Taylor & Francis
Taylor & Francis Group

Check for updates

Accelerated Bayesian Reciprocal LASSO

Erina Paul[a], Jingyu He[b] (iD), and Himel Mallick[c]

[a]Biostatistics and Research Decision Sciences, Merck & Co., Inc, Rahway, NJ, USA; [b]City University of Hong Kong, Kowloon Tong, China; [c]Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, NY, USA

ABSTRACT

Bayesian reciprocal LASSO (BRL) is a recently proposed nonlocal regularization method for Bayesian linear regression models. This paper develops a modified version of BRL, accommodating faster posterior sampling than published methods, by bypassing the use of auxiliary latent variables. We present a slice-within-Gibbs algorithm based on the elliptical slice sampler that matches the predictive accuracy of previous BRL implementations. Simulation studies and real data analyses show that the new method (XBRL) outperforms its Bayesian cousin (BRL) in out-of-sample prediction across a wide range of scenarios while offering the advantage of faster posterior computation. We have implemented the XBRL algorithm as part of the R package *BayesRecipe* available at: https://github.com/himelmallick/BayesRecipe.

1. Introduction

Bayesian statistics has long studied the problem of high-dimensional variable selection, which has led to a vast literature on Bayesian regularization priors with desirable statistical properties, exemplified by their successful utilization and omnipresence in a wide range of real-world application domains (Mallick and Yi 2013; Mallick 2015; Bhadra et al. 2019; Van Erp, Oberski, and Mulder 2019; Bai, Ročková, and George 2021). While traditional methods in this space have relied on spike and slab approaches and their computationally tractable extensions inspired by the popular class of global-local shrinkage priors, nonlocal priors have been recognized as a useful class of priors in recent years (Johnson and Rossell 2010, 2012; Nikooienejad, Wang, and Johnson 2016, 2020; Shin, Bhattacharya, and Johnson 2018; Sanyal et al. 2019).

Compared to local priors, which preserve a non-zero probability at null parameter values, nonlocal priors are identically zero whenever a model parameter is equal to its null value, leading to strong model selection and posterior consistency properties (Rossell and Telesca 2017; Shin, Bhattacharya, and Johnson 2018; Cao, Khare, and Ghosh 2020; Cao and Lee 2022). In particular, nonlocal priors discard spurious covariates faster as the sample size grows, while preserving exponential learning rates to detect nontrivial coefficients (Rossell and Telesca 2017; Mallick et al. 2021). Motivated by their appealing Bayesian model selection properties, global-local scale mixture representations of nonlocal priors have recently gained momentum, which enable simple MCMC sampling, facilitating a flexible framework for coherent posterior inference (Rossell and Telesca 2017; Mallick et al. 2021).

CONTACT Himel Mallick ✉ him4004@med.cornell.edu 🖥 Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, NY, USA
Second affiliation for Himel Mallick: Department of Statistics and Data Science, Cornell University, Ithaca, NY, USA

This paper proposes a computationally efficient posterior sampling scheme for Bayesian reciprocal LASSO (BRL) (Mallick et al. 2021), a recently proposed nonlocal prior, which is in turn a Bayesian analog of the frequentist reciprocal LASSO (rLASSO) regularization (Song and Liang 2015; Song 2018). In particular, we consider the slice-within-Gibbs sampler based on the elliptical slice sampler (Murray, Adams, and MacKay 2010; Hahn, He, and Lopes 2019), which was recently developed for Gaussian linear regression in the context of local priors. Our new approach is motivated by the fact that existing software implementations for Bayesian rLASSO do not readily scale well to problems with a large number of observations and predictors. Compared to previous Gibbs samplers that rely on case-specific latent variable representations (Mallick et al. 2021), the new algorithm does not rely on auxiliary variables, which leads to faster posterior computation than existing implementations while maintaining similar or better inferential accuracy. The rest of the paper is organized as follows. In Section XBRL methodology, we describe the algorithmic details of the proposed method, followed by the numerical results in Section Numerical Studies. In Section Conclusions, we conclude with further discussion in this context.

2. XBRL methodology

Without loss of generality, we consider the standard setup of a regression model that links a univariate mean response y and p candidate predictors X through its linear predictor $X\boldsymbol{\beta}$, where y is the $n \times 1$ vector of responses, X is the $n \times p$ matrix of standardized regressors, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of coefficients to be estimated (without the intercept). We consider the linear regression model

$$y = X\beta + \epsilon, \tag{1}$$

where $\epsilon \sim N(0, \sigma^2 I)$. We assume the prior of β as $\pi(\beta|\lambda)$, and a standard conjugate inverse Gamma (IG) prior with shape parameter $\alpha/2$ and scale parameter $\beta/2$ for the residual variance, i.e. $\sigma^2 \sim IG(\alpha/2, \gamma/2)$. For the prior on β, researchers are typically interested in regularization priors rather than the standard conjugate normal prior to induce sparsity or shrinkage of the coefficients for stable out-of-sample predictions. Traditional regularization priors include Bayesian LASSO or horseshoe prior; however, they usually assume the highest prior density around zero (Park and Casella 2008; Carvalho, Polson, and Scott 2010). Mallick et al. (2021) proposed an alternative nonlocal prior that imposes high prior densities away from zero, named Bayesian Reciprocal LASSO and this paper focuses on implementing an efficient sampling algorithm for BRL. The rest of this section is organized as follows. We first introduce the major ingredients such as the frequentist reciprocal LASSO (Section Reciprocal LASSO), BRL (Section BRL), and the slice-within-Gibbs sampler for Gaussian linear regression (Section Slice-within-Gibbs sampler for Gaussian linear regression). Then, we introduce our main algorithm (XBRL) for efficient sampling of Bayesian Reciprocal LASSO regression (Sections XBRL, Selection of the hyperparameter λ, Implementation).

