Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 199–209
http://www.ssca.org.in/journal

Horseshoe Prior for Bayesian Linear Regression with Hyperbolic Errors

Shamriddha De and Joyee Ghosh
Department of Statistics and Actuarial Science, The University of Iowa

Received: 31 May 2024; Revised: 18 July 2024; Accepted: 18 July 2024

Abstract

It is well known that squared error loss is not robust to outliers. As an alternative, Huber loss may be used for robust regression; however, Huber loss is not readily amenable to Bayesian computation. It has been shown that hyperbolic loss can be regarded as an approximation to Huber loss, and the hyperbolic distribution can be expressed as a scale mixture of normal distributions, which makes it appealing for Bayesian computation. The idea of Bayesian Huberized lasso was first proposed by Park and Casella (2008), and was formally developed and implemented by Kawakami and Hashimoto (2023). Since the Bayesian Huberized lasso cannot shrink regression coefficients to exactly zero, and has lighter tailed errors than a Cauchy distribution, De and Ghosh (2024) proposed a model that encompasses both hyperbolic and t-errors, with a mixture prior on regression coefficients consisting of two parts, a point mass at zero and a continuous distribution, that can shrink coefficients to exactly zero. The approach of De and Ghosh (2024) could be considered as a gold standard for Bayesian model averaging, but posterior computation with such a point mass mixture prior, popularly known as the spike and slab prior, can be challenging with many covariates. The horseshoe prior is known to mimic some of the desirable properties of spike and slab priors, while being computationally less intensive. Motivated by this attractive property of the horseshoe prior, in this article we develop an algorithm for Bayesian linear regression with hyperbolic errors, and horseshoe priors on the regression coefficients. We illustrate using simulation studies and an analysis of the famous Boston housing dataset, that posterior distributions under horseshoe priors can capture sparsity better than Bayesian lasso priors. For moderate dimensional regression problems, the spike and slab prior performs better than the horseshoe in capturing the sparsity of regression coefficients. However, we find that Markov chain Monte Carlo (MCMC) algorithms with horseshoe priors have improved mixing, which suggests that Bayesian shrinkage with the horseshoe prior and its generalizations, such as the regularized horseshoe prior, could be a promising direction to explore for high dimensional robust regression.

Key words: Bayesian lasso; Markov chain Monte Carlo; Model averaging; Robust regression; Spike and slab prior; Variable selection.

Corresponding Author: Joyee Ghosh
Email: joyee-ghosh@uiowa.edu

1. Introduction

The majority of Bayesian variable selection methods for linear regression have focused on normal errors, which is a challenging problem in its own right, especially for high dimensional problems. Since estimates derived under the normality assumption for errors can be sensitive to outliers, our goal is to robustify the error distribution. The Bayesian variable selection method for linear regression with normal errors can adapt to an unknown degree of sparsity by placing a prior on the unknown inclusion probability of variables. De and Ghosh (2024) developed a model with additional flexibility by allowing the likelihood to simultaneously adapt to an unknown degree of tail heaviness. They focused on the class of scale mixtures of normal densities for robust error distributions. The availability of the scale mixture of normal representation of the heavy tailed error models makes it convenient to implement MCMC sampling algorithms. Let the error distribution have the following form:

$$p(\epsilon) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\epsilon^2}{2\sigma^2}} dF(\sigma^2), \tag{1}$$

such that $F(.)$ is a cumulative distribution function (CDF). Then the random error ϵ is said to follow a scale (or variance) mixture of normals, and $F(.)$ is called a mixing distribution. Some popular distributions that can be represented by the scale mixtures of normal representation are the hyperbolic, Student-t, Laplace (double exponential), exponential power *etc.* (Andrews and Mallows (1974); West (1987); Gneiting (1997)).

In particular, the hyperbolic distribution forms a vital point of attention in this article. The said distribution has the following probability density function:

$$p_h(\epsilon; \eta, \rho^2) = \frac{1}{2\sqrt{\eta\rho^2} K_1(\eta)} e^{-\left(\eta\left(\eta + \frac{\epsilon^2}{\rho^2}\right)\right)^{1/2}}, \quad -\infty < \epsilon < \infty, \tag{2}$$

where $K_1(.)$ is a modified Bessel function, $\eta > 0$ is the shape parameter regulating the tail heaviness and $\rho^2 > 0$ is the scale parameter. Gneiting (1997) showed that the above distribution can be represented as a generalized inverse Gaussian (GIG) scale mixture of normally distributed random variables, and thus the hyperbolic distribution belongs to the family of scale mixture of normal distributions, defined in (1). We provide more detail in Section 2 about this representation. In a regression problem, using a hyperbolic error model is equivalent to using a hyperbolic loss function. Additionally, the hyperbolic loss has similarities with the Huber loss (Park and Casella (2008)). The Huber loss is popular for robust regression in the frequentist literature but it is computationally difficult to handle in a Bayesian set up. Accordingly, in a Bayesian setting, we focus on the hyperbolic loss as an alternative to the Huber loss, like previous authors (Park and Casella (2008); Kawakami and Hashimoto (2023); De and Ghosh (2024)).

