int_{\tau_0}^{\tau} \{\dot{\beta}_0(u) + X_i' \dot{\beta}(u)\}du$. They put Gaussian
process prior on $w = (w_1, \ldots, w_p)$ since they are not restricted. On $\zeta$, they put scaled Gaussian process prior similar to that used in Tokdar and Kadane (2012).
Chapter 2

Univariate Regression Using B-spline Series Prior

Introduction

A brief overview of the existing methods of quantile regression has been discussed in Chapter 1. As mentioned over there, Tokdar and Kadane (2012) obtained a very useful characterization of the monotonicity constraint (see Theorem 1.1). They used the characterization to propose a suitable prior on the quantile function in a Bayesian approach, and computed the posterior distribution of the quantile curves. More specifically, they used two (possibly) dependent Gaussian processes $\xi_1(\cdot)$ and $\xi_2(\cdot)$ to induce prior on $\beta_0(\tau)$ and $\beta_1(\tau)$ in such a way so that $Q(\tau|x)$ remains a monotonically increasing function of $\tau$.

Gaussian processes do not have any shape restriction. To induce monotonicity property on $\xi_1(\cdot)$ and $\xi_2(\cdot)$, Tokdar and Kadane (2012) took the running integral of the exponential transformation of two (possibly correlated) Gaussian processes. To evaluate the likelihood, it requires solving the equation $Q(\tau|x) = y$ (see Equation (2.6) below). Without having a convenient expression, the solution needs to be done completely numerically which involves computing integrals of transformed Gaussian process over a very fine grid. Substantial improvement in computing approach is possible using a finite random series prior (Shen and Ghosal (2015)). Especially using a B-spline basis, monotonicity can be ensured by choosing a prior only on coefficients in increasing order. Piece-wise polynomial representation of splines allows obtaining $\tau_x(y)$ solving analytically. Thus we can reduce computation time substantially using a random series prior based on the B-spline basis.
Model Assumptions

We observe $n$ independent random samples $(X_1, Y_1), \ldots, (X_n, Y_n)$ of an explanatory variable $X$ and a response variable $Y$, both of which are assumed to be univariate. By monotonic transformations, we transform both of them to the unit interval. From Theorem 1 of Tokdar and Kadane (2012), it follows that a linear specification $Q(\tau|x) = \beta_0(\tau) + x\beta_1(\tau)$, $\tau \in [0, 1]$, is monotonically increasing in $\tau$ for every $x \in [0, 1]$ if and only if

$$Q(\tau|x) = \mu + \gamma x + \sigma_1 x \xi_1(\tau) + \sigma_2 (1-x) \xi_2(\tau),$$

(2.1)

where $\sigma_1$ and $\sigma_2$ are positive constants and $\xi_1, \xi_2 : [0, 1] \mapsto [0, 1]$ are monotonically increasing in $\tau \in [0, 1]$. Because $Y$ also has domain $[0, 1]$, we have the boundary conditions $Q(0|x) = 0$ and $Q(1|x) = 1$. Now, putting $\tau = 0$ in Equation (2.1), and using $\xi_1(0) = \xi_2(0) = 0$, we obtain

$$\mu + \gamma x = 0 \quad \text{for all } x \in [0, 1]$$

which implies $\mu = \gamma = 0$ and hence (2.1) reduces to

$$Q(\tau|x) = \sigma_1 x \xi_1(\tau) + \sigma_2 (1-x) \xi_2(\tau).$$

(2.2)

Now putting $\tau = 1$ and using $\xi_1(1) = \xi_2(1) = 1$, we obtain

$$1 = \sigma_1 x + \sigma_2 (1-x) \quad \text{for all } x \in [0, 1],$$

which implies $\sigma_1 = \sigma_2 = 1$. Therefore in the present context, the quantile regression function has representation

$$Q(\tau|x) = x \xi_1(\tau) + (1-x) \xi_2(\tau) \quad \text{for } \tau \in [0, 1], \; x, y \in [0, 1].$$

(2.3)

Equation (2.3) can be re-framed as

$$Q(\tau|x) = \beta_0(\tau) + x\beta_1(\tau) \quad \text{for } \tau \in [0, 1], \; x, y \in [0, 1],$$

(2.4)

where $\beta_0(\tau) = \xi_2(\tau)$ and $\beta_1(\tau) = \xi_1(\tau) - \xi_2(\tau)$ denotes the slope and the intercept of the quantile regression, which are smooth functions of $\tau$. Provided that $Q(\tau|x)$ is strictly
increasing in $\tau$ for all $x$, the conditional density for $Y$ at $y$ given $X = x$ is given by

$$f(y|x) = \left(\frac{\partial}{\partial \tau} Q(\tau|x)\big|_{\tau = \tau_x(y)}\right)^{-1} = \left(\frac{\partial}{\partial \tau} \beta_0(\tau) + x \frac{\partial}{\partial \tau} \beta_1(\tau)\big|_{\tau = \tau_x(y)}\right)^{-1}, \quad (2.5)$$

where $\tau_x(y)$ solves the equation

$$x \xi_1(\tau) + (1 - x) \xi_2(\tau) = y. \quad (2.6)$$

Then the joint conditional density of $Y_1, \ldots, Y_n$ given $X_1, \ldots, X_n$ is given by $\prod_{i=1}^n f(Y_i|X_i)$.

### Regression with Spline

Function estimation on a bounded interval through B-spline basis expansion is one of the most convenient approaches. To construct a prior on the quantile function $Q(\tau|x)$, or equivalently on $\xi_1$ and $\xi_2$, we need to ensure their monotonically increasing properties. If the coefficients in a B-spline basis expansion are in increasing order, then the corresponding quantile function will be an increasing function of $\tau$ (de Boor (2001)). In fact, if the target function is strictly increasing, the quality of the approximation of spline expansions is maintained when the coefficients are restricted by increasing order (Shen and Ghosal (2015)). This motivates us to use B-splines with increasing coefficients as the basis of the quantile function. The degree of B-spline determines the degree of smoothness, in that piecewise quadratic splines are continuously differentiable, piecewise cubic splines are continuously twice differentiable and so on, and are able to optimally approximate functions of smoothness index only up to the degree of splines. Thus optimal estimation of smoother functions requires higher degree splines. On the other hand, computational complexity increases with the degree of the B-spline basis. Smoothness of order higher than two is typically not visually distinguishable from smoothness of order 2, and therefore splines of degree up to 3 suffices in most applications. In case of quadratic B-spline basis, $Q(\tau|x)$ is a linear combination of piecewise second degree polynomials in terms of $\tau$, and hence restricted to an interval between two knots, $Q(\tau|x)$ is an increasing quadratic function. Thus we can solve $Q(\tau|x_i) = Y_i$ for each data points $(X_i, Y_i)$ analytically, unlike the case of Gaussian process prior, where one has to use numerically integrate and apply iterative Newton-Raphson method to solve them for each realization of the quantile process from its posterior distribution. Due to the increasing complexity and higher computation time, results for B-splines of order higher than 3 have not been shown in this chapter. For a cubic B-spline, $Q(\tau|x)$ is linear combination of piecewise
third degree polynomials. Hence, for any given interval in the domain, \( Q(\tau | x) \) is a monotonically increasing third degree polynomial. Hence in this case also, \( \tau_x(y) \) can be solved analytically. However the analytic algorithm for solving cubic equation requires computing all three roots even though only one root falls in the admissible interval, which makes the cost of computing roots in a cubic equation substantial. Due to the monotonicity of \( Q(\tau | x) \), the bisection method of finding roots is an attractive alternative. In our numerical experiments, we found that for cubic splines the bisection method is slightly faster than the analytical approach.

Let \( 0 = t_0 < t_1 < \cdots < t_k = 1 \) be the equidistant knots on the interval \([0, 1]\) where \( t_i = i/k, \ i = 0, 1, \ldots, k \). For B-spline of \( m \)th degree, the number of basis functions is \( J = k + m \). Let \( \{B_{j,m}(t)\}_{j=1}^{k+m} \) be the basis functions of \( m \)th degree B-splines on \([0, 1]\) on the above mentioned equidistant knots. We consider the following basis expansion of the quantile functions through the relations

\[
\xi_1(\tau) = \sum_{j=1}^{k+m} \theta_j B_{j,m}(\tau) \quad \text{where} \quad 0 = \theta_1 < \theta_2 < \cdots < \theta_{k+m} = 1,
\]

\[
\xi_2(\tau) = \sum_{j=1}^{k+m} \phi_j B_{j,m}(\tau) \quad \text{where} \quad 0 = \phi_1 < \phi_2 < \cdots < \phi_{k+m} = 1.
\] (2.7)

Taking \( \theta_1 = \phi_1 = 0 \) ensures \( \xi_1(0) = \xi_2(0) = 0 \) and \( \theta_{k+m} = \phi_{k+m} = 1 \) ensures \( \xi_1(1) = \xi_2(1) = 1 \). The increasing values of the coefficients takes care on monotonicity (de Boor (2001)). Thus, \( \xi_1 \) and \( \xi_2 \) are monotonically increasing functions of \( \tau \) from \([0, 1]\) onto \([0, 1]\).

Let, \( \{\gamma_j\}_{j=1}^{k+m-1} \) and \( \{\delta_j\}_{j=1}^{k+m-1} \) be defined by

\[
\gamma_j = \theta_{j+1} - \theta_j, \quad \delta_j = \phi_{j+1} - \phi_j, \quad j = 1, \ldots, k + m - 1.
\] (2.8)

Hence

\[
\theta_{j+1} = \sum_{i=1}^{j} \gamma_i, \quad \phi_{j+1} = \sum_{i=1}^{j} \delta_i, \quad j = 1, \ldots, k + m - 1.
\] (2.9)

and the restrictions in (2.7) equivalently be written as

\[
\gamma_j, \delta_j \geq 0, \ j = 1, \ldots, k + m - 1, \quad \text{and} \quad \sum_{j=1}^{k+m-1} \gamma_j = \sum_{j=1}^{k+m-1} \delta_j = 1.
\] (2.10)

Therefore the spacings of the B-spline coefficients take value in the unit simplex. A natural
prior is thus given by the uniform distribution on the unit simplex, or the Dirichlet prior with parameter \((1, \ldots, 1)\). The Bayesian procedures based on quadratic splines (i.e. \(m = 2\)) will be called the Quadratic Spline Simultaneous Quantile regression (QSSQR) and that based on cubic splines (i.e. \(m = 3\)) will be called the Cubic Spline Simultaneous Quantile regression (CSSQR).

**Model Fitting**

To compute the log-likelihood function, we proceed like Tokdar and Kadane (2012). By (2.5), the log-likelihood is given by

\[
\sum_{i=1}^{n} \log f(Y_i|X_i) = -\sum_{i=1}^{n} \log \frac{\partial}{\partial \tau} Q(\tau_X(Y_i)|X_i) 
\]

\[
= -\sum_{i=1}^{n} \log \left\{ \frac{X_i}{\partial \tau} \xi_1(\tau_X(Y_i)) + (1 - X_i) \frac{\partial}{\partial \tau} \xi_2(\tau_X(Y_i)) \right\},
\]

(2.11)

where \(\tau_X(Y_i)\) is obtained from (2.6).

Due to the monotonicity of \(\xi_1\) and \(\xi_2\), any convex combination of them will have one and only one solution on the interval \([0, 1]\). To evaluate the log-likelihood we compute the derivatives of the basis splines (de Boor (2001)),

\[
\frac{d}{dt} \xi_1(t) = \sum_{j=2}^{k+m} \tilde{\theta}_j B_{j-1,m-1}(t), \quad \frac{d}{dt} \xi_2(t) = \sum_{j=2}^{k+m} \tilde{\phi}_j B_{j-1,m-1}(t),
\]

where

\[
\tilde{\theta}_j = (k + m)(\theta_j - \theta_{j-1}), \quad \tilde{\phi}_j = (k + m)(\phi_j - \phi_{j-1}), \quad j = 2, \ldots, k + m.
\]

The log-likelihood \(\sum_i \log f(Y_i|X_i)\) thus reduces to

\[
-\sum_i \log \left\{ X_i \sum_{j=2}^{k+m} \theta_j B_{j-1,m-1}(\tau_X(Y_i)) + (1 - X_i) \sum_{j=2}^{k+m} \phi_j B_{j-1,m-1}(\tau_X(Y_i)) \right\}.
\]

**MCMC and Transition Step**

In Equation (2.10), we note that \(\{\gamma_j\}_{j=1}^{k+m-1}\) and \(\{\delta_j\}_{j=1}^{k+m-1}\) are on the unit simplex. We shall use a Metropolis-Hastings algorithm to obtain Markov Chain Monte Carlo (MCMC)
samples from the posterior distribution. In MCMC, to move on the simplex, we generate independent sequences $U_j$ and $W_j$, $j = 1, \ldots, k + m - 1$, from $U(1/r, r)$ for some $r > 1$. It should be noted that smaller values of $r$ yields smaller jumps with higher acceptance probability; while larger values of $r$ yields bigger jumps with lower acceptance probability. Hereby its value can be tuned by monitoring the acceptance ratio.

Define $V_j = \gamma_j U_j$ and $T_j = \delta_j W_j$ for $j = 1, \ldots, k + m - 1$. Consider the proposal moves $\gamma_j \mapsto \gamma_j^*$ and $\delta_j \mapsto \delta_j^*$ given by

$$
\gamma_j^* = \frac{V_j}{\sum_{j=1}^{k} V_j}, \quad \delta_j^* = \frac{T_j}{\sum_{j=1}^{k} T_j}, \quad i = 1, \ldots, k + m - 1.
$$

We take the uniform prior on $\{\gamma_j\}_{j=1}^{k}$ and $\{\delta_j\}_{j=1}^{k}$ (i.e., Dirichlet(1,...,1)). We generate sequences $\{U_{j}\}_{j=1}^{k}$ and $\{W_{j}\}_{j=1}^{k}$ of independent random variables from $U(1/r, r)$ for some $r > 1$. Since $\{\gamma_j\}_{j=1}^{k}$ and $\{\delta_j\}_{j=1}^{k}$ are independent and $\{U_{j}\}_{j=1}^{k}$ and $\{W_{j}\}_{j=1}^{k}$ are also independent, it is enough to show the results for the first variable transition steps and following conditionals. For the second set of variables, the results will follow similarly. Below $p$ will stand for a generic (joint) density.

Define $V_j = \gamma_j U_j$, $j = 1, \ldots, k$. So we have,

$$
p(V_1, \ldots, V_k|\gamma_1, \ldots, \gamma_k) = \prod_{j=1}^{k} \left\{ \frac{r}{(r^2 - 1)\gamma_j} \right\} I \left[ \frac{\gamma_j}{r} \leq V_j \leq r\gamma_j \right]
$$

Now, define $V = \sum_{j=1}^{k} V_j$. After transforming the variables we get,

$$
p(V_1, \ldots, V_{k-1}, V|\gamma_1, \ldots, \gamma_k)
= \left( \frac{r}{r^2 - 1} \right)^k \left( \prod_{j=1}^{k} \gamma_j \right)^{-1} \prod_{j=1}^{k-1} I \left[ \frac{\gamma_j}{r} \leq V_j \leq r\gamma_j \right]
\times I \left[ \frac{\gamma_k}{r} \leq V - \sum_{j=1}^{k-1} V_j \leq r\gamma_k \right].
$$

We define $\gamma_j^* = V_j/V$, $j = 1, \ldots, k-1$ and set $\gamma_J = (1 - \sum_{j=1}^{k-1} \gamma_j)$. Then after transformation
of variables we get

\[
p(\gamma_1^*, \ldots, \gamma_{k-1}^*, V) = \left(\frac{r}{r^2 - 1}\right)^k V^{k-1} \left(\prod_{j=1}^{k} \gamma_j\right)^{-1} \prod_{j=1}^{k-1} I\left[\frac{\gamma_j}{rV} \leq \gamma_j^* \leq \frac{r\gamma_j}{V}\right] \\
\times I\left[\frac{\gamma_{k-1}^*}{r(1 - \sum_{j=1}^{k-1} \gamma_j^*)} \leq V \leq \frac{r\gamma_k}{(1 - \sum_{j=1}^{k} \gamma_j^*)}\right] \\
= \left(\frac{r}{r^2 - 1}\right)^k V^{k-1} \left(\prod_{j=1}^{k} \gamma_j\right)^{-1} \prod_{j=1}^{k-1} I\left[\frac{\gamma_j}{rV} \leq \gamma_j^* \leq \frac{r\gamma_j}{V}\right] \\
= \left(\frac{r}{r^2 - 1}\right)^k V^{k-1} \left(\prod_{j=1}^{k} \gamma_j\right)^{-1} \prod_{j=1}^{k-1} I\left[\frac{\gamma_j}{rV} \leq \gamma_j^* \leq \frac{r\gamma_j}{V}\right] \\
= \left(\frac{r}{r^2 - 1}\right)^k V^{k-1} \left(\prod_{j=1}^{k} \gamma_j\right)^{-1} I\left[\max_{0 \leq j \leq k} \frac{\gamma_j}{r\gamma_j^*} \leq \frac{V}{\min_{0 \leq j \leq k} \frac{r\gamma_j}{\gamma_j^*}}\right].
\]

Hence integrating over the range of \(V\), we get

\[
p(\gamma_1^*, \ldots, \gamma_{k-1}^*|\gamma_1, \ldots, \gamma_k) = \int_{\min_{0 \leq j \leq k} \frac{r\gamma_j}{\gamma_j^*}}^{\max_{0 \leq j \leq k} \frac{r\gamma_j}{\gamma_j^*}} \left(\frac{r}{r^2 - 1}\right)^k \left(\prod_{j=1}^{k} \gamma_j\right)^{-1} V^{k-1} dV
\]

Thus we get the conditional density to be

\[
p(\gamma^*|\gamma) = \left(\frac{r}{r^2 - 1}\right)^k \left(\prod_{j=1}^{k} \gamma_j\right)^{-1} \left[\left\{\min_{0 \leq j \leq k} \frac{r\gamma_j}{\gamma_j^*}\right\} - \left\{\max_{0 \leq j \leq k} \frac{\gamma_j}{\gamma_j^*}\right\}\right] \left[\prod_{j=1}^{k} \gamma_j\right]^{-1} / k.
\]

Similarly, we obtain,

\[
p(\delta^*|\delta) = \left(\frac{r}{r^2 - 1}\right)^k \left(\prod_{j=1}^{k} \delta_j\right)^{-1} \left[\left\{\min_{0 \leq j \leq k} \frac{r\delta_j}{\delta_j^*}\right\} - \left\{\max_{0 \leq j \leq k} \frac{\delta_j}{\delta_j^*}\right\}\right] \left[\prod_{j=1}^{k} \delta_j\right]^{-1} / k.
\]

Hence, replacing \(p\) with \(k + m - 1\), the conditional distribution of \(\{\gamma_j^*\}_{j=1}^{k+m-1}\) given
\{\gamma_j\}_{j=1}^{k+m-1} is

\[ p(\gamma^* | \gamma) = \left( \frac{r}{r^2 - 1} \right)^{k+m-1} \left( \prod_{j=1}^{k+m-1} \gamma_j \right)^{-1} (k + m - 1)^{-1} \]

\[ \times \left[ \left\{ \min_{0 \leq j \leq k+m-1} \left( r \gamma_j / \gamma^*_j \right) \right\}^{k+m-1} - \left\{ \max_{0 \leq j \leq k+m-1} \left( r \gamma_j / \gamma^*_j \right) \right\}^{k+m-1} \right]. \]

Similarly, we can find the conditional distribution of \{\delta^*_j\}_{j=1}^{k+m-1}. Now from the given set of values for \{\gamma^*_j\}_{j=1}^{k+m-1} and \{\delta^*_j\}_{j=1}^{k+m-1} the updated values \{\theta^*_j\}_{j=1}^{k+m-1} and \{\phi^*_j\}_{j=1}^{k+m-1} can be found using the following relation.

\[ \theta^*_j = \sum_{i=1}^{j} \gamma^*_i, \quad \phi^*_j = \sum_{i=1}^{j} \delta^*_i \quad j = 1, \ldots, k + m - 1. \quad (2.13) \]

While evaluating the likelihood, using the relations mentioned in Equation (2.8) and (2.9), we can find the likelihoods at the initial and destination points respectively in terms of \{\gamma_j\}_{j=1}^{k+m-1} and \{\delta_j\}_{j=1}^{k+m-1}, and \{\gamma^*_j\}_{j=1}^{k+m-1} and \{\delta^*_j\}_{j=1}^{k+m-1}. The acceptance probability in Metropolis-Hastings algorithm is then given by

\[ P_a = \min \left\{ \frac{L(\gamma^*, \delta^*) \pi(\gamma^*) \pi(\delta^*) f(\gamma | \gamma^*) f(\delta | \delta^*)}{L(\gamma, \delta) \pi(\gamma) \pi(\delta) f(\gamma | \gamma^*) f(\delta | \delta^*)}, 1 \right\} \]

\[ = \min \left\{ \frac{L(\gamma^*, \delta^*) f(\gamma | \gamma^*) f(\delta | \delta^*)}{L(\gamma, \delta) f(\gamma | \gamma^*) f(\delta | \delta^*)}, 1 \right\}, \quad (2.14) \]

where \( L(\cdot) \) denotes the likelihood and \( \pi \) is the uniform Dirichlet prior density on the corresponding parameters.

**Choosing value of \( k \) and Model Averaging**

In the previous sections, we discussed fitting spline basis functions for fixed number of partitions of equal length in [0, 1], the transformed domain of \( X \). In other words, we developed the methodology for fixed number of basis functions. The number of basis functions \( k + m \) controls the smoothness and hence the quality of the estimate. For smaller value of \( k \), the bias is high but variability is less, so there is a bias-variance trade-off. Moreover MCMC chain runs faster with better mixing if the value of \( k \) is smaller. It is desirable to determine the value of \( k \) based on the data to apply the right amount of smoothness. In the Bayesian setting, it is natural to put a prior on the smoothing parameter \( k \) and make inference based on the posterior distribution. For computational efficiency, the range of
should be a finite collection of consecutive integers. Once we fix the range of all possible values of \( k \), we can either consider empirical Bayes selection based on the marginal likelihood of \( k \), or a Bayesian model averaging. A common approach to posterior computation for a parameter space of varying dimension is through the reversible jump MCMC method, but its implementation can be challenging. An alternative approach is provided by the method described in Chib and Jeliazkov (2001). In this approach marginal probabilities corresponding to different values of \( k \) are obtained from separate MCMC output for each value of \( k \). Since all these chains are run independently of each other, parallel computing can be implemented, which can greatly offset the cost of running separate chains for all feasible values of \( k \). Both the empirical Bayes approach by identifying the value of \( k \) with highest posterior probability, and the hierarchical Bayes approach based on Bayesian model averaging, can be implemented using this method, and they have identical computing cost.

