Some Explorations of Bayesian Joint Quantile Regression

by

Wenli Shi

Department of Statistical Science
Duke University

Date: ____________________
Approved:

____
Surya Tokdar, Supervisor

____
Sayan Mukherjee

____
Alessandro Arlotto

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in the Department of Statistical Science
in the Graduate School of Duke University
2017

Abstract

Some Explorations of Bayesian Joint Quantile Regression

by

Wenli Shi

Department of Statistical Science
Duke University

Date: ____________________
Approved:

____
Surya Tokdar, Supervisor

____
Sayan Mukherjee

____
Alessandro Arlotto

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Master of Science in the Department of Statistical Science
in the Graduate School of Duke University
2017

Abstract

Although quantile regression provides a comprehensive and robust replacement for the traditional mean regression, a complete estimation technique is in blank for a long time. Original separate estimation could cause severe problems, which obstructs its popularization in methodology and application. A novel complete Bayesian joint estimation of quantile regression is proposed and serves as a thorough solution to this historical challenge. In this thesis, we first introduce this modeling technique and propose some preliminary but important theoretical development on the posterior convergence rate of this novel joint estimation, which offers significant guidance to the ultimate results. We provide the posterior convergence rate for the density estimation model induced by this joint quantile regression model. Furthermore, the prior concentration condition of the truncated version of this joint quantile regression model is proved and the entropy condition of the truncated model with any sphere predictor plane centered at $\mathbf{0}$ is verified. An application on high school math achievement is also introduced, which reveals some deep association between math achievement and socio-economic status. Some further developments about the estimation technique, convergence rate and application are discussed. Furthermore, some suggestions on school choices for minority students are mentioned according to the application.

To my parents.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Symbols

$\mathbb{R}$	Real number
$\log$	Natural logarithm
Q	Quantile function
q	Quantile density function
F	Cumulative distribution function
f	Probability density function
$\mathcal{X}$	Convex hull of predictors
d_{KL}	Kullback-Leibler divergence

Abbreviations

QR	Quantile Regression
GP	Gaussian Process
MCMC	Markov Chain Monte Carlo
KL	Kullback-Leibler
SES	Socio-Economic Status
RKHS	Reproducing Kernel Hilbert Space
WAIC	Watanabe-Akaike Information Criterion

Acknowledgements

I would like to convey my primary appreciation to the Department of Statistical Science where I recognize as a family and start my journey on Bayesian statistics. I would like to thank my advisor, professor Surya Tokdar, for his tremendously invaluable guidance on me. I obtained a more comprehensive and systematic understanding of Bayesian nonparametrics during this procedure. I cannot finish this work without his support.

I also want to express my thanks to all my committee members, professor Sayan Mukherjee and professor Alessandro Arlotto, for their support on my defense. I also want to thank professor David Dunson and professor Merlise Clyde to initiate my interest in Bayesian modeling and professor Mike West for his patience and fantastic teaching for leading me to the fascinating dynamic world. I imagine to continue on these paths successfully in future.

One most special gratitude to my family for supporting my master study and my girlfriend for her companion. Thanks to all that helped me achieve who I am.

1

Introduction

Modern statistics realizes the noticeable heterogeneity within some population and demands more attention on addressing this puzzle through more flexible modeling techniques. Quantile regression (QR; Koenker 2005; Koenker and Bassett Jr 1978) has provided a suitable and more comprehensive replacement of normal least squares regression and contributes to the research in epidemiology, climate science, marine biology and economics (Burgette et al., 2011; Elsner et al., 2008; McClain and Rex, 2001; Buchinsky, 1994). Not only reflecting the heterogeneity along different quantiles of the response distribution, QR also maintains the robust property over the least squared mean regression which could be easily driven by the outliers and allows researchers to understand the casual relationship between response and predictors at each percentile proportion of the population.

However, compared with the comprehensive model structure, the estimation technique of QR is still underdeveloped. The original inference approach mentioned in Koenker (2005) estimates the coefficients β_0 and β in $Q(\tau|x) = \beta_0(\tau) + x^T\beta(\tau)$ separately for any fixed quantile τ and therefore could be problematic. Lack of borrowing of information, this estimation technique could generate crossing and erratic regression curves for different quantiles especially when the quantile function is complicated (e.g. polynomial, wavelet or spline transform on the predictors), which violates the laws of probability and induces the research on joint estimation of the conditional quantile plane.

The joint estimation requires the specification of the conditional quantile plane as $Q(\tau|x) = \beta_0(\tau) + x^T\beta(\tau)$ and work on all $\tau \in (0, 1)$ simultaneously, where β_0: $(0, 1) \to \mathbb{R}$ and β: $(0, 1) \to \mathbb{R}^p$. Under these specifications of any smooth quantile function Q, for any $x \in \mathcal{X}$ and $\tau_1 > \tau_2$, $Q(\tau_1|x) > Q(\tau_2|x)$ which is equivalent to $\beta_0(\tau_1) + x^T\beta(\tau_1) > \beta_0(\tau_2) + x^T\beta(\tau_2)$ becomes an automatic and reasonable condition by the laws of probability. Many preceding research works on joint estimation of quantile regression are fruitful in some aspects but restrictive and unsatisfactory (Tokdar et al., 2012; Dunson and Taylor, 2005; He, 1997). Therefore, a further exploration for seeking a sounder statistical model is required.

In this thesis, Chapter 2 introduces a recent developed estimation method compared with former approaches on joint quantile regression and the remainder provides some auxiliary theoretical and applied explorations, which are organized as follows. In Chapter 3, we propose the crucial prior concentration condition and entropy condition for seeking the convergence rate of the posterior based on Ghosal et al. (2000) under some extra assumptions. In Chapter 4, we implement this novel estimation method on the math achievement gap between the minority students and non-minority students in high school. Chapter 5 summarizes the whole paper with a brief discussion on aforementioned work and future potential development.

2

Bayesian Joint Quantile Regression

The original estimation approach introduced in Koenker (2005) converts the statistical inference to a programming problem through minimizing the expectation of the loss function $\rho_\tau(y) = y \times (\tau - I(y < 0))$ as

$$\min_{u \in \mathbb{R}} E[\rho_\tau(Y - u)] = \min_{u \in \mathbb{R}} (\tau - 1) \int_{-\infty}^{u} (y - u) dF(y) + \tau \int_{u}^{\infty} (y - u) dF(y)$$

where the minimum is achieved at $y_\tau = F^{-1}(\tau)$ through basic calculus computation, which is the τth-quantile of random variable Y. By replacing u with some specific quantile function and minimizing the expected loss function according to the parameters in u, a quantile regression model is defined and estimated. However, it is noticeable that the estimation of parameters are proceeded for each quantile separately, which potentially could induce the crossing regression curves that are invalid and could also contain some irregular waviness and erratic behavior at different quantiles. Another unreasonable case is observed in Yang and Tokdar (2016, Figure 1(a)) that the curves aggregate to nearly same values at both ends, where the uncertainty should be largest because of the scarcest amount of data.

Since then, joint quantile regression attracts statisticians' attention and gains profound development. Although many works and publications are brilliant, they are unsatisfactory with some particular weakness. Wu and Liu (2009) proposed an algorithm to estimate the parameters sequentially, which highly depends on the order that the quantiles are fitted. It is also unrealistic to restrict the shape of $\beta(\tau)$ as proposed in He (1997) and Bondell et al. (2010). The approximate works in Dunson and Taylor (2005) and Lancaster and Jae Jun (2010) are based on the thoughts of substitution and empirical likelihoods, which could be recognized as partial or approximate solutions. Tokdar et al. (2012) proposes a univariate joint quantile regression model, which is general and comprehensive but reluctant to be generalized to the multivariate case. A recent work in Reich et al. (2011) requires the expansion of predictor domain $\mathcal{X}$ to form a hyper-rectangle $\mathcal{X}_0$, which could be problematic when the original domain $\mathcal{X}$ under transformation is a tiny fraction of $\mathcal{X}_0$. Under this situation, which is common in nonlinear quantile regression models, the estimated curves are tended to be parallel with each other as evidence shown in Yang and Tokdar (2016, Figure 1(e)). Therefore, a general and comprehensive joint quantile model without serious restriction on both the parameter and predictor spaces is being pursued.

Recently, another general simultaneous estimation approach is proposed in Yang and Tokdar (2016) for multivariate case and based on an arbitrary convex hull of $\mathcal{X}$. They first move an interior point of the predictor space to $\mathbf{0}$, generate a convex hull of $\mathcal{X}$ containing $\mathbf{0}$ as an interior point and define a map on specific predictors domain $\mathcal{X}$ from vector b to $a(b, \mathcal{X})$ as

$$a(b, \mathcal{X}) = \begin{cases} \sup_{x \in \mathcal{X}} \{-x^T b\}/||b||, & ||b|| \neq 0 \\ ||x||_\infty, & ||b|| = 0 \end{cases}$$

where $||x||_\infty = \sup_{x \in \mathcal{X}} ||x||$. Therefore, this function serves as the supremum of the projection of any vector x on $\mathcal{X}$ onto the direction of b to avoid the difficulty caused by the arbitrary direction of b. Based on the aforementioned joint quantile regression model $Q_Y(\tau|X) = \beta_0(\tau) + x^T\beta(\tau)$, they specify the differentiable $\beta_0(\tau)$ and $\beta(\tau)$ as

$$\dot{\beta}_0(\tau) > 0, \qquad \dot{\beta}(\tau) = \dot{\beta}_0(\tau)\frac{v(\tau)}{a(v(\tau), \mathcal{X})\sqrt{1 + ||v(\tau)||^2}} \tag{2.1}$$

for any p-variate function $v(\tau)$ defined on $(0, 1)$ and prove (2.1) as an equivalent condition as $\dot{\beta}_0(\tau) + x^T\dot{\beta}(\tau) > 0$ for monotonicity on any $x \in \mathcal{X}$.

Selecting a candidate probability density function f_0 with the same support as the conditional density of the response (Gaussian distribution or t distribution on $(-\infty, \infty)$ in most cases), they define the corresponding cumulative distribution function as F_0 (where $\tau_0 = F_0(0)$), quantile function as Q_0 and quantile density function as q_0. With an arbitrary monotonical stretching or shrinking bijection $\zeta(\tau)$ from $(0, 1)$ to $(0, 1)$, it is appropriate to define

$$\beta_0(\tau) - \gamma_0 = \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} q_0(u)du, \qquad \tau \in (0, 1) \tag{2.2a}$$

$$\beta(\tau) - \gamma = \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{\omega(u)}{a(\omega(u), \mathcal{X})\sqrt{1 + ||\omega(u)||^2}} q_0(u)du, \qquad \tau \in (0, 1) \tag{2.2b}$$

where $\gamma_0 = \beta_0(\tau_0) \in \mathbb{R}$ and $\gamma = \beta(\tau_0) \in \mathbb{R}^p$ are arbitrary scalar and p-dimensional vector; $\sigma \in \mathbb{R}^+$ controls the scale of the quantile function; $\omega:(0, 1) \to \mathbb{R}^p$ represents an unrestricted p-variate function and $\zeta:[0, 1] \to [0, 1]$ denotes an arbitrary monotonically increasing bijection controlling the shape of the quantile function transformed from $Q_0(\tau)$. Therefore, they propose the equivalence among (β_0, β) (2.1), $(\gamma_0, \gamma, \sigma, \omega, \zeta)$ (2.2) and the monotonicity constraint $\dot{\beta}_0(\tau) + x^T\dot{\beta}(\tau) > 0$.

From the condition on ζ, it is intuitive to define it as a logistic transform through $\zeta(\tau) = \int_0^\tau e^{\omega_0(u)}du / \int_0^1 e^{\omega_0(u)}du$ and assign ω_0 a Gaussian process independent with other random elements (Lenk, 1988). The prior specifications are

$$\omega_j \sim \text{GP}(0, \kappa_j^2 c^{SE}(\cdot, \cdot|\lambda_j)), \qquad j = 0, 1, 2, \cdots, p$$

$$(\kappa_j^2, \lambda_j) \sim \text{IG}(3/2, 3/2) \times \pi_\lambda(\lambda_j^2), \qquad j = 0, 1, 2, \cdots, p$$
$$(\gamma_0, \gamma, \sigma^2) \sim \pi(\gamma_0, \gamma, \sigma^2) \propto 1/\sigma^2$$

where $c^{SE}(\tau, \tau'|\lambda_j)) = \exp(-\lambda^2(\tau - \tau')^2)$ represents the scaling squared exponential covariance function; See van der Vaart and van Zanten (2009) for more features and properties of this GP. The λ_j is defined through a Beta distribution Beta(a_λ, b_λ) on $\rho_h(\lambda_j) = \exp(-h^2\lambda_j^2)$ where $h = 0.01$, $a_\lambda = 6$ and $b_\lambda = 4$; See Yang and Tokdar (2016) for more details about these choices.

