Al-Qadisiyah Journal of Pure Science

ISSN(Printed): 1997-2490 ISSN(Online): 2411-3514

DOI: 10.29350/jops

Al-Qadisiyah Journal of Pure Science

Robust Quantile Horseshoe algorithm for sparse data

Ahmed Alhamzawi
Department of Mathematics,
College of Science,
University of AL-Qadisiyah, Iraq
ahmed.alhamzawi@qu.edu.iq

Gorgees Shaheed Mohammad
Department of Mathematics,
College of Education,
University of AL-Qadisiyah, Iraq
gorgees.alsalamy@qu.edu.iq

Daniele Ettore Otera
Institute of Data Science and Digital Technologies,
Faculty of Mathematics and Informatics, Vilnius
University, 08412 Vilnius, Lithuania
daniele.otera@mif.vu.lt

Received: 13/04/2026; Revised: 07/05/2026; Accepted: 11/05/2026;
Published Online: 11/05/2026.

Abstract

Current Quantile approaches relay on the asymmetric laplace distribution, which requires solving large matrix equations at every step that beomes very complex as the data size grows. To fix this, a faster method for high-dimensional Bayesian quantile regression is introduced. This method works directly on the posterior distribution without extra latent variables. By using a Horseshoe prior as the proposal, this algorithm avoids costly matrix operations. This reduces the computational cost from cubic $\mathcal{O}(p^3)$ to linear $\mathcal{O}(np)$ complexity. Tests on real and simulation data show that this method is much faster and more efficient than existing techniques.

Keywords: Bayesian inference, Horseshoe prior, shrinkage priors, slice sampling.

1. Introduction

This paper develops a new algorithm for Bayesian quantile regression that is both computationally efficient and handles robust modeling. We focus on the horseshoe prior which has proven to be very useful for dealing with sparsity and high-dimensional data [1]. Quantile regression tends to be less sensitive to outliers but robust to extreme values [2] since it models the conditional quantiles. Thus, it provides a full view of the distribution and valuable in many fields, such as, ecology and economics [3]. However, Bayesian quantile regression has many computational challenges since the likelihood function is usually not smooth since it uses the check loss function which is not differentiable at zero. Most authors have adopted to using the Gibbs sampling by rewriting the error distribution from the Asymmetric Laplace Distribution (ALD) [4] as a mixture of a Normal distribution and an Exponential distribution. Howver, this data augmentation introduces one latent variable for every observation which doubles the parameter space and hence it slows down the mixing. This is problematic for for big data and a better way is needed. The goal is to find a methods which avoids these latent variables. One such method is the slice-within-Gibbs sampler based on the elliptical slice sampler [5].

The original elliptical slice sampler targets only a specific posterior form of two parts, a Gaussian prior and an arbitrary likelihood. We will adapt this method for the quantile regression by setting the shrinkage prior as the Gaussian part with ALD likelihood as the arbitrary part. This fits the framework perfectly and avoids the need to augment the data. Thus, we do not need to calculate the gradients and only need to evaluate the likelihood which is much faster process. We will focus on the horseshoe prior [1] as global-local shrinkage prior that is effective for sparse signals. Let β be the coefficient vector and λ_j be the local shrinkage parameter and τ be the global shrinkage parameter. The conditional prior is Gaussian:

$$\beta_j \mid \lambda_j, \tau \sim \mathcal{N}\left(0, \lambda_j^2 \tau^2\right) \tag{1}$$

This equation defines the Gaussian factor for our sampler. with the diagonal covariance matrix $\Sigma = \text{diag}(\lambda_1^2\tau^2, \ldots, \lambda_p^2\tau^2)$. Let y be the response, X be the predictors and τ represents the τ-th quantile. The error terms ϵ_i follows the ALD and the density is proportional to the exponential of the check loss:

$$L(\beta) \propto \exp\left(-\sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta)\right) \tag{2}$$

Here, $\rho_\tau(u)$ is the check loss function, defined as $u(\tau - \mathbb{I}(u < 0))$ serving as the arbitrary likelihood in our sampler that effectively replaces the likelihood in the original derivation [5]. We combine these two parts to get the posterior as proportional to the product:

$$\pi(\beta \mid y) \propto \mathcal{N}(\beta \mid 0, \Sigma) \times L(\beta) \tag{3}$$