Consider the \( m \)th degree B-spline with \( k \) many partitions of \([0, 1]\) of equal length. The equidistant knot sequence is given by \( 0 = t_0 < t_1 < \cdots < t_k = 1 \) such that \( t_i = i/k \) for \( i = 0, \ldots, k \). Let \( \omega = (\{ \gamma_j \}_{j=1}^{k+m-1}, \{ \delta_j \}_{j=1}^{k+m-1}) \) be the parameter of interest. We denote the whole data by \( z \), and with a slight abuse of notation, denote the joint density of \( z \) also by \( f \). The posterior density is then given by

\[
\pi(\omega | z) \propto \pi(\omega) f(z | \omega)
\]

over \( S \), a subset of \( \mathbb{R}^{2(k+m-1)} \), specified by the Equation (2.10). For any \( \omega^* \), the logarithm of the marginal likelihood is given by

\[
\log m(z) = \log f(z | \omega^*) + \log \pi(\omega^*) - \log \pi(\omega^* | z).
\] (2.15)

It is recommended to take \( \omega^* \) to be a point where it has high density under the posterior. For any given value of \( \omega^* \), we can easily calculate the first two terms of the Equation (2.15). Our goal is to estimate the posterior ordinate \( \pi(\omega^* | y) \) given the posterior sample \( \{\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(M)}\} \), where \( M \) denotes the number of draws from the posterior sample after burn-in. Let us denote the proposal density by \( q(\omega, \omega' | z) \) for the transition from \( \omega \) to \( \omega' \). Suppose the probability of move \( \alpha(\omega, \omega' | z) \) is given by

\[
\alpha(\omega, \omega' | z) = \min \left\{ 1, \frac{f(z | \omega') \pi(\omega') q(\omega', \omega | z)}{f(y | \omega) \pi(\omega) q(\omega, \omega' | z)} \right\}.
\]

Chib and Jeliazkov (2001) proposed a simulation consistent estimate of the posterior
ordinate given by
\[
\hat{\pi}(\omega^*|z) = \frac{M^{-1} \sum_{g=1}^{M} \alpha(\omega^{(g)}, \omega^*|z)q(\omega^{(g)}, \omega^*|z)}{L^{-1} \sum_{j=1}^{L} \alpha(\omega^*, \tilde{\omega}^{(j)}|z)},
\]  
(2.16)
where \{\omega^{(g)}\} are the sampled draws from the posterior distribution and \{\tilde{\omega}^{(j)}\} are the draws from \(q(\omega^*, \tilde{\omega}|z)\), given the fixed value \(\omega^*\). Then the estimated logarithm of the marginal is given by
\[
\log \hat{m}(z) = \log f(z|\omega^*) + \log \pi (\omega^*) - \log \hat{\pi}(\omega^*|z). 
\]  
(2.17)

As mentioned in Chib and Jeliazkov (2001), the computation time of the marginal likelihood of each of the probable model is small and is almost readily available after the MCMC chain finishes. After finding the marginal likelihood for each of those models, the value of \(k\) corresponding to the model with the highest marginal likelihood can be treated as the best possible value of \(k\). Suppose, we fix the domain of \(k\) to be \(D\) where \(D\) is a collection of discrete natural numbers. Let, \(\hat{m}_k(z)\) denotes the marginal likelihood of \(k \in D\). Now select the best possible value of \(k\) by
\[
\hat{K} = \arg\max_{k \in D} \log \hat{m}_k(z).
\]

Afterwards, in this chapter, we named the method of estimation with B-spline basis functions corresponding to the \(k = \hat{K}\) to be the empirical Bayes (EB) method.

Instead of choosing the value of \(k\) by the empirical Bayes method, another possibility is to perform model averaging with respect to the posterior distribution of \(k\) over \(D\). To find the weights of the corresponding models, we calculate the marginal likelihood by the Metropolis-Hastings algorithm for each \(k\) following the method described in Chib and Jeliazkov (2001). After obtaining the marginal likelihoods for each value of \(k\), we obtain the weighted average of the estimated quantile curves where weights are proportional to their marginal likelihoods. Let, \(w_k\) and \(\hat{m}_k(z)\) denote the weights and the marginal likelihoods for \(k \in D\). Let \(\hat{\xi}_{1,k}(\tau)\) and \(\hat{\xi}_{2,k}(\tau)\) be the posterior means of \(\xi_1(\tau)\) and \(\xi_2(\tau)\) respectively for given \(k\). Then we have
\[
w_k = \frac{\hat{m}_k(z)}{\sum_{k \in D} \hat{m}_k(z)}, \quad k \in D
\]
\[
\hat{\xi}_1(\tau) = \sum_{k \in D} w_k \hat{\xi}_{1,k}(\tau) \quad \text{and} \quad \hat{\xi}_2(\tau) = \sum_{k \in D} w_k \hat{\xi}_{2,k}(\tau).
\]
Monotonicity of $\hat{\xi}_{1,k}(\tau)$ for each $k \in D$ ensures that $\hat{\xi}_1(\tau)$ is also monotonic. Similarly, $\hat{\xi}_2(\tau)$ is also monotonic. We note that the estimated quantile functions for each $k$ is totally determined by the estimated coefficients of the corresponding basis functions. Since their relation is also linear, to calculate the overall posterior mean quantile function, it suffices to obtain the posterior mean of the coefficients of the basis functions for each $k$. With the knowledge of weights and posterior mean of basis coefficients for each $k$, we can derive the estimated slope, intercept and quantile functions easily without saving the quantile function values over grid points. For the rest of this chapter, we called this to be the hierarchical Bayes (HB) method.

**Large-sample Properties**

In this section, we give arguments that indicate the posterior distribution based on a random series of B-splines concentrates near the truth. We assume that the true quantile regression function $Q_0(\tau|x)$ (which necessarily has the linear structure of the form $\beta_{0,\text{true}}(u) + \beta_{1,\text{true}}(u)x$ is absolutely continuous with respect to the Lebesgue measure and $q_0(\tau|x)$ is the corresponding “quantile density function”. We assume that $q_0$ satisfies the following regularity condition:

(A) For any $\epsilon > 0$, there exists $\delta > 0$ such that for any strictly increasing, continuous, bijection $h : [0, 1] \rightarrow [0, 1]$ such that $\sup_u |h(u) - u| < \delta$, we have $\int \log q_0(h(u)|x)q_0(u|x)du < \epsilon$ for all $x$.

We also assume that the true conditional density function $f_0(y|x)$ is positive and continuous (and hence is bounded above and below) and that the distribution of $X$ is $G$. Let $F_0(y|x)$ stand for the true conditional distribution function.

Clearly the random quantile regression function (i.e., in the model following the B-spline basis expansion prior) $Q(\tau|x)$ is differentiable with quantile density function $q(\tau|x)$. Let the corresponding conditional density function be denoted by $f(y|x)$ and the cumulative distribution function $F(y|x)$. We assume that the prior on $k$ is positive for all values.

It follows from a well-known theorem of Schwartz (1995) that the posterior distribution of the joint distribution of $(X,Y)$ is consistent with respect to the weak topology if for every $\epsilon > 0$, the prior probability of

$$
\int \int f_0(y|x) \log \frac{f_0(y|x)}{f(y|x)} dy dG(x) < \epsilon
$$

(2.18)
is positive. Below we show that the condition holds at the true \( f_0 \) for the B-spline random series prior.

We follow the line of argument given in Hjort and Walker (2009) although in our case the presence of the conditioning variable \( X \) makes the situation considerably more complicated. Define \( h_x(u) = F(Q_0(u|x)|x) \). Note that if \( U \) is \( U(0,1) \), then \( Q_0(U|x) \) has conditional density \( f_0(y|x) \). Hence

\[
\int \int f_0(y|x) \log \frac{f_0(y|x)}{f(y|x)} \, dy \, dG(x) = \int \int \log \frac{p_0(Q_0(u|x)|x)}{p(Q_0(u|x)|x)} \, du \, dG(x) = \int \int \log \frac{q(h_x(u)|x)}{q_0(u|x)} \, du \, dG(x) = \int \int \log \frac{q_0(h_x(u)|x)}{q_0(u|x)} \, du \, dG(x),
\]

where we have used the relation between quantile density and probability density \( q(u|x) = 1/f(Q(u|x)|x) \) and \( q_0(u|x) = 1/f_0(Q_0(u|x)|x) \), and equivalently \( f(y|x) = 1/q(F(y|x)|x) \) and \( f_0(y|x) = 1/q_0(F_0(y|x)|x) \).

Note that the assumption implies that \( q_0(u|x) = \beta_{0,\text{true}}(u) + \beta_{1,\text{true}}(u)x \) and \( \beta_{0,\text{true}}(u) \) and \( \beta_{1,\text{true}}(u) \) are continuous functions, and hence for a sufficiently large number of knots \( k \), both functions can be approximated by a linear combinations of B-splines up to any desired degree of accuracy. Fixing \( k \) at a sufficiently large value (which has positive prior probability), we find the approximating coefficients, and then consider all vector of coefficients in a small neighborhood around the vector of approximating coefficients. By the choice of continuous and positive prior density on the coefficient vector in the random spline series, it follows that the event has positive prior probability. Thus neighborhoods of the true quantile function get positive prior probabilities. In view of the portmanteau theorem characterizing weak convergence in terms of convergence of quantile functions (see van der Vaart (1998), Lemma 21.2), it follows that weak neighborhood of the true distribution get positive prior probabilities, and so do uniform neighborhoods of the true distribution in view of Polya’s theorem. This means, in view of the Condition (A) on \( q_0 \), that the second term in (2.19) can be arbitrarily small with positive probability.

For the first term in Equation (2.19), by the same arguments we find that \( q \) is uniformly close to \( q_0 \) with positive probability, and hence \( \log(q(h_x(u)|x)/q_0(h_x(u|x)) \) (the
integrand) is small with positive probability since \(q_0\) is assumed to be bounded below.

Thus the prior puts positive mass in neighborhoods defined by Equation (2.18), and hence by Schwartz’s theorem, the posterior probability of all conditional distributions \(F(y|x)\) satisfying

\[
\left| \int \psi(x, y)dF(y|x)dG(x) - \int \psi(x, y)dF_0(y|x)dG(x) \right| < \epsilon
\]

tends to one for any continuous function \(\psi\) on \([0, 1] \times [0, 1]\) and \(\epsilon > 0\).

The above conclusion on posterior consistency is about the conditional distribution function in terms of the weak topology, which does not immediately show concentration of posterior for the quantile function in neighborhoods of the true quantile function. To this end, we further assume that the collection of true conditional distribution function \(\{F_0(\cdot|x) : x \in [0, 1]\}\) is equicontinuous, i.e., given \(\epsilon > 0\), there exists \(\delta > 0\) such that whenever \(|y - y'| < \delta\), we have that \(|F_0(y|x) - F_0(y'|x)| < \epsilon\) for all \(x\). Then by an easy modification of the proof of Polya’s theorem (see Lemma 2.11 of van der Vaart (1998)), it follows that \(\int F_n(y|x)dG(x) \to \int F_0(y|x)dG(x)\) for all \(y\) implies that \(\int \sup_{y \in [0, 1]} |F_n(y|x) - F_0(y|x)|dG(x) \to 0\). Thus the posterior probability of \(\int \sup_{y \in [0, 1]} |F(y|x) - F_0(y|x)|dG(x) < \epsilon\) tends to one for any \(\epsilon > 0\). Now for any \(\tau\), as \(F_0(Q_0(\tau|x)|x) = \tau\),

\[
\int |\tau - F_0(Q(\tau|x)|x)|dG(x) = \int |F_0(Q_0(\tau|x)|x) - F_0(Q(\tau|x)|x)|dG(x).
\]

As \(f_0\) is bounded below by a positive number, this implies that the posterior probability of \(\sup_{\tau \in [0, 1]} \int |Q(\tau|x) - Q_0(\tau|x)|dG(x) < \epsilon\) tends to one for any \(\epsilon > 0\).

**Simulation Study**

In Section 2.2, we note that after some monotonic transformation, every quantile regression function can be represented in the form mentioned in Equation (2.3). For simulation purposes, we consider two different true quantile functions structure. For each cases, we compare the true and the estimated values of the slope, intercept and the quantile regression function for \(x = 0.2, 0.5\) and \(0.7\), \(\tau = 0.25, 0.5\) and \(0.75\) for sample size \(n = 100\) by our proposed methods, Gaussian SQR (GSQR) proposed in Tokdar and Kadane (2012), Bernstein polynomial SQR (BPSQR) proposed in Reich et al. (2011) and \texttt{qreg} and \texttt{qreg_spline} functions under the \texttt{BSquare} package in R by Smith and Reich (2013) (see
Reich (2012); Reich and Smith (2013) for details).

For our proposed methods, suppose $0 = t_0 < t_1 < \ldots < t_k = 1$ be the equidistant knots on the interval $[0, 1]$ such that $t_i = 1/k$ for all $i = 0, 1, \ldots, k$. Since MCMC does not mix very well for large values of $k$, we considered only first 8 possible values of $k$ for either cases ($D = \{3, 4, \ldots, 10\}$ for QSSQR and $D = \{5, 6, \ldots, 12\}$ for CSSQR). The value of $r$ was taken to be 1.25 and it is observed that for this value of $r$, acceptance ratio stays between 0.25 to 0.45 for most of the cases. Separately, for quadratic and cubic B-spline methods, for each of the values of $k$, we run 20000 iterations with 5000 burn-in. For QSSQR, while evaluating likelihood, we can solve (2.6) analytically very fast. Although CSSQR can also be solved analytically, we note that the procedure is considerably slower than the bisection method. We apply the latter method with precision of $2^{-10} \approx 10^{-3}$, after finding the interval where $y_i$ is located for $i = 1, \ldots, n$, by linear search.

To calculate the marginal likelihood of each model, we need to calculate the expression given by Equation (2.16). To calculate the numerator of that term, we considered the last 5000 iterations and we take $\omega^* = (\{\gamma^{(l)}\}_{j=1}^{k+m-1}, \{\delta^{(l)}\}_{j=1}^{k+m-1})$ for $l = 15000$ where $\beta(l)$ denotes the $l$-th update of the parameter $\beta$ in MCMC. Thus the numerator term is evaluated by the end of MCMC chain. To calculate the denominator term, we consider 5000 draws of $\{\omega^{(j)}\}$ from $q(\omega^*, \omega | y)$. Then we computed the marginal likelihoods of each model. We also computed the estimates corresponding to weighted average quadratic and cubic spline models taking the weights proportional to the marginal likelihoods over the domain of values of $k$. We used parallel computing for different values of $k$ to obtain posterior probabilities.

To implement the GSQR (Gaussian process SQR) method, proposed in Tokdar and Kadane (2012), we took $M = 101$ equidistant knots over the interval $[0, 1]$, the starting and the ending knots being at 0 and 1 respectively. At each iteration step, we save the values of the Gaussian process at those knots and use those values for likelihood evaluations. We ran 20000 iterations with 5000 burn-in. We evaluate the likelihood function by solving (2.6) for each data-points using Newton-Raphson method, as mentioned in Tokdar and Kadane (2012). We took the precision for convergence criteria to be $10^{-3}$.

While estimating the quantile regression (QRF) functions using the functions `qreg` and `qreg_spline` functions under the `BSquare` package in R by Smith and Reich (2013), for each case we performed 20000 iterations with 5000 burn-in. For BPSQR (Bernstein polynomial SQR) method we also performed 20000 iterations with 5000 burn-in. The function (implemented in R by Reich et al. (2011)) for estimating the quantile functions
used for this method is available at the following link\(^1\).

For our proposed method, we calculate the uniform 95% posterior credible region for the quantile regression function (QRF) \( Q(\tau|x), \tau \in [0, 1] \) for three distinct co-variate values \( x = 0.2, 0.5, 0.7 \) for \( \tau \in [0, 1] \) for three different sizes of sample \( n = 50, 100 \) and 200. To find the 95% posterior credible region for the QRF for a given co-variate value using Empirical Bayes method, we calculate the posterior mean of the estimated QRF for all \( \tau \in [0, 1] \) for each value of \( k \in D \) at that given covariate value \( X = x \). Since we run 20000 iterations with burn-in 5000, only last \( H = 15000 \) are used for calculating the posterior mean. For a given \( k \in D \) and covariate value \( X = x \), define \( R^{EB}_{(x,k)} = \{a^1_{(x,k)}, \ldots, a^H_{(x,k)}\} \) such that

\[
a^i_{(x,k)} = \sup_{\tau \in [0,1]} |Q(i)(\tau|x,k) - \hat{Q}_y(\tau|x,k)|, \quad i = 1, \ldots, 15000
\]

where \( Q(i)(\tau|x,k) \) denotes the QRF for \( X = x \) in the \( i \)-th iteration after burn-in for B-spline method with \( k \) partitions and \( \hat{Q}_y(\tau|x,k) \) denotes the posterior mean of the QRF for \( X = x \) for B-spline method with \( k \) partitions. After that we calculate the 95\(^{th}\) percentile of \( R^{EB}_{(x,k)} \) for each \( k \in D \). Thus we find the size of the 95% posterior credible region for each value of \( k \in D \). To find the coverage, we repeat the simulation study 1000 times under different random number seeds and count the number of times the distance of the true QRF from the estimated EB estimate for all \( \tau \in [0, 1] \) is not more than the size of the 95% posterior credible region for EB approach of corresponding \( k \).

In the HB method, we again have 20000 iterations with burn-in 5000 and only last \( H = 15000 \) are used for calculating the posterior mean. To find the 95% posterior credible region for the QRF for a given co-variate value \( X = x \) in this case, we calculate \( R^{HB}_{(x)} = \{a^1_{(x)}, \ldots, a^H_{(x)}\} \) such that

\[
a^i_{(x)} = \sup_{\tau \in [0,1]} |Q(i)(\tau|x) - \hat{Q}_y(\tau|x)|, \quad i = 1, \ldots, 15000,
\]

where \( Q(i)(\tau|x) \) denotes the QRF for \( X = x \) in the \( i \)-th iteration after burn-in for B-spline method and \( \hat{Q}_y(\tau|x) \) denotes the posterior mean of the QRF for \( X = x \) for B-spline method. The weights corresponding to each \( k \in D \) to find \( Q(j)(\tau|x) \) and \( \hat{Q}_y(\tau|x) \) are the same and they are derived as mentioned in Section 2.3.3. Then we find the size of the 95% posterior credible region by calculating the 95\(^{th}\) percentile of \( R^{HB}_{(x)} \). To find the coverage, we repeat the simulation study 1000 times under different random number seeds and count the number of times the distance of the true QRF from the estimated HB estimate.

---

\(^1\)http://www4.stat.ncsu.edu/~reich/code/SpaceQRapprox.R
for all $\tau \in [0, 1]$ is not more than the size of the 95% posterior credible region for HB approach.

It is well-known that posterior credible band of smooth functions have an under coverage property (see Cox (1993); Knapik et al. (2011); Szabo et al. (2015); Yoo and Ghosal (2016)). To alleviate the problem, undersmoothing or modifying the credible region is needed. We inflate the obtained credible region by blowing the radius of the region by a slowly increasing factor. We choose the inflation factor $f(n) = 0.8\sqrt{\log n}$ in our examples. This works for all the sample sizes $n = 50, 100$ and $200$ under different simulation settings and different true quantile regression functions. We report two simulation studies in this chapter described in Section 2.5.1 and 2.5.2. We have also provided the posterior coverage with and without inflation of uniform 95% posterior credible interval for three different sample sizes under two different true quantile regression functions.

**First Study**

Consider $Q(\tau|x) = x\xi_1(\tau) + (1-x)\xi_2(\tau)$ where

$$\xi_1(\tau) = (1-A)\tau^2 + A\tau, \quad \xi_2(\tau) = (1-B)\tau^2 + B\tau$$

and $A = 0.3$, $B = 0.6$. We note that $\xi_1$ and $\xi_2$ are strictly increasing function from $[0, 1]$ to $[0, 1]$ satisfying $\xi_1(0) = \xi_2(0) = 0$ and $\xi_1(1) = \xi_2(1) = 1$. Note then the conditional quantile function is given by $Q(\tau|x) = a(x)\tau^2 + b(x)\tau$ where $a(x) = x(1-A) + (1-x)(1-B)$ and $b(x) = xA + (1-x)B$. Observe that since the quantile function is the inverse of the cumulative distribution function, for $U \sim U(0,1)$ the random variable $Q(U|x)$ has conditional quantile function $Q(\tau|x)$. We generate $n$ values $x_1, \cdots, x_n$ of the predictor variable $X$ independently from $U(0,1)$. Then we simulate $Y$ variable from the following equation

$$Y_i = a_iU_i^2 + b_iU_i \quad \text{for all} \quad i = 1, \ldots, n$$

where $a_i = x_i(1-A) + (1-x_i)(1-B)$, $b_i = x_iA + (1-x_i)B$ and $U_i$'s are i.i.d. $U(0,1)$, $i = 1, \ldots, n$. We simulated $n = 100$ observations from the distribution and compared the estimated results with the true ones for both Gaussian and B-spline methods. The highest and the lowest observed value of the acceptance ratios under in this simulation study are noted to be 0.24 and 0.58 respectively. For analyzing the mixing of the MCMC chain in the proposed method, we compute the value of the Gelman-Rubin (GR) statistics (see Gelman and Rubin (1992)) of $\xi_1(\tau)$ and $\xi_2(\tau)$ at $\tau = 0.25, 0.5, 0.75$ for both QSSQR and...
CSSQR approaches under all the considered values of $k$ using 10 identical chains under different random number generator seeds in each case. It is noted that out of 96 observed GR statistics (QSSQR and CSSQR for $D = \{3, \ldots, 10\}$ and $\{5, \ldots, 12\}$ respectively, 6 statistics in each case), in 64 cases (about 67%), the GR statistics stays between 1.00 and 1.10 and the highest observed value of the GR statistics is 1.48.