The derivation of the likelihood function is directly generalized from the univariate case within Tokdar et al. (2012). For the sampling method, they design a suitable procedure, which utilizes discretization and adaptive blocked MCMC algorithm. The consistency of the posterior distribution is guaranteed, reflecting that this approach fits both Frequentist and Bayesian tastes. This joint estimation quantile regression model is available in Rpackage as *qrjoint* (Tokdar, 2016).

3

Posterior Convergence Rate

A major difference between Frequentist and Bayesian perspectives is the viewpoint about true parameters. Frequentist world believes there exists a fixed truth and therefore always justifies Bayesian models through the asymptotic properties such as posterior consistency and posterior convergence rate (Ghosal et al., 2000). The consistency property is addressed in Yang and Tokdar (2016, Chapter 4) and here we conclude some preliminary results for pursuing the posterior convergence rate of this novel Bayesian joint quantile regression model.

3.1 Notation and Assumptions

We inherit most of the notations in Yang and Tokdar (2016) and represent the p-variate function with independent stationary GP equipped with squared exponential covariance function by $\omega(\tau) = (\omega_1(\tau), \omega_2(\tau), \cdots, \omega_p(\tau))$ with the truth as $\omega^*(\tau) = (\omega_1^*(\tau), \omega_2^*(\tau), \cdots, \omega_p^*(\tau))$, $\tau \in (0, 1)$. Therefore, the shaping bijection $\zeta(\tau)$

is controlled by ω_0 and has a truth as ζ^*. Then, the full model is described as

$$Q(\tau|x) = \gamma_0 + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} q_0(u)du + x^T\gamma + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{x^T\omega(u)}{a(\omega(u), \mathcal{X})\sqrt{1 + ||\omega(u)||^2}} q_0(u)du \tag{3.1a}$$

$$q(\tau|x) = \sigma q_0(\zeta(\tau))\dot{\zeta}(\tau)[1 + \frac{x^T\omega(\zeta(\tau))}{a(\omega(\zeta(\tau)), \mathcal{X})\sqrt{1 + ||\omega(\zeta(\tau))||^2}}], \quad \zeta(\tau) = \frac{\int_0^\tau e^{\omega_0(u)}du}{\int_0^1 e^{\omega_0(u)}du} \tag{3.1b}$$

Throughout this thesis, we use y^* to represent the truth for any random element y. For the l_2-norm of any multivariate function $\omega(\tau)$, we write it as $||\omega(\tau)||$ to indicate it as a function of τ and we denote the L_∞-norm through $||\omega||_\infty = \sup_{\tau\in(0,1)} ||\omega(\tau)||$ to indicate it as a constant for bounded ω. These representations are also applied to other vectors and functions. The notation $\lesssim$ denotes that the inequality is satisfied up to a multiplicative constant which is fixed throughout. We denote the Kullback-Leibler (KL) divergence as $d_{\mathrm{KL}}(f^*, f) = \int \log(f^*/f)dF^*$ and the "squared KL divergence" as $d_{\mathrm{L}^2}(f^*, f) = \int \{\log(f^*/f)\}^2 dF^*$.

Since the functional random element is the most crucial, we assume that scalers $\sigma = 1$, $\gamma_0 = \gamma_0^*$ and vector $\gamma = \gamma^*$, which would be handled in future research. For the base density f_0, we assume it has a standard logistic distribution, which has a lovely explicit quantile density function over any t distribution or Gaussian distribution. On the other hand, since logistic distribution maintains the support all over the real line $\mathbb{R}$, it also provides an "almost" perspective for the results of more general cases.

3.2 Density Estimation

From (3.1), it is easily noticed that this joint quantile regression model could be directly recognized as a density estimation approach through quantiles when using

$Q(\tau)$ and $q(\tau)$ for $x = 0$ as

$$Q(\tau) = \gamma_0 + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} q_0(u) du, \quad q(\tau) = \sigma q_0(\zeta(\tau))\dot{\zeta}(\tau), \quad \zeta(\tau) = \frac{\int_0^\tau e^{\omega_0(u)} du}{\int_0^1 e^{\omega_0(u)} du} \tag{3.2}$$

This density estimation model or intercept model maintains a nice convergence rate and neat proof procedure. The following lemma starts the whole proof and indicates some beautiful property within this model.

Lemma 1. *Given* $||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon$, *then we have* $||\log(\dot{\zeta}/\dot{\zeta}^*)||_\infty \leqslant 2\epsilon$, $||\log(\zeta/\zeta^*)||_\infty \leqslant 2\epsilon$, $||\log(\{1-\zeta\}/\{1-\zeta^*\})||_\infty \leqslant 2\epsilon$ *and* $||Q(\tau) - Q^*(\tau)||_\infty \leqslant 8\epsilon$.

Proof. Following *Theorem 4.1* in Tokdar and Ghosh (2007), we prove the first assertion, which is equivalent to that for any $\tau \in (0, 1)$, $\exp(-2\epsilon)\dot{\zeta}^*(\tau) \leqslant \dot{\zeta}(\tau) \leqslant \exp(2\epsilon)\dot{\zeta}^*(\tau)$. Through taking the integral from 0 to τ at two sides of the inequalities, the second statement is verified since ζ is monotonically increasing. The third assertion is straightforward according to the symmetry of $\zeta(\tau)$ and $\zeta^*(\tau)$. Since the base density $f_0 \sim$ Logistic$(0, 1)$, for any $\tau \in (0, 1)$,

$$\begin{aligned}
|Q(\tau) - Q^*(\tau)| &= |\log(\frac{\zeta(\tau)}{1-\zeta(\tau)}) - \log(\frac{\zeta(\tau_0)}{1-\zeta(\tau_0)}) - \log(\frac{\zeta^*(\tau)}{1-\zeta^*(\tau)}) + \log(\frac{\zeta^*(\tau_0)}{1-\zeta^*(\tau_0)})| \\
&\leqslant |\log(\frac{\zeta(\tau)}{\zeta^*(\tau)})| + |\log(\frac{\zeta(\tau_0)}{\zeta^*(\tau_0)})| + |\log(\frac{1-\zeta^*(\tau)}{1-\zeta(\tau)})| + |\log(\frac{1-\zeta^*(\tau_0)}{1-\zeta(\tau_0)})| \\
&\leqslant 8\epsilon
\end{aligned}$$

which is equivalent to the last assertion. □

Note that up to this point, we do not need any other information on ω_0 even the GP prior. Another feature worth to be pointed out is that these aforementioned formulation and restriction on the supremum distance between ω_0 and ω_0^* induce an amazing bound on the difference between Q and Q^* even though they diverge at two boundaries. With Lemma 1, we derive the supremum of $d_{\text{KL}}(f^*, f)$ as the following lemma with some extra condition on the truth, which is allowed and has no influence on the final assertion which is directly according to the probability measure of stationary Gaussian process.

Lemma 2. *Given* $||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon$*, then there exists some constant* C *only depends on the truth such that* $d_{KL}(f^*, f) \leqslant C\epsilon^2$ *and* $d_{L^2}(f^*, f) \leqslant C\epsilon^2$*.*

Actually, C only depending on the f^* and $\dot{\omega}_0^*$ suffices this lemma; see Appendix A.1 for more details. Then, we are well-prepared to provide the final theorem about the prior concentration condition of *Theorem 2.1* in Ghosal et al. (2000) of this density estimation method.

Theorem 3. *With* ω_0 *having a Gaussian process and all formulations in* (3.2)*, the prior concentration* $\Pi_n(f : d_{KL}(f^*, f) \leqslant \epsilon_n^2, d_{L^2}(f^*, f) \leqslant \epsilon_n^2) \geqslant \exp(-n\epsilon_n^2 C^*)$ *where* $\epsilon_n = n^{-\frac{1}{2+\beta}}(\log n)^t$ *for some* β *related to the smoothness of* ω_0^* *and* t *related to the prior distribution on the bandwidth parameter* λ*.*

Proof. Through Lemma 2, we calculate the prior concentration as

$$\Pi_n(f : d_{\mathrm{KL}}(f^*, f) \leqslant \epsilon_n^2, d_{\mathrm{L}^2}(f^*, f) \leqslant \epsilon_n^2) \geqslant \Pi(\omega_0 : ||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon_n^2) \geqslant \exp(-n\epsilon_n^2 C^*)$$

where $\epsilon_n = n^{-\frac{1}{2+\beta}}(\log n)^t$ according to van der Vaart and van Zanten (2009). □

Based on van der Vaart and van Zanten (2009), the RKHS of this GP, $\mathbb{H}$, consists of all the real functions ω_0 as

$$\omega_0(t) = \mathrm{Re}\{\int e^{-i\lambda t}\psi(\lambda)\mu(d\lambda)\}, \quad \text{where } \psi \in L^2(\mu), \text{ i.e.} \int |\psi|^2\mu(d\lambda) < \infty$$

According to Lemma 2, which only depends on the Lipschitz condition of F^* and ω_0^*, we prove that Lemma 2 is satisfied when the truth is changed to any random element through verifying the Lipschitz condition of any random F and ω_0, since

ω_0 is bounded within each open ball under the sieve construction. The Lipschitz condition on ω_0 is verified through definition as, for any closed enough t_1 and t_2,

$$|h(t_1) - h(t_2)| \leqslant \text{Re}\{\int |e^{-i\lambda(t_1-t_2)} - 1|\psi(\lambda)\mu(d\lambda)\} \leqslant K'_h|t_1 - t_2|$$

The Lipschitz condition on F is trivial since F is a continuous, smooth and monotonically increasing function from $\mathbb{R}$ to $[0, 1]$. Therefore we conclude that for any random element $\omega_0^\dagger$ and $\omega_0^\star$, if $||\omega_0^\dagger - \omega_0^\star||_\infty \leqslant \epsilon$, then $|| \log(f^\dagger/f^\star)||_\infty \leqslant C_0^*\epsilon$ and therefore $||f^\dagger - f^\star||_\infty \leqslant C^*\epsilon$ based on same proving procedure within Appendix A.1. Therefore, we prove that $\mathcal{K} : \mathbb{H} \to \mathbb{F}$ is an open map where $\mathbb{F}$ is the space of f induced by ω_0. Therefore, we conclude the following theorem to summarize the posterior convergence rate of the density estimation model.

Theorem 4. *With ω_0 having a Gaussian process and all formulations in* (3.2)*, there exists Borel measurable subsets A_n such that the entropy condition $P(f \notin A_n) \leqslant e^{-4n\epsilon_n^2}$ and $\log N(\bar{\epsilon}_n, A_n, ||\cdot||_\infty) \leqslant n\bar{\epsilon}_n^2$ where $\epsilon_n = n^{-\frac{1}{2+\beta}}(\log n)^t$ and $\bar{\epsilon}_n = K\epsilon_n(\log n)^{t^*}$ for some β related to the smoothness of ω_0^*, t and t^* related to the prior distribution on the bandwidth parameter λ and some sufficiently large constant K.*

3.3 Quantile Regression

When $x \in \mathcal{X}$ and $x \neq \mathbf{0}$ for some convex hull $\mathcal{X}$ containing $\mathbf{0}$ as an interior point, (3.1) could be approximately interpreted as an interaction model where the last term of $Q(\tau|x)$ represents the interaction of τ and x except that the computation for τ is an arbitrary monotonic bijection. This model specification indicates the gradual shape changing when τ goes from 0 to 1, which naturally reflects the the shrinkage property of the truth ω^* at both boundaries, which is $\omega^*(\tau) \to \mathbf{0}$ when $\tau \to 0$ or 1. With this tail condition and some revision on the proof procedure for density estimation model, we compute the prior concentration condition of this joint quantile regression model.