Bayesian variable selection with two component mixture priors used by De and Ghosh (2024) leads to a vast model space, when the number of covariates is large. An alternative strategy that has been shown to perform favorably is using a continuous shrinkage prior to replace the mixture priors. For example, the Bayesian lasso (Park and Casella (2008)) is a continuous shrinkage prior, which has been implemented by Kawakami and Hashimoto (2023), for regression models with hyperbolic errors. Another well known technique is to use the Bayesian horseshoe prior (Carvalho *et al.* (2010), Makalic and Schmidt (2015), Bhadra

et al. (2017)), which also belongs to the family of continuous shrinkage priors and has been demonstrated to perform very well for shrinking noise variables to practically zero, while keeping signals almost intact for the normal means problem. For a $p \times 1$ vector $\boldsymbol{\beta}$ of regression coefficients, the horseshoe prior is defined as

$$(\beta_j \mid \lambda_j, \tau^2, \rho^2) \overset{\text{ind}}{\sim} N(0, \lambda_j^2 \tau^2 \rho^2), \quad \lambda_j \overset{\text{ind}}{\sim} C^+(0,1), \quad \tau \sim C^+(0,1), \tag{3}$$

for $j = 1, 2, \ldots, p$, where ρ^2 is the scale of the error distribution, and $C^+(0,1)$ represents the standard half-Cauchy distribution with the density function

$$p(x) = \frac{2}{\pi(1+x^2)}, \quad x > 0.$$

In the context of regression models with heavy tailed errors, the horseshoe prior has been utilized by Hamura *et al.* (2022). However, the focus of their paper is on super-heavy tailed error distributions, in comparison to which even the Student-t distribution is regarded as a thin tailed distribution. Hamura *et al.* (2022) considered the horseshoe prior for illustration for some applications, but most of the paper focuses on multivariate normal priors for regression coefficients. In contrast, this article focuses on hyperbolic errors as a proxy to the Huberized loss function. The main question of interest that we try to investigate is how methods based on horseshoe priors compare with those based on lasso, and spike and slab priors, under varying levels of sparsity.

The article is organized as follows. In Section 2, we introduce the hyperbolic distribution and horseshoe priors, and develop an algorithm for posterior computation. In Section 3 we conduct simulation studies with the true model having hyperbolic errors, and compare the results of the posterior estimates from the horseshoe prior versus the spike and slab prior (De and Ghosh (2024)) and the Bayesian lasso prior (Kawakami and Hashimoto (2023)). In Section 4, we analyze the famous Boston housing data with the three aforementioned priors, after adding noise variables to the original dataset. Finally, in Section 5, we provide a summary of the results, and discuss some future directions.

2. Hyperbolic error model with horseshoe prior

Let $\mathbf{Y}$, $\mathbf{X}$ and $\boldsymbol{\beta}$ denote the $n \times 1$ vector of response variables, the $n \times p$ design matrix, and the $p \times 1$ vector of regression coefficients, respectively. We consider a regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{4}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ is the $n \times 1$ vector of errors, such that $\epsilon_i \overset{\text{iid}}{\sim} p_h(\epsilon_i; \eta, \rho^2)$, $i = 1, 2, \ldots, n$, where $p_h(.; \eta, \rho^2)$ is the hyperbolic density with parameters η and ρ^2 defined in (2). Park and Casella (2008) showed that the normal scale-mixture representation by Gneiting (1997) leads to the representation of (4) in a computationally convenient form as

$$\mathbf{Y} \mid \boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2 \;\sim\; \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{D}), \tag{5}$$

$$p(\sigma_1^2, \ldots, \sigma_n^2 | \eta, \rho^2) \;=\; \prod_{i=1}^{n} \frac{1}{2K_1(\eta)\rho^2} e^{-\frac{\eta}{2}\left(\frac{\sigma_i^2}{\rho^2} + \frac{\rho^2}{\sigma_i^2}\right)}, \tag{6}$$

where $\boldsymbol{D} = \mathsf{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$, and $\mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{D})$ denotes the multivariate normal distribution with mean and variance covariance matrix as $\mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{D}$, respectively. The diagonal elements of $\boldsymbol{D}$ have independent GIG distributions. Marginalizing out these scale parameters will yield a likelihood with independent hyperbolic errors with the form given in (2). The aforementioned normal scale-mixture representation of the hyperbolic error model is an important computational trick for developing a Gibbs sampling algorithm for posterior computation in our subsequent Bayesian analysis.