The comparative study of the performances in estimation of our method with other methods under this simulation study has been provided in Figure 2.1. We found that there is not much difference between the estimates given by quadratic and cubic B-spline approaches. Hence, for convenience, we only compared the output of our Hierarchical Bayes QSSQR method with other proposed methods in these figures. The root mean integrated squared error (RMISE) is given by the square root of the average of the square of the differences of the estimated and the true values of the curves at those grid points. In Table 2.1, we compared the RMISE for above mentioned estimated curves for all methods. It can be noted that in our simulation study, the estimated slope, intercept, quantile regression function for $x = 0.2, 0.5, 0.7$ and $\tau = 0.25, 0.5, 0.75$ are curves on the domain $[0, 1]$. To calculate the RMISE, we divide the interval $[0, 1]$ using partition $(t_0, t_1, \ldots, t_{100})$ such that $0 = t_0 < t_1 < \cdots < t_{100} = 1$ such that $(t_i - t_{i-1}) = 0.01$ for all $i = 1, \ldots, 100$. We note that our proposed methods have lower RMISE of estimation curves than that of GSQR. Among other methods, qreg function from BSquare worked quite good, though our method yields lower RMISE values for most of the cases. Among our set of proposed methods, we do not see any noticeable improvement by using CSSQR instead of QSSQR. We note that the HB approach has lower RMISE values than the corresponding EB alternative.

In Table 2.2, we present estimation accuracy and coverage with and without inflation of the proposed methods for three different sample sizes $n = 50, 100, 200$. We note that the posterior coverage of the inflated credible band for the Hierarchical Bayes are typically better than that of corresponding Empirical Bayes. No noticeable improvement is observed by using CSSQR instead of QSSQR.

**Second Study**

To check the performance of the proposed Bayesian method when the quantile function is not a polynomial, we consider,

$$
\xi_1(\tau) = \sin\left(\frac{\pi \tau}{2}\right), \quad \xi_2(\tau) = \frac{\log(1 + \tau)}{\log 2}
$$
Figure 2.1: (First Simulation study) Comparison of true and estimated intercept, slope and estimated quantile regression functions (QRF) at $x = 0.3, 0.5, 0.7, \tau = 0.25, 0.5, 0.75$ and $n = 100$ for different methods QSSQR(HB), GSQR, qreg, qreg.spline and BPSQR.
Figure 2.2: (First Simulation study) MCMC auto-correlation and trace plots of $\xi_1(\tau)$ at $\tau = 0.25, 0.5, 0.75$ and $\xi_2(\tau)$ at $\tau = 0.25, 0.5, 0.75$ respectively in row 1–6 for QSSQR ($k = 6$), CSSQR ($k = 9$) and GSQR.
Table 2.1: (First Simulation study) Comparison of RMISE of estimation of slope, intercept and quantile regression function at $x = 0.3, 0.5, 0.7$, $\tau = 0.25, 0.50, 0.75$ and $n = 100$ under different methods QSSQR(HB); QSSQR(EB); CSSQR(HB); CSSQR(EB); GSQR; BPSQR; BSquare 1 and BSquare 2 (denote qreg and qreg_spline function under BSquare package in R respectively)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Intercept</th>
<th>Slope</th>
<th>$x = 0.3$</th>
<th>$x = 0.5$</th>
<th>$x = 0.7$</th>
<th>$\tau = 0.25$</th>
<th>$\tau = 0.50$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSSQR(HB)</td>
<td>0.039</td>
<td>0.059</td>
<td>0.029</td>
<td>0.037</td>
<td>0.030</td>
<td>0.008</td>
<td>0.035</td>
<td>0.038</td>
</tr>
<tr>
<td>QSSQR(EB)</td>
<td>0.097</td>
<td>0.088</td>
<td>0.071</td>
<td>0.081</td>
<td>0.037</td>
<td>0.025</td>
<td>0.076</td>
<td>0.066</td>
</tr>
<tr>
<td>CSSQR(HB)</td>
<td>0.026</td>
<td>0.048</td>
<td>0.034</td>
<td>0.042</td>
<td>0.05</td>
<td>0.021</td>
<td>0.03</td>
<td>0.051</td>
</tr>
<tr>
<td>CSSQR(EB)</td>
<td>0.033</td>
<td>0.069</td>
<td>0.034</td>
<td>0.041</td>
<td>0.051</td>
<td>0.051</td>
<td>0.051</td>
<td>0.014</td>
</tr>
<tr>
<td>GSQR</td>
<td>0.075</td>
<td>0.052</td>
<td>0.091</td>
<td>0.101</td>
<td>0.112</td>
<td>0.121</td>
<td>0.14</td>
<td>0.089</td>
</tr>
<tr>
<td>BPSQR</td>
<td>0.267</td>
<td>0.353</td>
<td>0.177</td>
<td>0.118</td>
<td>0.063</td>
<td>0.181</td>
<td>0.168</td>
<td>0.145</td>
</tr>
<tr>
<td>BSquare 1</td>
<td>0.100</td>
<td>0.019</td>
<td>0.089</td>
<td>0.082</td>
<td>0.074</td>
<td>0.106</td>
<td>0.142</td>
<td>0.026</td>
</tr>
<tr>
<td>BSquare 2</td>
<td>0.101</td>
<td>0.003</td>
<td>0.093</td>
<td>0.084</td>
<td>0.077</td>
<td>0.103</td>
<td>0.118</td>
<td>0.019</td>
</tr>
</tbody>
</table>

again, we note that $\xi_1$ and $\xi_2$ are strictly increasing function from $[0, 1]$ to $[0, 1]$ satisfying $\xi_1(0) = \xi_2(0) = 0$ and $\xi_1(1) = \xi_2(1) = 1$. We generate a sample of size $n = 100$ using the quantile function given by

$$Q(\tau|x) = x\xi_1(\tau) + (1 - x)\xi_2(\tau).$$

In this case, the assumptions, precision, prior model parameter values, number of iterations and burn-in have been taken to be the same as the previous study for all above-mentioned models. The highest and the lowest observed value of the acceptance ratios under in this simulation study are noted to be 0.18 and 0.55 respectively. Similar to the first simulation study, GR statistics are calculated for $\xi_1(\tau)$ and $\xi_2(\tau)$ at $\tau = 0.25, 0.5, 0.75$ using 10 identical MCMC chains under different random number generator seeds for all the cases. For this simulation study, out of 96 cases, 66 observed GR statistics are between 1.00 and 1.10 and the highest observed value is 1.50.

Under this simulation study, the comparative study of the performances in estimation of our method with other methods has been provided in Figure 2.3. Like the previous study, in this case also, we found that estimates given by quadratic and cubic B-spline models are similar. So we only compared the output of our QSSQR(HB) with other methods in the above-mentioned figures. In Table 2.3, we compared the RMISE of the estimated curves for all methods. In Table 2.4, estimation performance and posterior
Table 2.2: (First Simulation study) Size and posterior coverage of inflated (in bold) and regular uniform 95% posterior credible interval of estimated quantile regression function for $x = 0.2, 0.5, 0.7$ for $\tau \in [0, 1]$ for three different sizes of sample $n = 50, 100, 200$ for QSSQR(HB), QSSQR(EB), CSSQR(HB) and CSSQR(EB).

<table>
<thead>
<tr>
<th>Degree</th>
<th>Sample size</th>
<th>Type</th>
<th>$x = 0.2$</th>
<th>$x = 0.5$</th>
<th>$x = 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Size</td>
<td>Coverage</td>
<td>Size</td>
</tr>
<tr>
<td>QSSQR</td>
<td>$n = 50$</td>
<td>HB</td>
<td>0.2394</td>
<td>99.7</td>
<td>0.1936</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1513</td>
<td>86.1</td>
<td>0.1224</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2327</td>
<td>95.9</td>
<td>0.1935</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1471</td>
<td>70.4</td>
<td>0.1223</td>
</tr>
<tr>
<td></td>
<td>$n = 100$</td>
<td>HB</td>
<td>0.1925</td>
<td>96.6</td>
<td>0.1734</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1121</td>
<td>57.2</td>
<td>0.1010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2391</td>
<td>97.2</td>
<td>0.2079</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1393</td>
<td>67.1</td>
<td>0.1211</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>HB</td>
<td>0.2094</td>
<td>98.8</td>
<td>0.1727</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1137</td>
<td>61.2</td>
<td>0.0938</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2296</td>
<td>96.9</td>
<td>0.1803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1247</td>
<td>65.0</td>
<td>0.0979</td>
</tr>
<tr>
<td>CSSQR</td>
<td>$n = 50$</td>
<td>HB</td>
<td>0.2475</td>
<td>99.8</td>
<td>0.2109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1564</td>
<td>86.5</td>
<td>0.1333</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2361</td>
<td>97.7</td>
<td>0.1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1492</td>
<td>68.7</td>
<td>0.1260</td>
</tr>
<tr>
<td></td>
<td>$n = 100$</td>
<td>HB</td>
<td>0.2316</td>
<td>98.9</td>
<td>0.1775</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1349</td>
<td>68.9</td>
<td>0.1034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2168</td>
<td>95.8</td>
<td>0.1768</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1263</td>
<td>61.5</td>
<td>0.1030</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>HB</td>
<td>0.1978</td>
<td>94.8</td>
<td>0.1749</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1074</td>
<td>34.2</td>
<td>0.0950</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.1921</td>
<td>92.9</td>
<td>0.1729</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1043</td>
<td>38.7</td>
<td>0.0939</td>
</tr>
</tbody>
</table>

Coverage with and without inflation using both quadratic and cubic splines are given.
In this case also, the proposed methods yield lower RMISE than other methods. Again, no noticeable improvement on using CSSQR over QSSQR has been observed in this scenario also. Like in the previous study, the HB method has lower RMISE values and higher posterior coverage than the corresponding EB method.

In Figure 2.2, the trace plots of $\xi_1(\tau)$ and $\xi_2(\tau)$ have been provided at $\tau = 0.25, 0.5, 0.75$ in the corresponding MCMC using QSSQR, CSSQR and GSQR methods for the first simulation study. For QSSQR and CSSQR, the trace plots have been provided for $k = 6$ and $k = 9$ respectively. In Table 2.5, we compare the computation time of our proposed method with the only existing method of simultaneous quantile regression (if we disregard other existing plug-in type methods of non-crossing quantile regression), i.e., GSQR. For QSSQR, the computation time for both HB and EB are same. The same is also true for CSSQR. We note that, computation time of both QSSQR and CSSQR are lower than that of GSQR, but the computation time of CSSQR is greater than that of QSSQR.
as expected. For simulation purposes, all these codes have been written in MATLAB. Simulations have been performed in a cluster with DELL R815 Quad Processor AMD Opteron 16 core 2.3 GHz machines with 512GB RAM, each running 64Bit Fedora Core 20.

Table 2.3: (Second Simulation study) Comparison of RMISE of estimation of slope, intercept and quantile regression function at \(x = 0.2, 0.5, 0.7, \tau = 0.25, 0.50, 0.75\) and \(n = 100\) under different methods. QSSQR(HB); QSSQR(EB); CSSQR(HB); CSSQR(EB); GSQR; qreg; qreg_spline and BPSQR.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Intercept</th>
<th>Slope</th>
<th>(x = 0.2)</th>
<th>(x = 0.5)</th>
<th>(x = 0.7)</th>
<th>(\tau = 0.25)</th>
<th>(\tau = 0.50)</th>
<th>(\tau = 0.75)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSSQR(HB)</td>
<td>0.022</td>
<td>0.060</td>
<td>0.020</td>
<td>0.029</td>
<td>0.038</td>
<td>0.037</td>
<td>0.027</td>
<td>0.043</td>
</tr>
<tr>
<td>QSSQR(EB)</td>
<td>0.044</td>
<td>0.078</td>
<td>0.032</td>
<td>0.023</td>
<td>0.028</td>
<td>0.011</td>
<td>0.039</td>
<td>0.034</td>
</tr>
<tr>
<td>CSSQR(HB)</td>
<td>0.025</td>
<td>0.055</td>
<td>0.029</td>
<td>0.040</td>
<td>0.049</td>
<td>0.031</td>
<td>0.036</td>
<td>0.060</td>
</tr>
<tr>
<td>CSSQR(EB)</td>
<td>0.035</td>
<td>0.064</td>
<td>0.042</td>
<td>0.055</td>
<td>0.066</td>
<td>0.060</td>
<td>0.059</td>
<td>0.068</td>
</tr>
<tr>
<td>GSQR</td>
<td>0.063</td>
<td>0.089</td>
<td>0.081</td>
<td>0.108</td>
<td>0.126</td>
<td>0.091</td>
<td>0.152</td>
<td>0.135</td>
</tr>
<tr>
<td>qreg</td>
<td>0.071</td>
<td>0.068</td>
<td>0.070</td>
<td>0.074</td>
<td>0.080</td>
<td>0.185</td>
<td>0.014</td>
<td>0.042</td>
</tr>
<tr>
<td>qreg_spline</td>
<td>0.849</td>
<td>0.091</td>
<td>0.860</td>
<td>0.868</td>
<td>0.873</td>
<td>0.551</td>
<td>0.286</td>
<td>0.082</td>
</tr>
<tr>
<td>BPSQR</td>
<td>0.317</td>
<td>0.328</td>
<td>0.235</td>
<td>0.112</td>
<td>0.035</td>
<td>0.169</td>
<td>0.196</td>
<td>0.182</td>
</tr>
</tbody>
</table>

Application to Hurricane Intensity Data

Elsner et al. (2008) argued that the strongest hurricanes in the North Atlantic basin have gotten stronger over the last couple of decades. We apply our method to the hurricane intensity data\(^2\) in the North Atlantic basin during the period 1981–2006. We use the weighted quadratic spline procedure, i.e., QSSQR(HB).

To use QSSQR(HB), we first mapped the explanatory variable years to the interval \([0, 1]\) by change of scale and origin. In this case, we map the year 1981 to 0 and the year 2006 to 1. To map the response variable, the wind speed of the hurricanes at their maximum to the interval \([0, 1]\), we assumed that the velocities of the cyclone are coming from a Pareto distribution. The form of the power-Pareto density is given by

\[
f(y) = \frac{ak(y/\sigma)^{k-1}}{\sigma(1 + (y/\sigma)^k)^{(a+1)}} \quad y > 0
\]

Similar to Tokdar and Kadane (2012) we fix the values of the parameters as \(a = 0.45\), \(\sigma = 1\) and \(k = 3\).

Figure 2.3: (Second Simulation study) Comparison of true and estimated intercept, slope and estimated quantile regression functions (QRF) at $x = 0.2, 0.5, 0.7, \tau = 0.25, 0.5, 0.75$ and $n = 100$ for different methods QSSQR(HB), GSQR, qreg, qreg_spline and BPSQR.
Table 2.4: (Second Simulation study) Size and posterior coverage of inflated (in bold) and regular uniform 95% posterior credible interval of estimated quantile regression function for \( x = 0.2, 0.5, 0.7 \) for \( \tau \in [0, 1] \) for three different sizes of sample \( n = 50, 100, 200 \) for QSSQR(HB), QSSQR(EB), CSSQR(HB) and CSSQR(EB).

<table>
<thead>
<tr>
<th>Degree</th>
<th>Sample size</th>
<th>Type</th>
<th>( x = 0.2 )</th>
<th>( x = 0.5 )</th>
<th>( x = 0.7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Size</td>
<td>Coverage</td>
<td>Size</td>
</tr>
<tr>
<td>QSSQR</td>
<td>( n = 50 )</td>
<td>HB</td>
<td>0.2755</td>
<td>100</td>
<td>0.2149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2514</td>
<td>97.6</td>
<td>0.2157</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>HB</td>
<td>0.2421</td>
<td>99.9</td>
<td>0.1894</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.1997</td>
<td>93.4</td>
<td>0.1815</td>
</tr>
<tr>
<td></td>
<td>( n = 200 )</td>
<td>HB</td>
<td>0.1701</td>
<td>95.1</td>
<td>0.1539</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.1921</td>
<td>93.8</td>
<td>0.1856</td>
</tr>
<tr>
<td>CSSQR</td>
<td>( n = 50 )</td>
<td>HB</td>
<td>0.2563</td>
<td>100</td>
<td>0.1880</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2609</td>
<td>98.8</td>
<td>0.1772</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>HB</td>
<td>0.2010</td>
<td>96.0</td>
<td>0.1636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2192</td>
<td>96.4</td>
<td>0.1755</td>
</tr>
<tr>
<td></td>
<td>( n = 200 )</td>
<td>HB</td>
<td>0.1897</td>
<td>94.6</td>
<td>0.1698</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>0.2051</td>
<td>95.1</td>
<td>0.1777</td>
</tr>
</tbody>
</table>

\( \sigma = 52 \) and \( k = 4.9 \). The distribution function is given by

\[
F(y) = 1 - \frac{1}{(1 + (y/\sigma)^k)^a} \tag{2.20}
\]

Using Equation (2.20), we transform the hurricane wind speeds to the percentile values. Now, transformed \( y \)-values are well in the \([0, 1]\) interval and then apply the QSSQR(HB) method. After we evaluate our estimated quantile functions \( \hat{\xi}_1(\tau) \) and \( \hat{\xi}_2(\tau) \), we find out the slope and intercept at functions of \( \tau \).

After we evaluate them, suppose we want to estimate the wind speed of hurricanes at any particular year of the period 1981–2006, at a given quantile. First we transform the given year to a value well within the interval \([0, 1]\) via linear transformation. Then we evaluate the value of transformed wind speeds. After that we use the inverse transformation of Equation (2.20) on estimated transformed speed to find the estimated wind speed at that desired quantile of that year.
Figure 2.4: 
(a) Comparison of $W_{\text{maxST}}$ during first and last 10 years: The dotted superimposed line is assumed power-Pareto density used for transforming velocities; the solid superimposed lines are approximate density of $W_{\text{maxST}}$ for corresponding decades. 
(b) Simultaneous Quantiles: Estimated quantile curves of velocities over the period 1981–2006 has been shown for the quantiles $\tau \in \{0.05, 0.10, 0.20, \ldots, 0.80, 0.90, 0.95\}$. 
(d) Posterior probability of slope being negative for $\tau \in [0.5, 1]$. 


Table 2.5: Computation time (in seconds) of QSSQR, CSSQR and GSQR for simulation study 1 and 2 with sample size \( n = 100 \).

<table>
<thead>
<tr>
<th>Methods</th>
<th>QSSQR</th>
<th>CSSQR</th>
<th>GSQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>672</td>
<td>776</td>
<td>1479</td>
</tr>
<tr>
<td>Study 2</td>
<td>666</td>
<td>902</td>
<td>1566</td>
</tr>
</tbody>
</table>

In the data, \( W_{\text{maxST}} \) stands for the velocity of the cyclones. In Figure 2.4a, we note that the higher velocity cyclones are more frequent in the period 1997–2006 period than 1981–1990 period. We show the estimates for different quantiles over the period 1981–2006 in Figure 2.4b by using QSSQR (HB). We note that, the the quantile regression curves (QRF) are more steeper for higher quantiles than the lower quantiles.

In order to check whether really the strongest tropical cyclones in the North Atlantic basin have gotten stronger over the last couple of decades (argued by Elsner et al. (2008)), we draw the estimated velocities as a function of quantiles for the years 1981, 1986, 1991, 1996, 2001 and 2006 in Figure 2.4c. We note that, for these 6 equidistant years, the estimated \( W_{\text{maxST}} \) corresponding to the lower quantiles are not pretty much different. While, at higher quantiles, the estimated \( W_{\text{maxST}} \) has an increasing trend with time, i.e., more recent years has estimated \( W_{\text{maxST}} \) more than the older years for higher quantiles.

In Figure 2.4d, we show the posterior probabilities of the slope to be negative at the higher quantiles, i.e., for \( \tau \in [0.5, 1] \).

**Application to US Population Data**

For last few decades, the population of the states of USA are changing. But the rate of change of population is not the same over all zones of USA. We can divide the whole USA mainly in 4 regions namely Northeast, Midwest, South and West. Due to the current trend of globalization, the rate of change of population over the time is different for all these regions of USA. We apply the QSSQR(HB) method of simultaneous quantile regression on the population data\(^3\) of USA over the period 1985–2010. We use the USGS data where we can found population of each county of USA for the years 1985, 1990, 1995, 2000, 2005 and 2010.

Before applying our method, we did a monotone transformation so that both predictor and response variables lie in between 0 and 1. We transform the years to the unit interval via linear transformation so that the year 1985 gets mapped to 0 and the year 2010

\(^3\)Source [http://water.usgs.gov/watuse/data/](http://water.usgs.gov/watuse/data/)
gets mapped to 1. For the transformation of the population, we considered, the county-wise population of each region follows log-normal density. We fit log-normal density to each of these regions separately. Then for each region, for each county, we use the cumulative distribution function of the population according to the corresponding log-normal distribution. After transforming both explanatory and response variables into the unit interval, we did our analysis. After our analysis we transform the results back to the original scale via the inverse transformation.

Figure 2.5: Estimated quantile curves of county wise population of Northeast, Midwest, South and West regions of USA for the quantiles \( \tau \in \{0.05, 0.10, 0.20, \ldots, 0.80, 0.90, 0.95\} \) over the years 1985–2010.

In Figure 2.5, we show the simultaneous quantiles of county-wise population for all 4 regions of USA over the period 1985–2010. In Figure 2.6 we plotted the estimated
quantile regression function curve of population for $\tau \in [0, 1]$ for the years 1985, 1990, 1995, 2000, 2005 and 2010 for all regions.