We first define another truncated model compared with (3.1) for some small δ as

$$\eta(\tau) = \rho^*(\tau)\omega(\tau), \qquad \rho^*(\tau) = \begin{cases} 1, & \tau \in (\delta, 1-\delta) \\ 0, & \text{otherwise} \end{cases}, \qquad \zeta(\tau) = \frac{\int_0^\tau e^{\omega_0(u)} du}{\int_0^1 e^{\omega_0(u)} du} \tag{3.3a}$$

$$Q_\eta(\tau|x) = \gamma_0 + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} q_0(u) du + x^T\gamma + \sigma \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1 + ||\eta(u)||^2}} q_0(u) du \tag{3.3b}$$

$$q_\eta(\tau|x) = \sigma q_0(\zeta(\tau))\dot{\zeta}(\tau)[1 + \frac{x^T \eta(\zeta(\tau))}{a(\eta(\zeta(\tau)), \mathcal{X})\sqrt{1 + ||\eta(\zeta(\tau))||^2}}] \tag{3.3c}$$

From (3.3), $\eta(\tau)$ is truncated to 0 at two ends. For the truth, we define $\eta^*(\tau) = \omega^*(\tau)$. According to the tail assumption, for any fixed ϵ, there exists a small enough δ such that $||\eta^*(\tau)||$ is monotonically decreasing towards the two ends and therefore $||\eta^*(\tau)|| \leqslant \epsilon$ when $\tau \in (0, \delta) \cup (1 - \delta, 1)$. This requirement indirectly refines the behavior of the truth ω^* and η^* around two boundaries. However, we will show later that δ could decrease at any polynomial rate, which will only influence the constant within the ultimate result. Based on the definition of η and η^*, we could know that η also has a stationary GP on $(\delta, 1-\delta)$ as $\eta_j \sim GP(0, \kappa_j^2 c^{SE}(*, \star|\lambda_j))$, $j = 1, 2, \cdots, p$ and $c^{SE}(\tau, \tau'|\lambda) = \exp(-\lambda^2(\tau - \tau')^2)$ and $||\eta - \eta^*||_\infty = ||\rho^*\omega - \rho^*\omega^*||_\infty \leqslant ||\omega - \omega^*||_\infty$. Therefore, we have prior concentration for η as $\Pi(||\eta - \eta^*||_\infty \leqslant \epsilon) \geqslant \Pi(||\omega - \omega^*||_\infty \leqslant \epsilon)$. Another feature of this construction is that $\lim_{\delta \to 0} \eta(\tau) = \omega(\tau)$, which provides an approximate proof of model (3.1). We begin the proof procedure with two straightforward but intriguing lemmas about prior concentration and prior probability measure Π.

Lemma 5. *Given $||\omega - \omega^*||_\infty \leqslant \kappa\epsilon$, then $||r - r^*||_\infty \leqslant \kappa\epsilon$ through the representations $r(\tau) = ||\eta(\tau)||$ and $r^*(\tau) = ||\eta^*(\tau)||$. And therefore, the prior concentration $\Pi(||r - r^*||_\infty \leqslant \kappa\epsilon) \geqslant \Pi(||\omega - \omega^*||_\infty \leqslant \kappa\epsilon)$ holds.*

Proof. Through aforementioned explanation, we have $||\eta - \eta^*||_\infty \leqslant ||\omega - \omega^*||_\infty$ and

$$||r - r^*||_\infty = ||(||\eta|| - ||\eta^*||)||_\infty \leqslant ||\eta - \eta^*||_\infty \leqslant ||\omega - \omega^*||_\infty \leqslant \kappa\epsilon$$

Therefore, we could conclude that $||r - r^*||_\infty \leqslant \kappa\epsilon$. The second result follows. □

Lemma 6. *Given* $||\omega - u^*||_\infty \leqslant \epsilon/4$ *for unit vector* $u^*(\tau)$ *where* $u^*(\tau) = \omega^*(\tau)/||\omega^*(\tau)||$, *when* $||\omega^*(\tau)|| \neq 0$, *then* $||u - u^*||_\infty \leqslant \epsilon$ *through the representation* $u(\tau) = \omega(\tau)/||\omega(\tau)||$. *Therefore, with* $u(\tau) = \eta(\tau)/||\eta(\tau)||$ *and* $u^*(\tau) = \eta^*(\tau)/||\eta^*(\tau)||$, *the prior concentration* $\Pi(||u - u^*||_\infty \leqslant \epsilon) \geqslant \Pi(||\omega - u^*||_\infty \leqslant \epsilon/4)$ *holds.*

Proof. Since $||\omega(\tau)|| \neq 0$ almost surely, we only require that condition for the truth $||\omega^*(\tau)|| \neq 0$. From $||\omega - u^*||_\infty \leqslant \epsilon/4$, for any τ, $|(||\omega(\tau)||) - 1| \leqslant \epsilon/4$ and therefore $||\omega(\tau)|| \geqslant 1 - \epsilon/4 \geqslant 1/2$ for small enough $\epsilon \leqslant 2$. Then, for any τ,

$$||u(\tau) - u^*(\tau)|| = \frac{||\omega(\tau) - u^*(||\omega(\tau)||)||}{||\omega(\tau)||} \leqslant 2||\omega(\tau) - u^*(\tau) + u^*(\tau) - u^*(\tau)(||\omega(\tau)||)||$$

$$\leqslant 2(||\omega - u^*||_\infty + ||u^*||_\infty[|(1 - ||\omega||_\infty)|]) \leqslant 2(\frac{\epsilon}{4} + \frac{\epsilon}{4}) \leqslant \epsilon$$

Therefore, we could conclude that $||u - u^*||_\infty \leqslant \epsilon$ and the second result follows. □

Assuming that u^* and ω_0^* both satisfies the Lipschitz condition, which does not influence the final probability measure on Gaussian process random element η or ω, the following powerful lemma is induced; See Appendix A.2 for detailed proof.

Lemma 7. *Given* $||r - r^*||_\infty \leqslant \kappa\epsilon$, $||u - u^*||_\infty \leqslant \epsilon$ *and* $||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon$ *where* $r(\tau) = ||\eta(\tau)||$, $r^*(\tau) = ||\eta^*(\tau)||$ *and* $u(\tau) = \eta(\tau)/||\eta(\tau)||$, $u^*(\tau) = \eta^*(\tau)/||\eta^*(\tau)||$, *then* $d_{KL}(f_\eta^*, f_\eta) \leqslant C(\log(1/\delta)\epsilon)^2$ *and* $d_{L^2}(f_\eta^*, f_\eta) \leqslant C(\log(1/\delta)\epsilon)^2$ *for some constant* C *only depending on the truth and the predictor convex hull* $\mathcal{X}$, *where* f_η *and* f_η^* *are defined through* (3.3).

Through Lemma 5 and Lemma 6, because of the independence between r and u as well as independence between r^* and u^* for the polar decomposition of Gaussian process, the prior concentration is $\Pi(||r - r^*||_\infty \leqslant \kappa\epsilon, ||u - u^*||_\infty \leqslant \epsilon) = \Pi(||r - r^*||_\infty \leqslant \kappa\epsilon) \times \Pi(||u - u^*||_\infty \leqslant \epsilon) \geqslant \Pi(||\omega - \omega^*||_\infty \leqslant \kappa\epsilon) \times \Pi(||\omega - u^*||_\infty \leqslant \epsilon/4)$. Plus Lemma 7, we could compute that $\Pi(d_{\mathrm{KL}}(f_\eta^*, f_\eta) \leqslant C(\log(1/\delta)\epsilon)^2, d_{\mathrm{L}^2}(f_\eta^*, f_\eta) \leqslant C(\log(1/\delta)\epsilon)^2) \geqslant \Pi(||r - r^*||_\infty \leqslant \kappa\epsilon) \times \Pi(||u - u^*||_\infty \leqslant \epsilon) \times \Pi(||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon) \geqslant \Pi(||\omega - \omega^*||_\infty \leqslant \kappa\epsilon) \times \Pi(||\omega - u^*||_\infty \leqslant \epsilon) \times \Pi(||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon)$. Then, the final theorem about the prior concentration condition of *Theorem 2.1* in Ghosal et al. (2000) with respect to model (3.3) is described as follows.

Theorem 8. *With ω_0 and ω having independent Gaussian processes and all formulations in* (3.3)*, the prior concentration $\Pi_n(f_\eta : d_{KL}(f_\eta^*, f_\eta) \leqslant \epsilon_n^2, d_{L^2}(f_\eta^*, f_\eta) \leqslant \epsilon_n^2) \geqslant \exp(-n\epsilon_n^2 C_1)$ where $\epsilon_n = kn^{-\frac{1}{2+\beta_\eta}}(\log n)^{t_\eta}$ when $\delta_n = 1/n^k$ for some β_η related to the smoothness of ω^* and t_η related to the prior distribution on bandwidth parameter λ.*

Proof. Assign the following quantities

$$\rho_n^*(\tau) = \begin{cases} 1, & \tau \in (\delta_n, 1 - \delta_n) \\ 0, & \text{otherwise} \end{cases}, \qquad \eta_n(\tau) = \rho_n^*(\tau)\omega(\tau)$$

and Π_n as the prior measure of Gaussian process corresponding with δ_n and $\rho_n^*(\tau)$. Let $\epsilon = \tilde{\epsilon}_n$ in Lemma 7, from *Theorem 1.3* in van der Vaart and van Zanten (2009), we know that $\tilde{\epsilon}_n = n^{-\frac{1}{2+1/\beta}}(\log(n))^t$ where β and t are some constant unrelated with n but related to the smoothness of the Gaussian process and the Gamma prior for the bandwidth λ. Therefore, supported by Lemma 7,

$$\begin{aligned} &\Pi_n(f_\eta : d_{\mathrm{KL}}(f_\eta^*, f_\eta) \leqslant (\log(\frac{1}{\delta_n})\tilde{\epsilon}_n)^2, d_{\mathrm{L}^2}(f_\eta^*, f_\eta) \leqslant (\log(\frac{1}{\delta_n})\tilde{\epsilon}_n)^2) \\ \geqslant &\Pi_n(||\omega - \omega^*||_\infty \leqslant \kappa\epsilon_n) \times \Pi_n(||\omega - u^*||_\infty \leqslant \epsilon_n) \times \Pi_n(||\omega_0 - \omega_0^*||_\infty \leqslant \epsilon_n) \geqslant \exp(-C_0 n\tilde{\epsilon}_n^2) \end{aligned}$$

Then we let $\epsilon_n = k\tilde{\epsilon}_n \log(n) \geqslant \tilde{\epsilon}_n$ for large n and we will get

$$\Pi_n(f_\eta : d_{\mathrm{KL}}(f_\eta^*, f_\eta) \leqslant \epsilon_n^2, d_{\mathrm{L}^2}(f_\eta^*, f_\eta) \leqslant \epsilon_n^2) \geqslant \exp(-C_0 n\tilde{\epsilon}_n^2) \geqslant \exp(-C_1 n\epsilon_n^2)$$

Therefore, we proved the prior concentration condition (condition 3) in Theorem 2.1 of Ghosal et al. (2000) with $\epsilon_n = kn^{-\frac{1}{2+1/\beta_\eta}}(\log(n))^{t_\eta}$ with same β and t as the aforementioned. $\beta_\eta = \beta$ and $t_\eta = t+1$ satisfies the result. $\square$

It is noticeable that the Lipschitz condition of u^*, ω^* and ω_0^* is required in Lemma 7 for the whole proof to go through. However, the Lipschitz condition on any random u is difficult to be verified. Therefore, at here, we require $\mathcal{X}$ to be a sphere. Without loss of generality, $\mathcal{X}$ is assumed to be the unit ball. Under this case, we only required the Lipschitz condition of ω and ω_0, which is true and maintains same proof procedure as the density estimation model. Therefore, we could conclude that

Theorem 9. *With $\mathcal{X}$ as any sphere centered at $\mathbf{0}$, ω_0 and ω having independent Gaussian processes and all formulations in (3.3), there exists Borel measurable subsets A_n^η such that the entropy condition $P(f_\eta \notin A_n^\eta) \leqslant e^{-4n\epsilon_n^2}$ and $\log N(\bar{\epsilon}_n, A_n^\eta, ||\cdot||_\infty) \leqslant n\bar{\epsilon}_n^2$ where $\epsilon_n = kn^{-\frac{1}{2+\beta_\eta}}(\log n)^{t_\eta}$ and $\bar{\epsilon}_n = K\epsilon_n(\log n)^{t_\eta^*}$ for some β_η related to the smoothness of ω^* and t_η and t_η^* related to the prior distribution on the bandwidth parameter λ and some sufficiently large constant K.*

Since $\mathcal{X}$ could be any irregular convex hull containing $\mathbf{0}$ rather than a sphere, we remain this generalization for further research.

