A Sandwich Likelihood for Bayesian Quantile Regression

Karthik Sriram*

Last revised on February 24, 2015

Abstract

The Bayesian approach to quantile regression as proposed by Yu and Moyeed (2001) based on the asymmetric Laplace distribution (ALD) has been extensively used in many applications. While the approach has many advantages, an inferential limitation is that the credible intervals are asymptotically invalid if ALD is a misspecification. This paper proposes a sandwich likelihood method to remedy the issue by deriving supporting theoretical results. The working of the method is demonstrated using simulations.

Key words and phrases. Bayesian; Asymmetric Laplace; Credible Interval, Misspecification.

1 Introduction

Quantile Regression models a particular quantile of the independently distributed responses \( \{Y_i\}_{i=1}^n \) as a function of a \( d- \) dimensional covariate vectors \( \{X_i\}_{i=1}^n \). The classical approach to modeling the \( \tau \)-th quantile \( (0 < \tau < 1) \) as proposed by Koenker and Basset (1978) is to solve the following minimization problem:

\[
\hat{\beta}_n = \text{arg min}_{\beta} \sum_{i=1}^{n} \rho_{\tau}(Y_i - X_i^T \beta), \tag{1}
\]

where \( \rho_{\tau}(u) = u(\tau - I(u \leq 0)) \) with \( I(\cdot) \) being the indicator function. This procedure is equivalent to Maximum Likelihood Estimation (MLE) if the responses are assumed to follow the asymmetric Laplace distribution (ALD), whose probability density function (p.d.f) is given by

\[
f_{\beta}(y) = \tau(1 - \tau) \exp \{ -\rho_{\tau}(y - \mu_{\tau}) \}, \quad \text{with } \mu_{\tau} = X_i^T \beta \text{ and } y \in (-\infty, \infty). \tag{2}
\]

*Karthik Sriram is Assistant Professor, Production and Quantitative Methods Area, Indian Institute of Management Ahmedabad, India (karthiks@iimahd.ernet.in)
It is easy to check that $\mu_i^\tau$ is the $\tau^{th}$ quantile of the p.d.f $f_i(\cdot)$.

The focus of this paper is a widely used Bayesian approach to the problem proposed by Yu and Moyeed (2001), where the Bayesian posterior is obtained by assuming $Y_i \sim f_i(\cdot)$ and a prior $\Pi$ on $\beta$. The approach is computationally attractive especially due to the location scale mixture normal representation of ALD (see Kozumi and Kobayashi (2011)), and is known to give posterior consistent estimates even if it is a misspecification (Sriram et al. 2013). Therefore, the approach has been useful in many applications (e.g. Yu et al. 2005; Yue and Rue 2011; Benoit and Van den Poel 2012; Alhamzawi and Yu 2013; Waldmann et al. 2013).

The aim of this paper is to highlight and remedy an inferential limitation of the approach. Since the ALD model is more often than not a misspecification of the true underlying likelihood, the resulting Bayesian credible intervals turn out to be asymptotically invalid. In other words, they do not possess the desired asymptotic “coverage property”, where a $100(1-\alpha)\%$ credible interval needs to contain the true parameter value approximately $100(1-\alpha)\%$ of the time, for large sample sizes.

Section 2 describes a method to remedy this issue followed by a theoretical justification in Section 3. Section 4 presents supporting simulations. Detailed proofs are deferred to an appendix.

2 The Sandwich Likelihood Method

Suppose $\{Y_i, i = 1, 2, \ldots, n\}$ are independent but non-identically distributed (i.n.i.d) with probability density function $Y_i \sim p_i(\cdot)$, such that the “true” $\tau^{th}$ quantile of $Y_i$ is given by $Q_\tau(X_i) = X_i^T\beta_0$, where $X_i$ is a vector of non-random covariates. We model the $\tau^{th}$ quantile as $Q_\tau(X_i) = X_i^T\beta$ along with a proper prior $\Pi$ (with p.d.f $\pi$) for $\beta$. We write the posterior distribution of $\beta$ given the data $Y_1, Y_2, \ldots, Y_n$ as

$$\Pi_n(B) = \frac{\int_B f_B^{(n)}(\pi(\beta))d\beta}{\int_{R^d} f_B^{(n)}(\pi(\beta))d\beta}, \quad \text{where} \quad f_B^{(n)} = \prod_{i=1}^n f_i(\beta_i(Y_i)).$$