In Figure 2.7 we plot the posterior probabilities of the slope of the simultaneous quantile regression estimates to be negative for $\tau \in [0, 1]$ for all 4 regions. We note that for the South region, the posterior probability of the slope being negative is below 0.05 for all quantiles. Therefore in this region, population of all types of counties have increased over the time. In the West, for $\tau$ greater than 0.35, the posterior probability of slope being negative is less than 0.05. That implies population of the more populated counties have increased at higher rate than less populated counties. In the Midwest, we note there is a big difference between the posterior probability of the slope being negative among the sparsely and highly populated counties. Here, the population of sparsely populated counties have decreased over years, on the other hand, the population of highly populated counties...
Figure 2.7: Estimated probability of negative slope for $\tau \in [0, 1]$ of Northeast, Midwest, South and West regions of USA for the years 1985, 1990, 1995, 2000, 2005 and 2010.

counties have increased. The possible reason is that as fewer and fewer people are now relying on agriculture for their livelihood, new jobs in the Midwest nowadays are available almost exclusively in urban areas. In the Northeast, the probability of the slope being negative is low for sparsely populated counties and relatively high for the highly populated counties. Since there are many highly populated cities in the Northeast, and most of these cities are over-populated and expensive, people tend to move to the outlying areas which still have all facilities and good connection to urban centers.

Overall, we note that except in the Northeast, population of all other regions are significantly increasing for the high-populated counties. In general, most people immigrating to USA move in urban centers where jobs are more easily available. Immigration is also a significant force for the population growth of USA. Due to globalization, people from different countries move in USA. Commonly they move in high population areas in
the urban areas because offices, business organizations, universities and other places of interests are mainly located in those regions. This explains the reason why the population of the counties with high population are generally increasing faster over time compared with sparsely populated counties.

**Conclusion**

In this chapter we have proposed a Bayesian non-parametric method for fitting simultaneous linear quantile regression using quadratic and cubic B-spline basis function. In order to estimate the coefficients of the B-spline basis functions, we use Metropolis-Hastings algorithm. Finally, to select the optimum number of B-spline basis functions, we calculate the model marginal likelihoods and choose the model with the highest marginal log-likelihood, which we refer as EB. An alternative approach is to use the hierarchical Bayes approach, denoted by HB, which is a combination of a reasonable range of numbers of B-spline basis function with weights proportional to their model posterior probabilities. From the simulation study on RMISE and posterior coverage of the credible bands, we note somewhat higher accuracy with HB approach over the EB. Since, computation time for both of these methods are the same, HB is preferred over the EB. Even though the use of cubic B-spline basis function is more time consuming than that of the quadratic B-spline basis, we do not find any noticeable improvement in accuracy by using the former over the latter. However the cubic B-spline approach produces estimates which are second order continuously differentiable, instead of only continuously differentiable estimates obtained by the quadratic B-spline approach.

Unlike most of the previous works on non-crossing quantile regression, the proposed method can estimate the slope and the intercept of the quantile regression equation as continuous function of \( \tau \). Most existing methods for non-crossing quantile regression depend on the chosen grid of \( \tau \) values where the quantile regression coefficients are estimated. Due to the use of characterization of required monotonicity of the simultaneous linear quantile regression, it can be avoided in our approach. Besides, estimating the quantile regression equation simultaneously gives further insight on the dependence structure of the predictor and the response variable instead of looking at a fixed and finite number of quantiles.

For the B-spline basis expansion approach, in the likelihood evaluation step, due the piece-wise polynomial structure of B-spline basis function, we can solve Equation (2.6) analytically. This is unlike the case of using a Gaussian process prior where only a nu-
numerical solution can be obtained after implementing numerical integration based on a chosen grid. In our proposed method, dependency structure of the slope, intercept and the quantile function of the regression equation with the spline coefficients being linear, it is enough to track the posterior mean of the spline coefficients for the estimation purpose. Evaluation and storage of the estimated quantile functions at grid points over the domain of estimation is unnecessary for our method since, those can be directly derived from the posterior mean of the estimated spline coefficients. So we can unfold each and every details of our estimation by only using posterior mean of a few number of parameters (and model weight vector for HB). This is sharply in contrast with using a Gaussian process prior, for which each realization of the quantile function from its posterior distribution needs to be stored on a grid. The grid also needs to be sufficiently fine to maintain accuracy and smoothness. Though taking more dense grid would improve the quality of estimation, on the other hand, it would increase the computation time and will lead to the problems of singularity of co-variance matrix of the underlying Gaussian process. Again for estimation, we need to evaluate the estimated quantile function at any point by kriging or interpolation. B-spline method does not need any subsequent kriging or interpolation step as the entire estimation is obtained from the estimated coefficients. In our simulation study, we noted that the proposed method based on B-spline basis expansion has slightly more accuracy than the Gaussian process based method for simultaneous quantile regression and several other non-crossing quantile regression methods.

Application of our method on the North Atlantic hurricane intensity data reveals that the intensity of the strongest hurricanes in the North Atlantic basin have gotten stronger whereas we did not notice any significant increase in the wind speeds of the hurricanes near median or lower quantiles over the time period 1981–2006. We also applied this method to analyze the rate of change of county level population in the four geographical regions of USA, namely, the Northeast, Midwest, South and West, based on the 5-yearly data over the period 1985–2010. It gives us a broader look into how the county population of different regions of USA are changing over time. We note that except sparsely populated counties of the Midwest and urban center of the Northeast, the population has generally increased over time in all regions.
Chapter 3

Single Index Model for Multivariate Predictors

Introduction

In this chapter we extend the proposed method to the case where the explanatory variable is multivariate. In general quantile regression with multivariate predictors has a cumbersome representation (Yang and Tokdar (2016)). However, if we assume that the response variable is dependent on the explanatory variable only through a linear projection, then the method proposed in Chapter 2 can be extended. In other words, we assume the quantile levels of $Y$ are dependent on $\eta^T X$ for some unknown vector $\eta = (\eta_1, \ldots, \eta_d)$. By monotone transformation, each coordinate of $X$ are transformed to $[-1,1]$ and $Y$ is transformed to unit interval. Hereby after the transformation, $X \in [-1,1]^d$, $Y \in [0,1]$ where $d$ is the dimension of $X$. $\eta$, the linear projection vector is taken on a unit $d$-dimensional sphere. Suppose $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and the infinity norms respectively. Now by the Cauchy-Schwarz inequality

$$|\eta^T X| \leq \|\eta\| \|X\| \leq \|\eta\| \sqrt{d} \|X\|_\infty \leq \sqrt{d}$$

as $\|\eta\| = 1$ and $\|X\|_\infty \leq 1$. Therefore,

$$0 \leq \left(\frac{1}{2} + \frac{\eta^T X}{2\sqrt{d}}\right) \leq 1.$$  \hspace{1cm} (3.1)
Then, similar to Equation (2.3), the quantile function is given by

\[
Q(\tau|x) = \left(\frac{1}{2} + \frac{\eta^T x}{2\sqrt{d}}\right)\xi_1(\tau) + \left[1 - \left(\frac{1}{2} + \frac{\eta^T x}{2\sqrt{d}}\right)\right]\xi_2(\tau)
= \left(\frac{1}{2} + \frac{\eta^T x}{2\sqrt{d}}\right)\xi_1(\tau) + \left(\frac{1}{2} - \frac{\eta^T x}{2\sqrt{d}}\right)\xi_2(\tau)
= (\xi_1(\tau) + \xi_2(\tau))/2 + \frac{1}{2\sqrt{d}}(\xi_1(\tau) - \xi_2(\tau))\eta^T x.
\] (3.2)

**Likelihood evaluation, MCMC and Transition**

From Equation (3.1) for any given \(\eta\), the likelihood can be evaluated in a similar fashion as mentioned in Section 2.3.1 by replacing \(X_i\) with \(\left(\frac{1}{2} + \frac{\eta^T X_i}{2\sqrt{d}}\right)\). The priors and the transition steps of the parameters corresponding to \(\xi_1(\tau)\) and \(\xi_2(\tau)\) are kept similar to that mentioned in Section 2.3.1. Unlike the previous case, here the projection vector \(\eta\) (which belongs to the surface of the unit sphere) also needs to be estimated. In \(\mathbb{R}^d\), spherical coordinate system consist of a radial coordinate \(R\) and \((d - 1)\) angular coordinates \(\Phi = (\phi_1, \ldots, \phi_{d-1})\) where \(\phi_{d-1}\) ranges over \([0, 2\pi]\) and other angles range over \([0, \pi]\) radians. Hence any point \((l_1, \ldots, l_d)\) on the surface of unit \(d\)-dimensional sphere can be given by (see Blumenson (1960))

\[
l_i = \cos \phi_i \prod_{j=1}^{i-1} \sin \phi_j, \text{ for } i = 1, \ldots, d - 1, \quad l_d = \prod_{j=1}^{d-1} \sin \phi_j,
\]

where \(\phi_1, \ldots, \phi_{d-2} \in [0, \pi]\) and \(\phi_{d-1} \in [0, 2\pi]\). Hereby, any point on the surface of unit sphere can be uniquely denoted by \(\Phi = (\phi_1, \ldots, \phi_{d-1}) \in [0, \pi]^{d-2} \times [0, 2\pi]\) and vice-versa. For \(\phi_1, \ldots, \phi_{d-2}\), we take uniform \(U(0, \pi)\) prior and for \(\phi_{d-1}\) we take \(U(0, 2\pi)\) prior. We consider normal proposal density with mean 0 and standard deviation \(\kappa\) for the movement in the Metropolis-Hastings algorithm. While moving over the parameter space of any \(\phi_i\) for \(i = 1, \ldots, d - 2\), if any point is selected outside the range \([0, \pi]\), its modulus value with respect to \(\pi\) is taken. Similarly in the case of \(\phi_{d-1}\), we consider the final movement as the modulus of the new point with respect to \(2\pi\). Inside each iteration, the update steps of \(\{\gamma_j\}_{j=1}^{k+m-1}, \{\delta_j\}_{j=1}^{k+m-1}\) (same interpretation as in Section 2.3) and \(\Phi\) are performed in the following order.

(i) Initialize \(\{\gamma_j\}_{j=1}^{k+m-1}, \{\delta_j\}_{j=1}^{k+m-1}, \Phi = (\phi_1^{(0)}, \ldots, \phi_{d-1}^{(0)})\).

(ii) For \(i = 1, 2, \ldots\) during the \(i\)-th iteration starting with \(\{\gamma_j\}_{j=1}^{k+m-1} = \{\gamma_j^{(i-1)}\}_{j=1}^{k+m-1}, \quad \{\delta_j\}_{j=1}^{k+m-1} = \{\delta_j^{(i-1)}\}_{j=1}^{k+m-1}, \quad \Phi = (\phi_1^{(i-1)}, \ldots, \phi_{d-1}^{(i-1)})\).
\[
\{\delta_j\}_{j=1}^{k+m-1} = \{\delta_j^{(i-1)}\}_{j=1}^{k+m-1}, \quad \Phi = (\phi_1^{(i-1)}, \ldots, \phi_d^{(i-1)})
\]

(a) update \(\{\gamma_j\}_{j=1}^{k+m-1}, \{\delta_j\}_{j=1}^{k+m-1}\) fixing \(\Phi = (\phi_1^{(i-1)}, \ldots, \phi_d^{(i-1)})\) and obtain the updated values \(\{\gamma_j^{(i)}\}_{j=1}^{k+m-1}, \{\delta_j^{(i)}\}_{j=1}^{k+m-1}\).

(b) update \(\Phi\) fixing \(\{\gamma_j\}_{j=1}^{k+m-1}, \{\delta_j\}_{j=1}^{k+m-1}\) as well as \(\Phi\) as mentioned in Section 3.1.1. The best possible value of \(k\) and model weights corresponding to each considered values of \(k\) are compared as in Section 2.3.3. For this case also, we take \(D = \{3, \ldots, 10\}\) and \(D = \{5, \ldots, 12\}\) for QSSQR and CSSQR respectively.

Note that for any given values of \(k\) and \(m\), the mixing of the MCMC depends on two tuning parameters which are \(r\) (see Section 2.3.2) and \(\kappa\). So there the value should be set based on corresponding acceptance probabilities.

### Choosing value of \(k\) and Model Averaging

Similar to the univariate case, in this case also choosing the value of \(k\) plays a crucial part in the estimation procedure. But unlike the previous case, here in each iteration, we need to update the B-spline coefficients \(\{\gamma_j\}_{j=1}^{k+m-1}, \{\delta_j\}_{j=1}^{k+m-1}\) as well as \(\Phi\) as mentioned in Section 3.1.1. The best possible value of \(k\) and model weights corresponding to each considered values of \(k\) are compared as in Section 2.3.3. For this case also, we take \(D = \{3, \ldots, 10\}\) and \(D = \{5, \ldots, 12\}\) for QSSQR and CSSQR respectively.

### Simulation Study

For simulation purpose, we consider three univariate predictors \(X_1, X_2, X_3\) and one response variable \(Y\) whose quantile levels are dependent on the predictors through a linear projection. Suppose the quantile function is given by

\[
Q(\tau|x) = \frac{1}{2d}(\xi_1(\tau) - \xi_2(\tau))\eta^T x.
\]

In our case \(d = 3\). We consider

\[
\xi_1(\tau) = \frac{\log(1 + \sqrt{\tau})}{\log 2}; \quad \xi_2(\tau) = 1 - \cos\left(\frac{\pi \tau}{2}\right); \quad \eta = (0.347, 0.776, -0.526).
\]

Note that \(\xi_1\) and \(\xi_2\) are strictly increasing function from \([0, 1]\) to \([0, 1]\) satisfying \(\xi_1(0) = \xi_2(0) = 0\) and \(\xi_1(1) = \xi_2(1) = 1\) and \(\eta^T \eta = 1\). We take sample size \(n\) to be 100. The variables \(X_1, X_2, X_3\) are independently generated from \(U(-1, 1)\). The response variable \(Y\) is generated from Equation (3.3). To make inference in our proposed model, we consider
both HB and EB alternatives of QSSQR and CSSQR like the univariate case. The possible values of \( k \) have been taken in the sets to be \( D = \{3, \ldots, 10\} \) and \( D = \{5, \ldots, 12\} \) respectively for QSSQR and CSSQR. We run 20000 iterations discarding the first 5000 of them. The values of the \( r \) and \( \kappa \) are taken to be 1.25 and 1 respectively yielding acceptance probabilities between 0.2 and 0.5 for all the cases. Following Chib and Jeliazkov (2001), the marginal likelihood of the models corresponding to different values of \( k \) have been calculated in a similar fashion to the univariate case. For GSQR, we used \texttt{qrjoint} function of the R package \texttt{qrjoint} by Tokdar (2016). We draw 15000 posterior samples for estimation using this function. It should be noted that due to the inability of this function to deal with univariate predictor, we could not use it for the previous simulation studies for univariate cases. We also estimated the quantile functions using \texttt{qreg} and \texttt{qreg.spline} functions of \texttt{BSquare} package and in each cases 20000 iterations are performed and first 5000 iterations were not considered. For BPSQR also 20000 posterior samples were drawn disregarding the first 5000 iterations as burn-in.

For comparison, RMISE of the QRF has been evaluated at 8 points given by \((X_1, X_2, X_3) = (\pm 0.5, \pm 0.5, \pm 0.5)\) using all the above-mentioned methods. In Table 3.1, it is noted that the proposed method performed better in the most of the cases. No noticeable differences have been seen among the performances of the proposed methods. Among other methods, GSQR and \texttt{qreg} function of \texttt{BSquare} package performed quite well.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Point 1</th>
<th>Point 2</th>
<th>Point 3</th>
<th>Point 4</th>
<th>Point 5</th>
<th>Point 6</th>
<th>Point 7</th>
<th>Point 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSSQR(HB)</td>
<td>0.056</td>
<td>0.039</td>
<td>0.122</td>
<td>0.069</td>
<td>0.041</td>
<td>0.060</td>
<td>0.095</td>
<td>0.049</td>
</tr>
<tr>
<td>QSSQR(EB)</td>
<td>0.052</td>
<td>0.037</td>
<td>0.120</td>
<td>0.061</td>
<td>0.043</td>
<td>0.056</td>
<td>0.103</td>
<td>0.051</td>
</tr>
<tr>
<td>CSSQR(HB)</td>
<td>0.056</td>
<td>0.050</td>
<td>0.115</td>
<td>0.063</td>
<td>0.048</td>
<td>0.070</td>
<td>0.094</td>
<td>0.053</td>
</tr>
<tr>
<td>CSSQR(EB)</td>
<td>0.065</td>
<td>0.046</td>
<td>0.122</td>
<td>0.065</td>
<td>0.048</td>
<td>0.066</td>
<td>0.095</td>
<td>0.050</td>
</tr>
<tr>
<td>GSQR</td>
<td>0.068</td>
<td>0.032</td>
<td>0.150</td>
<td>0.118</td>
<td>0.059</td>
<td>0.093</td>
<td>0.060</td>
<td>0.060</td>
</tr>
<tr>
<td>\texttt{qreg}</td>
<td>0.143</td>
<td>0.077</td>
<td>0.172</td>
<td>0.102</td>
<td>0.039</td>
<td>0.075</td>
<td>0.071</td>
<td>0.065</td>
</tr>
<tr>
<td>\texttt{qreg.spline}</td>
<td>0.286</td>
<td>0.253</td>
<td>0.299</td>
<td>0.260</td>
<td>0.217</td>
<td>0.215</td>
<td>0.222</td>
<td>0.212</td>
</tr>
<tr>
<td>BPSQR</td>
<td>0.626</td>
<td>0.172</td>
<td>0.340</td>
<td>0.125</td>
<td>0.458</td>
<td>0.054</td>
<td>0.175</td>
<td>0.293</td>
</tr>
</tbody>
</table>

Table 3.1: (Simulation study : Multivariate) Comparison of RMISE of estimation of quantile regression functions (QRF) at the Points 1-8 which are \((X_1, X_2, X_3) = (-0.5, -0.5, -0.5), (-0.5, -0.5, 0.5), (-0.5, 0.5, -0.5), (-0.5, 0.5, 0.5), (0.5, -0.5, -0.5), (0.5, 0.5, -0.5), (0.5, 0.5, 0.5)\) respectively under different methods QSSQR(HB); QSSQR(EB); CSSQR(HB); CSSQR(EB); GSQR; \texttt{qreg}; \texttt{qreg.spline} and BPSQR.
Figure 3.1: (Simulation study: Multivariate) Comparison of true and estimated quantile regression functions (QRF) at \((X_1, X_2, X_3) = (\pm 0.5, \pm 0.5, \pm 0.5)\) for \(n = 100\) using QSSQR(HB), GSQR, qreg, qreg_spline and BPSQR.
Chapter 4

Bayesian Spatio-temporal Quantile Regression

Introduction

Consider the scenario where time is the explanatory variable and the values of the response variable corresponding to time-points are available for different spatial locations. The methods of estimating a single quantile regression curve has been extended to analyze spatial data (Lee and Neocleous (2010), Oh et al. (2011), Sobotka and Kneib (2012)). Various methods for modeling spatially varying distribution function have been proposed in last decade (Dunson and Park (2008), Gelfand et al. (2005), Reich and Fuentes (2007), Griffin and Steel (2006)). As mentioned in Reich (2012), these methods treat the conditional distribution of the response at each spatial location as unknown quantities and are estimated from the data.

One of the first work addressing the non-crossing issue of the spatial quantile regression was proposed in Reich et al. (2011). They proposed a two-stage method and ensured monotonicity of the estimated quantile functions using Bernstein polynomial basis function. Reich (2012) assumed separate non-crossing quantile functions for each spatial location which evolves over time. Spatial smoothing was performed using Gaussian process priors. Separate analysis for each spatial location may be sensitive to small sample sizes at individual locations. While this method allows estimation at all quantile levels, post-estimation processing step makes the quantification of uncertainty difficult. Instead of estimating the quantile function for a set of grid-points, the estimation of the entire quantile function is more informative. To avoid the small area estimation problem, instead of assuming unrelated quantile functions for each spatial locations, we assume
the quantile function to be smooth over space and linear in time. Similar to the method proposed in Chapter 2, the monotonicity constraint on the curves $\xi_1$ and $\xi_2$ are obtained through B-spline basis expansion with coefficients (which is a function of spatial location) increasing and lying in the unit interval. The coefficients of the B-spline basis expansion over $d$-dimensional space are modeled as tensor products of univariate B-splines basis functions. A closed form likelihood is obtained in terms of the parameters in $d$ independent simplexes whose dimensions depend of the degree of the used B-spline basis functions.

Unlike the methods proposed in Reich et al. (2011) and Reich (2012), the main advantage of this method is that once we estimate the parameters of the model, estimated response variable can be found for any spatial location, time and quantiles without further kriging or interpolation, and hence allows proper uncertainty quantification within the Bayesian paradigm. Most importantly, the approach incorporates small area estimation issues automatically in its modeling approach.

**Ozone in troposphere**

Ozone can be naturally found in the troposphere and other parts of the atmosphere. While, more ozone concentration at the stratosphere is desirable as it helps absorb the ultra-violet radiation, higher ozone concentration level at the troposphere is harmful for living beings. Ozone occurs naturally in the troposphere in low concentration. There are mainly two sources of tropospheric ozone. A significant amount of ozone is released by plants and soil. Other than that, sometimes ozone migrates down to the troposphere from the stratosphere. However, the extent of ground-level naturally occurring ozone concentration is not considered a threat to living beings and the environment.

On the other hand, ozone is a byproduct of many human activities. Increasing automobiles and industries are the main source the “bad” ozone in the ground-level. As mentioned in an online article\(^1\), current ground-level ozone concentration has doubled since 1900. Although no single source emits ozone directly, it is generated when the hydrocarbons and nitrogen-oxides, emitted by automobiles, fuel power plants and other industrial machineries, interact with each other in the presence of sunlight, specifically the ultra-violet (UV) ray. In general, ozone level reaches its highest level during the summer season. Typically, in the span of a day, ozone level reaches its highest level during mid and late afternoon.

\(^1\)Source [http://www.windows2universe.org/earth/Atmosphere/ozone_tropo.html&edu=high](http://www.windows2universe.org/earth/Atmosphere/ozone_tropo.html&edu=high)
Recently tropospheric ozone has been related to many health problems and environmental issues. Exposure to higher level of ozone might cause respiratory problems. In plants, it slows down growth and photosynthesis and damages internal cells. High ground-level ozone also damages textile dyes, rubbers and fibers and a few genre of paints.