This form allows us to use the elliptical slice sampler directly. To sample β, we condition on the shrinkage parameters and propose a new β^*. We use an auxiliary variable ν drawn from the prior $\mathcal{N}(0, \Sigma)$ and define an ellipse:

$$\beta^* = \beta\cos\,\theta + \nu\sin\,\theta \tag{4}$$

The goal is to search for an acceptable θ using this slice sampling procedure by evaluating $L(\beta^*)$ and comparing it to a threshold. We accept if the likelihood is high enough and we do not differentiate the check loss and simply evaluate it. This approach has the benefit of being generic since it works for any conditionally Gaussian prior. It can also work for exotic priors too, such as, the asymmetric Cauchy prior. This method is much faster since does not need to derive new Gibbs steps and one only needs the prior's variance, thus avoids large matrix operations. This is in contrast to the standard Gibbs samplers which are slower because then invert large matrices at every step which costly for many predictors. Unlike, our method which precomputes the necessary quantities. We demonstrate this approach in our simulations. We show that this method is often much faster especially when handling higher-dimensional cases $p > n$ such as in many genomics. In section 2, we review the elliptical slice sampler and present in details the algorithm adapted it to quantile regression. Then, in section 3 and 4, we test

the perfomance of our model using simulations to other popular models. Section 4 concludes by discussing the limitations and future work.

2. Elliptical Slice Sampling for Quantile Horseshoe Regression

The Elliptical Slice Sampler is a Markov Chain Monte Carlo method that generates samples from a posterior distribution. It especially targets density make up of a product of two factors, a multivariate Gaussian distribution and an arbitrary scalar function. Let $\boldsymbol{\beta}$ be the parameter of interest and $p(\boldsymbol{\beta})$ the target density. We can write the density as follows:

$$p(\boldsymbol{\beta}) \propto \mathcal{N}(\boldsymbol{\beta}; \mathbf{0}, \boldsymbol{\Sigma}) \times \mathcal{L}(\boldsymbol{\beta}) \tag{5}$$

where $\mathcal{L}(\boldsymbol{\beta})$ is the likelihood function. It is clear that the Gaussian factor comes from the prior and the likelihood factor comes from the quantile model. The algorithm updates $\boldsymbol{\beta}$ by proposing a new state $\boldsymbol{\beta}^*$ which lies on an ellipse defined by the current state $\boldsymbol{\beta}$ and an auxiliary variable $\boldsymbol{\nu}$ drawn from the Gaussian prior $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ The proposal $\boldsymbol{\beta}^*$ is a linear combination with an angle parameter θ.

$$\boldsymbol{\beta}^*(\theta) = \boldsymbol{\beta}\cos\,\theta + \boldsymbol{\nu}\sin\,\theta \tag{6}$$

This transformation preserves the Gaussian distribution. Also, if $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are Gaussian, then $\boldsymbol{\beta}^*$ is Gaussian which is true for any angle θ. Thus, the proposal always respects the prior and can be accepted based on only the likelihood $\mathcal{L}(\boldsymbol{\beta})$ for the τ-th quantile with $\tau \in (0,1)$. The quantile regression model can be written as

$$Q_\tau(y_i \mid \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta} \tag{7}$$

Similaraly, Bayesian inference can be done using the Asymmetric Laplace Distribution [3], where the density for a single observation is simple.

$$f(y_i \mid \boldsymbol{\beta}) = \tau(1-\tau)\exp\left(-\rho_\tau\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right) \tag{8}$$

Similarly, Bayesian inference can be facilitated by expressing the Asymmetric Laplace Distribution as a location-scale mixture of normals [4]. While the density for a single observation is given by:

$$f(y_i \mid \boldsymbol{\beta}) = \tau(1-\tau)\exp\left(-\rho_\tau\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right) \tag{9}$$

it can be equivalently represented by introducing latent variables z_i. Specifically, we model the response y_i as:

$$y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \zeta_1 z_i + \zeta_2\sqrt{z_i}u_i \tag{10}$$

where $u_i \sim \mathcal{N}(0,1)$ is a standard normal error and $z_i \sim \text{Exp}(1)$ follows a standard exponential distribution. The scalar parameters θ and ψ are determined by the quantile level τ :

$$\zeta_1 = \frac{1-2\tau}{\tau(1-\tau)}, \zeta_2 = \sqrt{\frac{2}{\tau(1-\tau)}} \tag{11}$$

Conditional on the latent variable z_i, the distribution of y_i becomes Gaussian:

$$y_i \mid \boldsymbol{\beta}, z_i \sim \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta} + \zeta_1 z_i, \zeta_2^2 z_i\right) \tag{12}$$

This data augmentation strategy enables the use of standard Gibbs sampling steps for the regression coefficients.