4

Application: High School Math Achievement

Petscher and Logan (2014) introduces the application of quantile regression in the developmental science research through studying the association between math achievement and socio-economic status (SES) for high school students. High School and Beyond (HS&B) data (available within R package *nlme*) consists of 7185 observations with 11 variables from 160 schools. Among those variables, MathAch (math achievement), Minority (1=Minority) and SES (standardized socio-economic status) are individual-level variables and PRACAD (percentage of students on the academic track), HIMINTY (1=More than 40% minority students) and MEANSES (mean of SES) are school-level variables.

We analyze the casual relationship between MathAch and SES for both minority and non-minority students. In the initial exploratory data analysis, Figure 4.1 suggests that difference exists in SES and MathAch between two groups, especially along the middle range of SES. Some outliers are also found on the extreme SES values. Although quantile regression is robust for outliers in response, it cannot avoid the influence of outliers with erratic predictors. Therefore, we remove the two influential points with highest SES and lowest SES in minority group and the one observation with lowest SES in non-minority group. Furthermore, the correlation of SES and Minority is -0.27, which implies that the minority group tends to acquire a lower SES. This generates a confounding problem and we utilize a 5-degree basic spline (B-spline) model on SES for the adjustment. To address the relationship

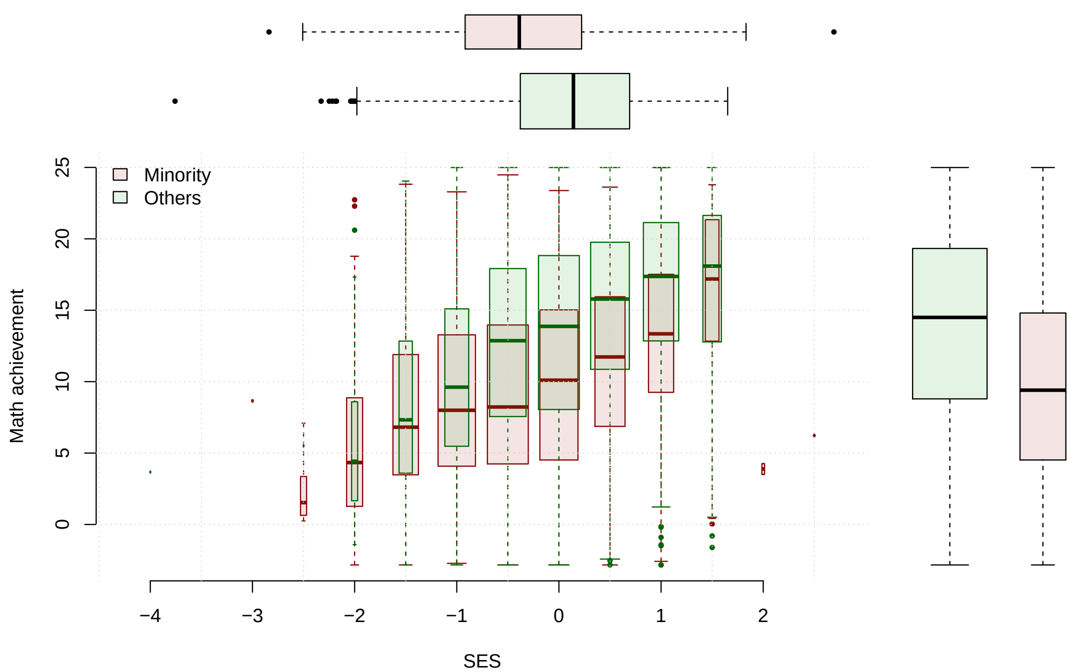

FIGURE 4.1: Boxplot of SES and MathAch of two groups

between SES and math achievement (MathAch), this analysis implements separate models for minority students and non-minority students and compares eight models with every combination of the aforementioned three school-level predictors through Watanabe-Akaike information criterion (WAIC). We fit these eight separate models with B-spline transformed SES and other components linearly for 200,000 posterior samples and thinning for every 10 samples. Table 4.1 lists their WAICs, including all combinations of school-level covariates, which indicates that the full model with bs(SES), PRACAD, HIMINTY and MEANSES leads to the lowest WAIC.

Table 4.1: WAIC of each model

Involved variables	Minority	Non-Minority	Total
bs(SES), PRACAD, HIMINTY and MEANSES	12509.67	32982.55	45492.22
bs(SES), PRACAD and HIMINTY	12490.88	33033.96	45524.84
bs(SES), PRACAD and MEANSES	12521.46	33104.00	45625.46
bs(SES), HIMINTY and MEANSES	12593.35	33142.21	45735.56
bs(SES) and PRACAD	12542.22	33053.73	45595.95
bs(SES) and HIMINTY	12643.21	33179.55	45822.76
bs(SES) and MEANSES	12610.57	33185.88	45796.45
bs(SES)	12663.16	33158.81	45821.97

One more revision we implement on the predictors is replacing SES with relative SES (RSES) by subtracting the MEANSES from SES for each observation; see Lazar and Zerbe (2011) and Raudenbush and Bryk (2002) for similar use in hierarchical modeling. This modification expresses the single influence of individual SES on MathAch without respect to the school-wise SES. Therefore, we fit separate models for minority and non-minority groups with 400,000 samples and thinning every 10 iteration to pursue the lower auto-correlation and better performance in the convergence diagnosis. Due to the difficulty for interpreting the coefficients of spline components, we extract the estimate (median) of $\beta \times$ bs(RSES), the inner product of the spline component coefficients and the spline transform of x, at different quantiles $\tau \in (0, 1)$. In both Figure 4.2 and 4.3, the left set of four graphs is the coefficient for each predictor and its 95% credible interval; The right plot is the prediction through RSES on math achievement along quantiles under all school-level variables equal to 0. From Figure 4.2 and Figure 4.3, the RSES dominates on mathematical achievement for both groups and such domination is more significant for the non-minority group. For minority group, PRACAD maintains a positive effect on students' math achievement and this effect goes up to the highest at the median range and down to the lowest at both tails; HIMINTY also displays positive impact and but shrinks to

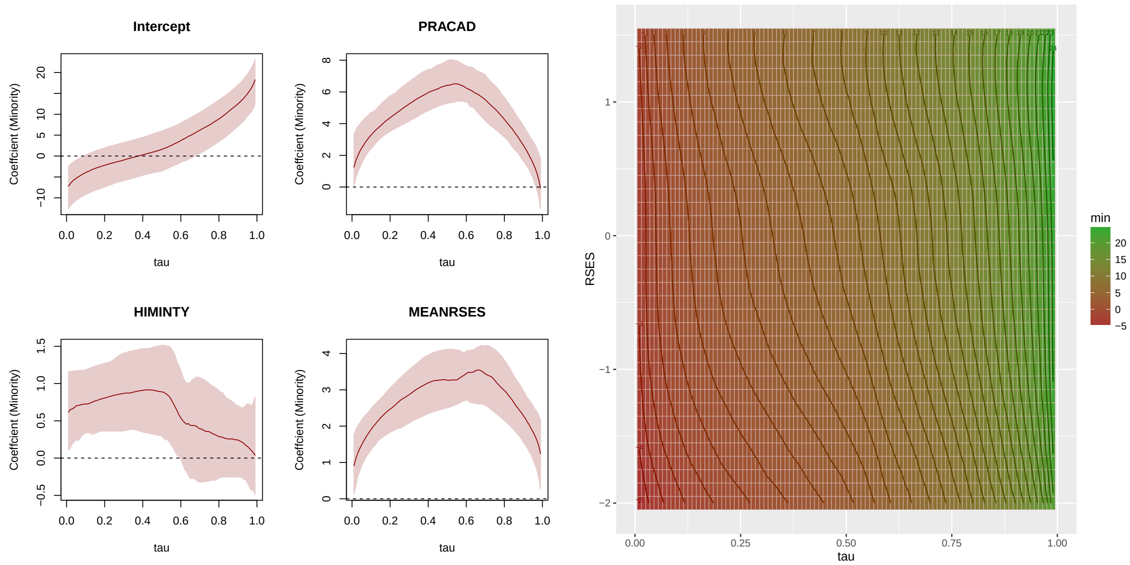

FIGURE 4.2: Coefficients curves of minority group

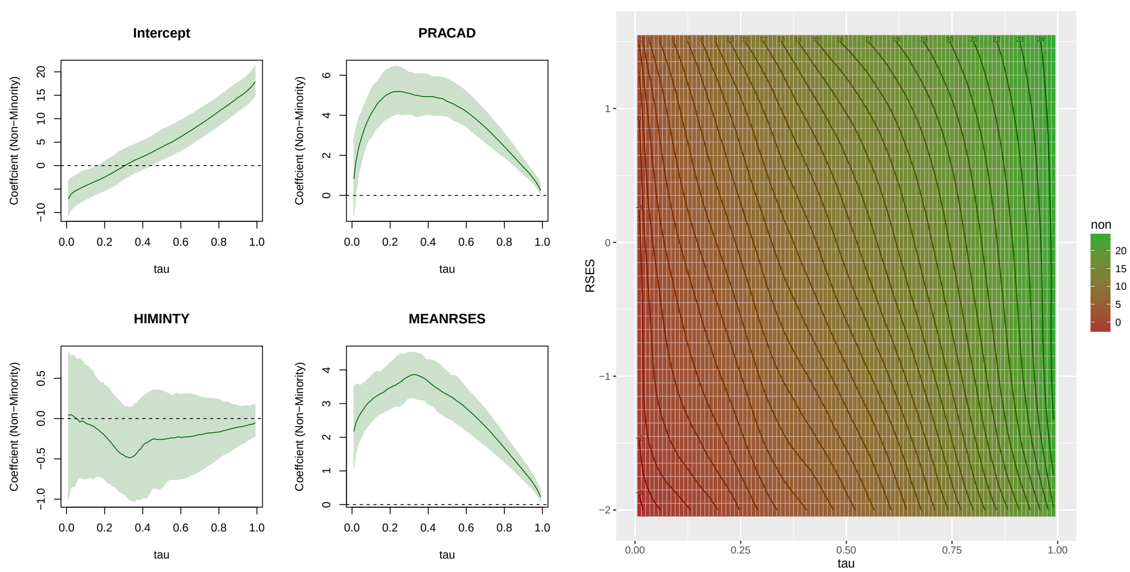

FIGURE 4.3: Coefficients curves of non-minority group

0 as τ increases; MEANSES sustains positive influence and contains the two models around the median. And for non-minority group, PRACAD also maintains a positive effect and this effect reaches maximum at approximately $\tau = 0.2$; HIMINTY has slightly negative impact and shrinks to 0 at right tail; MEANSES maintains positive influence and hits the peak round $\tau = 0.35$.

Since we are always intrigued by the difference between minority students and non-minority students on math achievement, the coefficients difference between two groups are also displayed in Figure 4.4. The left four plots presents the coefficients difference from the non-minority group to the minority group along quantiles τ except RSES. The right plot displays the contour plot narrowed RSES to $(-2, 1.5)$ for the median difference (difference estimator) of posterior predicted math achievement (non-minority $-$ minority) versus RSES and τ including intercept with PRACAD, HIMINTY and MEANSES equal to 0.