Because of its unstable nature, there is no existing way to move the tropospheric ozone to the stratosphere which can be thought as the ideal way to deal with this environmental hazard. Under Clean Air act, US Environment Protection Agency (EPA) considers tropospheric ozone as one of the six pollutants considered to be harmful to human health and environment. In this chapter, we develop a spatio-temporal quantile regression model for measuring the ozone concentration level over the US and California during the period 2006–2015.

Suppose \( Q(\tau|x, z) \) denotes the \( \tau \)-th quantile of the dependent variable i.e., ozone concentration level at location \( z \) and time-point \( x \). We assume linear dependence structure of the dependent variable on time and the quantile function to be smooth over the space. In our model, the quantile function is given by

\[
Q(\tau|x, z) = \beta_0(\tau, z) + x\beta_1(\tau, z) \quad \text{for} \quad \tau \in [0, 1].
\]  

(4.1)

**Proposed Bayesian Method**

Our explanatory variable \( X \) is time. Let \( \{X_{li}\}_{i=1}^{n_l} \) and \( \{Y_{li}\}_{i=1}^{n_l} \) denote the values of \( X \) and \( Y \) at location \( z_i = (z_{1i}, \ldots, z_{di}) \) for \( l = 1, \ldots, L \), and let \( n_l \) denote the sample size at location \( z_i \). By monotonic transformation, \( X, Y \) and each coordinates of the spatial locations are transformed to unit intervals. From Theorem 1 of Tokdar and Kadane (2012), it follows that a linear specification

\[
Q(\tau|x) = \beta_0(\tau) + x\beta_1(\tau), \quad \tau \in [0, 1]
\]

is monotonically increasing in \( \tau \) for every \( x \in [0, 1] \) if and only if

\[
Q(\tau|x) = x\xi_1(\tau) + (1 - x)\xi_2(\tau) \quad \text{for} \quad \tau \in [0, 1], \ x \in [0, 1]
\]  

(4.2)

where \( \xi_1, \xi_2 : [0, 1] \mapsto [0, 1] \) are monotonically increasing in \( \tau \in [0, 1] \). Let \( Z = (z_1, \ldots, z_d) \), \( z_i \in [0, 1] \) for \( i = 1, \ldots, d \), denote the spatial location (site) where the response variable is measured. We assume that the dependence structure of the quantile function at any site \( Z = z \) is given by

\[
Q(\tau|x, z) = x\xi_1(\tau, z) + (1 - x)\xi_2(\tau, z), \quad \tau \in [0, 1], \ x, y \in [0, 1], \ z \in [0, 1]^d.
\]  

(4.3)
where \( \xi_1(\cdot, z), \xi_2(\cdot, z) \) are monotonic functions from \([0, 1]\) to \([0, 1]\). For any fixed \( z \), Equation (4.3) can be reframed as

\[
Q(\tau | x, z) = \beta_0(\tau, z) + x\beta_1(\tau, z), \quad \tau \in [0, 1], \quad x, y \in [0, 1],
\]

(4.4)

where \( \beta_0(\tau, z) = \xi_2(\tau, z) \) and \( \beta_1(\tau, z) = \xi_1(\tau, z) - \xi_2(\tau, z) \) denote the slope and the intercept of type quantile regression.

A B-spline basis expansion is one of the most convenient approaches for estimating a function on bounded interval. To estimate \( \xi_1(\cdot, z) \) and \( \xi_2(\cdot, z) \) in Equation (4.3), we use a B-spline basis expansion. Taking the coefficients of the B-splines basis functions in increasing order we ensure the monotonicity of these two above-mentioned functions and any monotone increasing function can be approximated through a monotone linear combination of B-splines (de Boor (2001)). In Chapter 2, we argued the advantages of using B-spline over Gaussian process for estimating the quantile functions. More specifically, using B-splines allows their Equation (4.10) to be solved analytically, significantly reducing the cost of likelihood evaluation compared with Tokdar and Kadane (2012). We use quadratic B-spline because solving Equation (4.10), which is crucial for the likelihood evaluation, reduces to quadratic equation solving.

**Prior**

Let \( 0 = t_0 < t_1 < \cdots < t_{p_1} = 1 \) be the equidistant knots on the interval \([0, 1]\) such that \( t_i - t_{i-1} = 1/p_1 \) for \( i = 1, \ldots, p_1 \). For B-spline of degree \( m_1 \), the number of basis functions is \( J_1 = p_1 + m_1 \). Let \( \{B_{j,m_1}(\cdot)\}_{j=1}^{p_1+m_1} \) be the basis functions of B-splines of degree \( m_1 \) on \([0, 1]\) on the above-mentioned equidistant knots. The basis expansion of the quantile functions at site \( z \) is given by the relations

\[
\xi_1(\tau, z) = \sum_{j=1}^{p_1+m_1} \theta_j(z)B_{j,m_1}(\tau) \quad \text{where} \quad 0 = \theta_1(z) < \cdots < \theta_{p_1+m_1}(z) = 1,
\]

\[
\xi_2(\tau, z) = \sum_{j=1}^{p_1+m_1} \phi_j(z)B_{j,m_1}(\tau) \quad \text{where} \quad 0 = \phi_1(z) < \cdots < \phi_{p_1+m_1}(z) = 1,
\]

(4.5)

where the coefficients \( \{\theta_j(z)\}_{j=1}^{p_1+m_1} \) and \( \{\phi_j(z)\}_{j=1}^{p_1+m_1} \) are dependent on the spatial location of the data-points. To put priors on \( \{\theta_j(z)\}_{j=1}^{k_1+m_1} \) and \( \{\phi_j(z)\}_{j=1}^{k_1+m_1} \), we use \( d \)-fold tensor product of B-spline basis functions. Thus for the spatial location \( z = (z_1, \ldots, z_d) \),
the basis expansion is given by

\[
\theta_j(z) = \sum_{k_1=1}^{p_2+m_2} \cdots \sum_{k_d=1}^{p_2+m_2} \alpha_{j,k_1 \cdots k_d} B_{k_1,m_2}(z_1) \cdots B_{k_d,m_2}(z_d), \quad j = 1, \ldots, p_2 + m_2,
\]

\[
\phi_j(z) = \sum_{k_1=1}^{p_2+m_2} \cdots \sum_{k_d=1}^{p_2+m_2} \beta_{j,k_1 \cdots k_d} B_{k_1,m_2}(z_1) \cdots B_{k_d,m_2}(z_d), \quad j = 1, \ldots, p_2 + m_2,
\]

(4.6)

where \(\{B_{k,m_2}(\cdot)\}_{k=1}^{p_2+m_2}\) are B-spline basis functions of degree \(m_2\) on [0, 1] with equidistant knots \(0 = s_1 < \cdots < s_{p_2} = 1\) and \(s_i - s_{i-1} = 1/p_2\) for \(i = 1, \ldots, p_2\). To ensure the constraints of \(\{\theta_j(z)\}_{j=1}^{p_1+m_1}\) and \(\{\phi_j(z)\}_{j=1}^{p_1+m_1}\) mentioned in Equation (4.5), it is sufficient to ensure the following constraints

\[
0 = \alpha_{1,k_1 \cdots k_d} < \cdots < \alpha_{(p_1+m_1),k_1 \cdots k_d} = 1, \quad 0 = \beta_{1,k_1 \cdots k_d} < \cdots < \beta_{(p_1+m_1),k_1 \cdots k_d} = 1,
\]

(4.7)

for \(\{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d\).

Suppose that we have data for \(L\) spatial sites \(z_1, \ldots, z_L\) and at the \(l\)-th site, \(\{X_{il}, Y_{il}\}_{i=1}^{n_l}\) denote the values of the explanatory variable (time) and the response variable, \(n_l\) being the number of data-points for \(l = 1, \ldots, L\). Then log-likelihood can be derived in the following manner.

Then Equation (4.3) can be written as

\[
Q(\tau|x, z) = x \xi_1(\tau, z) + (1 - x) \xi_2(\tau, z)
\]

\[
= x \sum_{j=1}^{p_1+m_1} \theta_j(z) B_{j,m_1}(\tau) + (1 - x) \sum_{j=1}^{p_1+m_1} \phi_j(z) B_{j,m_1}(\tau)
\]

\[
= x \sum_{j=1}^{p_1+m_1} \sum_{k_1=1}^{p_2+m_2} \cdots \sum_{k_d=1}^{p_2+m_2} \alpha_{j,k_1 \cdots k_d} B_{k_1,m_2}(z_1) \cdots B_{k_d,m_2}(z_d)
\]

\[
+ (1 - x) \sum_{j=1}^{p_1+m_1} \sum_{k_1=1}^{p_2+m_2} \cdots \sum_{k_d=1}^{p_2+m_2} \beta_{j,k_1 \cdots k_d} B_{k_1,m_2}(z_1) \cdots B_{k_d,m_2}(z_d)
\]

(4.8)

We evaluate the likelihood in a similar approach shown in Tokdar and Kadane (2012) and Chapter 2. Suppose we have data for \(L\) spatial sites \(z_1, \ldots, z_L\) and at the \(l\)-th site, \(\{X_{il}, Y_{il}\}_{i=1}^{n_l}\) denote the values of the explanatory variable (time) and the response vari-
able, \( n_l \) being the number of data-points for \( l = 1, \ldots, L \), the likelihood is given by

\[
\prod_{l=1}^{L} \left\{ \prod_{i=1}^{n_l} f(Y_{li}|X_{li}, z_l) \right\}
\]

where

\[
f(Y_{li}|X_{li}, z_l) = \left( \frac{\partial}{\partial \tau} Q(\tau|X_{li}, z_l) \bigg|_{\tau=\tau_X(Y_{li}, z_l)} \right)^{-1}, \quad l = 1, \ldots, L, \quad i = 1, \ldots, n_l.
\]

Now,

\[
\log f(Y_{li}|X_{li}, z_l) = -\log \left( \frac{\partial}{\partial \tau} Q(\tau|X_{li}, z_l) \bigg|_{\tau=\tau_X(Y_{li}, z_l)} \right)
\]

\[
= -\log \left\{ X_{li} \frac{\partial}{\partial \tau} \xi_1(\tau, z_l) + (1 - X_{li}) \frac{\partial}{\partial \tau} \xi_2(\tau, z_l) \right\} \bigg|_{\tau=\tau_X(Y_{li}, z_l)} \quad (4.9)
\]

where \( \tau_{X_{li}}(Y_{li}, z_l) \) is the solution of

\[
Y_{li} = X_{li} \xi_1(\tau, z_l) + (1 - X_{li}) \xi_2(\tau, z_l). \quad (4.10)
\]

For any given location \( Z = z \), since \( \xi_1(\cdot, z) \) and \( \xi_2(\cdot, z) \) are strictly monotonic, their convex combination is also strictly monotonic. Hence there will exist a unique solution of Equation (4.10). Using the properties of derivative of B-spline (de Boor (2001)) we have

\[
\frac{d}{dt} \xi_1(t, z) = \frac{d}{dt} \sum_{j=1}^{p_1+m_1} \theta_j(z) B_{j,m_1}(t) = \sum_{j=2}^{p_1+m_1} \theta^*_j(z) B_{j-1,m_1-1}(t),
\]

\[
\frac{d}{dt} \xi_2(t, z) = \frac{d}{dt} \sum_{j=1}^{p_1+m_1} \phi_j(z) B_{j,m_1}(t) = \sum_{j=2}^{p_1+m_1} \phi^*_j(z) B_{j-1,m_1-1}(t), \quad (4.11)
\]

where

\[
\theta^*_j(z) = (p_1 + m_1)(\theta_j(z) - \theta_{j-1}(z)), \quad \phi^*_j(z) = (p_1 + m_1)(\phi_j(z) - \phi_{j-1}(z)),
\]

for \( j = 2, \ldots, p_1 + m_1 \). Now, using Equation (4.9) and (4.11), we have

\[
\log f(Y_{li}|X_{li}, z_l) = -\log \left\{ X_{li} \sum_{j=2}^{p_1+m_1} \theta^*_j(z_l) B_{j-1,m_1-1}(\tau_{X_{li}}(Y_{li}, z_l)) \right.
\]

\[
+ (1 - X_{li}) \sum_{j=2}^{p_1+m_1} \phi^*_j(z_l) B_{j-1,m_1-1}(\tau_{X_{li}}(Y_{li}, z_l)) \right\} \bigg|_{\tau_{X_{li}}(Y_{li}, z_l)} \quad (4.12)
\]

56
Hence the total log-likelihood is given by

\[ \sum_{l=1}^{L} \sum_{i=1}^{n_l} \log f(Y_{li}|X_{li}, z_l) = - \sum_{l=1}^{L} \sum_{i=1}^{n_l} \log \left\{ X_{li} \sum_{j=2}^{p_1+m_1} \theta_j^*(z_l) B_{j-1,m_1-1}(\tau X_{li}(Y_{li}, z_l)) \right\} \]

\[ + (1 - X_{li}) \sum_{j=2}^{p_1+m_1} \phi_j^*(z_l) B_{j-1,m_1-1}(\tau X_{li}(Y_{li}, z_l)) \right\}. \quad (4.13) \]

We note that the parameters of the likelihood are the coefficients of the B-spline basis expansion of \( \{\theta_j(z)\}_{j=1}^{p_1+m_1} \) and \( \{\phi_j(z)\}_{j=1}^{p_1+m_1} \) which are

\[
0 = \alpha_{1k_1\ldots k_d} < \cdots < \alpha_{(p_1+m_1)k_1\ldots k_d} = 1, \\
0 = \beta_{1k_1\ldots k_d} < \cdots < \beta_{(p_1+m_1)k_1\ldots k_d} = 1,
\]

for \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \).

The parameters of the log-likelihood are given by Equation (4.7). Hence if we fix the values of \( k_1, \ldots, k_d \), the vector of spacings of the coefficients \( \alpha_{1k_1\ldots k_d}, \ldots, \alpha_{(p_1+m_1)k_1\ldots k_d} \) lie on the unit simplex. The same is true for \( \beta_{1k_1\ldots k_d}, \ldots, \beta_{(p_1+m_1)k_1\ldots k_d} \). Define

\[
\gamma_{jk_1\ldots k_d} = \alpha_{(j+1)k_1\ldots k_d} - \alpha_{jk_1\ldots k_d}, \quad \delta_{jk_1\ldots k_d} = \beta_{(j+1)k_1\ldots k_d} - \beta_{jk_1\ldots k_d},
\]

for \( j = 1, \ldots, p_1 + m_1 - 1 \) and \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). Hence for each combination of \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \), we have

\[
\sum_{j=1}^{p_1+m_1-1} \gamma_{jk_1\ldots k_d} = 1 \quad \text{and} \quad \gamma_{jk_1\ldots k_d} \geq 0, \quad \sum_{j=1}^{p_1+m_1-1} \delta_{jk_1\ldots k_d} = 1 \quad \text{and} \quad \delta_{jk_1\ldots k_d} \geq 0. \quad (4.16)
\]

Thus we note that the parameters of the log-likelihood can be divided into \( 2 \times (p_2 + m_2)^d \) unit simplex blocks. We take uniform Dirichlet prior on the unit simplex blocks \( \{\gamma_{jk_1\ldots k_d}\}_{j=1}^{p_1+m_1-1} \) and \( \{\delta_{jk_1\ldots k_d}\}_{j=1}^{p_1+m_1-1} \) for \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). As mentioned in the earlier section, \( m_1 \) has been taken to be 2. The degree of the basis functions corresponding to each spatial coordinates has been considered to be cubic (i.e., \( m_2 = 3 \)).

**Block Metropolis-Hastings MCMC algorithm**

Recall,

\[
\gamma_{jk_1\ldots k_d} = \alpha_{(j+1)k_1\ldots k_d} - \alpha_{jk_1\ldots k_d}, \quad \delta_{jk_1\ldots k_d} = \beta_{(j+1)k_1\ldots k_d} - \beta_{jk_1\ldots k_d},
\]

\[
(4.17)
\]
for \( j = 1, \ldots, p_1 + m_1 - 1 \) and \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). It is noted that \( \{\gamma_{jk_1\cdots k_d}^{p_1+m_1-1}\}_{j=1}^{p_1+m_1-1} \) and \( \{\delta_{jk_1\cdots k_d}^{p_1+m_1-1}\}_{j=1}^{p_1+m_1-1} \) are on the unit-simplex for any given \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2+m_2)\}^d \). We use Block Metropolis-Hastings Monte Carlo Markov Chain (MCMC) algorithm (see Chib and Greenberg (1995)) for sampling from the posterior distribution.

Note that, the number of unit simplex blocks is \((p_2 + m_2)^d\). In MCMC, a movement is performed on each unit simplex block in a loop. During the updating stage of a single unit-simplex block, similar to the updating strategy used in Chapter 2, independent sequences \( \{U_j\}_{j=1}^{p_1+m_1-1} \) and \( \{W_j\}_{j=1}^{p_1+m_1-1} \) are generated from \( U(1/r, r) \) for some \( r > 1 \). For given \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \), define \( V_j = \gamma_{jk_1\cdots k_d} U_j \) and \( T_j = \delta_{jk_1\cdots k_d} W_j \) for \( j = 1, \ldots, p_1 + m_1 - 1 \). The proposal moves \( \gamma_{jk_1\cdots k_d} \mapsto \gamma_{jk_1\cdots k_d}^* \) and \( \delta_{jk_1\cdots k_d} \mapsto \delta_{jk_1\cdots k_d}^* \) are given by

\[
\gamma_{jk_1\cdots k_d}^* = \frac{V_j}{\sum_{i=1}^{p_1+m_1-1} V_i}, \quad \delta_{jk_1\cdots k_d}^* = \frac{T_j}{\sum_{i=1}^{p_1+m_1-1} T_i}, \quad j = 1, \ldots, p_1 + m_1 - 1. \tag{4.18}
\]

The conditional distribution of \( \{\gamma_{jk_1\cdots k_d}^*\}_{j=1}^{p_1+m_1-1} \) given \( \{\gamma_{jk_1\cdots k_d}\}_{j=1}^{p_1+m_1-1} \) is given by (see Section 2.3.2 for the derivation)

\[
f(\gamma_{jk_1\cdots k_d}^* | \gamma_{jk_1\cdots k_d}) = \left( \frac{r}{r^2 - 1} \right)^{p_1+m_1-1} \left\{ \prod_{j=1}^{p_1+m_1-1} \gamma_{jk_1\cdots k_d} \right\}^{-1} \left( \frac{D_1 - D_2}{(p_1 + m_1 - 1)} \right), \tag{4.19}
\]

where

\[
D_1 = \left( \min_{0 \leq j \leq p_1 + m_1 - 1} \frac{r \gamma_{jk_1\cdots k_d}^*}{\gamma_{jk_1\cdots k_d}} \right)^{p_1+m_1-1},
\]

\[
D_2 = \left( \max_{0 \leq j \leq p_1 + m_1 - 1} \frac{\gamma_{jk_1\cdots k_d}^*}{r \gamma_{jk_1\cdots k_d}} \right)^{p_1+m_1-1}.
\]

The conditional distribution of \( \{\delta_{jk_1\cdots k_d}^*\}_{j=1}^{p_1+m_1-1} \) can be found in a similar way. The updated values of \( \{\alpha_{jk_1\cdots k_d}^{p_1+m_1-1}\}_{j=1}^{p_1+m_1-1}, \{\beta_{jk_1\cdots k_d}^{p_1+m_1-1}\}_{j=1}^{p_1+m_1-1}, \) denoted by \( \{\alpha_{jk_1\cdots k_d}^{p_1+m_1-1}\}_{j=1}^{p_1+m_1-1} \) and \( \{\beta_{jk_1\cdots k_d}^{p_1+m_1-1}\}_{j=1}^{p_1+m_1-1} \), can be found from \( \{\gamma_{jk_1\cdots k_d}^*\}_{j=1}^{p_1+m_1-1} \) and \( \{\delta_{jk_1\cdots k_d}^*\}_{j=1}^{p_1+m_1-1} \) using the relation

\[
\alpha_{jk_1\cdots k_d}^* = \sum_{i=1}^{j} \gamma_{ik_1\cdots k_d}^*, \quad \beta_{jk_1\cdots k_d}^* = \sum_{i=1}^{j} \delta_{ik_1\cdots k_d}^*, \quad j = 1, \ldots, p_1 + m_1 - 1.
\]
The likelihood can be expressed as the function of \( \{ \alpha_{j_1 \ldots k_d}, \beta_{j_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \) assuming the values of the other unit simplex blocks to be fixed. Hence, it can be also expressed as a function of \( \{ \gamma_{j_1 \ldots k_d}, \delta_{j_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \). Suppose \( L(\gamma_{k_1 \ldots k_d}, \delta_{k_1 \ldots k_d}) \) and \( L(\gamma^*_{k_1 \ldots k_d}, \delta^*_{k_1 \ldots k_d}) \) denote the likelihood at \( \{ \gamma_{j_1 \ldots k_d}, \delta_{j_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \) and \( \{ \gamma^*_{j_1 \ldots k_d}, \delta^*_{j_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \) respectively fixing the values of the parameters of the other unit simplex blocks. The acceptance probability in the Block Metropolis-Hastings algorithm for the update step of the corresponding block is given by \( P_a = \min\{p, 1\} \) where

\[
p = \frac{L(\gamma^*_{k_1 \ldots k_d}, \delta^*_{k_1 \ldots k_d}) \pi(\gamma^*_{k_1 \ldots k_d}) \pi(\delta^*_{k_1 \ldots k_d}) f(\gamma_{k_1 \ldots k_d} | \gamma^*_{k_1 \ldots k_d}) f(\delta_{k_1 \ldots k_d} | \delta^*_{k_1 \ldots k_d})}{L(\gamma_{k_1 \ldots k_d}, \delta_{k_1 \ldots k_d}) \pi(\gamma_{k_1 \ldots k_d}) \pi(\delta_{k_1 \ldots k_d}) f(\gamma^*_{k_1 \ldots k_d} | \gamma_{k_1 \ldots k_d}) f(\delta^*_{k_1 \ldots k_d} | \delta_{k_1 \ldots k_d})}
\]

and \( \pi \) denotes the prior density. Since we take uniform Dirichlet prior on the unit simplex blocks \( \{ \gamma_{j_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \) and \( \{ \delta_{j_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \), each unit simplex block is updated one at a time in a loop.