2.1. Reciprocal LASSO

For a general model, the rLASSO solves the following regularization problem (Song and Liang 2015):

$$Q(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} \ L(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} \frac{1}{|\beta_j|} I\{\beta_j \neq 0\}, \tag{2}$$

where $L(\boldsymbol{\beta})$ denotes the negative log-likelihood, $I(.)$ denotes an indicator function, and $\lambda > 0$ is the tuning parameter that controls the degree of penalization. As compared to traditional penalization functions that are usually symmetric about 0, continuous and non-decreasing in $(0, \infty)$,
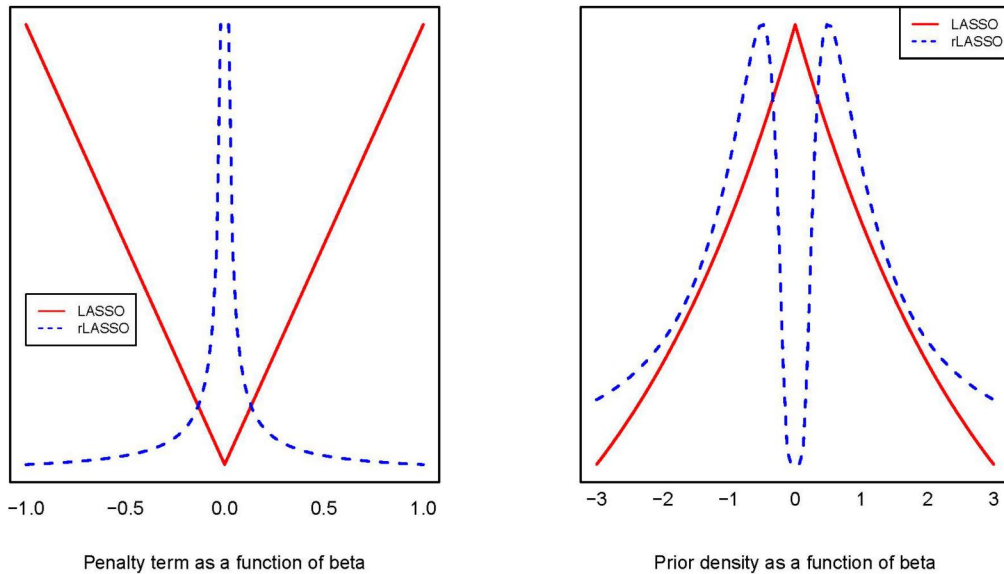
Figure 1. Comparison of the LASSO and rLASSO penalty functions (left) and LASSO and rLASSO prior densities (right) for a single regression coefficient. The rLASSO penalty function assigns nearly zero coefficients infinite penalties, whereas the conventional LASSO penalty assigns nearly zero coefficients almost zero penalties. Similarly, the LASSO prior density assigns a high non-zero probability near zero, whereas the rLASSO priors are exactly zero near zero.

the rLASSO penalty functions are decreasing in $(0, \infty)$, discontinuous at 0, and converge to infinity when the coefficients approach zero (Figure 1). Song and Liang (2015) proposed reciprocal LASSO as a new class of reciprocal regularization penalty functions and argued that these distinguishing properties of reciprocal LASSO make it a very attractive method for variable selection as it can produce sparser and more accurate coefficient estimates, and catch the true model with a higher probability. Moreover, reciprocal LASSO has a one-to-one correspondence with nonlocal priors (Johnson and Rossell 2010, 2012; Rossell and Telesca 2017), which have appealing Bayesian model selection properties for high-dimensional estimation. This motivated Mallick et al. (2021) to develop a Bayesian analog of the reciprocal LASSO method, namely, BRL, by formulating the rLASSO penalty as a nonlocal prior. Unlike classical rLASSO regression which summarizes inference using a single point estimate, BRL provides standard error estimates of the regression parameters, facilitating uncertainty quantification. We argue that this one-to-one correspondence makes reciprocal regularization particularly attractive as it facilitates complementary investigation of reciprocal penalty functions from two disparate but interconnected perspectives (i.e. frequentist and Bayesian), as has been the case with non-reciprocal regularization over the last two decades or so (Park and Casella 2008; Kyung et al. 2010; Mallick and Yi 2013; Mallick et al. 2021).

2.2. BRL

Mallick et al. (2021) noted that the rLASSO estimates for linear regression parameters can be interpreted as Bayesian posterior mode estimates when the priors on the regression parameters are assigned independent inverse Laplace distributions (with $\lambda > 0$ as the scale parameter):

$$\pi(\beta|\lambda) = \frac{\lambda}{2\beta^2} \exp\left\{-\frac{\lambda}{|\beta|}\right\} I\{\beta \neq 0\}. \tag{3}$$

Motivated by the close connection between frequentist rLASSO and nonlocal priors, Mallick et al. (2021) proposed a computationally efficient Gibbs sampler (BRL) to solve the rLASSO

problem taking advantage of the hierarchical global-local scale mixture representation of the inverse Laplace prior. In particular, Mallick et al. (2021) introduces the following latent variable formulation for BRL:

$$y^{n \times 1} | X, \boldsymbol{\beta}, \sigma^2 \sim N_n(X\boldsymbol{\beta}, \sigma^2 I_n),$$

$$\boldsymbol{\beta}^{p \times 1} | \boldsymbol{\tau}, \boldsymbol{u}, \sigma^2 \sim \prod_{j=1}^{p} N(0, \sigma^2 \tau_j^2) I\left\{ |\beta_j| > \frac{\sigma}{u_j} \right\},$$

$$\boldsymbol{\tau}^{p \times 1} | \boldsymbol{\zeta} \sim \prod_{j=1}^{p} \mathrm{Exp}\,(\zeta_j^2/2),$$

$$\boldsymbol{\zeta}^{p \times 1} | \boldsymbol{u} \sim \prod_{j=1}^{p} \mathrm{Exp}\left(\frac{1}{u_j}\right),$$

$$\boldsymbol{u}^{p \times 1} | \lambda \sim \prod_{j=1}^{p} \mathrm{Gamma}(2, \lambda),$$

$$\sigma^2 \sim \pi(\sigma^2).$$

The above hierarchical formulation leads to the following posterior distributions:

$$\boldsymbol{\beta} | \boldsymbol{y}, X, \boldsymbol{\tau}, \boldsymbol{\zeta}, \boldsymbol{u}, \lambda, \sigma^2 \sim N_p((X'X + T^{-1})^{-1} X'\boldsymbol{y}, \sigma^2 (X'X + T^{-1})^{-1}) \prod_{j=1}^{p} I\left\{ |\beta_j| > \frac{\sigma}{u_j} \right\},$$

$$\boldsymbol{\tau}^{-1} | \boldsymbol{y}, X, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{u}, \lambda, \sigma^2 \sim \prod_{j=1}^{p} \mathrm{Inverse\text{-}Gaussian}\left(\sqrt{\frac{\zeta_j^2 \sigma^2}{\beta_j^2}}, \zeta_j^2 \right),$$

$$\boldsymbol{\zeta} | \boldsymbol{y}, X, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{u}, \lambda, \sigma^2 \sim \prod_{j=1}^{p} \mathrm{Gamma}\left(2, \left(\frac{|\beta_j|}{\sigma} + \frac{1}{u_j} \right) \right),$$