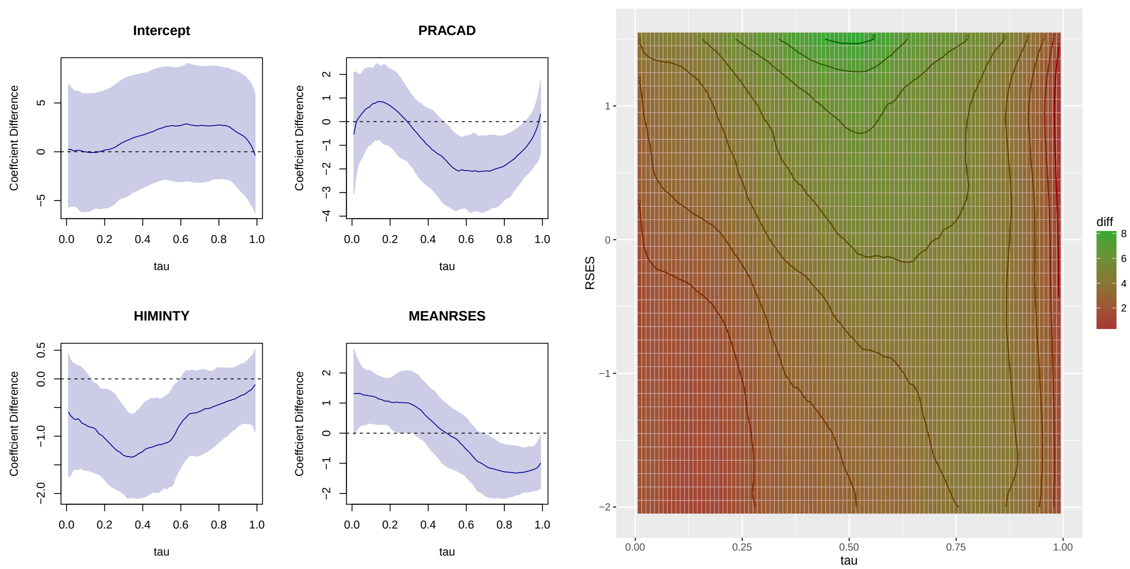

FIGURE 4.4: Difference of coefficients

Considering the influence by school-level variables, the PRACAD maintains a positive effect on the gap at the lower tail and negative effect from $\tau = 0.3$. Although the coefficients of PRACAD is quite considerable, it does not influence the gap significantly because of the lower scale of PRACAD (percentage). The MEANSES attains positive impact before hitting the median and sustains negative since then till the upper boundary. Unlike the PRACAD and MEANSES, the HIMINTY retains negative influence on the gap along all quantiles and is shrunk to 0 at the upper tail.

5

Discussion

This novel joint quantile regression model provides a more comprehensive understanding of the whole casual relationship structure and generalizes the quantile regression from an exploratory data analysis to broader theories and applications. Although the MCMC scheme requires quantities of sampling steps to converge to the stationary distribution and will be slower for more covariates case, this problem could potentially be improved by some parallel and GPU computing techniques.

From the convergence rate proof procedure, it is worth to notice that the whole proof is correspondingly parallel with van der Vaart and van Zanten (2009) and therefore we conjecture that the first two entropy conditions of *Theorem 2.1* in Ghosal et al. (2000) on any convex hull $\mathcal{X}$ should also maintain the similar proof procedure and results. Furthermore, all other random scalars involving the priors on them will be addressed in future and become more tractable based on current work. Generalization from the logistic distribution to any t distribution or Gaussian distribution does not intrigue us so much at this stage, which might potentially be achieved through polynomial approximation with $1/\tau^k$ and $1/(1-\tau)^k$ around two boundaries. Similarly, recovery of the original ω model could be explained as the limiting case of the η model with polynomial rates for any degree and therefore it also will not attract much attention and development.

The application on the math achievement is preliminary and could be improved with more casual inference techniques suited with this novel quantile regression model. However, according to the current results, we provide some recommendations on selecting school for minority students. Transferring to a school with more students on the academic track is beneficial to their math achievement, especially for the median achievers. The schools with high proportion of minority students could also upgrade the performance of minority students in math, where they would feel more engaged and obtain more opportunity for receiving assistance from others. A school with higher MEANSES could also improve the students' math achievement, which is also corresponding to lift the SES level of a family for increasing the RSES. This is beneficial for children's mathematical education in a long run, which is displayed by both estimates of MEANSES and RSES.

Appendix A

Proof Details

A.1 Proof of Lemma 2

Proof. With the direct expression of Q through Logistic distribution, we derive that $F(y) = \zeta^{-1}(1/\{1 + \exp(-(y + \log(\zeta(\tau_0)/(1 - \zeta(\tau_0)))))\})$ for any $y \in \mathbb{R}$. Then, we explore the supremum of $|\log(f^*(y)/f(y))|$, which is equivalent to the supremum of $|\log(f^*(Q^*(\tau))/f(Q^*(\tau)))|$ for any $\tau \in (0,1)$ since the support of F^* is $\mathbb{R}$.

For any $\tau \in (0,1)$ and correspondingly $y = Q^*(\tau) \in \mathbb{R}$,

$$
\begin{aligned}
|\log(\frac{f^*(Q^*(\tau))}{f(Q^*(\tau))})| &= |\log(\frac{q(F(Q^*(\tau)))}{q^*(\tau)})| \leqslant |\log(\frac{q(\tau)}{q^*(\tau)})| + |\log(\frac{q(F(Q^*(\tau)))}{q(\tau)})| \\
&= \alpha(\tau) + \beta(\tau)
\end{aligned}
$$

The supremums of $\alpha(\tau)$ and $\beta(\tau)$ are computed separately. Through Lemma 1, $\alpha(\tau)$ is controlled by 6ϵ as

$$
\alpha(\tau) = |\log(\frac{q_0(\zeta(\tau))}{q_0(\zeta^*(\tau))}) + \log(\frac{\dot{\zeta}(\tau)}{\dot{\zeta}^*(\tau)})| \leqslant |\log(\frac{\zeta^*(\tau)(1-\zeta^*(\tau))}{\zeta(\tau)(1-\zeta(\tau))})| + |\log(\frac{\dot{\zeta}(\tau)}{\dot{\zeta}^*(\tau)})| \tag{A.1a}
$$

$$
\leqslant |\log(\frac{\zeta^*(\tau)}{\zeta(\tau)})| + |\log(\frac{1-\zeta^*(\tau)}{1-\zeta(\tau)})| + |\log(\frac{\dot{\zeta}(\tau)}{\dot{\zeta}^*(\tau)})| \leqslant 2\epsilon + 2\epsilon + 2\epsilon = 6\epsilon \tag{A.1b}
$$

For $\beta(\tau)$,

$$\beta(\tau) = |\log(\frac{q(F(Q^*(\tau)))}{q(\tau)})| = |\log(\frac{q_0(\zeta(F(Q^*(\tau))))\dot{\zeta}(F(Q^*(\tau)))}{q_0(\zeta(\tau))\dot{\zeta}(\tau)})|$$

$$\leqslant |\log(\frac{q_0(\zeta(F(Q^*(\tau))))}{q_0(\zeta(\tau))})| + |\log(\frac{\dot{\zeta}(F(Q^*(\tau)))}{\dot{\zeta}(\tau)})| = \beta_1(\tau) + \beta_2(\tau)$$

Then we calculate the supremums for $\beta_1(\tau)$ and $\beta_2(\tau)$ separately as

$$\beta_1(\tau) = |\log(\frac{q_0(\zeta(F(Q^*(\tau))))}{q_0(\zeta(F(Q(\tau))))})| = |\log(\frac{q_0(\frac{1}{1+\exp(-(Q^*(\tau)+\log(\zeta(\tau_0)/(1-\zeta(\tau_0)))))})}{q_0(\frac{1}{1+\exp(-(Q(\tau)+\log(\zeta(\tau_0)/(1-\zeta(\tau_0)))))})})|$$

$$= |\log(\frac{\frac{(1+\exp(-(Q^*(\tau)+\log(\zeta(\tau_0)/(1-\zeta(\tau_0))))))^2}{\exp(-(Q^*(\tau)+\log(\zeta(\tau_0)/(1-\zeta(\tau_0)))))}}{\frac{(1+\exp(-(Q(\tau)+\log(\zeta(\tau_0)/(1-\zeta(\tau_0))))))^2}{\exp(-(Q(\tau)+\log(\zeta(\tau_0)/(1-\zeta(\tau_0)))))}})|$$

$$\leqslant |Q^*(\tau) - Q(\tau)| + 2|\log(\frac{1 + \frac{1-\zeta(\tau_0)}{\zeta(\tau_0)}\exp(-Q^*(\tau))}{1 + \frac{1-\zeta(\tau_0)}{\zeta(\tau_0)}\exp(-Q(\tau))})|$$

$$\leqslant 8\epsilon + 2|\log(\frac{\exp(-Q^*(\tau))}{\exp(-Q(\tau))})| \leqslant 8\epsilon + 2 \times 8\epsilon \leqslant 24\epsilon$$

Before computing the supremum of $\beta_2(\tau)$, we derive a Lipschitz condition on F, which should be satisfied by intuition. Since F^* is the true cumulative distribution function and bounded on $[0, 1]$, there must exist a $K_F'^*$ such that for any y_1 and y_2, $|F^*(y_1) - F^*(y_2)| \leqslant K_F'^*|y_1 - y_2|$. Here $K_F'^*$ could be the supremum of true density function f^* on $[y_1, y_2]$ if $y_1 < y_2$ or on $[y_2, y_1]$ if $y_1 > y_2$. By replacing $y_1 = Q^*(\tau_1)$ and $y_2 = Q^*(\tau_2)$, we could derive that $|\tau_1 - \tau_2| \leqslant K_F'^*|Q^*(\tau_1) - Q^*(\tau_2)| \leqslant K_F'^*|Q^*(\tau_1) - Q(\tau_1)| + K_F'^*|Q(\tau_1) - Q(\tau_2)| + K_F'^*|Q(\tau_2) - Q^*(\tau_2)| \leqslant 16K_F'^*\epsilon + K_F'^*|Q(\tau_1) - Q(\tau_2)|$. Therefore, since y_1 and y_2 at the beginning are arbitrary, we could fix $\tau_1 = F(Q(\tau))$ and $\tau_2 = F(Q^*(\tau))$ and acquire a Lipschitz condition of F as $|F(Q(\tau)) - F(Q^*(\tau))| \leqslant 16K_F'^*\epsilon + K_F'^*|Q^*(\tau) - Q(\tau)| \leqslant 24K_F'^*\epsilon$ through Lemma 1.

Besides, assuming the truth ω_0^* is smooth and satisfies the Lipschitz condition, we could conclude the Lipschitz condition on $\log(\dot{\zeta}^*)$ as that for any τ_1 and τ_2,

$|\log(\dot{\zeta}^*(\tau_1)) - \log(\dot{\zeta}^*(\tau_2))| = |\omega_0^*(\tau_2) - \omega_0^*(\tau_1)| \leqslant K_\omega'^*|\tau_1 - \tau_2|$, where $K_\omega'^*$ is the supremum of the absolute value of $\dot{\omega}_0^*$. Then we could derive the upper bound of $\beta_2(\tau)$.

$$\begin{aligned}\beta_2(\tau) &= \log(\frac{\dot{\zeta}(F(Q^*(\tau)))}{\dot{\zeta}^*(F(Q^*(\tau)))}) + \log(\frac{\dot{\zeta}^*(F(Q^*(\tau)))}{\dot{\zeta}^*(F(Q(\tau)))}) + \log(\frac{\dot{\zeta}^*(F(Q(\tau)))}{\dot{\zeta}(F(Q(\tau)))}) \\ &\leqslant 2\epsilon + 24K_F'^*K_\omega'^*\epsilon + 2\epsilon \leqslant (24K_F'^*K_\omega'^* + 4)\epsilon\end{aligned}$$

Overall, we conclude that $|\log(f^*(y)/f(y))| \leqslant (24K_F'^*K_\omega'^* + 34)\epsilon$ for any $y = Q^*(\tau) \in \mathbb{R}, \tau \in [0,1]$ and therefore $||\log(f^*/f)||_\infty \leqslant C_0\epsilon$, where $C_0 = 24K_F'^*K_\omega'^* + 34$ only depends on the f^* and $\dot{\omega}_0^*$.