### Simulation study

In this section we compare the performance of estimation by the proposed method, i.e., Spline Spatio-temporal Quantile Regression (SSTQR) with piece-wise Gaussian Basis function Spatio-temporal Quantile Regression (GBSTQR) (Reich (2012)) based on simulation. For the proposed SSTQR method, we consider both likelihood based and Bayesian approaches. Finding the maximum likelihood estimate (MLE) can be seen as an optimization problem of maximizing an objective function of a constrained parameter space which is given by a collection of unit simplex blocks. Since the evaluation of the likelihood involves integration and linear search, it is hard to check whether the likelihood function is convex or not. Instead of using convex optimization algorithms, global or non-convex optimization algorithms are more reasonable in this scenario. Das (2016b) proposed an efficient global optimization technique on a hyper-rectangular parameter space which has been shown to work faster and than existing global optimization techniques, namely Genetic Algorithm (Fraser (1957), Bethke (1980), Goldberg (1989)) and Simulated annealing (Kirkpatrick et al. (1983), Granville et al. (1994)) yielding better solutions. Das (2016c) modified that algorithm for the case where the sample space is given by an unit simplex. Following that paper, Das (2016a) proposed the ‘Greedy Coordinate Descent of Varying Step sizes on Multiple Simplexes’ (GCDVMS) algorithm which efficiently min-
imizes any non-convex (or, maximizes any non-concave) objective function of parameters
given by a collection of unit simplex blocks. The main idea of the GCDVSMS algorithm
is making jumps of varying step-sizes within each unit simplex blocks parallelly and
searching for the most favorable direction of movement. In this algorithm, every time a
local solution is found, co-ordinate wise jumps of various step-sizes are performed until
a better solution is found. To find the MLE in this case, the GCDVSMS algorithm has
been used. The values of the tuning parameters in the GCDVSMS algorithm have been
taken to be as follows; initial global step size \( s_{\text{initial}} = 1 \), step decay rate for the first run
\( \rho_1 = 2 \), step decay rate for other runs \( \rho_2 = 1.5 \), step size threshold \( \phi = 10^{-1} \), sparsity
threshold \( \lambda = 10^{-2} \), the convergence criteria controlling parameters \( \text{tol}_\text{fun} = 10^{-1} \), maximum number of iterations inside each run \( \text{max}_\text{iter} = 5000 \), maximum number of allowed runs \( \text{max}_\text{runs} = 200 \).

We consider the spatial locations are uniformly distributed over the unit square i.e.,
\([0, 1]^2\). Number of spatial locations is taken to be \( L = 50 \). For simulation purpose, three
different numbers of equidistant temporal data-points for each site have been considered
which are \( n = 5, 10, 20 \) (note that, \( n_l = n \) for \( l = 1, \ldots, 50 \)). Let \( \{z_l\}_{l=1}^{L} \) denote the
spatial locations of the available sites of data where \( z_l = (z_{l1}, z_{l2}) \) is a two-tuple such
that \( 0 \leq z_{l1}, z_{l2} \leq 1 \) for \( l = 1, \ldots, L \). Consider the spatial quantile function
\( Q(\tau|x, z) = x\xi_1(\tau, z) + (1 - x)\xi_2(\tau, z) \) where

\[
\begin{align*}
\xi_1(\tau, z) &= (1 - \frac{z_{l1} + z_{l2}}{2})\tau^2 + \frac{z_{l1}}{2}\log(1 + \tau) + \frac{z_{l2}}{2}\tau^3 \\
\xi_2(\tau, z) &= (1 - \frac{z_{l2}^2}{2})\sin\left(\frac{\pi\tau}{2}\right) + \frac{z_{l2}^2}{(e - 1)}(e^\tau - 1)
\end{align*}
\]  

(4.20)

Note that for any given \( z \in [0, 1]^2 \), \( \xi_1(\cdot, z), \xi_1(\cdot, z) \) are strictly increasing function from
\([0, 1]\) to \([0, 1]\) satisfying \( \xi_1(0, z) = \xi_2(0, z) = 0 \) and \( \xi_1(1, z) = \xi_2(1, z) = 1 \). Since the
quantile function is the inverse of the cumulative distribution function, taking \( U \sim
U(0, 1) \), \( Q(U|x, z) \) has conditional quantile function \( Q(\tau|x, z) \). For each locations \( \{z_l\}_{l=1}^{L} \),
we considered the same set of equidistant time-points such that \( 0 = x_{l1} < \cdots < x_{ln} = 1 \)
and \( (x_{li} - x_{li-1}) = 1/(n - 1) \) for \( i = 2, \ldots, n \). We simulate the response variable \( Y \) using
the equation

\[
y_{li} = x_{li}\xi_1(U_{li}, z_l) + (1 - x_{li})\xi_2(U_{li}, z_l), \ l = 1, \ldots, L, \ i = 1, \ldots, n,
\]

where \( U_{li} \sim U(0, 1) \) for \( l = 1, \ldots, L, \ i = 1, \ldots, n \). Using both methods, for each of the
locations with number of data-points \( n = 5, 10, 20 \), the quantile curves are estimated
Table 4.1: Average MSE (based on $S = 50$ repetitions of simulation study) of estimated slope, intercept and estimated quantile function value at three time-points $X = 0.2, 0.5, 0.8$ using SSTQR (both MLE based and Bayesian approach) and GBSTQR over all $L = 50$ locations and $T = 19$ quantile levels $\tau = 0.05, 0.10, \ldots, 0.95$.

<table>
<thead>
<tr>
<th>Sample size (per site)</th>
<th>Methods</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0(\tau, z)$</td>
<td>$\beta_1(\tau, z)$</td>
</tr>
<tr>
<td>$n = 5$</td>
<td>SSTQR (ML)</td>
<td>0.0265</td>
</tr>
<tr>
<td></td>
<td>SSTQR (Bayes)</td>
<td>0.0410</td>
</tr>
<tr>
<td></td>
<td>GBSTQR</td>
<td>0.2014</td>
</tr>
<tr>
<td>$n = 10$</td>
<td>SSTQR (ML)</td>
<td>0.0248</td>
</tr>
<tr>
<td></td>
<td>SSTQR (Bayes)</td>
<td>0.0290</td>
</tr>
<tr>
<td></td>
<td>GBSTQR</td>
<td>0.6000</td>
</tr>
<tr>
<td>$n = 20$</td>
<td>QSSTQR (ML)</td>
<td>0.0186</td>
</tr>
<tr>
<td></td>
<td>SSTQR (Bayes)</td>
<td>0.0154</td>
</tr>
<tr>
<td></td>
<td>GBSTQR</td>
<td>0.0685</td>
</tr>
</tbody>
</table>

For the quantiles $\{\tau_t\}_{t=1}^T$ where $\tau_t = 0.05 t$ and $T = 19$. The above-mentioned simulation study has been repeated $S = 50$ times under different random number generating seeds. Let at the $s$-th simulation study with fixed number of data-points ($n = 5, 10, 20$) at each site, the estimated intercept and slope at $\tau_t$-th quantile and location $z_l$ are $\hat{\beta}_0(s)(\tau_t, z_l)$ and $\hat{\beta}_1(s)(\tau_t, z_l)$ respectively for $s = 1, \ldots, S$, $t = 1, \ldots, T$ and $l = 1, \ldots, L$. $Q(\tau_t|x, z_l)$ denotes the true value of the $\tau_t$-th quantile at location $z_l$ at time-point $X = x$ and $\hat{Q}_1(s)(\tau_t|x, z_l)$ is the estimated value of it based on $s$-th simulation study. Then the average of mean squared error (MSE) of the slope, intercept and the quantile value at a given time-point $X = x$ are given by

$$T_1 = \frac{1}{SLT} \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{t=1}^{T} (\beta_0(\tau_t, z_l) - \hat{\beta}_0(s)(\tau_t, z_l))^2$$

$$T_2 = \frac{1}{SLT} \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{t=1}^{T} (\beta_1(\tau_t, z_l) - \hat{\beta}_1(s)(\tau_t, z_l))^2$$

$$T_x = \frac{1}{SLT} \sum_{s=1}^{S} \sum_{l=1}^{L} \sum_{t=1}^{T} (Q(\tau_t|x, z_l) - \hat{Q}_1(s)(\tau_t|x, z_l))^2.$$

For the SSTQR (Bayes) method, 10000 iterations have been performed and the first 1000 iterations has been disregarded as burn-in. In the Table (4.1), the comparison of the average MSE of estimated slope, intercept and estimated quantile curve value at three
time-points $X = 0.2, 0.5, 0.8$ over all locations has been provided for both of the above-mentioned methods. Since we are using quadratic B-spline for the quantile basis functions and cubic B-spline for the coordinate-wise spatial basis functions, $m_1 = 2, m_2 = 3$. To select the optimum number of knots, we use the Akaike information criterion (AIC). In this simulation study, we consider the cases $p_1 = p_2 = 3, 4, 5, 6$ and selected the model with best value of AIC.

It is noted that both SSTQR (Bayes) and SSTQR (ML) perform generally better than GBSTQR. SSTQR (Bayes) performs slightly better than SSTQR (ML) which beats the GBSTQR method in performance. It should be also noted that the optimization in the ML estimation can be considerably more challenging than obtaining samples since the negative likelihood is possibly non-convex with multiple local maxima. To find the ML estimate, we solve 250-1024 dimensional constrained black-box optimization problem (for $p_1, p_2 = 3, 4, 5, 6$).

**Analysis of ozone concentration data of the US**

Ozone concentration data of the US over the last several years can be found at the US EPA website\(^2\). The yearly averages of the daily maximum of observed hourly ozone concentration values between 9:00 AM and 8:00 PM (Daily 1-hour maximum average ozone concentration) and the daily maximum of 8 hour running average of observed hourly ozone concentration values (Daily 8-hour maximum average ozone concentration) have been collected at around 1629 sites all over the US in the last several years. As mentioned in the EPA website\(^3\), in 1997, EPA replaced the previous 1-hour ozone standard with 8-hour standard (which is supposed to be more protective) at a level of 84 parts per billion (ppb). 8-hour primary standard is considered to be met at any given site if the 3-year mean of the annual fourth-highest daily maximum 8-hour average ozone concentration is less than or equal to the proposed level, i.e., 84 ppb as per EPA standard announced in 1997. Later in 2008, EPA changed the 8-hour ozone concentration standard to 75 ppb and in 2015, it was decreased to 70 ppb due to emergence of more scientific evidence regarding the effects of ozone on public health and welfare. As mentioned in a news article Barboza (2015), environmentalists and health advocacy groups including the American Lung Association endorsed 8-hour ozone standard to be 60 ppb. But Barboza (2015) mentioned as per EPA data from 2014, more than 40 million people, or in other words, about 1 in every 8 people of the US lives in the counties with air pollution levels

---

\(^2\)Source [https://www3.epa.gov/airquality/airdata/ad_data.html](https://www3.epa.gov/airquality/airdata/ad_data.html)

\(^3\)Source [https://www3.epa.gov/region1/airquality/avg8hr.html](https://www3.epa.gov/region1/airquality/avg8hr.html)
Here an analysis of the spatio-temporal trend of daily 1-hour maximum average and daily 8-hour maximum average ozone concentration for the last 10 years, i.e., during the period 2006-2015, has been provided using the proposed method. It should be noted that in the given dataset, each site does not have the data for all the years considered while some of the sites have multiple observations recorded at the same year measured using different measuring instruments. With the proposed method, these cases can be handled automatically without any further required modifications. For the analysis, the Markov Chain Monte Carlo (MCMC) has been initialized at the SSTQR (ML) estimate obtained using the GCDVSMS algorithm proposed in Das (2016a) with the values of the tuning parameters as mentioned in the earlier section. For this analysis, 10000 iterations have been performed disregarding the first 1000 iterations as burn-in. The number of knots of the B-spline basis functions has been chosen based on the AIC criterion as mentioned in Section 4.4.

For the years 2006, 2010 and 2015 the daily 1-hour maximum and 8-hour maximum average ozone concentration are plotted over the US at the median level. It is noted that in the most of the parts of the US the daily 1-hour maximum average ozone concentration has decreased over time. Specifically the decreasing trend of 1-hour concentration is noticeable in the West North Central, Upper Midwest, Northeast and southern part of the South climate regions (see Figure 4.1, 4.2, 4.3). The daily 8-hour maximum average ozone concentration level also has an overall decreasing trend over the time period considered. The 8-hour concentration has decreased quite a bit in the upper Northwest and
West North Central regions. Besides, it has also decreased in the Northeast, South and Southeast regions (see Figures 4.4, 4.5). In Figure 4.6 we also show the quantile levels across the US where the 4th highest daily 8-hour maximum ozone concentration is 70 ppb. It is noted that over the last ten years the 4th highest daily 8-hour maximum ozone concentration in the US has generally a decreasing trend since 70 ppb ozone concentration is attained at relatively higher quantiles in the latter years.

Analysis of California ozone concentration data

As mentioned in Barboza (2015), California has the worst smog in the US and it does not meet the existing smog limits. It also says that the air quality is the worst in the inland valleys, where pollution from vehicles and factories yields ozone in the presence of sunlight and that ozone is blown and trapped against the mountains. Some of the areas in California are so polluted that as per EPA, they are expected to meet the standards by 2037, getting 12 more years to recover unlike the rest of the nation. According to the state Air Resources Board, about one-third people of California live in communities where the pollution level exceeds federal standards. As mentioned in the California Environmental Protection Agency site, in April 2005, the Air Resources Board retained the previous 1-hour ozone standard of 90 ppb and set a new 8-hour standard of 70 ppb. This ozone standard review was also mandated by the Children’s Environmental Health Protection Act.

In this section, the spatio-temporal trend of daily 1-hour maximum and 8-hour maximum average ozone concentration data of California are analyzed over the time period 2006-2015 and spatial plots are shown for the years 2006, 2010 and 2015 at 50-th quantile level. For those three years, we also show the quantile levels across California where the 4th highest daily 8-hour maximum ozone concentration is 70 ppb and 4th highest daily 1-hour maximum ozone concentration is 90 ppb. Similar to the previous section, the SSTQR (Bayes) method is used for the analysis and SSTQR (ML) estimate is used as the starting point of the MCMC chain. In this case also 10000 iterations are performed and the first 1000 iterations are disregarded as burn-in. It is noted that overall there is a decreasing trend of daily 1-hour maximum average concentration level at both the quantile levels considered. Specifically in the Northern California, Sacramento Region, Central Valley, High Sierra Desert, Los Angeles and San Diego (see Figures 4.7, 4.8).

4Source: https://www.arb.ca.gov/research/aaqs/ccaqs/ozone/ozone.htm
Figure 4.2: Daily 1-hour maximum average ozone concentration (in ppb) of the US in 2006, 2010 and 2015 at $\tau = 0.5$. The dots denote weather stations where data have been collected.
Figure 4.3: Yearly rate of change of daily 1-hour maximum average ozone concentration (in ppb/year) of the US at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected.
Figure 4.4: Daily 8-hour maximum average ozone concentration (in ppb) of the US in 2006, 2010 and 2015 at $\tau = 0.5$. The dots denote weather stations where data have been collected.
Figure 4.5: Yearly rate of change of daily 8-hour maximum average ozone concentration (in ppb/year) of the US at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected.
Figure 4.6: Quantiles at which 4th highest daily 8-hour maximum ozone concentration (in ppb) of the US is 70 ppb in 2006, 2010 and 2015.
the 1-hour maximum average ozone concentration has noticeably decreased over time. It is also noted that in the Inland Empire region, it has a slightly increasing trend over time. As mentioned earlier, unlike in other parts of the US, Air Resources Board retained 1-hour ozone standard in California and its 8-hour standard has been 70 ppb for the last ten years while till September 2015, 8-hour standard for the rest of the country has been greater than or equal to 75 ppm. Thus to comply with the stricter rules, tighter controls on factories, vehicles, power plants and other emitters of smog-forming pollutants, the overall daily 1-hour maximum average ozone concentration of California shows a decreasing trend.

The change in daily 8-hour maximum average ozone concentration varies a lot spatially over the time and there no particular temporal trend is observed for California as whole (see Figure 4.9). In the High Sierra Desert and north-east part of Northern California, the 8-hour maximum average ozone concentration has a decreasing trend over time at both the quantiles. On the other hand, in Sacramento Region, Inland Empire and San Diego, it has an increasing trend over the time. Not much temporal variations have been observed in the Bay Area and the Central Coast for both 1-hour maximum and 8-hour maximum average ozone concentration levels. In High Sierra Desert and Inland Empire, this ozone concentration level is changing rapidly.

In Figures 4.10a, 4.10b and 4.10c it is noted that in Northern California, Bay Area, Sacramento Region and in the upper part of the Central Valley, the 4th highest daily 8-hour maximum ozone concentration threshold, i.e., 70 ppb is achieved at very lower
Figure 4.8: (a-c) Daily 1-hour maximum average ozone concentration (in ppb) of California in 2006, 2010 and 2015 at \( \tau = 0.5 \). (d-f) Yearly rate of change of daily 1-hour maximum average ozone concentration (in ppb/year) of California at \( \tau = 0.25, 0.5, 0.75 \). The dots denote weather stations where data have been collected.
Figure 4.9: (a-c) Daily 8-hour maximum average ozone concentration (in ppb) of California in 2006, 2010 and 2015 at $\tau = 0.5$. (d-f) Yearly rate of change of daily 8-hour maximum average ozone concentration (in ppb/year) of California at $\tau = 0.25, 0.5, 0.75$. The dots denote weather stations where data have been collected.
quantile levels. In Los Angeles and San Diego area this threshold value is achieved at a higher quantile level. In Figures 4.10d, 4.10e and 4.10f we note that at the junction of Bay Area, Sacramento Region, south Northern California and upper Central Valley the 4th highest daily 1-hour maximum ozone concentration threshold, i.e., 90 ppb is achieved at a lower quantile level compared to other parts of California. It is also noted that with time, the ozone concentration is decreasing in this region. In southern part of California, the threshold of daily 1-hour maximum ozone concentration is met at higher quantile level indicating that southern California is performing better in maintaining the ozone concentration level below recommended threshold level. But in San Diego area a strictly increasing pattern of ozone concentration is noted.

**Discussion**

As long as spatio-temporal regression is concerned, one of the major issues is the presence of missing temporal data at some locations. One way to solve this problem is to replace the missing data by estimated value based on the available data. It makes more sense to fit the model on the available data without using extrapolation or estimated missing values of data. In our regression model, the number of time-points of available data at various sites can be different (see equation (4.13)). This problem does not arise in the proposed method.

When the sample size is small at a location, in the method of Reich (2012), the estimated separate temporal quantile regression curve at that location will be widely affected by sampling fluctuations. If a lot of temporal data are missing at some locations, or when our data size is available only at a few temporal points, in spite of fitting separate quantile curves for all sites, it is desirable to fit a quantile curve considering the available data of all the sites together. Thus, estimation at the neighborhood of the locations with small temporal data can be improved by borrowing strength from neighboring sites. Hence the proposed method yields better estimates and is less affected by missing data and scarcity often.

In some approaches to spatio-temporal regression and time-series analysis, the temporal data are required to be obtained periodically. However in the proposed method, the temporal data are not required to be periodic. Thus we conclude that the proposed approach gives a natural Bayesian method for estimation and uncertainty quantification for simultaneous spatial quantile regression. The proposed method allows the data to be flexible with respect to missingness, monotonicity and low-sample size at individual
Figure 4.10: (a-c) Quantiles at which 4th highest daily 8-hour maximum ozone concentration (in ppb) of California is 70 ppb in 2006, 2010 and 2015. (d-f) Quantiles at which 4th highest daily 1-hour maximum ozone concentration (in ppb) of California is 90 ppb in 2006, 2010 and 2015.
locations. The method improves the quality of the inference by borrowing strength across all locations and all quantile levels by incorporating a natural non-parametric smoothing in its prior construction. By using B-spline functions in its smoothing, the method allows a relatively more efficient approach to likelihood evaluation compared with an analogous procedure based on Gaussian procedures.

As mentioned in Barboza (2015), according to the EPA ozone level has decreased by about one-third over the US since 1980 by imposing regulations targeting emissions from cars, factories, consumer products and other sources of pollutants. In this analysis, it is observed that overall there is a decreasing trend in the daily 1-hour and 8-hour maximum average ozone concentration levels in the US during the period 2006-2015. Specifically, in the least ten years, the ozone concentration has decreased considerably in the northern part of the US. In California, it is noted that there is an overall decreasing trend in the 1-hour concentration while there is no specific trend for the 8-hour concentration and the trend varies a lot spatially. According to Barboza (2015), except for a few places in California, the rest of the US is expected to comply with the EPA ozone standards by 2025. Some of the most polluted areas in Southern California and the San Joaquin Valley are expected to comply with it by 2037.
Chapter 5

Bayesian Non-parametric
Simultaneous Quantile Regression
for Complete and Grid Data

Introduction

One of the shortcomings of using linear quantile regression method is that it is not able to reveal a higher degree polynomial trend in the quantile curves. For example, while regressing household income data of a country over time, lower quantile levels may be linear with time but the upper quantile levels may evolve very differently. In that case, using linear quantile regression will not be appropriate since it can only give a linear approximation of the quantile curves. Quantile regression methods addressing this issue were proposed in Chaudhuri (1991b), Chaudhuri (1991a). Yu and Jones (1998) proposed the local linear quantile regression method based on the techniques proposed in Fan et al. (1994), Fan et al. (1996), Chaudhuri (1991b). Chaudhuri and Loh (2002) and Honda (2010) proposed alternative methods of quantile regressions. Koenker (2015) provided a detailed description of local linear quantile regression method using software R. Alongside, he also proposed an alternative method called local spline quantile regression.

As long as non-linear quantile regression is concerned, most of the above-mentioned methods do not take care of the non-crossing issues. Although, Bondell et al. (2010) proposed a non-parametric quantile regression method addressing the non-crossing issue, it can only estimate a set of non-crossing curves for a given grid of quantiles and it does not estimate the whole quantile regression function. Secondly, the quantile regression estimates obtained by this method is sensitive to the number and location of chosen quantile
grids. For example, using this method, the estimate of \( \tau = 0.5 \) would come different for the two cases where quantiles are estimated for quantile grids \( T_1 = \{0.25, 0.5, 0.75\} \) and \( T_2 = \{0.2, 0.5, 0.8\} \) which is not desirable. Instead of estimating the quantile curves at a given set of quantiles, it is more desirable to estimate the whole quantile curve simultaneously to emerge the broader picture.