$$\boldsymbol{u} | \boldsymbol{y}, X, \boldsymbol{\beta}, \boldsymbol{\zeta}, \lambda, \lambda, \sigma^2 \sim \prod_{j=1}^{p} \mathrm{Exp}\,(\lambda) I\left\{ u_j > \frac{\sigma}{|\beta_j|} \right\},$$

$$\sigma^2 | \boldsymbol{y}, X, \boldsymbol{\beta}, \boldsymbol{\zeta}, \lambda, \boldsymbol{u}, \lambda \sim \mathrm{Inverse\text{-}Gamma}\left(\frac{n-1+p}{2}, \frac{R + \boldsymbol{\beta}' T^{-1} \boldsymbol{\beta}}{2} \right) I\left\{ \sigma^2 < \mathrm{Min}_j(\beta_j^2 u_j^2) \right\},$$

where $R = (\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})$ and $T = \mathrm{diag}(\tau_1, ..., \tau_p)$.

While BRL is an effective nonlocal regularization method for small-scale settings (Mallick et al. 2021), the associated MCMC approach can be computationally inefficient for large-scale problems. This is particularly because the BRL algorithm relies on sampling from a mid-truncated

multivariate normal distribution (Kim 2007) and the use of latent variables, which can lead to suboptimal performance in high dimensions.

2.3. Slice-within-Gibbs sampler for Gaussian linear regression

Hahn, He, and Lopes (2019) proposed a slice-within-Gibbs sampler (SGS) based on the elliptical slice sampler (ESS) (Murray, Adams, and MacKay 2010) for posterior inference of Bayesian linear regression models with Gaussian errors and priors that can be evaluated up to a constant. The original ESS assumes a normal prior and an arbitrary likelihood and operates by drawing proposals from the Gaussian prior and then accepting or rejecting them based on the non-Gaussian likelihood. However, the constraint of a Gaussian prior limited the wide adoption of the original elliptical slice sampler. Additionally, drawing proposals from the prior tends to result in a higher rejection rate when the likelihood is strong.

To address these issues, Hahn, He, and Lopes (2019) leverage the fact that the posterior of the coefficients for linear regression with Gaussian errors and arbitrary priors has a specific form. They propose the slice-within-Gibbs sampler, which combines the Gibbs sampler for updating the precision parameter with the ESS for updating the coefficients. This approach reduces the computational burden and allows for the efficient sampling of arbitrary priors. By using a slice sampler within the Gibbs sampler, the SGS can generate proposals that are more likely to be accepted, leading to better mixing and faster convergence. The model has the following form

$$p(\beta \mid y, X, \sigma^2) \propto f(y \mid \beta, X, \sigma^2 I)\pi(\beta|\lambda) \tag{4}$$

which is exactly a product of Gaussian and non-Gaussian components. This setting is ideal for the application of ESS, however, a few practical problems may exist.

First, the ESS samples all coefficients jointly. However, if the number of regressors is large, the ESS may reject the proposal, leading to high autocorrelation of the posterior samples. To overcome this issue, SGS samples each of the coefficients β^k conditional on all other coefficients, which we denote β^{-k} in a Gibbs style, while each step implements the elliptical slice sampler. In specific, notice that the $\beta = (\beta^k, \beta^{-k})$ has the following joint multivariate normal distribution

$$\begin{bmatrix} \beta^k \\ \beta^{-k} \end{bmatrix} \sim N_p\left(\begin{bmatrix} \hat{\beta}^k \\ \hat{\beta}^{-k} \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{\Sigma}_{k,k} & \boldsymbol{\Sigma}_{k,-k} \\ \boldsymbol{\Sigma}_{-k,k} & \boldsymbol{\Sigma}_{-k,-k} \end{bmatrix}\right), \tag{5}$$

where $\begin{bmatrix} \hat{\beta}^k \\ \hat{\beta}^{-k} \end{bmatrix} = \hat{\beta}$ (the OLS estimator) and $\begin{bmatrix} \boldsymbol{\Sigma}_{k,k} & \boldsymbol{\Sigma}_{k,-k} \\ \boldsymbol{\Sigma}_{-k,k} & \boldsymbol{\Sigma}_{-k,-k} \end{bmatrix} = (X^TX)^{-1}$. Then the conditional distribution of any β^k given all other β^{-k} has the form $N_k(\tilde{\beta}^k, \widetilde{\boldsymbol{\Sigma}}^k)$ where

$$\tilde{\beta}^k = \hat{\beta}^k + \boldsymbol{\Sigma}_{k,-k}\boldsymbol{\Sigma}_{-k,-k}^{-1}(\beta^{-k} - \hat{\beta}^{-k}), \tag{6}$$

$$\widetilde{\boldsymbol{\Sigma}}^k = \sigma^2\left(\boldsymbol{\Sigma}_{k,k} - \boldsymbol{\Sigma}_{k,-k}\boldsymbol{\Sigma}_{-k,-k}^{-1}\boldsymbol{\Sigma}_{-k,k}\right). \tag{7}$$

Although the choice of the subset k can be arbitrary, we prefer to sample one coefficient at a time for better mixing of the Markov chains.

Second, it is common to analyze data with more predictors than data observations, where the matrix X^TX can be rank-deficient and thus the posterior

$$\begin{aligned} p(\beta \mid y, X, \sigma) &\propto N_Y(X\beta, \sigma^2)\pi(\beta|\lambda) \\ &\propto N_\beta(\hat{\beta}, \sigma^2(X^TX)^{-1})\pi(\beta|\lambda), \end{aligned} \tag{8}$$

which relies on the inverse of the rank-deficient matrix. The algorithm tackles this case by

observing a trick that multiplies and divides the joint posterior by the same normal density,

$$p(\beta|y, X, \sigma^2) \propto N_p(X\beta, \sigma^2)N(0, c\sigma^2)\frac{\pi(\beta|\lambda)}{N(0, c\sigma^2 I)}$$

$$\propto N_p(\bar{\beta}, \sigma^2(X^TX + c^{-1}I)^{-1})\frac{\pi(\beta|\lambda)}{N(0, c\sigma^2 I)}. \tag{9}$$

where $\bar{\beta} = (X^TX + c^{-1}I)^{-1}X^Ty$. Therefore, the covariance matrix $\sigma^2(X^TX + c^{-1}I)^{-1}$ is full rank and thus the ellipse is well-defined. Rather than evaluating the prior density $\pi(\beta|\lambda)$, we consider $\frac{\pi(\beta|\lambda)}{N(0, c\sigma^2 I)}$ instead to adjust for the extra term. Hahn, He, and Lopes (2019) recommended that small c near one works fine in most cases.