The Horseshoe prior represents one of the most popular models used for regularization in the present of sparsity [1]. It is a hierarchical prior that uses local and global shrinkage parameters. Let λ_j be the local shrinkage parameter for β_j, τ_g be the global shrinkage parameter and the coefficient β_j follows a Normal distribution as

$$\beta_j \mid \lambda_j, \tau_g \sim \mathcal{N}\left(0, \lambda_j^2\tau_g^2\right) \tag{13}$$

where the scales follow half-Cauchy distributions.

$$\lambda_j \sim \mathcal{C}^+(0,1) \tag{14}$$
$$\tau_g \sim \mathcal{C}^+(0,1) \tag{15}$$

and the full conditional prior is multivariate Gaussian.

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{\beta}; \mathbf{0}, \boldsymbol{\Lambda}) \tag{16}$$

with

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1^2\tau_g^2, \ldots, \lambda_p^2\tau_g^2\right) \tag{17}$$

Thus, we have obtained the Gaussian component for the sampler with the diagonal covariance matrix $\boldsymbol{\Lambda}$ which makes computations very fast since sampling from this Gaussian is trivial and also inverting this matrix is also trivial. This contrasts with standard regression samplers which often require inverting dense matrices. Now, we can combine the model and prior using a Gibbs sampler by updating the scales $\boldsymbol{\lambda}$ and τ_g separately [6]. Our focus is the update of $\boldsymbol{\beta}$ by conditioning on the current scales $\boldsymbol{\Lambda}$. The conditional posterior is the product of two terms

$$p(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\Lambda}) \propto \mathcal{N}(\boldsymbol{\beta}; \mathbf{0}, \boldsymbol{\Lambda}) \times \exp\left(-\sum_{i=1}^{n} \rho_\tau\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right) \tag{18}$$

This matches the Elliptical Slice Sampler form and we can perform the following two steps to update $\boldsymbol{\beta}$. Firstly, we draw $\boldsymbol{\nu}$ from the prior. $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$. represents the search direction, then secondly we evaluate the current likelihood value by drawing a uniform random variable $u \sim$ Uniform [0,1]. We define the threshold y_{slice} .

$$\log\, y_{\text{slice}} = \log\, L_{curr} + \log\, u \tag{19}$$

where $L_{\text{curr}} = \mathcal{L}(\boldsymbol{\beta})$ and the log scale is used for numerical stability. The new likelihood must exceed y_{slice} for acceptance. We can pick an initial angle θ drawen uniformly from a range, $\theta \sim$ Uniform $[0,2\pi]$ defined on a bracket $[\theta_{\min}, \theta_{\max}]$, where $\theta_{\min} = \theta - 2\pi$ and $\theta_{\max} = \theta$. The last step is to evaluate $\boldsymbol{\beta}^* = \boldsymbol{\beta}\cos\,\theta + \boldsymbol{\nu}\sin\,\theta$ and check if the candidate is acceptable. If yes, we accept $\boldsymbol{\beta}^*$ and exit the loop, then update the state. If no, we shrink the bracket and repeat the loop until acceptance. While this algorithm is computationally efficient, its main cost is the likelihood evaluation since calculating $\mathcal{L}(\boldsymbol{\beta})$

takes $O(np)$ operations. Since sampling $\boldsymbol{\nu}$ takes only $O(p)$ operations then, one slice step costs $O(K \cdot np)$, where K is the number of slice evaluations which is usually is small. Also, this scaling is linear in n and p and hence it allows us to handle large datasets.

We utilize the Horseshoe prior for the coefficients $\boldsymbol{\beta}$ [1], augmented with auxiliary variables for efficient sampling [6].