A Bayesian approach requires putting suitable priors on all unknown parameters. For the above model, this requires putting priors on the vector of regression coefficients, $\boldsymbol{\beta}$, as well as on the two other model parameters, namely, η and ρ^2, which correspond to the error distribution. In this article, our goal is to use the horseshoe prior on regression coefficients in conjunction with an hyperbolic error model. To that end, we use the following hierarchical representation of the horseshoe prior in (3), proposed by Makalic and Schmidt (2015), which facilitates posterior computation via Gibbs sampling. In particular, this hierarchical representation leads to closed form full conditional distributions for all unknown parameters, which is a crucial step for the subsequent Bayesian analysis.

$$\begin{aligned}
p(\boldsymbol{\beta}|\lambda_1^2, \lambda_2^2, \ldots, \lambda_p^2, \tau^2, \rho^2) &= \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\rho^2\tau^2\lambda_j^2}} e^{-\frac{1}{2}\left(\frac{\beta_j^2}{\rho^2\tau^2\lambda_j^2}\right)}, \\
p(\lambda_1^2 \ldots, \lambda_p^2|\nu_1, \ldots \nu_p) &= \prod_{j=1}^{p} \frac{(1/\nu_j)^{1/2}}{\Gamma(1/2)} e^{-\frac{1}{\lambda_j^2\nu_j}} \left(\frac{1}{\lambda_j^2}\right)^{\frac{1}{2}+1}, \\
p(\nu_1, \ldots, \nu_p) &= \prod_{j=1}^{p} \frac{1}{\Gamma(1/2)} e^{-\frac{1}{\nu_j}} \left(\frac{1}{\nu_j}\right)^{\frac{1}{2}+1}, \\
p(\tau^2|\xi) &= \frac{(1/\xi)^{1/2}}{\Gamma(1/2)} e^{-\frac{1}{\tau^2\xi}} \left(\frac{1}{\tau^2}\right)^{\frac{1}{2}+1}, \\
p(\xi) &= \frac{1}{\Gamma(1/2)} e^{-\frac{1}{\xi}} \left(\frac{1}{\xi}\right)^{\frac{1}{2}+1}. \qquad (7)
\end{aligned}$$

The hierarchical prior structure in (7) is equivalent to the horseshoe prior in (3) on the regression coefficients, upon marginalization over ν_j's ($j = 1, 2, \ldots, p$) and ξ. As far as the scale parameter ρ^2 and the shape parameter η of the error density are concerned, we put the following priors:

$$\begin{aligned}
p(\rho^2) &= \frac{b^a}{\Gamma(a)} (\rho^2)^{-(a+1)} e^{-b/\rho^2}, \\
p(\eta) &= \frac{1}{K}, \text{ for } \eta \in \{\eta_1, \ldots, \eta_K\}. \qquad (8)
\end{aligned}$$

To reduce ambiguity about different forms of parametrizations, we have directly specified the probability density function (pdf) or probability mass function (pmf) in the above prior specification. In particular, we have specified a conditional normal prior on the regression coefficients, inverse gamma priors on the scale parameters, and a discrete uniform prior on the tail heaviness parameter η. The full conditional distributions corresponding to the above priors lead to standard distributions from which sampling is straightforward. We use Gibbs sampling to approximately sample from the joint posterior distribution.

3. Simulation study

In this section, we generate data from models with hyperbolic errors, and compare the performances of posterior distributions under the horseshoe, lasso, and spike and slab priors. We consider two cases as follows.

3.1. Sparse true model

We first consider an example with $n = 100$ observations and $p = 15$ (excluding the intercept). We generate the errors from a hyperbolic distribution with $\eta = 0.5$ and $\rho^2 = 2$. We set the intercept equal to 2 and specify a relatively sparse model with 5 nonzero regression coefficients, all equal to 3. We generate 100 datasets from this model. We set the priors and hyperparameters for lasso and spike and slab priors following De and Ghosh (2024), denoted by them as HBL (Bayesian Huberized lasso) and HEM (hyperbolic error model), respectively. For the priors proposed by us in this article, given in (7) and (8), we use the same hyperparameters for the tail heaviness and scale parameters, η and ρ^2, respectively, as the other two priors. In particular, we standardize the response variables and each column of the design matrix, to have mean and standard deviation equal to 0 and 1, respectively. We set $a = 2.1$ and $b = 0.1$ for the hyperparameters of the inverse gamma prior on ρ^2, to have most of the prior mass between 0 and 1. This choice is not unreasonable as the response variables have been standardized to have variance equal to 1. For the tail heaviness parameter η, we specify the support points as $\{0.05, 0.1, 0.2, 0.3, \ldots, 1, 2, 5, 10, 20, 50\}$, following De and Ghosh (2024), to have a wide range of tail heaviness parameters. We run the MCMC algorithms for 100,000 iterations, after a burnin of 10,000 iterations. We estimate the regression coefficients using posterior medians of the MCMC samples.