2.4. XBRL

In order to build a more efficient posterior sampling algorithm for BRL, we implemented the SGS sampler described in Section Slice-within-Gibbs sampler for Gaussian linear regression by taking $\pi(\beta|\lambda)$ as the rLASSO prior, described in Section BRL. The new sampler, which we refer to as Accelerated Bayesian Reciprocal LASSO (XBRL), takes advantage of the special structure of the linear regression likelihood, allowing it to lower the time for iteration than the original BRL method that relies on latent variable formulation of the reciprocal LASSO prior as described in Section BRL.

2.5. Selection of the hyperparameter λ

For both BRL and XBRL, we extend the procedure of Shin, Bhattacharya, and Johnson (2018) for nonlocal priors to the rLASSO prior and select λ such that the L_1 distance between the posterior distribution on the regression parameters under the null distribution (i.e. $\beta = 0$) and the rLASSO prior distributions on these parameters is constrained to be less than a specified value (e.g. $\frac{1}{\sqrt{p}}$). By choosing an optimal λ so that the intersection of these two null distributions falls below a specified threshold, this procedure approximately bounds the probability of false positives in the model, while maintaining sensitivity to detect large effects (Nikooienejad, Wang, and Johnson 2016). For brevity, we skip the technical details of the algorithm and refer the readers to Shin, Bhattacharya, and Johnson (2018); Mallick et al. (2021) and references therein. We use the R package *BayesRecipe* (Mallick et al. 2021) to carry out these calculations. With this estimation strategy, λ is estimated prior to posterior sampling, as opposed to empirical Bayes or hyperprior-based estimation, making this step identical for both BRL and XBRL (Mallick et al. 2021).

2.6. Implementation

We have implemented the XBRL algorithm as part of the R package *BayesRecipe* (Mallick et al. 2021). *BayesRecipe* is open source and publicly available with source code, documentation, and tutorial data for end users at: https://github.com/himelmallick/BayesRecipe.

3. Numerical Studies

3.1. Simulation Studies

To assess the performance of our new algorithm (XBRL), we compare our approach with published BRL methods with respect to both computing time and performance. To verify that different BRL samplers are comparable with respect to prediction accuracy, we report the average mean squared

error (MSE) of the associated posterior point estimates. For both BRL and XBRL, we run the corresponding sampler for 11,000 iterations, discarding the first 1,000 as burn-in and estimate λ by the Apriori Estimation method described in Section 2.5 (Mallick et al. 2021). This choice of tuning parameters appears to work satisfactorily based on the convergence diagnostics. We use the posterior mean as our point estimator. Results are summarized over 100 simulation runs. We simulate data from the true model $y = X\beta_0 + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n)$, and consider both $n \leq p$ and $n > p$ settings as well as a range of sparse and dense models with diverse effect sizes (Table 1). The design matrix X is generated from the multivariate normal distribution $N_p(0, \Sigma)$, where Σ has AR(1) (Autoregressive correlated design, where $\Sigma_{ij} = 0.5^{|i-j|}$ for all $1 \leq i \leq j \leq p$) covariance structure. For each parameter combination, we generate 100 datasets and each synthetic dataset is further partitioned into a training set and a test set.

3.1.1. Results

The results in Table 2 reveal that XBRL has a substantially better average MSE in all simulation settings. It is to be noted that the posterior sampling of BRL relies on sampling from a truncated multivariate normal distribution whereas no such step is required for XBRL. When the design matrix is highly collinear, the inefficiency of sampling from a constrained multivariate normal distribution is known to be severe (Polson, Scott, and Windle 2014; Rossell and Telesca 2017). The efficiency of XBRL thus can be attributed to the fact that, unlike BRL, it avoids sampling from a truncated multivariable normal distribution and bypasses the use of auxiliary variables. This is particularly eminent from the posterior histograms in the prostate cancer data (Section Results), where the bimodality of posterior marginals is not captured by BRL although it was captured well by XBRL.

The efficiency of a simpler and faster algorithm for the same Bayesian method has been reported previously in the literature. For example, Johndrow, Orenstein, and Bhattacharya (2020) recently proposed scalable approximate MCMC algorithms for the Horseshoe Prior and showed that the new algorithm yields estimates with lower mean squared error, intervals with better coverage, and elucidates features of the posterior that was often missed by previous algorithms in high dimensions, including the bimodality of posterior marginals indicating uncertainty about which covariates belong in the model. A similar conclusion was found by Mallick and Yi (2014), Mallick and Yi (2018), Polson, Scott, and Windle (2014), and Bon (2019), among others.

In light of this, we conclude that although both BRL and XBRL have good mixing in general, XBRL leads to faster sampling and better prediction accuracy due to the simplicity of the slice sampling algorithm. In terms of computing time, XBRL is significantly faster and less memory-intensive than BRL, further confirming the clear and significant benefits of XBRL over BRL in the context of both sparse and dense problems with important practical implications.

3.2. Real data applications

3.2.1. Datasets

To illustrate parameter estimation in a real data application, we pay a revisit to the prostate cancer (PC) dataset (Stamey et al. 1989). This dataset has been analyzed by many authors

Table 1. Benchmarking configurations for the simulation study.

Model	β_0	(n, p)
Fairly Sparse	tibsA: $(3, 0, 0, 1.5, 0, 0, 2, 0, 0)^T$	(100, 20), (200, 20), (500, 20), (50, 50), (20, 50), (50, 100)
Highly Dense	tibsB: $(0.85, ..., 0.85)^T$	(100, 20), (200, 20), (500, 20), (50, 50), (20, 50), (50, 100)
Highly Sparse	tibsC: $(5, 0, ..., 0)^T$	(100, 20), (200, 20), (500, 20), (50, 50), (20, 50), (50, 100)

Table 2. Simulation results for tibsA, tibsB, and tibsC. The boldface indicates the best performance.