Our sandwich likelihood approach is motivated from Müller (2013) and is based on two key observations. First, $\Pi_n$ is asymptotically equivalent to a normal distribution centered at $\beta_n^{ML}$, which is a solution to equation $\Pi_n$, and with covariance matrix $\frac{1}{n}V^{-1}$, where

$$V = \left( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n p_i(X_i^T\beta_0)X_iX_i^T \right).$$

$$2$$
This fact is established later in Theorem 1. Second, it is known from Koenker 2005 (page 74) that \( \beta_n^M \) is asymptotically normal with mean \( \beta_0 \) and (“sandwich”) covariance matrix \( \frac{1}{n} \Sigma \), where

\[
\Sigma = \tau(1 - \tau)V^{-1}S V^{-1} \quad \text{with} \quad S = \lim_{n \to \infty} S_n, \quad S_n = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T.
\]  

In this paper, we will assume that \( S \) and \( V \) are well defined. Since \( V^{-1} \neq \Sigma \) in general, the credible intervals from \( \Pi_n \) will be asymptotically invalid, i.e. they do not asymptotically match the normal confidence intervals obtained from the classical approach. The sandwich likelihood approach remedies this issue by constructing a normal likelihood for \( \beta_n^M \) centered at \( \beta \) and with covariance matrix \( \Sigma/n \). The method can be described in two steps.

**Step 1.** Obtain \( V^{-1} \) from the posterior distribution \( \Pi_n \) based on \( Y_i \sim f_{i\beta}, \beta \sim \Pi \).

**Step 2.** Recompute a new posterior using the “Sandwich Likelihood” as follows:

\[
\Pi^{\text{prop}}(\beta|Y_1,Y_2,\ldots,Y_n) \propto \frac{1}{|\Sigma/n|^{\frac{1}{2}}} \cdot e^{-n(\beta_n^M - \beta)^T \Sigma^{-1}(\beta_n^M - \beta)} \times \Pi(\beta)
\]  

where \( \Sigma = \tau(1 - \tau)V^{-1}S V^{-1} \).

Computationally, the matrix \( \Sigma \) needed in Step 2 can be easily obtained. This is because \( V^{-1} \) can be approximated by the covariance matrix computed based on the MCMC simulations from Step 1 and then \( S \) can be approximated by \( S_n \). It is worth noting that estimation of the \( \Sigma \) in the \textit{i.n.i.d.} case is in general a challenging problem (see section 3.4.2 of Koenker 2005) since we do not have knowledge of the true underlying densities \( p_i(\cdot) \) for all \( i \). This method is a simple alternative. It will be formalized in Theorem 2 of the next section that Step 2 indeed leads to asymptotically valid credible intervals.

### 3 Theoretical Results

While the sandwich likelihood approach is motivated from Müller 2013, our results do not directly follow from there. For example, Müller 2013 requires the log-likelihood to be twice continuously differentiable with respect to the parameters, which is not the case with ALD. Also, verifying some of the other conditions would largely involve establishing the kind of results derived in this paper. Hence the required results for ALD are directly derived in this section.
Recall that the true underlying p.d.f of $Y_i$ is $p_i$. By way of notation, let $P^{(n)}$ denote the product probability $p_1 \times p_2 \times \cdots p_n$ and $P$ denote the infinite product probability $\prod_{i=1}^{\infty} p_i$. Let $P(\cdot)$ and $E[\cdot]$ denote the probability and expectation with respect to the true product probability. We define $Z_i := Y_i - X_i^T \beta_0$ and note that $P(Z_i \leq 0 | X_i) = \tau$.

Our Assumptions 1 through 4 are essentially those needed for Theorem 2 part (b) of Sriram et al. (2013) and lead up to showing $\sqrt{n}$-posterior consistency for $\Pi_n$. The first assumption is on the prior and the second requires that the covariates be bounded.

**Assumption 1.**

$\Pi$ is proper with a bounded and continuous p.d.f $\pi$ that is positive around $\beta_0$.

**Assumption 2.**

$\exists M > 0$ such that $\|X_i\| \leq M, \forall i$.

The first part of the next assumption essentially says that the non-intercept covariates (if need be after appropriate centering) take values in all quadrants of the Euclidean plane. In particular, this implies that they cannot be collinear. The second part of the assumption requires that the true underlying likelihood put positive mass around the true quantile, in particular ensuring that it is unique.