The results are summarized in Figures 1 and 2. The top left panel in Figure 1 shows the root mean squared error (RMSE) for signals (nonzero regression coefficients, excluding the intercept term), that is

$$\sqrt{\sum_{\substack{j=1 \\ \beta_j \neq 0}}^{p} \left(\beta_j - \hat{\beta}_j\right)^2 / 5},$$

where $\hat{\beta}_j$ is the estimate of β_j, $j = 1, \ldots p$, and there are 5 nonzero regression coefficients. This RMSE is similar for the horseshoe and lasso, and somewhat better for the spike and slab prior. The top right panel shows the RMSE for 10 noise variables (zero regression coefficients), that is

$$\sqrt{\sum_{\substack{j=1 \\ \beta_j = 0}}^{p} \left(\beta_j - \hat{\beta}_j\right)^2 / 10}.$$

This is where the spike and slab prior shines, and the horseshoe is significantly better than the lasso, though not as good as the spike and slab prior. The bottom left panel shows the overall RMSE in estimating all regression coefficients (including the intercept β_0), given by

$$\sqrt{\sum_{j=0}^{p} \left(\beta_j - \hat{\beta}_j\right)^2 / (p+1)}.$$

The bottom right panel shows the overall RMSE for each method, relative to the RMSE of the method that has the smallest RMSE for that dataset. The relative RMSE for the spike and slab prior is concentrated around 1, which shows it is the best method overall, followed by the horseshoe, which also seems significantly better than the lasso prior.

Figure 2 shows the effective sample size (ESS), for the MCMC samples of the regression coefficients. ESS can be used to quantify the mixing in the Markov chain, and larger values are preferable. For example, for independent Monte Carlo sampling, the values of ESS would be equal to 100,000, the actual Monte Carlo sample size. For both signals and noise variables, the lasso has the largest ESS, followed by the horseshoe, and the spike and slab priors. Spike and slab priors are known to have slow mixing, so the results are in agreement with this well known fact.

3.2. Non-sparse true model

We next turn to a non-sparse data generating model, with many nonzero regression coefficients. The spike and slab and horseshoe priors are not expected to have as much of an advantage over lasso, in this set up, as they had enjoyed in the previous sparse set up. Here we set $n = 200$ observations and $p = 30$. We set the intercept equal to 2 as earlier, and specify a non-sparse model with 20 nonzero regression coefficients, all equal to 0.8. Everything else is specified as in the earlier simulation study.

The results are presented in Figures 3 and 4. The advantage of the horseshoe prior over the lasso prior disappears in this example, and while the spike and slab prior seems to be the best overall from the bottom right panel in Figure 3, its gains over the other methods is much reduced in this example. This is expected, due to the relatively non-sparse nature of this example. Figure 4 shows that lasso still has the largest values of ESS for both signals and noise variables. Thus this example illustrates scenarios where the lasso prior could be preferable, compared to spike and slab and horseshoe priors.

4. Application to boston housing dataset

We use the Boston housing dataset, available from the `MASS` package in `R`. This dataset is known to be heavy tailed compared to a normal distribution, and has been extensively used as a benchmark dataset in the literature, to illustrate the performance of methods in robust regression. The dataset has $n = 506$ observations and $p = 13$ covariates. The response variable is the median value of occupied homes in Boston, and the covariates are crime rate, property tax, distance to Boston employment centers, access to highways *etc.* We use a log transformation on the response variable, so that the distribution of residuals is roughly symmetric.

A preliminary frequentist linear regression analysis with the usual assumption of normal errors shows that most of the variables have significant p-values; or in other words, the vector of regression coefficients is not sparse. We have illustrated in the second example of the simulation study, that spike and slab and horseshoe priors are are not expected to have much of an advantage in such non-sparse scenarios. To make the application more interesting, we add 30 noise variables, generated from a normal distribution with mean 0 and standard deviation 1, to the original dataset, to have a total of $p = 43$ covariates.