Prior Hierarchy:

$$\beta_j \mid \lambda_j, \tau_g \sim \mathcal{N}\left(0, \lambda_j^2 \tau_g^2\right) \tag{20}$$

$$\lambda_j \sim \mathcal{C}^+(0,1) \Longrightarrow \lambda_j^2 \mid \nu_j \sim \mathcal{IG}\left(1/2, 1/\nu_j\right), \nu_j \sim \mathcal{IG}(1/2,1) \tag{21}$$

$$\tau_g \sim \mathcal{C}^+(0,1) \tag{22}$$

$$\tau_g^2 \mid \xi \sim \mathcal{IG}(1/2, 1/\xi), \xi \sim \mathcal{IG}(1/2,1) \tag{22}$$

The conditional posterior for λ_j^2 depends on β_j and the auxiliary variable ν_j :

$$p\left(\lambda_j^2 \mid \cdot\right) \propto \left(\lambda_j^2\right)^{-1/2} e^{-\frac{\beta_j^2}{2\lambda_j^2 \tau_g^2}} \times \left(\lambda_j^2\right)^{-3/2} e^{-\frac{1}{\nu_j \lambda_j^2}} \tag{23}$$

Combining the terms yields:

$$\lambda_j^2 \,\Big|\, \cdot \sim \mathcal{IG}\left(1, \frac{1}{\nu_j} + \frac{\beta_j^2}{2\tau_g^2}\right) \tag{24}$$

The auxiliary variable ν_j depends only on λ_j :

$$\nu_j \,\Big|\, \cdot \sim \mathcal{IG}\left(1, 1 + \frac{1}{\lambda_j^2}\right) \tag{25}$$

The conditional posterior for τ_g^2 pools information from all p coefficients:

$$p\left(\tau_g^2 \mid \cdot\right) \propto \left(\tau_g^2\right)^{-p/2} e^{-\frac{1}{2\tau_g^2} \Sigma \frac{\beta_j^2}{\lambda_j^2}} \times \left(\tau_g^2\right)^{-3/2} e^{-\frac{1}{\xi \tau_g^2}} \tag{26}$$

Combining terms yields:

$$\tau_g^2 \Big| \cdot \sim \mathcal{IG}\left(\frac{p+1}{2}, \frac{1}{\xi} + \frac{1}{2}\sum_{j=1}^{p} \frac{\beta_j^2}{\lambda_j^2}\right) \tag{27}$$

The auxiliary variable ξ follows:

$$\xi \Big| \cdot \sim \mathcal{IG}\left(1{,}1 + \frac{1}{\tau_g^2}\right) \tag{28}$$

Algorithm 1: Quantile Horseshoe Regression via ESS

Require: Data $\mathbf{y}, \mathbf{X}$, Quantile τ, Iterations N_{iter}.
Initialize: $\beta^0 = 0$, $\lambda_j^2 = 1$, $\nu_j = 1$, $\tau_g^2 = 1$, $\xi = 1$.
Define log-likelihood: $\ell(\beta) = -\sum_{i=1}^{n} \rho_\tau\left(y_i - x_i^T \beta\right)$.
for $t = 1$ to N_{iter} do
    // 1. Update Global Shrinkage
    Sample $\xi \sim \text{IG}\left(1{,}1 + \frac{1}{\tau_g^2}\right)$.
    Sample $\tau_g^2 \sim \text{IG}\left(\frac{p+1}{2}, \frac{1}{\xi} + \frac{1}{2}\sum_{j=1}^{p} \frac{\beta_j^2}{\lambda_j^2}\right)$.
    // 2. Update Local Shrinkage
    for $j = 1$ to p do
        Sample $\nu_j \sim \text{IG}\left(1{,}1 + \frac{1}{\lambda_j^2}\right)$.
        Sample $\lambda_j^2 \sim IG\left(1, \frac{1}{\nu_j} + \frac{\beta_j^2}{2\,\tau_g^2}\right)$.
    end for
    // 3. Update Coefficients via Elliptical Slice Sampler
    Set $f = \beta^{t-1}$.
    Set covariance $\Sigma = diag(\lambda_1^2\,\tau_g^2, \ldots, \lambda_p^2\,\tau_g^2)$.
    Sample proposal $\nu \sim N(0, \Sigma)$.
    Sample $u \sim \text{Uniform}[0,1]$ and set $y_{log} = \ell(f) + \log(u)$.
    Sample $\theta \sim \text{Uniform}[0, 2\pi]$.
    Set $[\theta_{\min}, \theta_{\max}] = [\theta - 2\pi, \theta]$.
    loop
        $f^* = f\cos\theta + \nu\sin\theta$.