**Assumption 3.**

(a) Let the first coordinate of $X_i$ be identically 1 representing the intercept. After appropriate centering (if need be) of the other co-ordinates, $\exists \epsilon_0 > 0$ such that $\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_{X_i \in D} > 0, \forall D \subset \mathbb{R}^d$ of the form $\{1\} \times U_2 \times \cdots U_d$, where for some $j$, $U_j$ is either $(\epsilon_0, \infty)$ or $(-\infty, -\epsilon_0)$ and for $k \neq j$, $U_k$ is either $(0, \infty)$ or $(-\infty, 0)$.

(b) For some $C > 0$ and all sufficiently small $\Delta > 0$, $P(0 < Z_i < \Delta) > C\Delta$ and $P(-\Delta < Z_i < 0) > C\Delta \forall i$.

Assumption 4 is a technical condition required to apply Strong Law of Large Numbers for i.n.i.d random variables $\{Z_i\}_{i=1}^{n}$.

**Assumption 4.**

$$\limsup_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} E|Z_i| < \infty \text{ and } \sum_{i=1}^{\infty} \frac{EZ_i^2}{i^2} < \infty.$$
The lemma below is essentially Theorem 2 part (b) of Sriram et al. (2013). A minor difference is that the convergence below is stated “in expectation” instead of “in probability”. Suffices to note that the same arguments in their paper can be used to conclude the former.

**Lemma 1** (\(\sqrt{n} - \) Posterior consistency). Suppose assumptions 1 to 4 hold. Then for any \(M_n \to \infty\),

\[
E \left[ \Pi_n \left( \|\beta - \beta_0\| > M_n \sqrt{n} \right) \right] \to 0.
\]

The asymptotic normality of the posterior is derived by applying the Bernstein-von-Mises theorem for misspecified models given by Kleijn and van der Vaart (2012). A key requirement for their result is the “Local Asymptotic Normality” (LAN) property. For the ALD model specification for i.n.i.d data, we establish the LAN property by making the following assumption on the boundedness and continuity of the true underlying densities \(p_i\).

**Assumption 5.**

For some \(C, \eta > 0\) and for all \(\beta\) in some small enough neighborhood of \(\beta_0\), the p.d.fs \(p_i\) satisfy the following conditions:

(a) \(\{p_i(X_i^T \beta_0), i \geq 1\}\) are uniformly bounded away from \(\infty\).

(b) \(|p_i(X_i^T \beta) - p_i(X_i^T \beta_0)| \leq C |X_i^T \beta - X_i^T \beta_0|^\eta \forall i\).

**Lemma 2** (LAN property). Under assumptions 2 and 5, the LAN property holds, i.e., for any compact set \(K \subset \mathbb{R}^d\),

\[
\sup_{\delta \in K} \left| \log \frac{f_{\hat{\beta}_0 + \frac{\delta}{n}}}{f_{\hat{\beta}_0}} - V \Delta_n, \beta_0 - \frac{1}{2} \delta^T V \delta \right| \to 0 \text{ in probability } [P]. \tag{8}
\]

where, \(\Delta_n, \beta_0 = -V^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\gamma - I(Y_i \leq X_i^T \beta_0))X_i\). \tag{9}

In the interest of flow, the proof of the lemma is deferred to the appendix.

The next lemma, which establishes the connection between the posterior probability and the normal distribution is an immediate consequence of Lemma 2. Let \(\Phi(B, \mu, \Sigma)\) denote the probability of a set \(B\) under the multivariate normal distribution with mean \(\mu\) and covariance matrix \(\Sigma\). Then, we have the following result.
Lemma 3. Under Assumptions 1 to 5,

$$\sup_{B} \left| \Pi_{n} \left( \sqrt{n}(\beta - \beta_{0}) \in B \right) - \Phi \left( B, \Delta_{n,\beta_{0}}, V^{-1} \right) \right| \rightarrow 0 \text{ in probability } [P]. \quad (10)$$

Proof. Under the given assumptions, conclusions of Lemmas 1 and 2 hold. The result follows since these are precisely the conditions required for Theorem 2.1 of Kleijn and van der Vaart (2012).

It remains to show that such a normal distribution can indeed be centered at $\beta_{n}^{M}$. The next lemma is a result from Koenker (2005) (see page 122, equation 4.4).