Rewrite $f = \exp(\log(f))/\int_{\mathbb{R}}\exp(\log(f))dy$ and $f^* = \exp(\log(f^*))/\int_{\mathbb{R}}\exp(\log(f^*))dy$. Following the second assertion of *Lemma 3.1* in van der Vaart and van Zanten (2008), since $\log(f)$ and $\log(f^*)$ are both measurable functions, we have

$$\begin{aligned}d_{\mathrm{KL}}(f^*, f) &= \int_y f^*(y)\log(\frac{f^*(y)}{f(y)})dy \\ &\leqslant C_1||\log(f^*(y)/f(y))||_\infty^2 e^{||\log(f^*(y)/f(y))||_\infty}(1 + ||\log(f^*(y)/f(y))||_\infty) \\ &\leqslant C_1C_0^2\epsilon^2 e^{C_0\epsilon}(1 + C_0\epsilon) \leqslant C\epsilon^2\end{aligned}$$

where C_1 is referred to van der Vaart and van Zanten (2008) and irrelevant with the model. Also, since $||\log(f^*/f)||_\infty \leqslant C_0\epsilon$, it is trivial to show that $d_{\mathrm{L}^2}(f^*, f) \leqslant d_{\mathrm{KL}}(f^*, f) \leqslant C\epsilon^2$, where C only depends on the true f^* and $\dot{\omega}_0^*$. □

A.2 Proof of Lemma 7

Proof. First, we have $||\eta - \eta^*||_\infty \leqslant ||ru - ru^*||_\infty + ||ru^* - r^*u^*||_\infty \leqslant (||r^*||_\infty + \kappa)\epsilon + ||u^*||_\infty\kappa\epsilon \leqslant \kappa_\eta^*\epsilon$ for some constant κ_η^* and correspondingly, $\Pi(||\eta - \eta^*||_\infty \leqslant \kappa_\eta^*\epsilon) \geqslant \Pi(||r - r^*||_\infty \leqslant \kappa\epsilon, ||u - u^*||_\infty \leqslant \epsilon)$.

Denoting the supremum of any conditional function g with respective to some random variable and x as $||g(\cdot|x)||_{\infty,\infty} := \sup_{x\in\mathcal{X}}||g(\cdot|x)||_\infty$, then specifically, we have

$$||\log(\frac{f_\eta^*(\cdot|x)}{f_\eta(\cdot|x)})||_{\infty,\infty} := \sup_{x\in\mathcal{X}}||\log(\frac{f_\eta^*(\cdot|x)}{f_\eta(\cdot|x)})||_\infty, \quad ||\log(\frac{q_\eta^*(\cdot|x)}{q_\eta(\cdot|x)})||_{\infty,\infty} := \sup_{x\in\mathcal{X}}||\log(\frac{q_\eta^*(\cdot|x)}{q_\eta(\cdot|x)})||_\infty$$

Similar as the density estimation model, we first compute the upper bound of $||\log(f^*_\eta(\cdot|x)/f_\eta(\cdot|x))||_{\infty,\infty}$ as

$$||\log(\frac{f^*_\eta(\cdot|x)}{f_\eta(\cdot|x)})||_{\infty,\infty} = ||\log(\frac{f^*_\eta(Q^*_\eta(\cdot|x)|x)}{f_\eta(Q^*_\eta(\cdot|x)|x)})||_{\infty,\infty} = ||\log(\frac{q_\eta(F_\eta(Q^*_\eta(\cdot|x)|x)|x)}{q^*_\eta(\cdot|x)})||_{\infty,\infty}$$

$$\leqslant ||\log(\frac{q_\eta(\cdot|x)}{q^*_\eta(\cdot|x)})||_{\infty,\infty} + ||\log(\frac{q_\eta(F_\eta(Q^*_\eta(\cdot|x)|x)|x)}{q_\eta(\cdot|x)})||_{\infty,\infty}$$

$$= ||\alpha(\cdot|x)||_{\infty,\infty} + ||\beta(\cdot|x)||_{\infty,\infty}$$

For any $\tau \in (0,1)$ and $x \in \mathcal{X}$, we have

$$\alpha(\tau|x) = \log(\frac{q_\eta(\tau|x)}{q^*_\eta(\tau|x)}) = \log(\frac{q_0(\zeta(\tau))\dot{\zeta}(\tau)[1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}]}{q_0(\zeta^*(\tau))\dot{\zeta}^*(\tau)[1+\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}]})$$

$$\leqslant ||\log(\frac{q_0(\zeta)\dot{\zeta}}{q_0(\zeta^*)\dot{\zeta}^*})||_\infty + \log(\frac{1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}}{1+\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}}) = \alpha_1 + \alpha_2(\tau|x)$$

From (A.1), $\alpha_1 \leqslant 6\epsilon$. Before exploring the upper bound of $\alpha_2(\tau|x)$, several useful inequalities need to be addressed. First, for any j and small enough ϵ, there exists C^*_η such that $||\eta(\zeta) - \eta^*(\zeta^*)||_\infty \leqslant ||\eta(\zeta) - \eta^*(\zeta) + \eta^*(\zeta) - \eta^*(\zeta^*)||_\infty \leqslant ||\eta(\zeta) - \eta^*(\zeta)||_\infty + ||\eta^*(\zeta) - \eta^*(\zeta^*)||_\infty \leqslant \kappa\epsilon + K'^*_\eta||\zeta - \zeta^*||_\infty \leqslant \kappa\epsilon + K'^*_\eta(e^{2\epsilon}-1)||\zeta^*||_\infty \leqslant \kappa\epsilon + K'^*_\eta(e^{2\epsilon}-1) \leqslant C^*_\eta\epsilon$, where $K'^*_\eta = \sup_{j\in\{1,2,\cdots,p\}}\{\sup_{\tau\in(0,1)}\dot{\eta}^*_j\}$ and $C^*_\eta = \kappa + 2K'^*_\eta$. And therefore, $||\eta(\zeta)||_\infty \leqslant C^*_\eta\epsilon + ||\eta^*(\zeta^*)||_\infty$. Writing diam($\mathcal{X}$) as $||x||_\infty$, since $a(b,\mathcal{X}) = \begin{cases}\sup_{x\in\mathcal{X}} x^Tb/||b||, & \text{if } ||b|| \neq 0\\ ||x||_\infty, & \text{if } ||b|| = 0\end{cases}$ and $\mathcal{X}$ is bounded with $\mathbf{0}$ as an interior point, there must exist some fixed small number $r_\mathcal{X}$ such that $0 < r_\mathcal{X} \leqslant a(b,\mathcal{X}) \leqslant ||x||_\infty < \infty$.

When $\tau \in \{\tau : ||\eta^*(\zeta^*(\tau))|| = 0\}$, then

$$\alpha_2(\tau|x) = \log(1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}) \leqslant |\log(1-\frac{||x||_\infty C^*_\eta\epsilon}{r_\mathcal{X}})| \leqslant M^0_2\epsilon$$

where M^0_2 is some positive constant only depending on $\mathcal{X}$ and η^*.

When $\tau \in \{\tau : ||\eta^*(\zeta^*(\tau))|| \neq 0\}$, then

$$||u(\zeta) - u^*(\zeta^*)||_\infty \leqslant ||u(\zeta) - u^*(\zeta)||_\infty + ||u^*(\zeta) - u^*(\zeta^*)||_\infty \leqslant \epsilon + K_u'^*\epsilon \leqslant C_u^*\epsilon$$

where $K_u'^* = \sup_{j\in\{1,2,\cdots,p\}}\{\sup_{\tau\in(0,1)} \dot{u}_j^*\}$ and $C_u^* = K_u'^* + 1$. Especially, using these representations,

$$s(\tau) = a(\eta(\zeta(\tau)), \mathcal{X}) = \sup_{x\in\mathcal{X}} x^T u(\zeta(\tau)), \quad s^*(\tau) = a(\eta^*(\zeta^*(\tau)), \mathcal{X}) = \sup_{x\in\mathcal{X}} x^T u^*(\zeta^*(\tau))$$

$$t(\tau) = \frac{x^T\eta(\zeta(\tau))}{||\eta(\zeta(\tau))||} = x^T u(\zeta(\tau)), \quad t^*(\tau) = \frac{x^T\eta^*(\zeta^*(\tau))}{||\eta^*(\zeta^*(\tau))||} = x^T u^*(\zeta^*(\tau))$$

$$v(\tau) = \frac{||\eta(\zeta(\tau))||}{\sqrt{1 + ||\eta(\zeta(\tau))||^2}}, \quad v^*(\tau) = \frac{||\eta^*(\zeta^*(\tau))||}{\sqrt{1 + ||\eta^*(\zeta^*(\tau))||^2}}$$

we have the following results

$$\begin{aligned}
|s(\tau) - s^*(\tau)| &= |\sup_{x\in\mathcal{X}} x^T u(\zeta(\tau)) - \sup_{x\in\mathcal{X}} x^T u^*(\zeta^*(\tau))| \leqslant \sup_{x\in\mathcal{X}} |x^T(u(\zeta(\tau)) - u^*(\zeta^*(\tau)))| \\
&\leqslant ||x||_\infty C_u^*\epsilon \leqslant \kappa_s^*\epsilon \\
|t(\tau) - t^*(\tau)| &= |x^T(u(\zeta(\tau)) - u^*(\zeta^*(\tau)))| \leqslant \sup_{x\in\mathcal{X}} |x^T(u(\zeta(\tau)) - u^*(\zeta^*(\tau)))| \\
&\leqslant ||x||_\infty C_u^*\epsilon \leqslant \kappa_t^*\epsilon \\
|v(\tau) - v^*(\tau)| &\leqslant |(||\eta(\zeta(\tau))||\sqrt{1 + ||\eta^*(\zeta^*(\tau))||^2} - ||\eta^*(\zeta^*(\tau))||\sqrt{1 + ||\eta(\zeta(\tau))||^2})| \\
&\leqslant |(||\eta(\zeta(\tau))||\sqrt{1 + ||\eta^*(\zeta^*(\tau))||^2} - ||\eta^*(\zeta^*(\tau))||\sqrt{1 + ||\eta^*(\zeta^*(\tau))||^2})| \\
&+ |(||\eta^*(\zeta^*(\tau))||\sqrt{1 + ||\eta^*(\zeta^*(\tau))||^2} - ||\eta^*(\zeta^*(\tau))||\sqrt{1 + ||\eta(\zeta(\tau))||^2})| \\
&\leqslant (\sqrt{1 + ||\eta^*(\zeta^*(\tau))||_\infty^2} + ||\eta^*(\zeta^*(\tau))||_\infty)|(||\eta(\zeta(\tau))|| - ||\eta^*(\zeta^*(\tau))||)| \\
&\leqslant (\sqrt{1 + K_\eta^{*2}} + K_\eta^*)C_\eta^*\epsilon \leqslant \kappa_v^*\epsilon
\end{aligned}$$

Based on the above inequalities and the fact that all s, t, v are bounded above, we have

$$||\frac{t(\tau)v(\tau)}{s(\tau)} - \frac{t^*(\tau)v^*(\tau)}{s^*(\tau)}||_\infty \leqslant \frac{1}{r_\mathcal{X}^2}||t(\tau)v(\tau)s^*(\tau) - t^*(\tau)v^*(\tau)s(\tau)|| \leqslant M_2^*\epsilon$$

Then, we are well-prepared to prove the upper bound for $\alpha_2(\tau|x)$. Since the distance between $\frac{t(\tau)v(\tau)}{s(\tau)}$ and $\frac{t^*(\tau)v^*(\tau)}{s^*(\tau)}$ is fairly small and $||\eta||$ and $||\eta^*||$ are bounded above, we have the upper bound for α_2 as

$$\alpha_2(\tau|x) = \log(\frac{1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}}{1+\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}})$$

$$\leqslant \log(1+\frac{\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}-\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}}{1+\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}})$$

$$\leqslant M_2' \times \frac{\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}-\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}}{1+\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}}$$

$$\leqslant M_2'M_2^*\epsilon \times \frac{||\eta^*(\zeta^*(\tau))||_\infty}{\sqrt{1+||\eta^*(\zeta^*(\tau))||_\infty^2}-||\eta^*(\zeta^*(\tau))||_\infty}$$

$$\leqslant M_2\epsilon$$

Therefore, the supremum of α is addressed as

$$||\alpha(\cdot|x)||_{\infty,\infty} \leqslant \alpha_1 + ||\alpha_2(\cdot|x)||_{\infty,\infty} \leqslant 6\epsilon + \max(M_2^0, M_2)\epsilon = M_\alpha\epsilon$$