n	p	β_0	Method	Mean of MSE	SD of MSE	Time (in minutes)
100	10	tibsA	XBRL	**86.8512**	8.2525	**0.01574**
			BRL	89.0356	8.9841	0.31522
		tibsB	XBRL	**85.9272**	8.3611	**0.01653**
			BRL	89.39	9.1909	0.28122
		tibsC	XBRL	**86.8986**	8.3278	**0.0161**
			BRL	88.3155	8.8503	0.2812
100	20	tibsA	XBRL	**97.6608**	11.1864	**0.02884**
			BRL	99.8023	12.6965	0.50606
		tibsB	XBRL	**98.1799**	11.5668	**0.0283**
			BRL	101.7056	12.9992	0.51134
		tibsC	XBRL	98.4052	11.1722	**0.02886**
			BRL	**98.3858**	12.5697	0.49999
200	20	tibsA	XBRL	**87.7459**	8.6053	**0.02853**
			BRL	88.6962	8.9393	0.50058
		tibsB	XBRL	**87.4255**	8.7853	**0.0282**
			BRL	89.4659	9.1265	0.48854
		tibsC	XBRL	88.0418	8.677	**0.02801**
			BRL	88.18	8.9148	0.48418
500	20	tibsA	XBRL	**81.8023**	8.5937	**0.02829**
			BRL	81.9671	8.7003	0.49178
		tibsB	XBRL	**81.2682**	8.676	**0.02812**
			BRL	82.1677	8.7202	0.48848
		tibsC	XBRL	**81.7219**	8.6714	**0.02816**
			BRL	81.855	8.7395	0.4914
50	50	tibsA	XBRL	**214.5135**	40.4324	**0.06122**
			BRL	324.4722	214.8135	0.88149
		tibsB	XBRL	**238.9341**	47.0732	**0.06153**
			BRL	379.7010	259.9224	0.88426
		tibsC	XBRL	**214.4992**	39.9290	**0.05488**
			BRL	260.0297	188.9276	0.76474
20	50	tibsA	XBRL	**128.7433**	16.6998	**0.09365**
			BRL	221.126	97.5224	1.11219
		tibsB	XBRL	**147.3917**	20.3451	**0.10036**
			BRL	212.1596	71.5677	1.46137
		tibsC	XBRL	**134.8029**	17.4516	**0.10159**
			BRL	204.5799	86.1511	1.40746
50	100	tibsA	XBRL	**152.7628**	21.4779	**0.12329**
			BRL	213.9409	74.4236	1.68540
		tibsB	XBRL	**185.3526**	29.7679	**0.19823**
			BRL	256.5213	41.5307	2.64111
		tibsC	XBRL	**157.0475**	21.1609	**0.16525**
			BRL	185.2690	54.2456	2.29940

Table 3. Predictive measures based on repeated holdout evaluation. The boldface indicates the best performance.

Dataset	Method	Mean of MSE	SD of MSE	Time (in minutes)
PC	XBRL	**0.4892**	0.763	**0.003**
	BRL	0.5037	0.803	0.058
BH	XBRL	26.6556	87.776	**0.003**
	BRL	**26.0417**	83.347	0.081
CC	XBRL	**1.0673**	1.730	**1.031**
	BRL	1.4685	2.070	37.587

including Tibshirani (1996), Park and Casella (2008), Zou and Hastie (2005), Li and Lin (2010), and Kyung et al. (2010), and is available from the R package ElemStatLearn. Briefly, it contains $n = 97$ prostate-specific antigen (PSA) measurements from prostate cancer patients who were about to receive radical prostatectomy. PSA is a protein that is produced by the prostate gland, with a higher level indicating a greater chance of having prostate cancer. The goal is to predict the log of PSA (*lpsa*) from a number of clinical measurements ($p = 8$) including (i) log

Table 4. Gelman-Rubin and Geweke diagnostics for the PC data. The point estimates of the Gelman-Rubin potential scale reduction factor (labelled Point est.) and their upper confidence limits (labelled Upper CI) are reported. Approximate convergence is diagnosed when the upper limit is close to 1. The occurrence of the Geweke scores well within 2 standard deviations of zero gives does not indicate lack of convergence, while deviations exceeding 2 standard deviations suggests that additional samples are required to achieve convergence.

| | Gelman-Rubin | | | | Geweke | |
| | BRL | | XBRL | | | |
Feature	Point Est.	Upper CI	Point Est.	Upper CI	BRL	XBRL
lcavol	1	1	1	1	−2.3523	1.2930
lweight	1	1	1	1	−0.8037	0.3674
age	1	1	1	1	2.1168	−0.3190
lbph	1	1	1	1	−2.5680	−0.7100
svi	1	1	1	1	−2.4449	−0.5527
lcp	1	1	1	1	2.3642	−0.9140
gleason	1	1	1	1	2.2179	1.2642
pgg45	1	1	1	1	−2.2471	−0.9587

Table 5. Gelman-Rubin and Geweke diagnostics for the BH data. The point estimates of the Gelman-Rubin potential scale reduction factor (labelled Point est.) and their upper confidence limits (labelled Upper CI) are reported. Approximate convergence is diagnosed when the upper limit is close to 1. The occurrence of the Geweke scores well within 2 standard deviations of zero gives does not indicate lack of convergence, while deviations exceeding 2 standard deviations suggests that additional samples are required to achieve convergence.

| | Gelman-Rubin | | | | Geweke | |
| | BRL | | XBRL | | | |
Feature	Point Est.	Upper CI	Point Est.	Upper CI	BRL	XBRL
crim	1	1	1	1.00	−0.2862	0.9295
zn	1	1	1	1.00	0.2861	−0.2214
indus	1	1	1	1.00	0.1271	0.6763
chas	1	1	1	1.00	0.2782	−1.3072
nox	1	1	1	1.00	−0.2746	−1.2312
rm	1	1	1	1.01	−0.2960	0.7738
age	1	1	1	1.00	−0.2880	−0.4071
dis	1	1	1	1.00	−0.2841	−0.9864
rad	1	1	1	1.01	0.2812	−0.1377
tax	1	1	1	1.00	−0.2761	0.3696
ptratio	1	1	1	1.00	−0.2718	−1.6289
b	1	1	1	1.00	0.2890	−0.2061
lstat	1	1	1	1.01	−0.0851	0.3243

Table 6. Gelman-Rubin and Geweke diagnostics for the CC data (top 10 predictors based on posterior median). The point estimates of the Gelman-Rubin potential scale reduction factor (labelled Point est.) and their upper confidence limits (labelled Upper CI) are reported. Approximate convergence is diagnosed when the upper limit is close to 1. The occurrence of the Geweke scores well within 2 standard deviations of zero gives does not indicate lack of convergence, while deviations exceeding 2 standard deviations suggests that additional samples are required to achieve convergence.