Lemma 4. Let $\beta_{n}^{M}$ be as in equation (7). Under assumptions 2 and 5

$$\sqrt{n}(\beta_{n}^{M} - \beta_{0}) - \Delta_{n,\beta_{0}} \rightarrow 0 \text{ in probability } [P].$$

The first main theorem given below is immediate from Lemmas 3 and 4.

Theorem 1. Let Assumptions 1 to 5 hold and $\beta_{n}^{M}$ be a solution to (1). Then,

$$\sup_{B} \left| \Pi_{n} \left( \sqrt{n}(\beta - \beta_{0}) \in B \right) - \Phi \left( B, \sqrt{n}(\beta_{n}^{M} - \beta_{0}), V^{-1} \right) \right| \rightarrow 0 \text{ in probability } [P].$$

Theorem 2 formalizes Step 2.

Theorem 2. Let Assumptions 1 to 5 hold, $\beta_{n}^{M}$ as in (7) and $\Pi_{n}^{prop}$ as in (6). Then,

$$\sup_{B} \left| \Pi_{n}^{prop} \left( \sqrt{n}(\beta - \beta_{0}) \in B \right) - \Phi \left( B, \sqrt{n}(\beta_{n}^{M} - \beta_{0}), \Sigma \right) \right| \rightarrow 0 \text{ in probability } [P].$$

Proof of Theorem 2 is presented in the appendix.

4 Simulation Study

In this section, we study the performance of the sandwich likelihood approach on simulated data from two different “true” underlying models and across sample sizes. The method is checked for the coverage property of the 95% confidence interval (COV), length of the confidence interval (LEN) and the mean absolute deviation (MAD) of the estimated parameter value from the true value.
We compare the sandwich likelihood approach (denoted by SLBA) with the frequentist quantile regression approach (denoted QR) and the Bayesian approach based purely on ALD (denoted ALD).

To construct the simulated data, we consider two covariates $X_1$ and $X_2$ simulated from $N(3,1)$ (truncated between 1 and 1000) and $\text{Bernoulli}(0.3)$ respectively. The first model is obtained as location shifted normal and the second model as a scaled gamma random variable. In both the models, the true $75^{th}$-quantile (i.e. $\tau = 0.75$) is given by $q_\tau(X) = (1 + 2X_1 + 3X_2)$. The models are described below:

**Model 1.** $Y = q_\tau(X) + \epsilon$, where $\epsilon = Z - q$, $Z \sim N(0,1)$ and $q = 75^{th}$ quantile of $N(0,1)$.

**Model 2.** $Y \sim \text{Gamma} \left(\text{shape} = 2, \text{scale} = \frac{q}{q_\tau(X)}\right)$, where $q = 75^{th}$ quantile of $\text{Gamma}(2,1)$.

The specified model for the $75^{th}$ quantile is $Q_\tau(X) = \alpha + \beta_1 X_1 + \beta_2 X_2$. Further, we separately consider a relatively flat prior for the parameters (i.e. a product of three $N(0,100)$ distributions), and an informative prior (i.e., a product of $N(0.9,1)$, $N(2.1,1)$ and $N(2.9,1)$). The parameters and the metrics (viz. COV, LEN, MAD) are estimated for all the methods (viz. QR, ALD, SLBA), for different sample sizes viz $N \in \{50, 100, 500, 2000\}$, and for both the relatively flat as well as informative priors. For the QR method, we use the bootstrap method of computing confidence intervals and for the other methods we compute the posterior credible intervals using 1000 MCMC simulations after a burn-in of 2000 simulations. The coverage is computed by generating 200 different random samples of $Y|\{X_1, X_2\}$ each of size $N$ and by calculating the percentage of times the true parameter value is contained in the 95% confidence/credible intervals. Ideally, 95% of the 200 replications should result in intervals containing the true parameter values. Table 1 shows computed results for the case of a relatively flat prior and Table 2 for the case of an informative prior.

We know from our results that the credible intervals resulting from a Bayesian analysis using ALD are asymptotically incorrect. For simulated model 1 in table 1 the coverage ratio for ALD is consistently higher than the expected 95%, across all sample sizes. Correspondingly, the coverage ratio from the QR and SLBA methods are comparatively closer to 95%. Further, for model 1, the length of the credible intervals from ALD are larger. However, it is not necessary that the ALD method always leads to larger coverage and bigger intervals. For model 2 in the same table, the coverage is much lower than 95% for ALD, while the other methods again tend to be closer to 95%. In this case, we see the lengths of credible intervals from ALD are smaller. Model 1 is simpler with i.i.d. errors. The coverage for the ALD method although higher than 95% is just higher by
Table 1: Comparison of methods for simulated models using an Non-informative Prior.