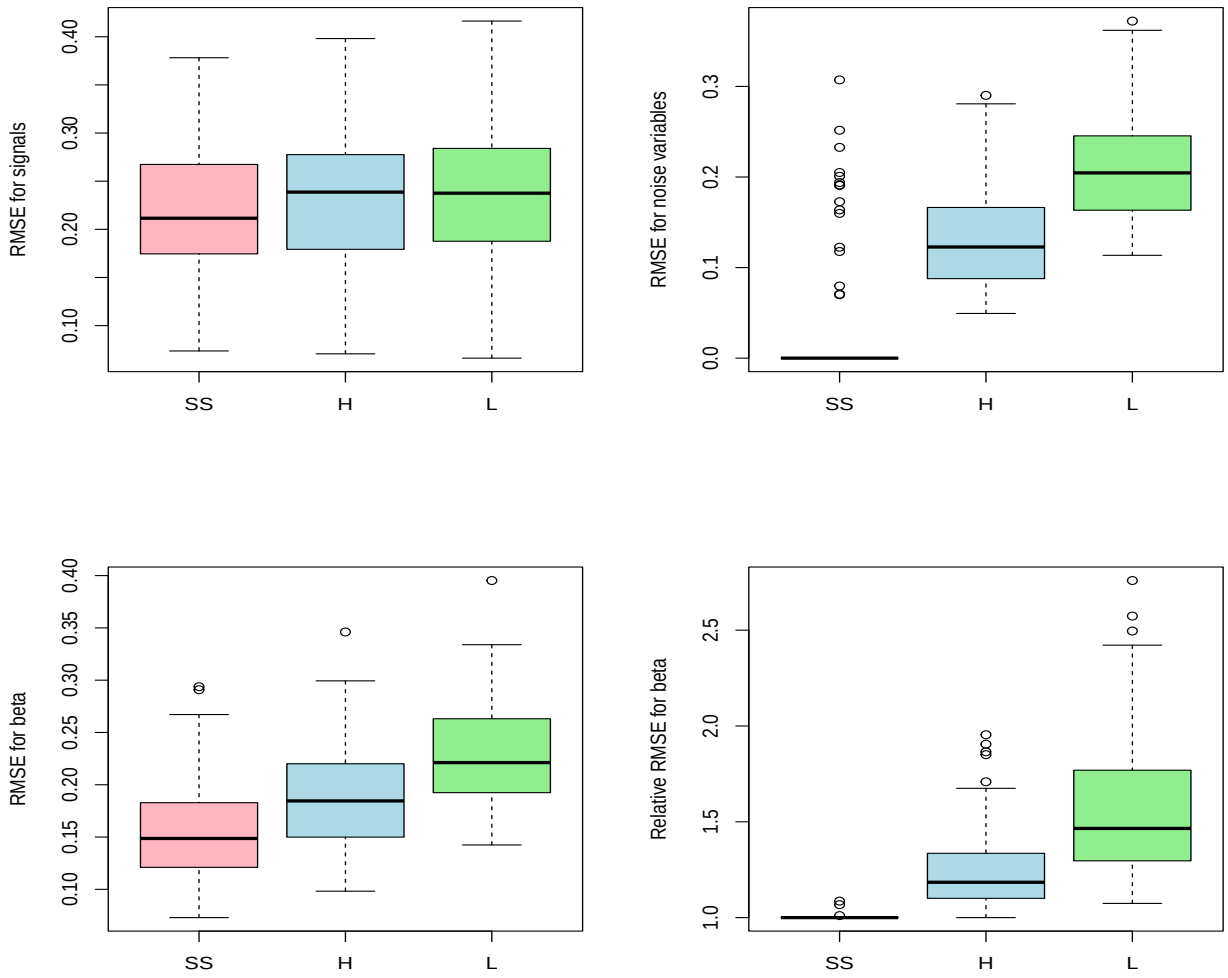

Figure 1: Results for simulation study under sparse true model with $p = 15$ and $n = 100$. Box plots in the top panel show the root mean squared error (RMSE) corresponding to spike and slab (SS), horseshoe (H), and lasso (L) priors for 100 datasets, for signals and noise variables respectively. Box plots in the bottom left panel show the overall RMSE in estimating all the regression coefficients, including the intercept. Box plots in the bottom right panel show the overall RMSE relative to the RMSE of the best method; values of relative RMSE close to 1 indicate that the method is frequently the best.