| | Gelman-Rubin | | | | Geweke | |
| | BRL | | XBRL | | | |
Feature	Point Est.	Upper CI	Point Est.	Upper CI	BRL	XBRL
X227058_at	1	1	1	1	−0.3489	1.3572
X217428_s_at	1	1	1	1	−1.8786	0.5877
X225803_at	1	1	1	1	1.8443	−0.5218
X200906_s_at	1	1	1	1	1.8233	−0.6768
X206026_s_at	1	1	1	1	0.4998	−1.4483
X211924_s_at	1	1	1	1	0.3627	0.3720
X227140_at	1	1	1	1	−0.2438	1.1590
X209101_at	1	1	1	1	−1.8731	0.6881
X202627_s_at	1	1	1	1	−2.1534	0.9113
X223392_s_at	1	1	1	1	1.2453	1.7800
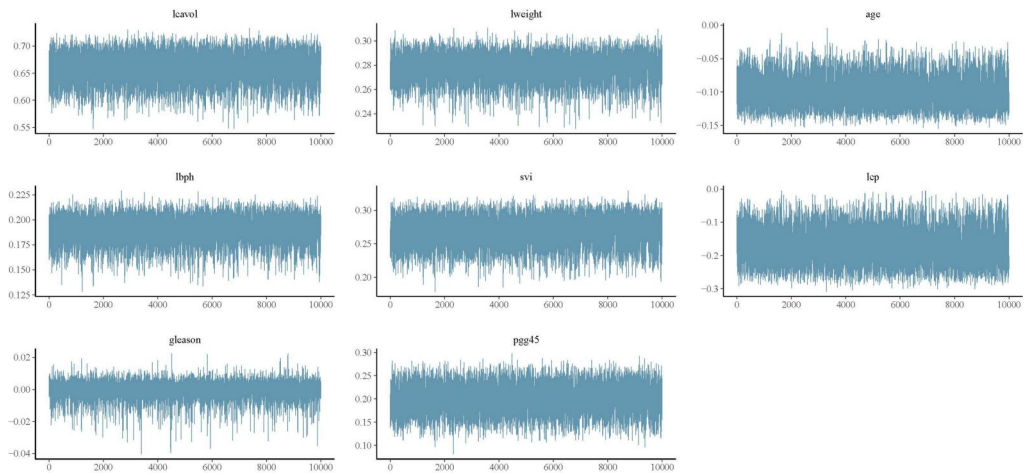
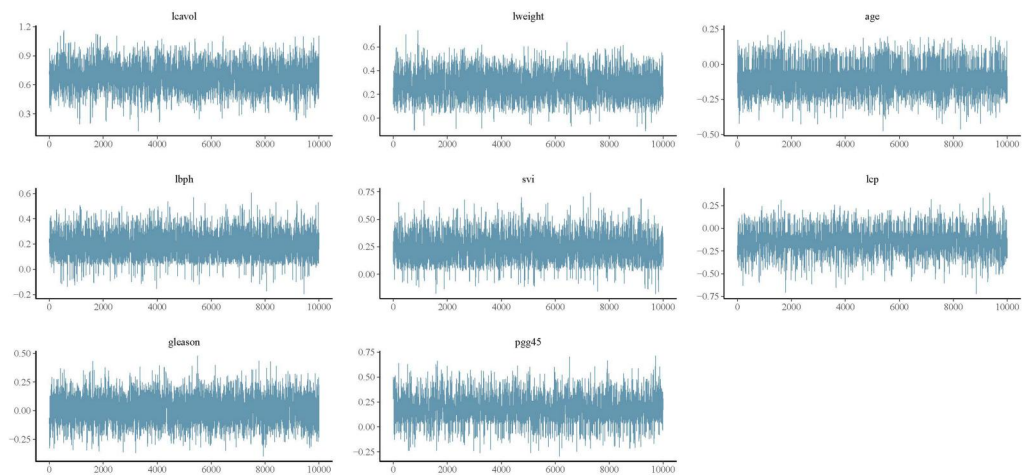**Figure 2.** Trace plots using BRL method for PC data.

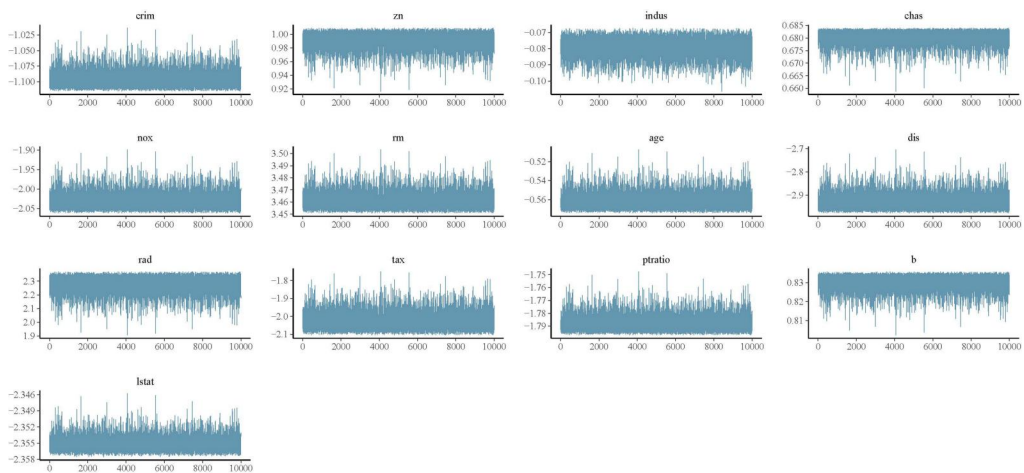**Figure 3.** Trace plots using XBRL method for PC data.