Simulated Model 1

| N  | Metric | \(\alpha\)  | \(\beta_1\) | \(\beta_2\) |
|----|--------|-------------|-------------|-------------|
|    |        | QR | ALD | SLBA | QR | ALD | SLBA | QR | ALD | SLBA |
| 50 | COV    | 0.94 | 1.00 | 0.96 | 0.96 | 1.00 | 0.98 | 0.94 | 1.00 | 0.99 |
| 50 | LEN    | 2.67 | 3.32 | 2.73 | 0.78 | 0.98 | 0.82 | 1.68 | 2.11 | 1.74 |
| 50 | MAD    | 0.55 | 0.46 | 0.45 | 0.15 | 0.13 | 0.13 | 0.33 | 0.29 | 0.29 |
| 100| COV    | 0.96 | 1.00 | 0.97 | 0.97 | 0.99 | 0.97 | 0.95 | 1.00 | 0.98 |
| 100| LEN    | 1.93 | 2.40 | 1.93 | 0.59 | 0.74 | 0.60 | 1.47 | 1.81 | 1.49 |
| 100| MAD    | 0.36 | 0.33 | 0.33 | 0.11 | 0.10 | 0.10 | 0.28 | 0.22 | 0.22 |
| 500| COV    | 0.94 | 0.98 | 0.95 | 0.94 | 0.99 | 0.96 | 0.94 | 1.00 | 0.96 |
| 500| LEN    | 0.86 | 1.07 | 0.84 | 0.28 | 0.35 | 0.27 | 0.53 | 0.66 | 0.52 |
| 500| MAD    | 0.17 | 0.16 | 0.16 | 0.05 | 0.05 | 0.05 | 0.10 | 0.10 | 0.10 |
| 2000| COV    | 0.94 | 0.98 | 0.94 | 0.94 | 0.98 | 0.94 | 0.94 | 1.00 | 0.94 |
| 2000| LEN    | 0.42 | 0.53 | 0.41 | 0.13 | 0.16 | 0.13 | 0.26 | 0.34 | 0.26 |
| 2000| MAD    | 0.09 | 0.08 | 0.08 | 0.03 | 0.02 | 0.02 | 0.05 | 0.05 | 0.05 |

Simulated Model 2

| N  | Metric | \(\alpha\)  | \(\beta_1\) | \(\beta_2\) |
|----|--------|-------------|-------------|-------------|
|    |        | QR | ALD | SLBA | QR | ALD | SLBA | QR | ALD | SLBA |
| 50 | COV    | 0.96 | 0.82 | 0.95 | 0.94 | 0.74 | 0.92 | 0.98 | 0.72 | 0.95 |
| 50 | LEN    | 12.84| 6.57 | 11.05| 4.43 | 2.10 | 3.88 | 9.17 | 4.56 | 8.24 |
| 50 | MAD    | 2.32 | 2.06 | 1.84 | 0.84 | 0.73 | 0.66 | 1.64 | 1.56 | 1.44 |
| 100| COV    | 0.97 | 0.84 | 0.94 | 0.97 | 0.75 | 0.94 | 0.96 | 0.78 | 0.96 |
| 100| LEN    | 7.91 | 4.57 | 7.23 | 2.82 | 1.51 | 2.58 | 8.11 | 4.02 | 7.47 |
| 100| MAD    | 1.46 | 1.36 | 1.30 | 0.54 | 0.50 | 0.48 | 1.35 | 1.26 | 1.19 |
| 500| COV    | 0.97 | 0.81 | 0.94 | 0.96 | 0.74 | 0.95 | 0.96 | 0.71 | 0.95 |
| 500| LEN    | 3.40 | 2.07 | 3.27 | 1.30 | 0.73 | 1.25 | 3.01 | 1.57 | 3.01 |
| 500| MAD    | 0.62 | 0.60 | 0.60 | 0.24 | 0.24 | 0.24 | 0.59 | 0.57 | 0.56 |
| 2000| COV    | 0.95 | 0.83 | 0.94 | 0.97 | 0.78 | 0.96 | 0.94 | 0.70 | 0.92 |
| 2000| LEN    | 1.75 | 1.07 | 1.72 | 0.64 | 0.36 | 0.63 | 1.51 | 0.79 | 1.48 |
| 2000| MAD    | 0.34 | 0.32 | 0.32 | 0.12 | 0.12 | 0.12 | 0.30 | 0.29 | 0.29 |