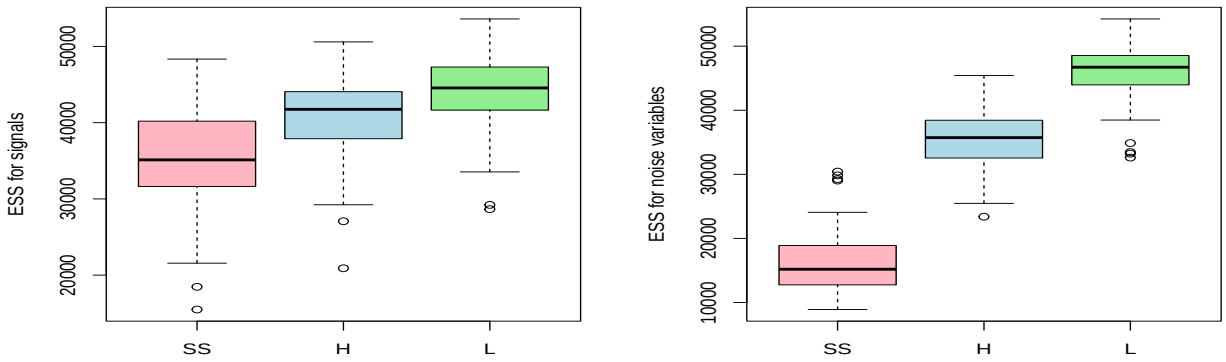

Figure 2: Results for simulation study under sparse true model with $p = 15$ and $n = 100$. Box plots show the effective sample size (ESS) of the MCMC samples for the regression coefficients, corresponding to spike and slab (SS), horseshoe (H), and lasso (L) priors for 100 datasets, for signals and noise variables respectively.

We randomly choose 50% of the observations to include in a training dataset, and use the remaining 50% as a test dataset, to evaluate out of sample predictive performance of the methods with different priors. To reduce sensitivity of the results to a specific choice of training and test data split, we repeat the process 100 times, to create 100 different training and test datasets.

We evaluate the predictive performance using both point and interval estimates. For point estimate for prediction, we use the median of the posterior predictive distribution. For each of the 253 observations in a test dataset, we compute the absolute difference between the observed value of the response variable and its predicted value (both on log scale), and then compute the median of these differences, which we refer to as Median absolute deviation (MAD) for prediction. For 100 test datasets, we get 100 values of MAD. For each test dataset, the method with the smallest MAD is deemed to have the best MAD, and the MAD for the other methods are compared relative to the best MAD. This is repeated 100 times, and presented in the left panel of Figure 5. If a method has values close to 1, that indicates the method has the smallest MAD frequently. The difference between the methods based on different priors is not large, but overall, the spike and slab prior seems to be the best with smallest values of MAD, followed by the horseshoe, and then by the lasso prior. We next consider interval estimates for prediction by estimating 90% equal-tailed prediction intervals for each observation in the test datasets. The resulting frequentist coverage of the prediction intervals is shown in the right panel of Figure 5. All methods seem to have coverage close to 90%, shown by the dashed line, though there is some variability around 90%. The diamond in each box plot shows the overall coverage across 100 test datasets, which seems fairly close to 90%.

5. Discussion

In this article, we have introduced an algorithm based on the horseshoe prior, for robust regression with hyperbolic errors, as an alternative to existing methods that rely on