21: if $\ell(f^*) > y_{log}$ then
22: $\beta^t = f^*$.
23: break loop.
24: else
25: if $\theta < 0$ then
26: $\theta_{\min} = \theta$
27: Else
28: $\theta_{\max} = \theta$
29: end if
30: Sample $\theta \sim$ Uniform$[\theta_{\min}, \theta_{\max}]$.
31: end if
32: end loop
33: end for

3. Simulation Studies and Real Data Application

In this section, we will show the attractive properties of our presentation using a simulation study to evaluate the effectiveness of our proposed algorithm. The primary goal is to show how our method does against other established Bayesian methods. We will compare the Elliptical Slice Sampler to standard methods like the quantile Lasso, quantile Ridge, and quantile Elastic Net. The main competitor in many high-dimensional settings is Gibbs sampler which is a traditional approach for Bayesian quantile regression [4]. The focus will be to test our implementation on the Horseshoe prior [1]. We want to show that this robust and highly sparse prior is an ideal choice for modern high-dimensional scenarios. We aim to demonstrate that our method generates independent samples much faster than existing samplers. Also, our method scales better with increasing dimension through testing the varying of the number of predictors in our experiments.

We compare several competing algorithms in our study. The first is our proposed Quantile Horseshoe with ESS. This is a slice-within-Gibbs sampler. It uses the Horseshoe prior for global and local shrinkage. It uses the Elliptical Slice Sampler to update the coefficient vector $\boldsymbol{\beta}$ in a single block. We use standard slice samplers for the remaining hyperparameters and implemented this code in C++ using the Rcpp interface for speed. The second algorithm is

the Quantile Lasso using a standard Gibbs sampler. This is the primary benchmark method in the literature, where uses a Laplace prior, which is the Bayesian version of the Lasso [7] through a data augmentation method [4] by introducing latent variables z_i to make the problem tractable. However, this solution requires inverting X at every iteration. We use the R package bayesQR for this comparison [8] and examine the quantile Ridge and Elastic Net options. These techniques use distinct priors but frequently encounter the same computing delays. The computing time for every Gibbs-based approach needs massive matrix inversions which makes our comparison fair regarding raw execution speed. We will asses the performance using the estimation error, then measure how accurately each algorithm recovers the true coefficients using the MSE by

$$\text{MSE} = \frac{1}{p}\sum_{j=1}^{p}\left(\hat{\beta}_j - \beta_{j,\,\text{true}}\right)^2 \tag{29}$$

where $\hat{\boldsymbol{\beta}}$ is the posterior mean. MCMC samples are naturally correlated with one another, thus the Effective Sample Size (ESS) is used to estimate the number of independent samples in the chain. Let ρ_k be the autocorrelation at lag k. The formula for ESS is:

$$\text{ESS} = \frac{M}{1 + 2\sum_{k=1}^{\infty}\rho_k} \tag{30}$$

Here M is the total number of iterations after burn-in. We report the median ESS across all coefficients using the method of [9]. The third metric is computational efficiency, defined as ESS per second:

$$\text{Efficiency} = \frac{\text{ESS}}{T} \tag{31}$$

Firstly, we will simulate the predictor matrix **X** using two distinct structures. The first structure involves independent predictors. We draw each entry from a standard Normal distribution. This means $x_{i,j} \sim \mathcal{N}(0,1)$. In this scenario, the predictors are uncorrelated. This is generally the easiest scenario for MCMC chains to mix well. The results show that all methods eventually

converge to the true posterior. The MSE values were very comparable in small dimensions ($p = 50$). However, for $p = 500$, the Horseshoe with ESS was slightly more accurate. This result is not surprising as the Horseshoe prior is better at shrinking noise than the Lasso prior, which tends to over-shrink large signals.