For $\beta(\tau|x)$, the proof procedure is similar as the α's. At here, to simplify the expression, we use $F_\eta(Q_\eta^*)$ to represent $F_\eta(Q_\eta^*(\tau|x)|x)$. Therefore,

$$\beta(\tau|x) = \log(\frac{q_0(\zeta(F_\eta(Q_\eta^*)))\dot{\zeta}(F_\eta(Q_\eta^*))[1+\frac{x^T\eta(\zeta(F_\eta(Q_\eta^*)))}{a(\eta(\zeta(F_\eta(Q_\eta^*))),\mathcal{X})\sqrt{1+||\eta(\zeta(F_\eta(Q_\eta^*)))||^2}}]}{q_0(\zeta(\tau))\dot{\zeta}(\tau)[1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}]})$$

$$\leqslant || \log(\frac{\dot{\zeta}(F_\eta(Q_\eta^*))}{\dot{\zeta}(\tau)})||_{\infty,\infty} + || \log(\frac{q_0(\zeta(F_\eta(Q_\eta^*)))}{q_0(\zeta(\tau))})||_{\infty,\infty} + \log(\frac{1+\frac{x^T\eta(\zeta(F_\eta(Q_\eta^*)))}{a(\eta(\zeta(F_\eta(Q_\eta^*))),\mathcal{X})\sqrt{1+||\eta(\zeta(F_\eta(Q_\eta^*)))||^2}}}{1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}})$$

$$=\beta_1 + \beta_2 + \beta_3(\tau|x)$$

Before proving any of the aforementioned terms, several useful results need to be established. Using $Q^*(\tau)$ and $Q(\tau)$ representing the quantile functions of intercept models as before, when $\tau \geqslant \delta$, by definition,

$$
\begin{aligned}
&|Q^*_\eta(\tau|x) - Q_\eta(\tau|x)| \\
\leqslant &| \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta^*(u)}{a(\eta^*(u), \mathcal{X})\sqrt{1+||\eta^*(u)||^2}} q_0(u) du - \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du| + 8\epsilon \\
\leqslant &| \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta^*(u)}{a(\eta^*(u), \mathcal{X})\sqrt{1+||\eta^*(u)||^2}} q_0(u) du - \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du| \\
+ &| \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du - \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du| + 8\epsilon \\
\leqslant &|| \frac{x^T \eta^*(u)}{a(\eta^*(u), \mathcal{X})\sqrt{1+||\eta^*(u)||^2}} - \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} ||_{\infty,\infty} | \int_{\zeta^*(\tau_0)}^{\zeta^*(\delta)} q_0(u) du| \\
+ &|| \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} ||_{\infty,\infty} ||Q^*(\tau) - Q(\tau)||_\infty + 8\epsilon \\
\leqslant &\log(\frac{\zeta^*(\tau_0)(1-\zeta^*(\delta))}{(1-\zeta^*(\tau_0))\zeta^*(\delta)}) M_2^* \epsilon + || \frac{x^T(\eta(u) - \eta^*(u)) + x^T \eta^*(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} ||_{\infty,\infty} \times 8\epsilon + 8\epsilon \lesssim \log(1/\delta)\epsilon
\end{aligned}
$$

When $\tau < \delta$, the proof procedure is likely to above case.

$$
\begin{aligned}
&|Q^*_\eta(\tau|x) - Q_\eta(\tau|x)| \\
\leqslant &| \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta^*(u)}{a(\eta^*(u), \mathcal{X})\sqrt{1+||\eta^*(u)||^2}} q_0(u) du - \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du| \\
+ &| \int_{\zeta^*(\tau_0)}^{\zeta^*(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du - \int_{\zeta(\tau_0)}^{\zeta(\tau)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du| + 8\epsilon \\
\leqslant &| \int_{\zeta^*(\tau_0)}^{\zeta^*(\delta)} \frac{x^T \eta^*(u)}{a(\eta^*(u), \mathcal{X})\sqrt{1+||\eta^*(u)||^2}} q_0(u) du - \int_{\zeta^*(\tau_0)}^{\zeta^*(\delta)} \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} q_0(u) du| \\
+ &| \int_{\zeta^*(\delta)}^{\zeta^*(\tau)} \frac{x^T \eta^*(u)}{a(\eta^*(u), \mathcal{X})\sqrt{1+||\eta^*(u)||^2}} q_0(u) du| + 8 || \frac{x^T \eta(u)}{a(\eta(u), \mathcal{X})\sqrt{1+||\eta(u)||^2}} ||_{\infty,\infty} \epsilon + 8\epsilon \\
\leqslant &\log(\frac{\zeta^*(\tau_0)(1-\zeta^*(\delta))}{(1-\zeta^*(\tau_0))\zeta^*(\delta)}) M_2^* \epsilon + \frac{1}{r_\mathcal{X}} \int_{\zeta^*(\delta)}^{\zeta^*(\tau)} \frac{x^T \eta^*(u)}{u} du + 16\epsilon \\
\lesssim &\log(\frac{1}{\zeta^*(\delta)})\epsilon + K'^*_\eta |\zeta^*(\epsilon) - \zeta^*(0)| + 16\epsilon \lesssim \log(1/\delta)\epsilon + K'^*_\eta K'^*_\zeta \epsilon + 16\epsilon \lesssim \log(1/\delta)\epsilon
\end{aligned}
$$

Therefore, there exists some universal constant κ_Q^* such that $||Q_\eta^*(\tau|x)-Q_\eta(\tau|x)||_{\infty,\infty} \leqslant \kappa_Q^* \log(1/\delta)\epsilon$, where κ_Q^* is some constant and not related to ϵ or δ. Another result based on this is verified as follows.

$$
\begin{aligned}
&||F_\eta(Q_\eta^*) - \tau||_{\infty,\infty} = ||F_\eta(Q_\eta^*(\tau|x)|x) - F_\eta(Q_\eta(\tau|x)|x)||_{\infty,\infty} \\
=&|| \int_{Q_\eta(\tau|x)}^{Q_\eta^*(\tau|x)} f_\eta(y|x)dy||_{\infty,\infty} = || \int_{Q_\eta(\tau|x)}^{Q_\eta^*(\tau|x)} f_\eta^*(Q_\eta^*(F_\eta(y|x)|x)|x)\frac{q_\eta^*(F_\eta(y|x)|x)}{q_\eta(F_\eta(y|x)|x)}dy||_\infty \\
=&||\frac{q_\eta^*(F_\eta(y|x)|x)}{q_\eta(F_\eta(y|x)|x)}||_{\infty,\infty} \times ||f_\eta^*(Q_\eta^*(F_\eta(y|x)|x)|x)||_{\infty,\infty} \times ||Q_\eta(\tau|x) - Q_\eta^*(\tau|x)||_{\infty,\infty} \\
\leqslant& e^{M_\alpha \epsilon} \times K_f^* \times \kappa_Q^* \log(1/\delta)\epsilon = \kappa_F^* \log(1/\delta)\epsilon
\end{aligned}
$$

where $K_f^* = ||f_\eta^*(\cdot|x)||_{\infty,\infty}$. For β_1, we have

$$
\begin{aligned}
\beta_1 \leqslant& || \log(\frac{\dot\zeta(F_\eta(Q_\eta^*))}{\dot\zeta^*(F_\eta(Q_\eta^*))})||_{\infty,\infty} + || \log(\frac{\dot\zeta^*(F_\eta(Q_\eta^*))}{\dot\zeta^*(\tau)})||_{\infty,\infty} + || \log(\frac{\dot\zeta^*(\tau)}{\dot\zeta(\tau)})||_{\infty,\infty} \\
\leqslant& 2\epsilon + K_\omega'^*||F_\eta(Q_\eta^*) - \tau||_{\infty,\infty} + 2\epsilon \leqslant 4\epsilon + K_\omega'^*\kappa_F^* \log(1/\delta)\epsilon \leqslant N_1 \log(1/\delta)\epsilon
\end{aligned}
$$

where $K_\omega'^* = \sup_{\tau\in(0,1)} \dot\omega_0^*(\tau)$. For β_2, by symmetry, we have

$$
\begin{aligned}
\beta_2 =& || \log(\frac{\zeta(\tau)}{\zeta(F_\eta(Q_\eta^*))}\frac{1-\zeta(\tau)}{1-\zeta(F_\eta(Q_\eta^*))})||_{\infty,\infty} \leqslant 2|| \log(\frac{\zeta(\tau)}{\zeta(F_\eta(Q_\eta^*))}||_{\infty,\infty} \\
\leqslant& 2|| \log(\frac{\zeta(\tau)}{\zeta^*(\tau)}||_{\infty,\infty} + 2|| \log(\frac{\zeta^*(\tau)}{\zeta^*(F_\eta(Q_\eta^*))}||_{\infty,\infty} + 2|| \log(\frac{\zeta^*(F_\eta(Q_\eta^*))}{\zeta(F_\eta(Q_\eta^*))}||_{\infty,\infty} \\
\leqslant& 8\epsilon + 2|| \log(\frac{\zeta^*(\tau)}{\zeta^*(F_\eta(Q_\eta^*))}||_{\infty,\infty} \leqslant 8\epsilon + 2(\sup_\tau \frac{\dot\zeta^*(\tau)}{\zeta^*(\tau)})||F_\eta(Q_\eta^*) - \tau||_{\infty,\infty} \\
\leqslant& 8\epsilon + 2e^{2\epsilon}H_\zeta^*\kappa_F^* \log(1/\delta)\epsilon \leqslant N_2 \log(1/\delta)\epsilon
\end{aligned}
$$

where $H_\zeta^* = \sup_\tau \frac{\dot\zeta^*(\tau)}{\zeta^*(\tau)}$. We applied the similar procedure of proving $\alpha_2(\tau|x)$ to get the upper bound for $\beta_3(\tau|x)$. We first prove the upper bound for

$$
\xi(\tau|x) = |\frac{x^T\eta(\zeta(F_\eta(Q_\eta^*)))}{a(\eta(\zeta(F_\eta(Q_\eta^*))),\mathcal{X})\sqrt{1+||\eta(\zeta(F_\eta(Q_\eta^*)))||^2}} - \frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}|
$$

which is bounded by the sum of these three terms

$$|\frac{x^T\eta(\zeta(F_\eta(Q^*_\eta)))}{a(\eta(\zeta(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta(\zeta(F_\eta(Q^*_\eta)))||^2}}-\frac{x^T\eta^*(\zeta^*(F_\eta(Q^*_\eta)))}{a(\eta^*(\zeta^*(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2}}| \tag{A.4a}$$

$$|\frac{x^T\eta^*(\zeta^*(F_\eta(Q^*_\eta)))}{a(\eta^*(\zeta^*(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2}}-\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}| \tag{A.4b}$$

$$|\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}-\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}| \tag{A.4c}$$

The first and third terms are controlled by $M_2^*\epsilon$ from aforementioned derivation on α. Denoting the second term as $\xi^*(\tau|x)=|\frac{x^T\eta^*(\zeta^*(F_\eta(Q^*_\eta)))}{a(\eta^*(\zeta^*(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2}}-\frac{x^T\eta^*(\zeta^*(\tau))}{a(\eta^*(\zeta^*(\tau)),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}|$, when $\tau\in\{\tau:||\eta^*(\zeta^*(\tau))||=0\}$, we have

$$\begin{aligned}\xi^*(\tau|x)&=|\frac{x^T\eta^*(\zeta^*(F_\eta(Q^*_\eta)))}{a(\eta^*(\zeta^*(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2}}|\\&=|\frac{x^T[\eta^*(\zeta^*(F_\eta(Q^*_\eta)))-\eta^*(\zeta^*(\tau))]}{a(\eta^*(\zeta^*(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2}}|\\&\leqslant\frac{1}{r_{\mathcal{X}}}||x||_\infty K'^*_\eta K'^*_\zeta\kappa^*_F\log(1/\delta)\epsilon\leqslant M_3^0\log(1/\delta)\epsilon\end{aligned}$$

where $K'^*_\zeta=\sup_{\tau\in(0,1)}\dot{\zeta}^*(\tau)$.