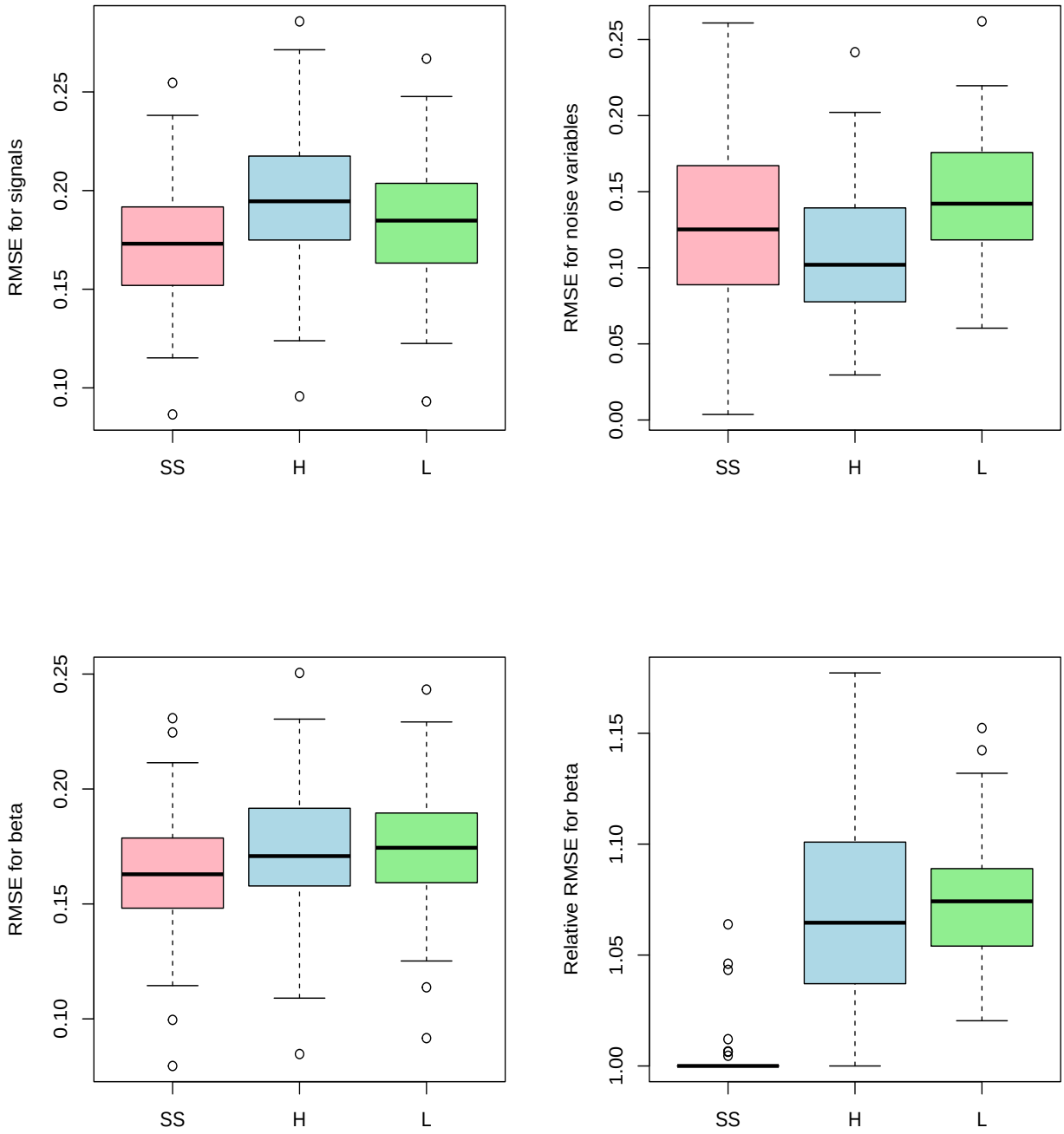

Figure 3: Results for simulation study under non-sparse true model with $p = 30$ and $n = 200$. Box plots in the top panel show the root mean squared error (RMSE) corresponding to spike and slab (SS), horseshoe (H), and lasso (L) priors for 100 datasets, for signals and noise variables respectively. Box plots in the bottom left panel show the overall RMSE in estimating all the regression coefficients, including the intercept. Box plots in the bottom right panel show the overall RMSE relative to the RMSE of the best method; values of relative RMSE close to 1 indicate that the method is frequently the best.

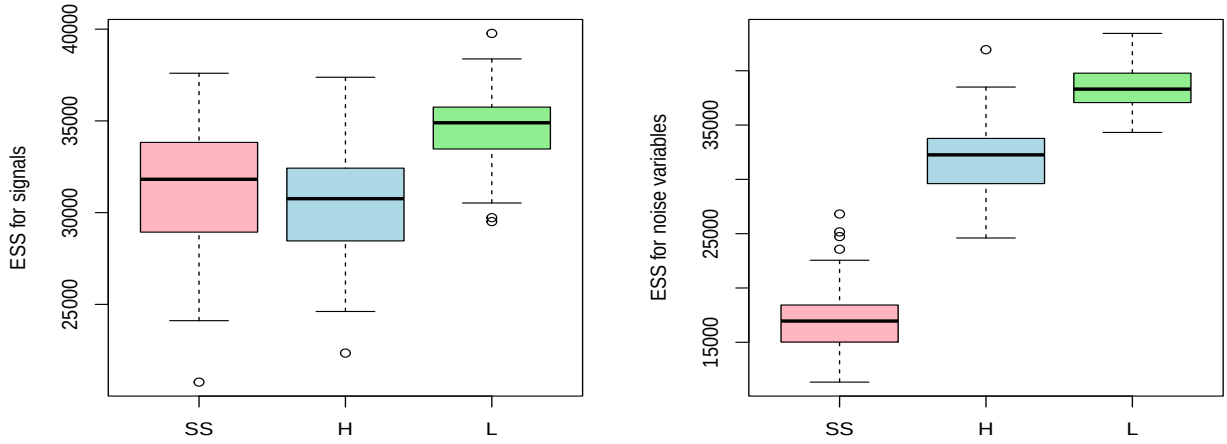

Figure 4: Results for simulation study under non-sparse true model with $p = 30$ and $n = 200$. Box plots show the effective sample size (ESS) of the MCMC samples for the regression coefficients, corresponding to spike and slab (SS), horseshoe (H), and lasso (L) priors for 100 datasets, for signals and noise variables respectively.