For our second structure, will involve correlated predictors since real-world datasets often have highly correlated features. We will generate these rows from a multivariate

	p	MSE (sd)	MSE (sd)	MSE (sd)
HS.Slice	50	0.0164 (0.0105)	0.0000 (0.0000)	0.0000 (0.0000)
QRIDGE	50	0.0689 (0.0382)	0.1000 (0.3162)	0.0000 (0.0000)
QLASSO	50	0.0599 (0.0311)	0.1000 (0.3162)	0.0000 (0.0000)
QENT	50	0.0637 (0.0353)	0.1000 (0.3162)	0.0000 (0.0000)
HS.Slice	500	0.0163 (0.0145)	0.0000 (0.0000)	0.0000 (0.0000)
QRIDGE	500	0.1697 (0.0499)	1.4000 (1.1738)	0.0000 (0.0000)
QLASSO	500	0.1357 (0.0443)	1.1000 (1.1972)	0.0000 (0.0000)
QENT	500	0.1408 (0.0438)	1.0000 (0.9428)	0.0000 (0.0000)

Normal distribution, specifically, $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. The covariance matrix $\mathbf{\Sigma}$ has a specific autoregressive form. The entry Σ_{jk} depends on the distance between

indices $|j-k|$, with $\Sigma_{jk} = 0.5^{|j-k|}$. This creates strong local correlations such that correlations make variable selection much harder for any algorithm because it slows down the mixing of standard MCMC chains significantly. We define the true coefficient vector $\boldsymbol{\beta}_{\text{true}}$ to simulate a sparse signal. This sparseness is the exact reason we select the Horseshoe prior. The sample size n remains at 200 . The vast majority of the values in $\boldsymbol{\beta}_{\text{true}}$ are fixed at zero. Only a small subset of the predictors is truly active. We let S be the set of active indices. We choose the size of S to be exactly 10 . For any $j \in S$, we set $\beta_j = 3$. For any $j \notin S$, we set $\beta_j = 0$. This setup creates a very clear signal-to-noise separation.

	p	MSE (sd)	MSE (sd)	MSE (sd)
HS.Slice	50	0.0281 (0.0241)	0.0000 (0.0000)	0.0000 (0.0000)
QRIDGE	50	0.1027 (0.0449)	0.4000 (0.6992)	0.0000 (0.0000)
QLASSO	50	0.0869 (0.0394)	0.3000 (0.4830)	0.0000 (0.0000)
QENT	50	0.0915 (0.0417)	0.3000 (0.4830)	0.0000 (0.0000)
HS.Slice. 1	500	0.0292 (0.0169)	0.0000 (0.0000)	0.0000 (0.0000)
QRIDGE. 1	500	0.2074 (0.0610)	0.7000 (1.0593)	0.0000 (0.0000)
QLASSO. 1	500	0.1579 (0.0462)	0.4000 (0.8433)	0.0000 (0.0000)
QENT. 1	500	0.1678 (0.0495)	0.5000 (1.0801)	0.0000 (0.0000)

In the high-dimensional case ($p = 500$), the difference becomes noticable since the sampler becomes extremely slow because it must invert a 500×500 matrix at every iteration, a $O(p^3)$ operation. Our ESS method stays remarkably quick, possessing a complexity of $O(np)$. We demonstrate out our method to the diabetes dataset [10] which contains data for 442 patients. We construct a complex, high-dimensional design matrix X by including squared terms and pairwise interactions. This expands our feature space to 64 predictors, to which we then add 500 noise variables, resulting in a total of $p = 564$. This specific $p > n$ condition causes many traditional statistical methods to fail. The shrinkage factor κ_j will be defined as:

$$\kappa_j = \frac{1}{1 + n\tau_g^2\lambda_j^2} \tag{32}$$

where the variable j is select if the average $\kappa_j < 0.5$. The Quantile Horseshoe reached the best result in comparison with other methods. Crucially, the method rejected the noise variables, confirming the Horseshoe prior here deals best with sparisty. This application proves the value of our method in handling high-dimensional, messy data efficiently.