When $\tau\in\{\tau:||\eta^*(\zeta^*(\tau))||\neq0\}$, we defined the following three terms

$$\phi(\tau)=a(\eta^*(\zeta^*(F_\eta(Q^*_\eta))),\mathcal{X}),\qquad \phi^*(\tau)=a(\eta^*(\zeta^*(\tau)),\mathcal{X})$$

$$\psi(\tau)=\frac{x^T\eta^*(\zeta^*(F_\eta(Q^*_\eta)))}{||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||}=x^Tu^*(\zeta^*(F_\eta(Q^*_\eta))),\qquad \psi^*(\tau)=\frac{x^T\eta^*(\zeta^*(\tau))}{||\eta^*(\zeta^*(\tau))||}=x^Tu^*(\zeta^*(\tau))$$

$$\gamma(\tau)=\frac{||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||}{\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2}},\qquad \gamma^*(\tau)=\frac{||\eta^*(\zeta^*(\tau))||}{\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}}$$

and we have the following results

$$|\phi(\tau)-\phi^*(\tau)| = |\sup_{x\in\mathcal{X}} x^T u^*(\zeta^*(F_\eta(Q^*_\eta))) - \sup_{x\in\mathcal{X}} x^T u^*(\zeta^*(\tau))|$$

$$\leqslant \sup_{x\in\mathcal{X}} |x^T(u^*(\zeta^*(F_\eta(Q^*_\eta))) - u^*(\zeta^*(\tau)))|$$

$$\leqslant ||x||_\infty K'^*_u K'^*_\zeta \kappa^*_F \log(1/\delta)\epsilon \leqslant \kappa^*_\phi \log(1/\delta)\epsilon$$

$$|\psi(\tau)-\psi^*(\tau)| = |x^T(u^*(\zeta^*(F_\eta(Q^*_\eta))) - u^*(\zeta^*(\tau)))| \leqslant \sup_{x\in\mathcal{X}} |x^T(u^*(\zeta^*(F_\eta(Q^*_\eta))) - u^*(\zeta^*(\tau)))|$$

$$\leqslant ||x||_\infty K'^*_u K'^*_\zeta \kappa^*_F \log(1/\delta)\epsilon \leqslant \kappa^*_\psi \log(1/\delta)\epsilon$$

$$|\gamma(\tau)-\gamma^*(\tau)| \leqslant |(||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||\sqrt{1+||\eta^*(\zeta^*(\tau))||^2} - ||\eta^*(\zeta^*(\tau))||\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2})|$$

$$\leqslant |(||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||\sqrt{1+||\eta^*(\zeta^*(\tau))||^2} - ||\eta^*(\zeta^*(\tau))||\sqrt{1+||\eta^*(\zeta^*(\tau))||^2})|$$

$$+ |(||\eta^*(\zeta^*(\tau))||\sqrt{1+||\eta^*(\zeta^*(\tau))||^2} - ||\eta^*(\zeta^*(\tau))||\sqrt{1+||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))||^2})|$$

$$\leqslant |\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}(||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))|| - ||\eta^*(\zeta^*(\tau))||)|$$

$$+ ||\eta^*(\zeta^*(\tau))|| \times |(||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))|| - ||\eta^*(\zeta^*(\tau))||)|$$

$$\leqslant (\sqrt{1+||\eta^*(\zeta^*(\tau))||^2_\infty} + ||\eta^*(\zeta^*(\tau))||_\infty)|(||\eta^*(\zeta^*(F_\eta(Q^*_\eta)))|| - ||\eta^*(\zeta^*(\tau))||)|$$

$$\leqslant (\sqrt{1+{K^*_\eta}^2} + K^*_\eta) K'^*_\eta K'^*_\zeta \kappa^*_F \log(1/\delta)\epsilon \leqslant \kappa^*_\gamma \log(1/\delta)\epsilon$$

Therefore, since all $\phi^*(\tau)$, $\psi^*(\tau)$ and $\gamma^*(\tau)$ are bounded above, we have

$$\xi^*(\tau|x) = |\frac{\psi^*(\tau)\gamma^*(\tau)}{\phi^*(\tau)} - \frac{\psi(\tau)\gamma(\tau)}{\phi(\tau)}| \leqslant \frac{1}{r^2_{\mathcal{X}}}|\psi^*(\tau)\gamma^*(\tau)\phi(\tau) - \psi(\tau)\gamma(\tau)\phi^*(\tau)| \leqslant M^*_3 \log(1/\delta)\epsilon$$

$$\xi(\tau|x) \leqslant 2M^*_2\epsilon + \max(M^0_3, M^*_3)\log(1/\delta)\epsilon \leqslant N^*_3 \log(1/\delta)\epsilon$$

Since $\xi(\tau|x)$ is fairly small and $||\eta||$ and $||\eta^*||$ are bounded above, we have the upper bound for $|\beta_3(\tau|x)|$ as

$$|\beta_3(\tau|x)| = |\log(\frac{1+\frac{x^T\eta(\zeta(F_\eta(Q^*_\eta)))}{a(\eta(\zeta(F_\eta(Q^*_\eta))),\mathcal{X})\sqrt{1+||\eta(\zeta(F_\eta(Q^*_\eta)))||^2}}}{1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}})| \leqslant |\log(1-\frac{\xi(\tau|x)}{1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}})|$$

$$\leqslant M'_3 N^*_3 \log(1/\delta)\epsilon \times |\frac{1}{1+\frac{x^T\eta(\zeta(\tau))}{a(\eta(\zeta(\tau)),\mathcal{X})\sqrt{1+||\eta(\zeta(\tau))||^2}}}|$$

$$\leqslant M_3' N_3^* \log(1/\delta)\epsilon \times \left|\frac{1}{1 - \frac{||\eta^*(\zeta^*(\tau))||}{\sqrt{1+||\eta^*(\zeta^*(\tau))||^2}} - \kappa_v^*\epsilon}\right| \leqslant N_3 \log(1/\delta)\epsilon$$

Overall, wrap the aforementioned proof

$$||\beta(\cdot|x)||_{\infty,\infty} \leqslant \beta_1 + \beta_2 + ||\beta_3(\cdot|x)||_{\infty,\infty} \leqslant (N_1 + N_2 + N_3)\log(1/\delta)\epsilon = M_\beta \log(1/\delta)\epsilon$$

and therefore the final inequality for log ratio is

$$||\log(\frac{f_\eta^*(\cdot|x)}{f_\eta(\cdot|x)})||_{\infty,\infty} = ||\alpha(\cdot|x)||_{\infty,\infty} + ||\beta(\cdot|x)||_{\infty,\infty} \leqslant M_\alpha\epsilon + M_\beta \log(1/\delta)\epsilon \leqslant M\log(1/\delta)\epsilon$$

Write $f_\eta(y|x) = \frac{\exp(\log(f_\eta(y|x)))}{\int_y \exp(\log(f_\eta(y|x)))dy}$ and $f_\eta^*(y|x) = \frac{\exp(\log(f_\eta^*(y|x)))}{\int_y \exp(\log(f_\eta^*(y|x)))dy}$. Following the second assertion of *Lemma 3.1* in van der Vaart and van Zanten (2008), since $\log(f_\eta(y|x))$ and $\log(f_\eta^*(y|x))$ are both measurable functions, we have

$$\begin{aligned} d_{\mathrm{KL}}(f_\eta^*, f_\eta) &= \int_y f_\eta^*(y|x)\log(\frac{f_\eta^*(y|x)}{f_\eta(y|x)})dy \\ &\lesssim ||\log(\frac{f_\eta^*(y|x)}{f_\eta(y|x)})||_{\infty,\infty}^2 e^{||\log(\frac{f_\eta^*(y|x)}{f_\eta(y|x)})||_{\infty,\infty}}(1 + ||\log(\frac{f_\eta^*(y|x)}{f_\eta(y|x)})||_{\infty,\infty}) \\ &\lesssim (\log(1/\delta)\epsilon)^2 \end{aligned}$$

For really small ϵ, there exists some constant C such that $e^{||\log(f_\eta^*(y|x)) - \log(f_\eta(y|x))||_{\infty,\infty}}(1 + ||\log(f_\eta^*(y|x)) - \log(f_\eta(y|x))||_{\infty,\infty})$ is bounded by C. Therefore, we could find the quadratic upper bound of $|d_{\mathrm{KL}}(f_\eta^*, f_\eta)|$ as $|d_{\mathrm{KL}}(f_\eta^*, f_\eta)| \leqslant C(\log(1/\delta)\epsilon)^2$. When ϵ is such small that $M\log(1/\delta)\epsilon \leqslant 1$, we also have $d_{\mathrm{L}^2}(f_\eta^*, f_\eta) = |\int_y f_\eta^*(y|x)(\log(\frac{f_\eta^*(y|x)}{f_\eta(y|x)}))^2 dy| \leqslant |d_{\mathrm{KL}}(f_\eta^*, f_\eta)| \leqslant C(\log(1/\delta)\epsilon)^2$, which is a straightforward corollary from the Kullback-Leibler divergence we derived before. □

Bibliography

Bondell, H. D., Reich, B. J., and Wang, H. (2010), "Noncrossing quantile regression curve estimation," *Biometrika*, pp. 825–838.

Buchinsky, M. (1994), "Changes in the US wage structure 1963-1987: Application of quantile regression," *Econometrica: Journal of the Econometric Society*, pp. 405–458.

Burgette, L. F., Reiter, J. P., and Miranda, M. L. (2011), "Exploratory quantile regression with many covariates: an application to adverse birth outcomes," *Epm*, 22, 859–866.

Dunson, D. B. and Taylor, J. A. (2005), "Approximate Bayesian inference for quantiles," *Nonparametric Statistics*, 17, 385–400.

Elsner, J. B., Kossin, J. P., and Jagger, T. H. (2008), "The increasing intensity of the strongest tropical cyclones," *Nature*, 455, 92–95.

Ghosal, S., Ghosh, J. K., and Van Der Vaart, A. W. (2000), "Convergence rates of posterior distributions," *Annals of Statistics*, pp. 500–531.

He, X. (1997), "Quantile curves without crossing," *The American Statistician*, 51, 186–192.

Koenker, R. (2005), *Quantile regression*, no. 38, Cambridge university press.

Koenker, R. and Bassett Jr, G. (1978), "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50.

Lancaster, T. and Jae Jun, S. (2010), "Bayesian quantile regression methods," *Journal of Applied Econometrics*, 25, 287–307.

Lazar, A. A. and Zerbe, G. O. (2011), "Solutions for determining the significance region using the Johnson-Neyman type procedure in generalized linear (mixed) models," *Journal of Educational and Behavioral Statistics*, 36, 699–719.

Lenk, P. J. (1988), "The logistic normal distribution for Bayesian, nonparametric, predictive densities," *Journal of the American Statistical Association*, 83, 509–516.

McClain, C. and Rex, M. (2001), "The relationship between dissolved oxygen concentration and maximum size in deep-sea turrid gastropods: an application of quantile regression," *Marine Biology*, 139, 681–685.

Petscher, Y. and Logan, J. A. (2014), "Quantile regression in the study of developmental sciences," *Child development*, 85, 861–881.

Raudenbush, S. W. and Bryk, A. S. (2002), *Hierarchical linear models: Applications and data analysis methods*, vol. 1, Sage.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011), "Bayesian spatial quantile regression," *Journal of the American Statistical Association*, 106, 6–20.

Tokdar, S. (2016), *qrjoint: Joint Estimation in Linear Quantile Regression*, R package version 1.0-3.

Tokdar, S. T. and Ghosh, J. K. (2007), "Posterior consistency of logistic Gaussian process priors in density estimation," *Journal of Statistical Planning and Inference*, 137, 34–42.

Tokdar, S. T., Kadane, J. B., et al. (2012), "Simultaneous linear quantile regression: a semiparametric Bayesian approach," *Bayesian Analysis*, 7, 51–72.

van der Vaart, A. W. and van Zanten, J. H. (2008), "Rates of contraction of posterior distributions based on Gaussian process priors," *The Annals of Statistics*, pp. 1435–1463.

van der Vaart, A. W. and van Zanten, J. H. (2009), "Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth," *The Annals of Statistics*, pp. 2655–2675.

Wu, Y. and Liu, Y. (2009), "Stepwise multiple quantile regression estimation using non-crossing constraints," *Statistics and Its Interface*, 2, 299–310.

Yang, Y. and Tokdar, S. T. (2016), "Joint estimation of quantile planes over arbitrary predictor spaces," *Journal of the American Statistical Association*.