Figure 4. Trace plots using BRL method for BH data.
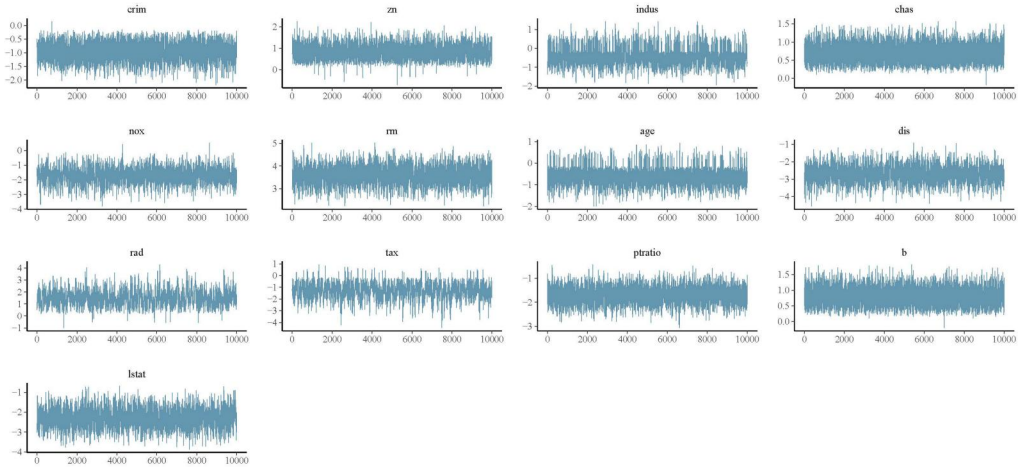
Figure 5. Trace plots using XBRL method for BH data.
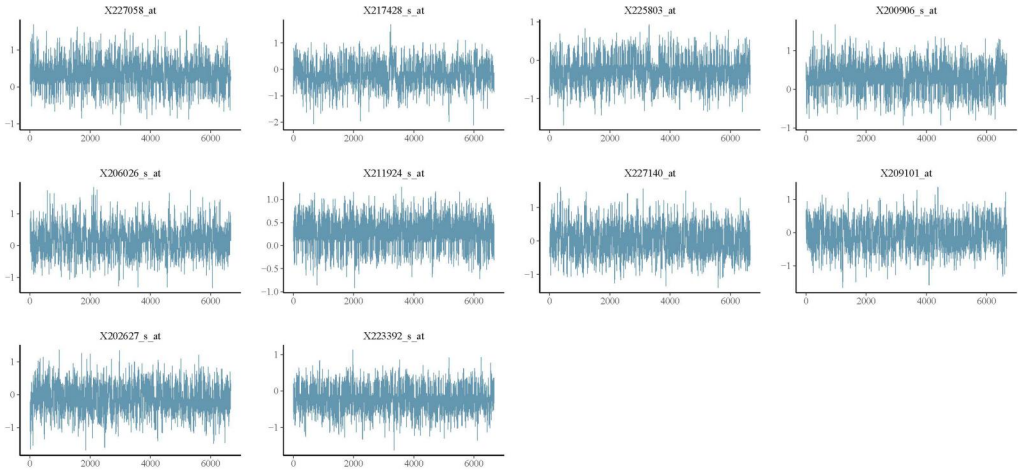

Figure 6. Trace plots using BRL method for CC data (top 10 predictors based on posterior median).

cancer volume (*lcavol*), (ii) log prostate weight (*lweight*), (iii) clinical age of the patient (*age*), (iv) log of benign prostatic hyperplasia amount (*lbph*), (v) seminal vesicle invasion (*svi*), (vi) log of capsular penetration (*lcp*), (vii) Gleason score (*gleason*), and (viii) percent of Gleason scores 4 or 5 (*pgg45*).

We also assess predictive performance in a high-dimensional gene expression data related to the gene TGFB, which encodes a secreted ligand of the Transforming growth factor beta (TGFB) superfamily of proteins that control proliferation, differentiation, and other functions in many cell types (Calon et al. 2012). Recently, Calon et al. (2012) used mice experiments to identify $p = 172$ TGFB-related genes potentially related to colon cancer progression and further validated these genes in an independent data set with $n = 262$ human patients. The response variable of interest is the overall TGFB level (average log-transformed expression of TGFB1, TGFB2 and TGFB3 mRNAs) in a given sample (a surrogate for colon cancer progression), which we want to predict based on the gene expression of other TGFB-related
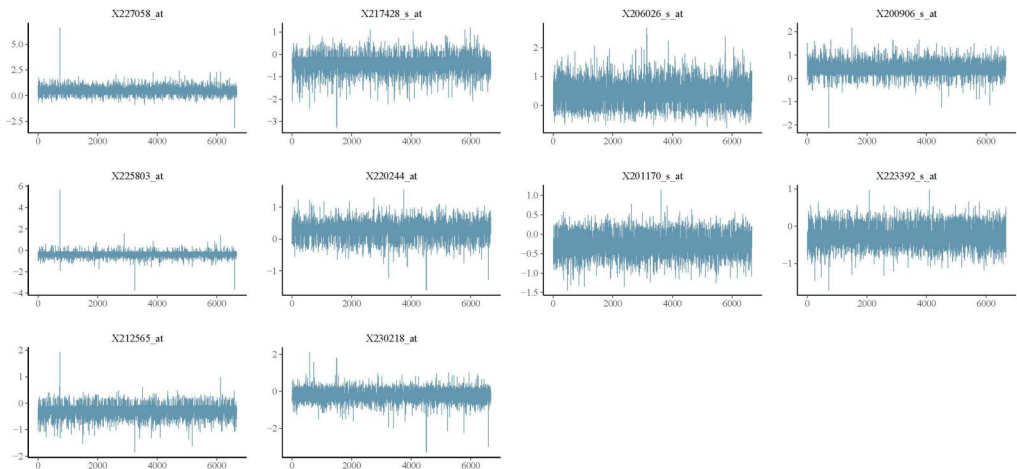
Figure 7. Trace plots using XBRL method for CC data (top 10 predictors based on posterior median).
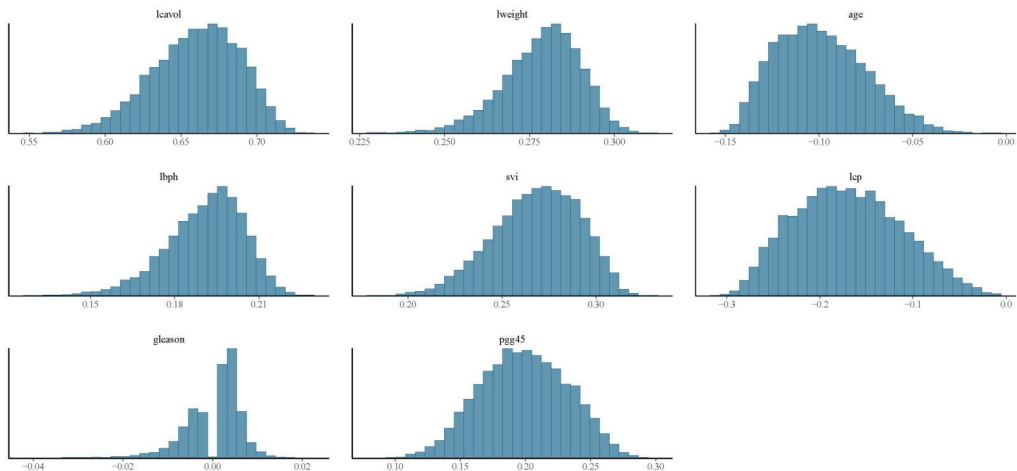