	τ	MSE (sd)	FPR (sd)	FNR (sd)
HS.Slice	0.1	0.0179 (0.0102)	0.0000 (0.0000)	0.0000 (0.0000)
QRIDGE	0.1	0.1191 (0.0658)	0.8000 (0.8367)	0.0000 (0.0000)
QLASSO	0.1	0.0994 (0.0474)	0.6000 (0.5477)	0.0000 (0.0000)
QENT	0.1	0.1072 (0.0534)	0.8000 (0.4472)	0.0000 (0.0000)
HS.Slice. 1	0.5	0.0175 (0.0057)	0.0000 (0.0000)	0.0000 (0.0000)

QRIDGE. 1	0.5	0.3013 (0.0398)	1.8000 (1.3038)	0.0000 (0.0000)
QLASSO. 1	0.5	0.2341 (0.0389)	1.0000 (0.7071)	0.0000 (0.0000)
QENT. 1	0.5	0.2453 (0.0382)	1.4000 (0.8944)	0.0000 (0.0000)
BP Slice	0.9	0.0173 (0.0050)	0.0000 (0.0000)	0.0000 (0.0000)
QRIDGE	0.9	0.0955 (0.0242)	0.4000 (0.5477)	0.0000 (0.0000)
QLASSO	0.9	0.0807 (0.0176)	0.2000 (0.4472)	0.0000 (0.0000)
QENT	0.9	0.0849 (0.0192)	0.2000 (0.4472)	0.0000 (0.0000)

4. Discussion and Future Directions

We have provided a practical tool that can be used to solve real computational problems. Bayesian Quantile Regression is now faster and accessible for high-dimensional data.We have presented a new algorithm to address the scenario of Bayesian quantile regression with special focus on high-dimensional variable selection. We utilized the Elliptical Slice Sampler (ESS) [5] by combining it with the Horseshoe prior [1] to produce a powerful tool. This method works in contrast standard methods that rely on Gibbs sampling by data augmentation of [4] which requires sampling latent variables inverting a large covariance matrix that scales cubically $O(p^3)$ which is problematic for large p. We have treated the prior as the proposal distribution and evaluated the likelihood directly where the cost is linear in dimensions, i.e. $O(np)$ making it much easier for scalable inference.

The success of this method implies that one should decouple the prior from the likelihood, unlike in Gibbs sampling where they are coupled leading restriction on the choice of priors, such as forcing the use of conjugate priors. In this method, the prior determines the proposal ellipse and the likelihood determines the acceptance making it particularly useful for quantile regression [11] since it works for any prior with a Gaussian form works including the entire class of scale mixtures. We have used the Horseshoe, but one could use other priors, such as, the Dirichlet-Laplace [12], the Normal-Gamma [13,14], since the algorithm remains exactly the same. Future work may focus on studying different types of priors, in the scenario of high-dimensionality and its respective perfomance with respect the the method presented here.

References

[1] Carlos M Carvalho, Nicholas G Polson, and James G Scott. The horseshoe estimator for sparse signals. Biometrika, 97(2):465-480, 2010.

[2] Roger Koenker. Quantile Regression. Cambridge University Press, 2005.

[3] Keming Yu and Rana A Moyeed. Bayesian quantile regression. Statistics & Probability Letters, 54(4):437-447, 2001.

[4] Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. Journal of Statistical Computation and Simulation, 81(11):1565-1578, 2011.

[5] Iain Murray, Ryan P Adams, and David JC MacKay. Elliptical slice sampling. In Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, volume 9, pages 541-548, 2010.

[6] Enes Makalic and Daniel F Schmidt. High-dimensional bayesian linear regression with the horseshoe prior. IEEE Signal Processing Letters, 23(1):1-5, 2016.

[7] Trevor Park and George Casella. The bayesian lasso. Journal of the American Statistical Association, 103(482):681-686, 2008.

[8] Dries F Benoit and Dirk Van den Poel. bayesqr: A bayesian quantile regression package for r. Journal of Statistical Software, 76(1):1-32, 2017.

[9] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. Bayesian Data Analysis, Third Edition. CRC Press, 2013.

[10] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. The Annals of Statistics, 32(2):407-499, 2004.

[11] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. arXiv preprint arXiv:1701.02434, 2017.

[12] Anirban Bhattacharya, Debdeep Pati, Natesh S Pillai, and David B Dunson. Dirichlet-laplace priors for optimal shrinkage. Journal of the American Statistical Association, 110(512):1479-1490, 2015.

[13] Jim E Griffin, Philip J Brown, et al. Inference with normal-gamma prior distributions in regression problems. Bayesian Analysis, 5(1):171-188, 2010.

[14] Alhamzawi, A., 2020, November. A new Bayesian elastic net for Tobit regression. In Journal of Physics: Conference Series (Vol. 1664, No. 1, p. 012047). IOP Publishing.