In this chapter, we propose two Bayesian methods for quantile regression using B-splines. In the first method, the entire quantile function is modeled by a B-spline series expansion. For each of the explanatory variables, a corresponding B-spline basis function is considered. The whole quantile function is obtained via tensor product of B-spline basis functions corresponding to each explanatory variable and one corresponding to the quantile level. The prior on the B-spline coefficients is put in such a way that the monotonicity of the quantile curves is maintained. We name this method ‘Non-parametric Simultaneous Quantile Regression (NPSQR)’. In the second method, instead of the quantile function, the conditional distribution function is estimated non-parametrically using B-spline basis expansion. Similar to NPSQR, in this method also, corresponding to each explanatory variable, a B-spline basis is considered. Again, the prior of the coefficients of the B-spline basis functions is put in such a way that the monotonicity of the distribution function is maintained. In this case also the whole distribution function is given by tensor product of the B-spline basis functions. The conditional distribution function is inverted to obtain the quantile regression function. The use of splines, which are piece-wise polynomials, allow efficient inversion through a combination of analytical and numerical technique. We name this method to be ‘Non-parametric Distribution Function Simultaneous Quantile Regression (NPDFSQR)’. Further using both of these two approaches, we propose the method of estimating the quantile curves when only the data with frequencies of the observations in each quantile range are available.

**Proposed Bayesian Method**

Suppose \( \{(X_{1i}, \ldots, X_{di})\}_{i=1}^{n} \) and \( \{Y_i\}_{i=1}^{n} \) denote the \( d \)-dimensional explanatory variable and the response variable respectively. Using monotonic transformation, each coordinates of the explanatory variable and the response variable are transformed into unit interval.

**Non-parametric Modeling of Quantile Function**

Suppose \( Q(\tau | \bm{x}) \) denotes the conditional quantile function of \( Y \) given \( \bm{X} = \bm{x} = (x_1, \ldots, x_d) \). A B-spline function of degree \( m_1 \) (i.e., degree of piece-wise polynomial is \( m \)) with knot
sequence \(0 = t_0 < t_1 < \cdots < t_{p_1} = 1\) has \((p_1 + m_1)\) basis functions. Let \(\{B_{j,m_1}(\cdot)\}_{j=1}^{p_1+m_1}\) denote B-spline basis functions of degree \(m_1\) on the above-mentioned knot sequence. For simplicity we consider equidistant knots, i.e., \((t_i - t_{i-1}) = 1/p_1\) for \(i = 1, \ldots, p_1\). Hence the quantile function is given by

\[
Q(\tau|\mathbf{x}) = \sum_{j=1}^{p_1+m_1} \theta_j(\mathbf{x})B_{j,m_1}(\tau), \quad 0 = \theta_1(\mathbf{x}) < \cdots < \theta_{p_1+m_1}(\mathbf{x}) = 1 \tag{5.1}
\]

where \(\theta_j(\mathbf{x}), j = 1, \ldots, p_1+m_1,\) are the coefficients of B-splines basis expansion of \(Q(\tau|\mathbf{x})\). Thus in the above-mentioned equation it is noted that the coefficients of the basis functions used to expand the quantile function are dependent on the explanatory variable \(\mathbf{X}\).

Secondly, the imposed monotonicity condition on the B-spline coefficients \(\{\theta_j(\mathbf{x})\}_{j=1}^{p_1+m_1}\) ensures the monotonicity of the quantile levels (de Boor (2001)).

Now to put a prior, the functions \(\{\theta_j(\mathbf{x})\}_{j=1}^{p_1+m_1}\), it is expanded using \(d\)-dimensional tensor product of the B-spline basis functions of degree \(m_2\). We use the knot sequence \(\{s_i\}_{i=1}^{p_2}\) such that \(0 = s_0 < s_1 < \cdots < s_{p_2} = 1, (s_i - s_{i-1}) = 1/p_2\) for \(i = 1, \ldots, p_2\) for all the coordinates of the explanatory variable \(\mathbf{X}\). Then \(\theta_j(\mathbf{x})\) is given by

\[
\theta_j(\mathbf{x}) = \sum_{k_1=1}^{p_2+m_2} \cdots \sum_{k_d=1}^{p_2+m_2} \alpha_{k_1\cdots k_d}B_{k_1,m_2}(x_1)\cdots B_{k_d,m_2}(x_d). \tag{5.2}
\]

Then the parameters which need to be estimated are given by

\[
0 = \alpha_{1k_1\cdots k_d} < \cdots < \alpha_{(p_1+m_1)k_1\cdots k_d} = 1, \quad (k_1, \ldots, k_d) \in \{1, \ldots, (p_2 + m_2)\}^d. \tag{5.3}
\]

**Non-parametric Modeling of Distribution Function**

For modeling the distribution function with B-spline basis functions, we adopt a similar technique. Suppose \(F(y|\mathbf{x})\) denotes the conditional distribution function of \(Y\) at \(\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_d)\). Suppose \(\{B_{j,m_1}(\cdot)\}_{j=1}^{p_1+m_1}\) denotes the B-spline coefficients of degree \(m_1\) on the knot sequence \(\{t_i\}_{i=0}^{p_1}\) as mentioned earlier. Then the conditional distribution \(F(y|\mathbf{x})\) is given by

\[
F(y|\mathbf{x}) = \sum_{j=1}^{p_1+m_1} \phi_j(\mathbf{x})B_{j,m_1}(y) \quad \text{where} \quad 0 = \phi_1(\mathbf{x}) < \cdots < \phi_{p_1+m_1}(\mathbf{x}) = 1 \tag{5.4}
\]
It is noted that again the coefficients \( \{\phi_j(x)\}_{j=1}^{p_1+m_1} \) are taken in such a way that the monotonicity of the distribution function is preserved. To put a prior on \( \{\phi_j(x)\}_{j=1}^{p_1+m_1} \), we use the same technique as mentioned in Section 5.2.1. Then \( \phi_j(x) \) is given by

\[
\phi_j(x) = \sum_{k_1=1}^{p_2+m_2} \cdots \sum_{k_d=1}^{p_2+m_2} \beta_{jk_1 \cdots k_d} B_{k_1,m_2}(x_1) \cdots B_{k_d,m_2}(x_d).
\]  

(5.5)

Hence the parameters to be estimated are given by

\[
0 = \beta_{1k_1 \cdots k_d} < \cdots < \beta_{(p_1+m_1)k_1 \cdots k_d} = 1, \ (k_1, \ldots, k_d) \in \{1, \ldots, (p_2 + m_2)\}^d.
\]  

(5.6)

### Likelihood Evaluation

In this section we describe the likelihood evaluation for both complete data and grouped data where only the frequencies of observations at each range of quantiles are given with the corresponding values of the explanatory variables.

#### Complete Data

Suppose that the explanatory and the response variables in the data are given by \( \{Y_i\}_{i=1}^n \) and \( \{X_i\}_{i=1}^n = \{(X_{i1}, \ldots, X_{id})\}_{i=1}^n \) where \( n \) denotes the sample size and \( d \) denotes the dimension of the explanatory variable. In the case of NPSQR, the likelihood derived from the quantile function is given by \( \prod_{i=1}^n f(Y_i|X_i) \) where \( f(Y_i|X_i) \) is given by

\[
f(Y_i|X_i) = \left( \frac{\partial}{\partial \tau} Q(\tau|X_i) \bigg|_{\tau=\tau_{X_i}(Y_i)} \right)^{-1}, \ i = 1, \ldots, n;
\]

here \( \tau_{X_i}(Y_i) \) solves the equation

\[
Y_i = Q(\tau|X_i) = \sum_{j=1}^{p_1+m_1} \theta_j(X_i) B_{j,m_1}(\tau).
\]  

(5.7)

As described in Section 5.2.1, \( Q(\tau|X_i) \) is constructed in such a way that it is monotonically increasing in \( \tau \). Hence Equation (5.7) has a unique solution. In case we consider piece-wise quadratic B-spline (i.e., \( m_1=2 \)), an advantage of our proposed method is that Equation (5.7) reduces to a quadratic equation and hence it can be solved analytically.
Solving it analytically provides the exact solution in lesser time. Now to compute the likelihood, using the properties of derivative of B-spline (de Boor (2001)), we get

$$\frac{\partial}{\partial t} Q(\tau | X_i) = \frac{\partial}{\partial \tau} \sum_{j=1}^{p_1+m_1} \theta_j(X_i) B_{j,m_1}(\tau) = \sum_{j=2}^{p_1+m_1} \theta^*_j(X_i) B_{j-1,m_1-1}(\tau),$$  \hspace{1cm} (5.8)

where

$$\theta^*_j(X_i) = (p_1 + m_1)(\theta_j(X_i) - \theta_{j-1}(X_i)), \hspace{0.5cm} j = 2, \ldots, p_1 + m_1.$$

Thus the log-likelihood in case of NPSQR is given by

$$\sum_{i=1}^{n} \log f(Y_i | X_i) = - \sum_{i=1}^{n} \log \left\{ \sum_{j=2}^{p_1+m_1} \theta^*_j(X_i) B_{j-1,m_1-1}(\tau X_i(Y_i)) \right\}. \hspace{1cm} (5.9)$$

In case of NPDFSQR, the log-likelihood function is given by

$$\sum_{i=1}^{n} \log f(Y_i | X_i) = \sum_{i=1}^{n} \log \left( \frac{\partial}{\partial y} F(y | X_i) \bigg|_{y=Y_i} \right)$$

$$= \sum_{i=1}^{n} \log \left( \frac{\partial}{\partial y} \sum_{j=1}^{p_1+m_1} \phi_j(X_i) B_{j,m_1}(y) \bigg|_{y=Y_i} \right)$$

$$= \sum_{i=1}^{n} \log \left( \sum_{j=2}^{p_1+m_1} \phi^*_j(X_i) B_{j-1,m_1-1}(Y_i) \right), \hspace{1cm} (5.10)$$

where

$$\phi^*_j(X_i) = (p_1 + m_1)(\phi_j(X_i) - \phi_{j-1}(X_i)), \hspace{0.5cm} j = 2, \ldots, p_1 + m_1.$$

**Quantile Grid Data**

In the case of grid data suppose the partition of the quantiles of the response are given by $0 = \rho_0 < \rho_1 < \cdots < \rho_c = 1$ and for each observation it is given between which two consecutive quantile divisions it belongs. Define

$$I_{Y_i}(l) = \begin{cases} 1 & \text{if } Y_i \text{ is in between } \rho_{l-1} \text{ and } \rho_l \text{-th quantiles} \\ 0 & \text{otherwise} \end{cases} \hspace{1cm} (5.11)$$
for $i = 1, \ldots, n$. Given the value of the explanatory variable for the $i$-th subject $X_i$, the probability of an observation $Y_i$ belonging between $\rho_{l-1}$ and $\rho_l$-th quantile is given by $(F(q_Y(\rho_l)|X_i) - F(q_Y(\rho_{l-1})|X_i))$. Here $q_Y(g)$ denotes the $g$-th quantile ($0 \leq g \leq 1$) of $Y$. Hence the total likelihood is given by

$$L = \prod_{i=1}^{n} \left( \prod_{l=1}^{c} (F(q_Y(\rho_l)|X_i) - F(q_Y(\rho_{l-1})|X_i))^{I_{Y_i}(l)} \right). \quad (5.12)$$

We assume that the values $\{q_Y(\rho_l)\}_{l=1}^{c-1}$ are provided and with $F(q_Y(\rho_0)|X_i) = 0, F(q_Y(\rho_c)|X_i) = 1$ for all $i = 1, \ldots, n$. Now for NPSQR, it should be noted that $F(q_Y(\rho_l)|X_i)$ is the solution of the equation $Q(\tau|X_i) = q_Y(\rho_l)$ since $Q(\tau|X_i) = q_Y(\rho_l)$ implies $F(q_Y(\rho_l)|X_i) = \tau$. Hence $F(q_Y(\rho_l)|X_i)$ can be obtained solving the following equation in terms of $\tau$

$$q_Y(\tau) = \sum_{j=1}^{p_1+m_1} \theta_j(X_i)B_{j,m_1}(\tau). \quad (5.13)$$

In case of NPDFSQR, once the values of $\{Y_{\rho_l}\}_{l=0}^{c}$ are evaluated, likelihood evaluation is straightforward and can be easily obtained using Equation (5.12).

**Block Metropolis-Hastings MCMC Algorithm**

To estimate the parameters for NPSQR and NPDFSQR methods (given by Equations (5.3) and (5.6) respectively) we use Block Metropolis-Hastings Markov Chain Monte Carlo algorithm (Chib and Greenberg (1995)). It should be noted that the parameter space which needs to be estimated are of the same form for NPSQR and NPDFSQR and hence we use the similar steps and same prior distribution for both the cases. Recall that for NPSQR the parameter space is given by

$$0 = \alpha_{1k_1 \cdots k_d} < \cdots < \alpha_{(p_1+m_1)k_1 \cdots k_d} = 1, \,(k_1, \ldots, k_d) \in \{1, \ldots, (p_2 + m_2)\}^d.$$ 

Now define

$$\gamma_{jk_1 \cdots k_d} = \alpha_{(j+1)k_1 \cdots k_d} - \alpha_{jk_1 \cdots k_d}, \quad j = 1, \ldots, p_1 + m_1 - 1$$
for \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). Then it can be noted that \( \{\gamma_{jk_1 \ldots k_d}\}_{j=1}^{p_1 + m_1 - 1} \) belongs to the unit-simplex for any given \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \) since
\[
\sum_{j=1}^{p_1 + m_1} \gamma_{jk_1 \ldots k_d} = 1, \quad \gamma_{jk_1 \ldots k_d} \geq 0, \quad j = 1, \ldots, p_1 + m_1 - 1,
\]
for \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). We put the uniform prior on each simplex block. The number of simplex blocks is given by \( (p_2 + m_2)^d \) where \( \{s_i\}_{i=0}^{p_2} \) is the equidistant knot sequence used for each coordinate of the explanatory variables and \( m_2 \) is the degree of the piece-wise polynomials use for B-spline regression (as mentioned in Section 5.2.1). Within an iteration, one simplex block is updated at a time. Hence there will be \( (p_2 + m_2)^d \) updates performed one at a time during a single iteration.

To make movements on the simplex block, we use the same strategy as explained in Chapters 2 and 4. First we fix \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). We generate independent sequence \( \{U_j\}_{j=1}^{p_1 + m_1 - 1} \) from \( U(1/r, r) \) for some \( r > 1 \). It should be noted that \( r \) works as a tuning parameter of the MCMC. Smaller value of \( r \) yields sticky movement with higher acceptance probability while larger value of \( r \) would result in bigger jumps with less acceptance probability. Define \( V_j = \gamma_{jk_1 \ldots k_d} U_j \). Hence the proposal move \( \gamma_{jk_1 \ldots k_d} \mapsto \gamma_{jk_1 \ldots k_d}^* \) is given by
\[
\gamma_{jk_1 \ldots k_d}^* = \frac{V_j}{\sum_{i=1}^{p_1 + m_1 - 1} V_i}, \quad j = 1, \ldots, p_1 + m_1 - 1.
\]
The conditional distribution of \( \{\gamma_{jk_1 \ldots k_d}^*\}_{j=1}^{p_1 + m_1 - 1} \) given \( \{\gamma_{jk_1 \ldots k_d}\}_{j=1}^{p_1 + m_1 - 1} \) is given by (see Section 2.3.2 for the derivation)
\[
f(\gamma_{k_1 \ldots k_d}^* | \gamma_{k_1 \ldots k_d}) = \left(\frac{r}{r^2 - 1}\right)^{p_1 + m_1 - 1} \left\{ \prod_{j=1}^{p_1 + m_1 - 1} \gamma_{jk_1 \ldots k_d} \right\}^{-1} \frac{(D_1 - D_2)}{(p_1 + m_1 - 1)}, \tag{5.14}
\]
where
\[
D_1 = \min_{0 \leq j \leq p_1 + m_1 - 1} \frac{r \gamma_{jk_1 \ldots k_d}}{\gamma_{jk_1 \ldots k_d}^*}^{p_1 + m_1 - 1},
\]
\[
D_2 = \max_{0 \leq j \leq p_1 + m_1 - 1} \frac{\gamma_{jk_1 \ldots k_d}}{r \gamma_{jk_1 \ldots k_d}^*}^{p_1 + m_1 - 1}.
\]
Suppose \( L(\gamma_{k_1 \ldots k_d}) \) and \( L(\gamma_{k_1 \ldots k_d}^*) \) denote the likelihood for NPSQR (either corresponding to complete data or quantile grid data) for the parameter values \( \gamma_{k_1 \ldots k_d} = \{\gamma_{jk_1 \ldots k_d}\}_{j=1}^{p_1 + m_1 - 1} \) and \( \gamma_{k_1 \ldots k_d}^* = \{\gamma_{jk_1 \ldots k_d}^*\}_{j=1}^{p_1 + m_1 - 1} \) respectively.
and \( \gamma^*_{k_1 \ldots k_d} = \{ \gamma_{jk_1 \ldots k_d} \}_{j=1}^{p_1+m_1-1} \) respectively. Then a single block update for \((k_1, \ldots, k_d) \in \{1, \ldots, (p_2 + m_2)\}^d\) is given by \( P_{k_1 \ldots k_d} = \min\{p_{k_1 \ldots k_d}, 1\} \) where

\[
P_{k_1 \ldots k_d} = \frac{L(\gamma^*_{k_1 \ldots k_d}) \pi(\gamma^*_{k_1 \ldots k_d}) f(\gamma_{k_1 \ldots k_d} | \gamma^*_{k_1 \ldots k_d})}{L(\gamma_{k_1 \ldots k_d}) \pi(\gamma_{k_1 \ldots k_d}) f(\gamma^*_{k_1 \ldots k_d} | \gamma_{k_1 \ldots k_d})},
\]

\[
= \frac{L(\gamma^*_{k_1 \ldots k_d}) f(\gamma_{k_1 \ldots k_d} | \gamma^*_{k_1 \ldots k_d})}{L(\gamma_{k_1 \ldots k_d}) f(\gamma^*_{k_1 \ldots k_d} | \gamma_{k_1 \ldots k_d})},
\]

and \( \pi(\cdot) \) denotes the uniform prior density.

For NPDFSQR we define

\[
\delta_{jk_1 \ldots k_d} = \beta_{j+1} k_1 \ldots k_d - \beta_{j} k_1 \ldots k_d, \ j = 1, \ldots, p_1 + m_1 - 1
\]

for \( \{k_1, \ldots, k_d\} \in \{1, \ldots, (p_2 + m_2)\}^d \). Hence in this case also the parameter space is given by a collection of \((p_2 + m_2)^d\) simplex blocks. Hence in case of NPDFSQR (for complete data and quantile grid data), the update steps are performed similar to that of NPSQR (as mentioned above). In this case also, inside each iteration step of MCMC, \((p_2 + m_2)^d\) simplex blocks are updated one by one.

**Warm Start**

In case of very large parameter space, the strategy of warm-start in general helps in reducing the burn-in for Metropolis-Hastings MCMC. In the proposed method, instead of using a randomly generated starting point, we use the maximum likelihood estimator (MLE) as the starting point. In Section 5.4, it is noted that for both NPSQR and NPDFSQR methods, the parameter space is given by a collection of simplex blocks. One of the challenging aspect of finding MLE under this scenario is that the parameter space is constrained and for NPSQR, the likelihood function does not have any closed form. It is almost impossible to verify whether the negative of the likelihood of NPSQR is convex or not (in case negative likelihood is convex, there exists only one global maximum and convex optimization techniques can be used to find it). Thus optimization should be performed assuming the possibility of existence of multiple local maximums of the likelihood function. Secondly, due to the absence of a closed form likelihood in the case of NPSQR, the derivative of the likelihood function does not have any closed form. In this case, one of the disadvantages of using derivative based optimization methods is that derivatives can be evaluated only numerically which is computationally intensive. Hence, this is an ideal
scenario to use black-box optimization technique since black-box optimization technique does not use analytically derivative and is used to optimize any function with (possibly) multiple maximums or minimums.

Recently Das (2016b) proposed an black-box optimization technique on a hyper-rectangular parameter space which has been shown to perform better (in terms of computing time, accuracy and successful convergence) than black-box optimization techniques Genetic Algorithm (Fraser (1957), Bethke (1980), Goldberg (1989)) and Simulated Annealing (Kirkpatrick et al. (1983), Granville et al. (1994)) yielding better solutions. Following that strategy, Das (2016c) modified that algorithm to optimize any function on an unit simplex. Later Das (2016a) extended that method and proposed ‘Greedy Co-ordinate Descent of Varying Step sizes on Multiple Simplexes’ (GCDVSMS) algorithm which efficiently minimizes (or maximizes) any black-box function of parameters given by a collection of unit simplex blocks. The main idea of the this algorithm is to make jumps of varying step-sizes within each unit simplex blocks parallelly and searching for the most favorable direction of movement. We use GCDVSMS algorithm to find the warm starting point before initializing the MCMC.

**Automatic Controlling of Acceptance Probability**

As mentioned Section 5.4, \( r \) plays a critical role in controlling the acceptance probability. Instead of fixing the value of \( r \), we propose an adaptive strategy so that during the MCMC iterations, acceptance probability is maintained within a desirable range of 0.15 to 0.45.

For both NPSQR and NPDFSQR, we start the first iteration with \( r = 1.05 \). It should be noted that in each iteration, \((p_2 + m_2)^d\) simplex blocks are updated; hence, \((p_2 + m_2)^d\) acceptance-rejection decisions are taken. After each iteration, the cumulative acceptance ratio is calculated. At the end of an iteration, if the cumulative acceptance probability drops below 0.15, \( r \) is updated to \( 1 + (r - 1)/2 \) and if the cumulative acceptance probability goes above 0.45, the value of \( r \) is updated to \( 1 + 2(r - 1) \). Note that in this way, the value of \( r \) will always be greater than 1.