a few percentage points. In contrast, for model 2, which is more complex with i.n.i.d errors, the inadequacy of coverage for the ALD method is more pronounced. For example, the coverage for the ALD method in model 2 is 70 to 85% across all sample sizes and for all parameters. We know that although the ALD method leads to asymptotically incorrect credible intervals, the parameter estimates are posterior consistent. In line with this, we see that the mean absolute deviation of the parameter estimates from the ALD method are comparable to the other methods. In summary, the simulation results in Table 1 are supportive of our results in section 3, i.e. in the case of relatively flat priors and large samples, the proposed sandwich likelihood and the classical approaches lead to similar inferences. Further, they highlight the fact that using Bayesian inference based on ALD
could be highly misleading especially when the true data generating model is complex.

Table 2 shows the simulation results when an informative prior is used. In this case, we observe that the SLBA method consistently has a lower mean absolute deviation than the QR method. As the sample size gets larger, the difference between the mean absolute deviations from the QR and SLBA methods decreases. However, for the more complex model 2, the SLBA method still tends to outperform QR method in terms of mean absolute deviation, even for larger sample sizes. This should not be surprising since the QR method is a classical approach, which does not utilize the prior information, whereas the SLBA method does. In summary, while dealing with possibly complex underlying models, small sample sizes and informative priors, the SLBA method can be expected to perform better than the classical approach and the Bayesian approach based on ALD.

Appendix

Proof of Lemma 2. Recall $Z_i = Y_i - X_i^T \beta_0$. Let $P_1(\cdot)$ denote the cumulative distribution function for the p.d.f $p_i(\cdot)$. In the lines of proof of Theorem 4.1 in Koenker [2005], we write

$$U_n(\delta) := \log \frac{f^{(n)}_{\beta_0 + \delta}}{f^{(n)}_{\beta_0}} = U_{1n}(\delta) + E[U_{2n}] + (U_{2n}(\delta) - E[U_{2n}(\delta)])$$

where,

$$U_{1n}(\delta) := -\delta^T \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tau - I(Z_i \leq 0))X_i, \text{ and}$$

$$U_{2n}(\delta) := \sum_{i=1}^{n} \int_{X_i}^{X_i + \delta} (I(Z_i \leq s) - I(Z_i \leq 0))ds = \sum_{i=1}^{n} U_{2ni}(\delta).$$

Further, by Taylor’s formula,

$$E(U_{2n}(\delta)) = \sum_{i=1}^{n} \int_{0}^{\frac{X_i^T \delta}{\sqrt{n}}} (P_i(X_i^T \beta_0 + s) - P_i(X_i^T \beta_0))ds = \sum_{i=1}^{n} \int_{0}^{\frac{X_i^T \delta}{\sqrt{n}}} p_i(X_i^T \beta_0 + r_{in}(s)) \cdot s ds,$$

where $0 \leq r_{in}(s) \leq s \leq \frac{|X_i^T \delta|}{\sqrt{n}}$. Therefore, by Assumptions 2 and 5, we have

$$\left| E(U_{2n}(\delta)) - \frac{1}{2n} \sum_{i=1}^{n} p_i(X_i^T \beta_0) \delta^T X_i X_i^T \delta \right|$$

$$\leq \sum_{i=1}^{n} \int_{0}^{\frac{X_i^T \delta}{\sqrt{n}}} C |r_{in}(s)|^{\eta} \cdot s ds \leq C \left( \max_{i \geq 1} \frac{|X_i^T \delta|}{\sqrt{n}} \right)^{\eta} \cdot \delta^T \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \delta.$$
Table 2: Comparison of methods for simulated models using an Informative Prior.