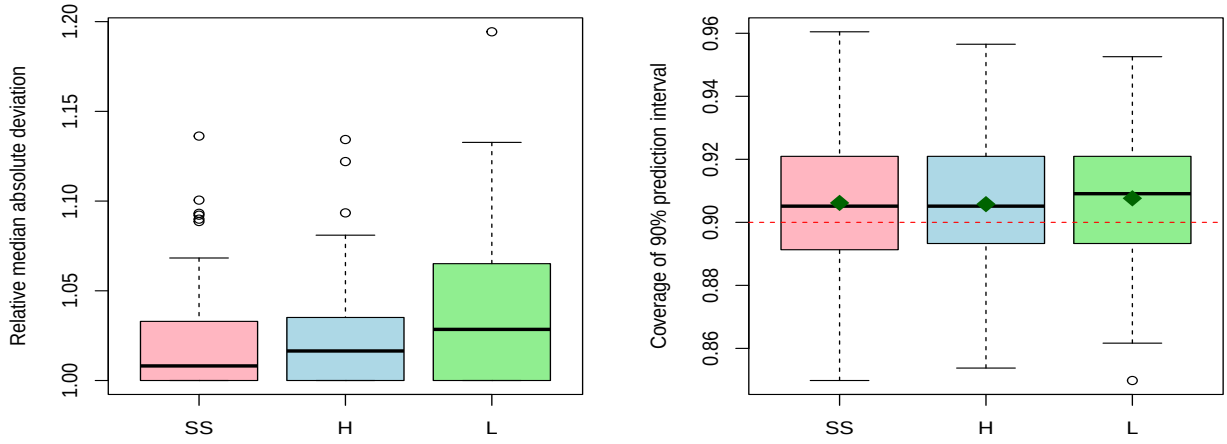

Figure 5: Results for the Boston housing dataset, after adding 30 noise variables to the original dataset with $p = 13$ covariates. Box plots in the left panel show the median absolute deviation (MAD) for out of sample prediction, corresponding to spike and slab (SS), horseshoe (H), and lasso (L) priors relative to the method with the least MAD for that dataset, for 100 test datasets. The right panel shows the corresponding coverage for 90% prediction intervals; the dashed line is at 0.9 and the diamonds represent the overall mean coverage over 100 test datasets.

the spike and slab or lasso priors. Our results based on simulation studies suggest that the horseshoe prior can improve upon the lasso prior in estimating sparsity. The horseshoe prior seems to be outperformed by the spike and slab prior, in terms of accuracy in recovering true parameters, for moderate dimensional problems that we investigated in this article.

We found that the mixing in the Markov chain for the horseshoe prior is consistently better than that of spike and slab priors. It is well known that posterior computation for Bayesian variable selection with spike and slab priors, does not scale well with high dimensions. So for large p, the horseshoe prior could offer an alternative approach, given its improved mixing. For hyperbolic regression, we found computation under the horseshoe prior to be somewhat unstable for large p, due to having to invert large $p \times p$ matrices. Further investigation is needed regarding how to make the computation more stable. One possible direction is using the regularized horseshoe prior of Piironen and Vehtari (2017).

Acknowledgements

Joyee Ghosh's research was supported by NSF Grant DMS-1612763. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. The authors thank the Editors for helpful suggestions that improved the quality of the paper.

References

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.

Bhadra, A., Datta, J., Polson, N. G., and Willard, B. (2017). The horseshoe estimator of ultra-sparse signals. *Bayesian Analysis*, **12**, 1105–1131.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**, 465–480.

De, S. and Ghosh, J. (2024). Robust Bayesian model averaging for linear regression models with heavy-tailed errors. Technical report.

Gneiting, T. (1997). Normal scale mixtures and dual probability densities. *Journal of Statistical Computation and Simulation*, **59**, 375–384.

Hamura, Y., Irie, K., and Sugasawa, S. (2022). Log-regularly varying scale mixture of normals for robust regression. *Computational Statistics and Data Analysis*, **173**, 107517.

Kawakami, J. and Hashimoto, S. (2023). Approximate Gibbs sampler for Bayesian huberized lasso. *Journal of Statistical Computation and Simulation*, **93**, 128–162.

Makalic, E. and Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, **23**, 179–182.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, **103**, 681–686.

Piironen, J. and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, **11**, 5018 – 5051.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646–648.