**Simulation study**

For simulation purpose, we consider the following true models and we generate sample of sizes \( n = 50, 100, 200 \) for each case.
(A) **First Simulation Study**: We consider the explanatory variable $X$ in coming from $U(0, 5)$. The dependence of the response variable $Y$ on $X$ is given by

$$y_i = x_i + \sin(2x_i) + 3\epsilon_i, \ i = 1, \ldots, n. \quad (5.15)$$

where $\epsilon_i$ follows $SN(4; 0, 1)$ (i.e., skewed normal distribution with scale parameter 4, mean 0 and standard deviation 1).

(B) **Second Simulation Study**: We consider the explanatory variable $X$ in coming from $U(-100, 100)$. The dependence of the response variable $Y$ on $X$ is given by

$$y_i = -x_i^3/100000 + (\sin(\pi x_i/100) + 4)U_i \epsilon_i \text{ for } i = 1, \ldots, n. \quad (5.16)$$

where $U_i$ follows discrete uniform $\{-1, 1\}$ and $\epsilon_i$ follows gamma distribution with shape and scale parameters 5 and 1 respectively.

**Case of Complete Data**

For each of the cases given by Equation (5.15) and (5.16), the quantiles are estimated using Non-parametric Simultaneous Quantile Regression (NPSQR), Non-parametric Distribution Function Simultaneous Quantile Regression (NPDFSQR), Local Linear Quantile Regression (LLQR) (see Yu and Jones (1998), Koenker (2015)) and Local Spline Quantile Regression (LSQR) (see Koenker (2015)).

For NPSQR and NPDFSQR, first the explanatory variable and the response variables are transformed into unit interval separately by linear transformation. We use piece-wise quadratic B-spline for expanding both the quantile function as well as the explanatory variable (i.e., $m_1 = m_2 = 2$). One of the biggest advantage of quadratic B-spline is that Equations (5.7) and (5.13) reduce to quadratic equation which can be solved analytically very easily. Suppose $\{t_i\}_{i=0}^{p_1}$ and $\{s_i\}_{i=0}^{p_2}$ denote the equidistant knots on unit interval such that

$$0 = t_0 < t_1 < \cdots < t_{p_1} = 1, (t_i - t_{i-1}) = \frac{1}{p_1} \text{ for } i = 1, \ldots, p_1;$$

$$0 = s_0 < s_1 < \cdots < s_{p_2} = 1, (s_i - s_{i-1}) = \frac{1}{p_2} \text{ for } i = 1, \ldots, p_2.$$

In case of NPSQR, for B-spline expansion, we consider $\{t_i\}_{i=0}^{p_1}$ and $\{s_i\}_{i=0}^{p_2}$ knot sequences for expanding the quantile function and transformed explanatory variable respectively.
Table 5.1: (First Simulation Study (Complete Data)) Prediction Mean Squared Errors using NPSQR, NPDFSQR, LLQR and LSQR based on simulation studies for sample sizes $n = 50, 100, 200$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>NPSQR</th>
<th>NPDFSQR</th>
<th>LLQR</th>
<th>LSQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>4.90</td>
<td>4.66</td>
<td>5.13</td>
<td>5.42</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>4.40</td>
<td>4.51</td>
<td>4.59</td>
<td>4.50</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>4.34</td>
<td>4.47</td>
<td>4.65</td>
<td>4.56</td>
</tr>
</tbody>
</table>

We consider $p_1 = p_2 = 3, \ldots, 10$ and choose the best model using the Akaike Information Criteria (AIC) criterion. For NPDFSQR, suppose that knot sequences $\{t_i\}_{i=0}^{p_1}$ and $\{s_i\}_{i=0}^{p_2}$ are used for expanding the distribution function and transformed explanatory variable respectively. For NPDFSQR, it is noted that taking $p_1 = p_2 = 3, 4$, yields poorly estimated posterior quantile curves under different simulation studies. The possible reason is that in case of NPDFSQR, a lot of estimated B-spline coefficients (which are used to estimate the distribution function of the response variable) are coming to be nearly zero which might be an indicator of usage of less number of knots to represent the variability of the distribution function. Therefore, for NPDFSQR, we omit the cases $p_1 = p_2 = 3, 4$ and consider the cases $p_1 = p_2 = 5, \ldots, 10$. Then we choose the best model via the AIC criterion.

As mentioned in Section 5.4.1, for all considered MCMC schemes in this chapter, we use GCDVSMS algorithm to find the starting point. The values of the tuning parameters in the GCDVSMS algorithm have been taken to be as follows: initial global step size $s_{\text{initial}} = 1$, step decay rate for the first run $\rho_1 = 2$, step decay rate for other runs $\rho_2 = 1.05$, step size threshold $\phi = 10^{-2}$, sparsity threshold $\lambda = 10^{-3}$, the convergence criteria controlling parameters $\text{tol} \_\text{fun} \_1 = \text{tol} \_\text{fun} \_2 = 10^{-2}$, maximum number of iterations inside each run $\text{max} \_\text{iter} = 5000$, maximum number of allowed runs $\text{max} \_\text{runs} = 200$. For both NPSQR and NPDFSQR, we perform 10000 iterations discarding first 1000 iterations as burn-in. After the quantile curves are estimated, inverse linear transformations are applied on the response and the explanatory variables to get them back to their original scales.

In NPDFSQR, once the whole distribution function is estimated non-parametrically, the quantile function is obtained evaluating numerically. We take a grid of length 1000 on transformed $Y$ variable which is a unit interval. For any given value of $X = x$, the distribution function is evaluated at these 1000 equidistant grid-points. Then $Q(\tau|x)$
Figure 5.1: (First Simulation Study (Complete Data)) True and estimated quantiles at \( \tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\} \) for \( n = 100 \) using NPSQR, NPDFSQR, LLQR and LSQR with the data points.
Figure 5.2: (Second Simulation Study (Complete Data)) True and estimated quantiles at $\tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR, NPDFSQR, LLQR and LSQR with the data points.
Table 5.2: (Second Simulation Study (Complete Data)) Prediction Mean Squared Errors using NPSQR, NPDFSQR, LLQR and LSQR based on simulation studies for sample sizes $n = 50, 100, 200$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>NPSQR</th>
<th>NPDFSQR</th>
<th>LLQR</th>
<th>LSQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>564.81</td>
<td>576.18</td>
<td>576.84</td>
<td>641.17</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>559.14</td>
<td>539.68</td>
<td>592.83</td>
<td>645.58</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>504.22</td>
<td>517.63</td>
<td>532.97</td>
<td>554.40</td>
</tr>
</tbody>
</table>

is estimated using interpolation from the values of the distribution function at those aforementioned 1000 points.

For any given quantile level, LSQR (Koenker (2015)) fits a piecewise cubic polynomial with any given number of knots (breakpoints in the third derivative) arranged at that quantile of the $X$. However, no explicit way for deciding the number of knots for a given data set has been provided in that article. For fair comparison, while fitting LSQR, we consider the same number of knots as used in estimating the quantile levels with NPSQR for that data-set. (Note that, as mentioned earlier, the number of knots in NPSQR is selected based on the AIC criterion). While estimating any given quantile level with LLQR, to select the bandwidth, we follow the technique mentioned in the section 2 of Yu and Jones (1998). We use ‘quantreg’ (Koenker et al. (2016)) R-package for LLQR and LSQR. Except for the bandwidth selection for LLQR, rest of the codes have been followed as provided in Koenker (2015).

For comparing the performances of the proposed methods, LLQR and LSQR, 1000 pairs of observations $\{(X_i, Y_i)\}_{i=1}^{1000}$ are generated from Equation (5.15). Let $\hat{Q}(\tau|X)$ denote the estimated value of the $\tau$-th quantile at $X = x$. Then the Prediction Mean Squared Error (PMSE) is given by

$$PMSE = \frac{1}{1000} \sum_{i=1}^{1000} (Y_i - \hat{Q}(0.5|X_i))^2.$$  

Note that $\hat{Q}(0.5|X_i)$ denotes the estimated median estimate at $X = X_i$. Except for LSQR, it is straightforward to find $\hat{Q}(0.5|X_i)$ for any given $X_i$. The way LSQR is performed in Koenker (2015), the quantile curves are evaluated only at those points where $X$ is given (in the data). So in this case, we use linear interpolation with \texttt{interp} function in R to find the approximate values of $\{\hat{Q}(0.5|X_i)\}_{i=1}^{1000}$ from the estimated median values at the
points given in the data.

In the simulation study, in Tables 5.1 and 5.2, it is noted that the performances of NPSQR and NPDFSQR are generally better than LLQR and LSQR in terms of PMSE. For both NPSQR and NPDFSQR, we note an overall decreasing trend of PMSE with increasing sample size. We note that unlike the case of the proposed methods, using LLQR and LSQR the estimated quantile lines cross each other. It is also noted that the estimated quantile curves using LSQR in the simulation studies (and also in the example provided in Koenker (2015)) have a tendency to pass through the data points which may not be desirable specially for estimating quantile curves with small sample.

Case of Grid data

In case of grid data, we obtain the grid data by coarsening the data generated in Section 5.5.1 into grid data. While transforming a sample of given size into grid data, we consider three types of grid data generated from each sample which are 5, 10 and 20 percentile gap grid data. For example, in case of 5 percentile grid data, the values of \( \{q_Y(\rho_l)\}_{l=1}^{19} \) are given where \( \rho_l = 0.05 \cdot l \) for \( l = 1, \ldots, 19 \). The values of \( \{q_Y(\rho_l)\}_{l=1}^{19} \) are computed non-parametrically from the given sample using \texttt{quantile} function in MATLAB (version R2014a). Then for each given value of \( X_i \), it is noted that between which two consecutive quantile grids the corresponding observation \( Y_i \) belongs for \( i = 1, \ldots, n \). Thus corresponding to each value of \( X_i \), the position of \( Y_i \) with respect to the quantile grids is given in the grid data. 10 and 20 percentile grid data are also generated in the similar way.

After the grid data is generated, we transform the values of explanatory and the response values into unit intervals separately using linear transformations. Once they are transformed into the unit intervals, the likelihood can be computed as described in Section 5.3.2. As mentioned earlier, except the likelihood evaluation part, the remaining part of the Block Metropolis-Hastings MCMC algorithm is similar to that of the case of complete data. Before starting the MCMC, we compute a warm starting point using GCDVSM algorithm (Das (2016a)) with the values of the tuning parameters as mentioned in Section 5.5.1. We estimate the quantile curves for each cases using NPSQR and NPDFSQR methods. We also compute the PMSE (as defined in Section 5.5.1) for comparison under each scenario. For each cases, we perform 10000 MCMC iterations discarding the first 1000 iterations as burn-in. Similar to the case of complete data, for NPSQR we consider \( p_1 = p_2 = 3, \ldots, 10 \) and for NPDFSQR we consider \( p_1 = p_2 = 5, \ldots, 10 \). The best possible value of \( p_1 \) and \( p_2 \) in either cases are selected based on AIC criteria. After the quantile
curves are estimated, inverse transformations on the response and explanatory variables are performed to return back to the original scale.

We note that NPSQR performs slightly better than NPDFSQR in terms of PMSE for both the simulation studies considered. It is also noted that with increasing sample size, there is a decreasing trend of PMSE for both the cases. Also, with smaller percentile gap data, PMSE comes out to be smaller in most of the cases for both NPSQR and NPDFSQR. To study the relative performance of estimating the quantile curves with proposed methods for complete and grid data, readers can compare the PMSE values in Tables 5.1 and 5.3 (for the first simulation study) and in Tables 5.2 and 5.4 (for second simulation study).

### Application to Hurricane Data

Elsner et al. (2008) made an argument that the hurricanes with higher velocities in the North Atlantic basin have got stronger in the last couple of decades. We apply NPSQR method to estimate the simultaneous quantiles of the hurricane velocities in the North Atlantic basin using the hurricane intensity data\(^1\) during the period 1981–2006. First the

\(^1\)Source [http://weather.unisys.com/hurricane/atlantic/](http://weather.unisys.com/hurricane/atlantic/)
Figure 5.3: (First Simulation Study (Grid Data)) True and estimated quantiles at \( \tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\} \) for \( n = 100 \) using NPSQR and NPDFSQR with the data points.
Figure 5.4: (Second Simulation Study (Grid Data)) True and estimated quantiles at \( \tau = \{0.05, 0.10, 0.15, \ldots, 0.90, 0.95\} \) for \( n = 100 \) using NPSQR and NPDFSQR with the data points.
Table 5.4: (Second Simulation Study (Grid Data)) Prediction Mean Squared Errors using NPSQR and NPDFSQR based on simulation studies with 5, 10, 20 distant percentile grids for sample sizes $n = 50, 100, 200$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Percentile gap for grid data</th>
<th>NPSQR (pred error)</th>
<th>NPDFSQR (pred error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>5</td>
<td>510.31</td>
<td>537.40</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>524.79</td>
<td>568.91</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>559.10</td>
<td>580.08</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>5</td>
<td>550.09</td>
<td>548.82</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>578.09</td>
<td>571.07</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>573.44</td>
<td>575.94</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>5</td>
<td>497.74</td>
<td>541.71</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>535.20</td>
<td>539.61</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>540.32</td>
<td>549.71</td>
</tr>
</tbody>
</table>

The explanatory variable time is linearly transformed to unit interval such that the years 1981 and 2006 are mapped to 0 and 1 respectively. To transform the hurricane velocities into the unit interval, we assume the hurricane velocities follow the power-Pareto distribution. The power-Pareto density is given by

$$f(y) = \frac{ak(y/\sigma)^{k-1}}{\sigma(1+(y/\sigma)^k)^{(a+1)}} \quad y > 0$$

The distribution function is given by

$$F(y) = 1 - \frac{1}{(1+(y/\sigma)^k)^a} \quad (5.17)$$

Tokdar and Kadane (2012) proposed the parameter values as $a = 0.45$, $\sigma = 52$ and $k = 4.9$ in the same context. We transform the hurricane velocities into unit interval using Equation (5.17).

After transforming both $X$ and $Y$ values into unit interval, we use NPSQR method to estimate the simultaneous quantiles of the hurricane wind velocities. Similar to the approach considered in Section 5.5.1, we start the Block-Metropolis Hastings algorithm with a warm starting point found using GCDVSMS algorithm. We consider 10000 posterior samples discarding the first 1000 samples as burn-in. The number of equidistant
knots to be used for B-spline basis expansion is selected using the AIC. After the quantile curves are estimated, corresponding inverse transformations are performed on the response and the explanatory variables before plotting them. In Figure 5.5 we note that unlike the upper quantiles, the lower quantiles of the hurricane velocities have changed little over the time. We note a periodic pattern in the upper quantiles. Specially this pattern becomes more prominent with increasing values of the quantile. An increasing pattern of the higher quantile curves is noted during the period 1987-1994 and 2002-2005 while a decreasing pattern is prominent during 1994-2002.

Application to US household income data

Historical tables for US household income data can be found in this site. In the data tables, the 20, 40, 60, 80 and 95-th quantiles of the household income in current dollars (accessed 11-25-2016) of all population (combined), Asian, Black, Hispanic, White and White non-Hispanic population have been provided for a few years. Along with that, the total population of each category at each considered year has been given. A snapshot of the data table showing the household income distribution of all population during 2010-2015 is given in Table 5.5. For analysis, we transform the years linearly to the the unit interval. For example, for Hispanic population since the data is available during 1972-2015, we transform it linearly to the unit interval such that 1972 and 2015 get mapped

\[\text{Source: } \text{http://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html}\]
Table 5.5: A snapshot of US household income data table showing the income distribution of all population during 2010 – 2015.

<table>
<thead>
<tr>
<th>Years</th>
<th>Population (thousands)</th>
<th>20 percentile</th>
<th>40 percentile</th>
<th>60 percentile</th>
<th>80 percentile</th>
<th>95 percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>125,819</td>
<td>22,800</td>
<td>43,511</td>
<td>72,001</td>
<td>117,002</td>
<td>214,462</td>
</tr>
<tr>
<td>2014</td>
<td>124,587</td>
<td>21,432</td>
<td>41,186</td>
<td>68,212</td>
<td>112,262</td>
<td>206,568</td>
</tr>
<tr>
<td>2013</td>
<td>123,931</td>
<td>21,000</td>
<td>41,035</td>
<td>67,200</td>
<td>110,232</td>
<td>205,128</td>
</tr>
<tr>
<td>2012</td>
<td>122,952</td>
<td>20,900</td>
<td>40,187</td>
<td>65,501</td>
<td>105,910</td>
<td>196,000</td>
</tr>
<tr>
<td>2011</td>
<td>121,084</td>
<td>20,262</td>
<td>38,520</td>
<td>62,434</td>
<td>101,582</td>
<td>186,000</td>
</tr>
<tr>
<td>2010</td>
<td>119,927</td>
<td>20,000</td>
<td>38,000</td>
<td>61,500</td>
<td>100,029</td>
<td>180,485</td>
</tr>
</tbody>
</table>

To transform the incomes into unit interval, we use a log-linear transformation. First we take logarithms of all income values of all races. After the log transformation, suppose \( a_1 \) and \( a_2 \) denote the smallest and the biggest values. Define \( L = a_1 - 0.01 \) and \( U = a_2 + 0.01 \). We use the transformation \( f(y) = (\log y - U)/(L - U) \) to transform all the incomes to unit interval and the values of \( L \) and \( U \) come out to be 7.47 and 12.55 respectively. After the analysis, inverse transformations are performed before plotting the quantile curves to return to the original scale.

It is noted that this data is somewhat different from the data considered for simulation study in Section 5.5.2. Firstly, the value of quantile grids (i.e., \( q_Y(\rho_l) \)) in this data is different for each year (or time-point). For example, as seen in Table 5.5, the values of income at different quantile levels in 2015 is different than that of 2014. Secondly, unlike the data considered in Section 5.5.2, here for each value of \( X \) (i.e., time-point), there are multiple observations. The quantile grid considered here is given by \( \{\rho_l\}_{l=0}^6 \) where \( \rho_0 = 0, \rho_1 = 0.2, \rho_2 = 0.4, \rho_3 = 0.6, \rho_4 = 0.8, \rho_5 = 0.95 \) and \( \rho_6 = 1 \). So if at \( n \) time-points \( \{X_i\}_{i=1}^n \), the number of observations (i.e., population) are \( \{V_i\}_{i=1}^n \) then the likelihood is given by

\[
L = \prod_{i=1}^n (F(q_Y(0.2|X_i)))^{0.2V_i} \cdot (F(q_Y(0.4|X_i)) - F(q_Y(0.2|X_i)))^{0.2V_i} \cdot (F(q_Y(0.6|X_i)) - F(q_Y(0.4|X_i)))^{0.2V_i} \cdot (F(q_Y(0.8|X_i)) - F(q_Y(0.6|X_i)))^{0.2V_i} \cdot (F(q_Y(0.95|X_i)) - F(q_Y(0.8|X_i)))^{0.15V_i} \cdot (1 - F(q_Y(0.95|X_i)))^{0.05V_i}.
\]
Clearly, the value of $\{F(q_Y(\rho_l|X_i))\}_{l=0}^6$ can be found using the same technique as used described in Section 5.5.2.

To estimate the quantile curves we use the NPSQR method. We start the Block Metropolis-Hastings MCMC algorithm with warm starting point found using the GCD-VSMS algorithm. Unlike all the previous studies, this data represent the population, not the sample. Hereby, instead of choosing the optimal number of knots for fitting the B-spline, we can fix their values anywhere depending on desired smoothness level. We set $p_1 = p_2 = 5$ for the whole analysis.

In Figure 5.6 we plot the estimated simultaneous quantiles of household income of all population (during 1967-2015), Asian (during 2002-2015), Black (during 1967-2015), Hispanic (during 1972-2015), White (during 1967-2015) and White non-Hispanic (during 1972-2015) population. Irrespective of the races, it is noted that the higher quantile curves increase at higher rates while the lower quantile curves are roughly constant over the years. Household incomes seem to be more evenly distributed among Asian people compared with other races. Among other races White and White non-Hispanic population have lesser gaps across different quantile levels of income distribution compared with that of Black and Hispanic population. The estimated quantile levels of household income of Asian population is greater than other populations. It should be also noted that the gaps between the quantile levels are more or less increasing over time.

**Conclusion**

In this chapter two novel methods for non-parametric simultaneous quantile regression methods have been proposed. In the first method, the quantile function is estimated non-parametrically using tensor products of quadratic B-splines basis expansion and in the second method the distribution function is estimated by a non-parametric approach using tensor product of quadratic B-splines basis expansion. These methods have been further developed for estimating the quantiles for the quantile grid data. We consider the Block Metropolis-Hastings MCMC algorithm to estimate the coefficients of the B-spline basis functions. Before initializing the MCMC, the Maximum Likelihood Estimator (MLE) is evaluated using GCDVSMS algorithm and it is used as the starting point. The optimal number of knots for the B-spline basis expansion is selected using the AIC criterion. Unlike the existing popular methods of non-linear quantile regression, e.g., local linear and local spline quantile regression, the monotonicity of the quantile curves are maintained using the proposed methods. In the simulation studies it is shown that both
Figure 5.6: Estimated simultaneous quantiles of household income of (a) All population (1967–2015) (b) Asian (2002–2015) (c) Black (1967–2015) (d) Hispanic (1972–2015) (e) White (1967–2015) and (f) White Non-Hispanic (1972–2015) population at $\tau = \{0.05, 0.10, 0.20, \ldots, 0.90, 0.95\}$ for $n = 100$ using NPSQR.
of the proposed methods generally perform better than LLQR and LSQR in terms of the PMSE. It is also observed that for the quantile grid data, NPSQR performs slightly better than NPDFSQR in terms of PMSE.

The NPSQR method has been used to analyze the hurricane intensity data of North Atlantic region for the years 1981-2006. A periodic nature is noted at higher quantile levels while estimated lower quantile curves are relatively stable and with respect to time. NPSQR method has been also used to analyze the historical household income data of different races in US given in the form of quantile grids. It is noted that the higher quantile curves are increasing generally at higher rates than the lower quantile curves. The differences between the household income levels tend to increase over time.
REFERENCES