**Simulated Model 1**

| N  | Metric | \(\alpha\) | \(\beta_1\) | \(\beta_2\) |
|----|--------|-------------|-------------|-------------|
| 50 | COV    | 0.97        | 1.00        | 0.99        |
|    | LEN    | 2.88        | 2.54        | 1.39        |
|    | MAD    | 0.54        | 0.25        | 0.22        |
| 100| COV    | 0.96        | 1.00        | 0.98        |
|    | LEN    | 2.11        | 2.11        | 1.31        |
|    | MAD    | 0.40        | 0.23        | 0.20        |
| 500| COV    | 0.98        | 1.00        | 0.98        |
|    | LEN    | 0.86        | 1.03        | 0.76        |
|    | MAD    | 0.15        | 0.13        | 0.12        |
| 2000| COV   | 0.95        | 0.98        | 0.94        |
|    | LEN    | 0.42        | 0.53        | 0.40        |
|    | MAD    | 0.08        | 0.08        | 0.08        |

**Simulated Model 2**

| N  | Metric | \(\alpha\) | \(\beta_1\) | \(\beta_2\) |
|----|--------|-------------|-------------|-------------|
| 50 | COV    | 0.96        | 1.00        | 0.99        |
|    | LEN    | 11.36       | 3.18        | 2.01        |
|    | MAD    | 2.04        | 0.44        | 0.30        |
| 100| COV    | 0.97        | 0.98        | 0.98        |
|    | LEN    | 8.77        | 3.00        | 2.28        |
|    | MAD    | 1.68        | 0.53        | 0.33        |
| 500| COV    | 0.96        | 0.87        | 0.97        |
|    | LEN    | 3.44        | 1.84        | 2.07        |
|    | MAD    | 0.63        | 0.47        | 0.32        |
| 2000| COV   | 0.96        | 0.84        | 0.96        |
|    | LEN    | 1.79        | 1.04        | 1.48        |
|    | MAD    | 0.35        | 0.31        | 0.26        |

Therefore, \(\sup_{\delta \in K} \left| E(U_{2n}(\delta)) - \frac{1}{2n} \sum_{i=1}^{n} p_i (X_i^T \beta_0) \delta^T X_i X_i^T \delta \right| \to 0\) in probability \([P]\). \hspace{1cm} (11)

To complete the proof, it is enough to show \(\sup_{\delta \in K} |U_{2n}(\delta) - E[U_{2n}(\delta)]| \to 0\) in probability \([P]\). To show this, let \(\delta_i = arg \max_{\delta \in K} |U_{2ni}(\delta) - E[U_{2ni}(\delta)]|\). \(\delta_i\) is possibly random since it can depend on \(Z_i\), but is well defined since \(U_{2ni}(\delta)\) is a continuous (random) function on a compact set \(K\). Then,

\[
P \left( \sup_{\delta \in K} |U_{2n}(\delta) - E[U_{2n}(\delta)]| > \epsilon \right) \leq P \left( \sum_{i=1}^{n} \sup_{\delta \in K} |U_{2ni}(\delta) - E[U_{2ni}(\delta)]| > \epsilon \right)
= P \left( \sum_{i=1}^{n} |U_{2ni}(\delta_i) - E[U_{2ni}(\delta_i)]| > \epsilon \right) \leq \sum_{i=1}^{n} \frac{E(U_{2ni}^2(\delta_i))}{\epsilon^2}. \hspace{1cm} (12)\]
The last step uses Chebychev’s inequality. Further, let \( G = \sup_{\delta \in K} \sup_{i \geq 1} |X_i^T \delta| \). In particular, \(|X_i^T \delta| \leq G\). Without loss of generality, if \( X_i^T \delta > 0, \forall i \), the right hand side of Equation (12) is

\[
\leq \sum_{i=1}^{n} \frac{n}{2} \left| \frac{\mathbb{E} \left( \int_{0}^{\pi} (I(Z_i \leq s) - I(Z_i \leq 0)) ds \right)^2}{\epsilon^2} \right| \leq \sum_{i=1}^{n} \frac{n}{2} \left| \frac{\mathbb{E} \left( \int_{0}^{\pi} (I(Z_i \leq s) - I(Z_i \leq 0)) ds \right)^2}{\epsilon^2} \right|
\]

\[
\leq \sum_{i=1}^{n} \frac{G^2}{n} \left( \frac{\mathbb{E} \left( I(Z_i \leq \delta_n^\pi) - I(Z_i \leq 0) \right)^2}{\epsilon^2} \right) \leq \frac{G^2}{n} \sum_{i=1}^{n} \left( P(X_i^T/\beta_0 + \frac{G}{\sqrt{n}}) - P(X_i^T/\beta_0) \right)
\]

\[
\leq \frac{G^2}{n} \sum_{i=1}^{n} |p_i(X_i^T/\beta_0 + r_i)| \frac{G}{\sqrt{n}}, \text{ for some } |r_i| \leq \frac{G}{\sqrt{n}} \text{ using Taylor’s formula.}
\]

A consequence of assumptions 2 and 5 is that \( \{p_i(X_i^T/\beta_0 + r_i), \ i \geq 1\} \) are uniformly bounded. Hence, we can conclude that \( P(\sup_{\delta \in K} |U_{2n}(\delta)| - E[U_{2n}(\delta)] > \epsilon) \to 0 \).

**Proof of Theorem 2** First, note that the LAN condition is trivially satisfied because

\[
\log \frac{p_{\beta_n}}{p_{\beta_0}} = -\delta^T \Sigma^{-1} \sqrt{n} (\beta_n^M - \beta_0) + \frac{1}{2} \delta^T \Sigma^{-1} \delta.
\]

The result is immediate from Theorem 2.1 of Kleijn and van der Vaart (2012), provided

\[
E \left[ \Pi_n^{prop} (\sqrt{n} ||\beta - \beta_0|| > M_n) \right] \to 0 \text{ for any sequence } M_n \to \infty.
\] (13)

Let \( B_n = \{\beta : \sqrt{n} ||\beta - \beta_0|| > M_n\} \). Making a change of variable \( \sqrt{n} (\beta - \beta_n) = t \), we can write

\[
\Pi_n^{prop}(B_n) = \int_{\mathbb{R}^d} \left( \frac{1}{\sqrt{2\pi \cdot \det(\Sigma)}} \right)^d \cdot e^{-\frac{t^T \Sigma^{-1} t}{2}} \cdot I(||\beta_n + t/\sqrt{n} - \beta_0|| > M_n) \cdot \pi(\beta_n + t/\sqrt{n}) dt
\]

\[
= \int_{\mathbb{R}^d} \left( \frac{1}{\sqrt{2\pi \cdot \det(\Sigma)}} \right)^d \cdot e^{-\frac{t^T \Sigma^{-1} t}{2}} \cdot \pi(\beta_n + t/\sqrt{n}) dt
\]

Since \( \beta_n \to \beta_0 \) in probability \( [P] \), \( M_n \to \infty \), and \( \pi(\cdot) \) is bounded and continuous by Assumption 1, an application of Skorohod representation theorem for the sequence \( \{\beta_n\}_{n \geq 1} \) along with the dominated convergence theorem implies that the numerator of the above expression converges to zero and the denominator to \( \pi(\beta_0) \). This in turn implies that \( E \left[ \Pi_n^{prop} (\sqrt{n} ||\beta - \beta_0|| > M_n) \right] \to 0 \).
References

Alhamzawi, R. and Yu, K. (2013), “Conjugate priors and variable selection for Bayesian quantile regression,” Computational Statistics and Data Analysis, 64, 209–219.

Benoit, D. F. and Van den Poel, D. (2012), “Binary Quantile Regression: a Bayesian Approach based on the Asymmetric Laplace Distribution,” Journal of Applied Econometrics, 1174–1188.

Kleijn, B. J. K. and van der Vaart, A. W. (2012), “The Bernstein-Von-Mises theorem under misspecification.” Electronic Journal of Statistics, 6, pp. 354–381.

Koenker, R. (2005), Quantile Regression (Econometric Society Monographs), Cambridge University Press.

Koenker, R. and Bassett, Gilbert, J. (1978), “Regression Quantiles,” Econometrica, 46, 33–50.

Kozumi, H. and Kobayashi, G. (2011), “Gibbs sampling methods for Bayesian quantile regression,” Journal of statistical computation and simulation, 81, 1565–1578.

Müller, U. K. (2013), “Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix.” Econometrica, 81, 1805–1849.

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013), “Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density,” Bayesian Analysis, 8, 479–504.

Waldmann, E., Kneib, T., Yue, Y. R., Lang, S., and Flexeder, C. (2013), “Bayesian semiparametric additive quantile regression,” Statistical Modelling, 13, 223–252.

Yu, K. and Moyeed, R. A. (2001), “Bayesian Quantile Regression,” Statistics and Probability Letters, 54, 437–447.

Yu, K., van Kerm, P., and Zhang, J. (2005), “Bayesian quantile regression: an application to the wage distribution in 1990s Britain,” Sankhyā, 57, 359–377.

Yue, Y. R. and Rue, H. (2011), “Bayesian Inference for Additive Mixed Quantile Regression Models,” Computational Statistics and Data Analysis, 55, 84–96.