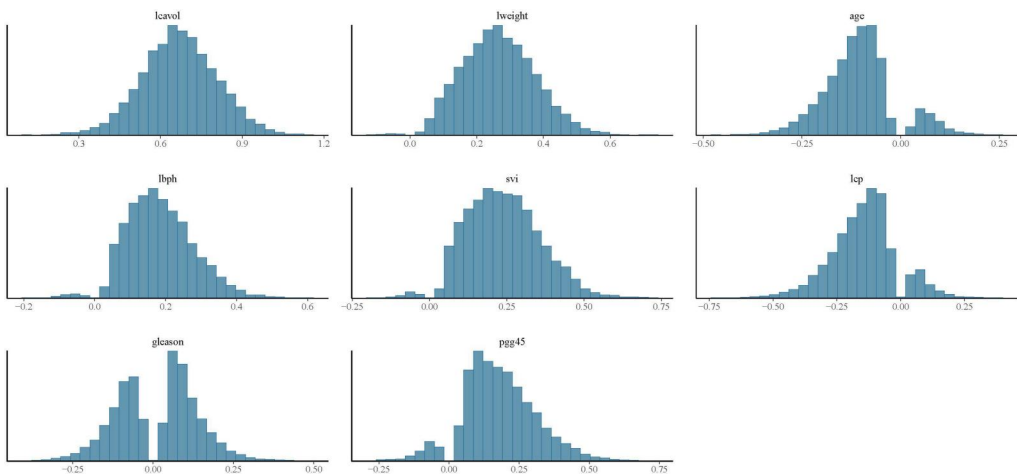**Figure 8.** Histograms using BRL method for PC data.

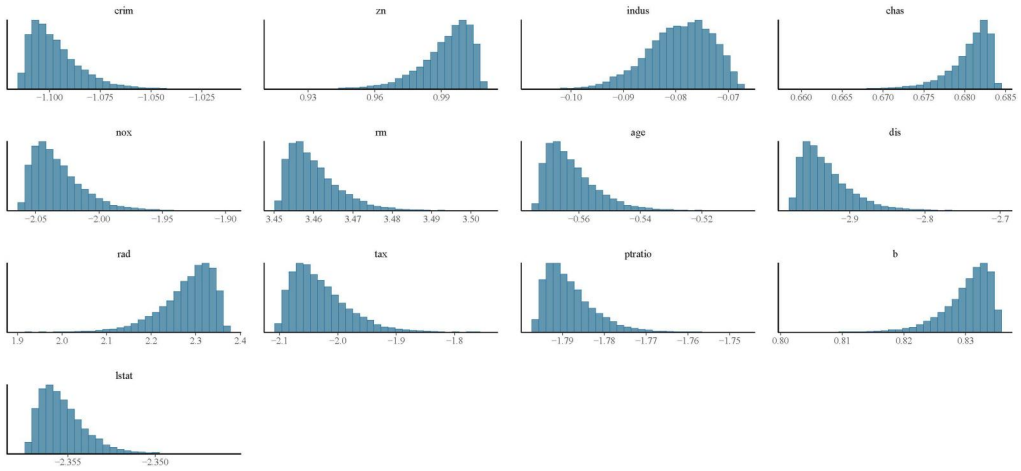**Figure 9.** Histograms using XBRL method for PC data.

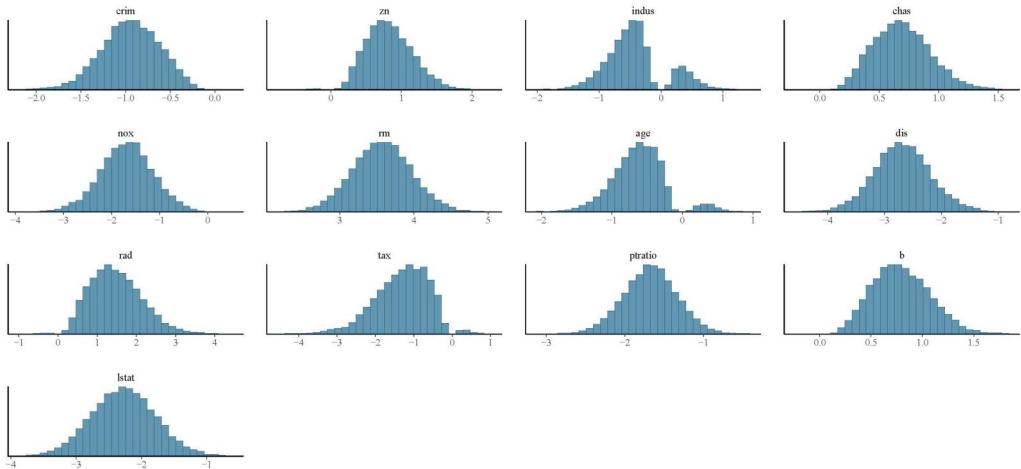**Figure 10.** Histograms using BRL method for BH data.

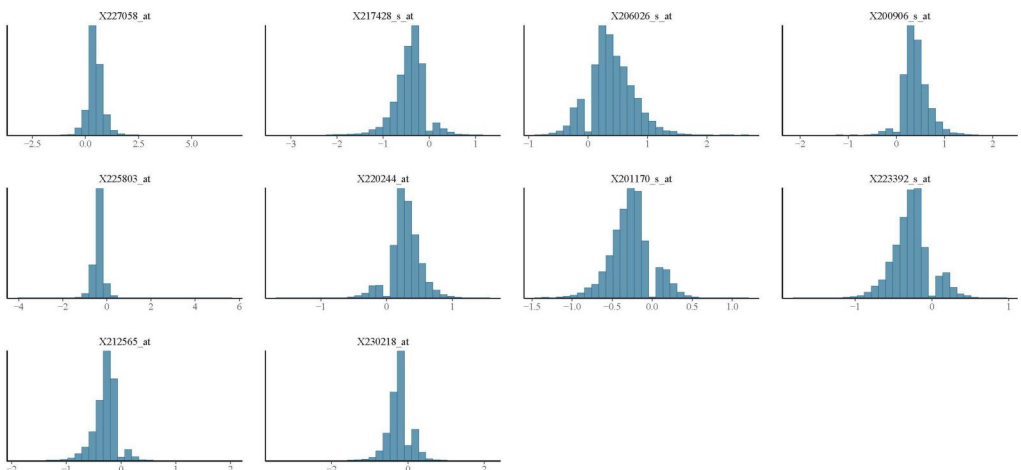**Figure 11.** Histograms using XBRL method for BH data.

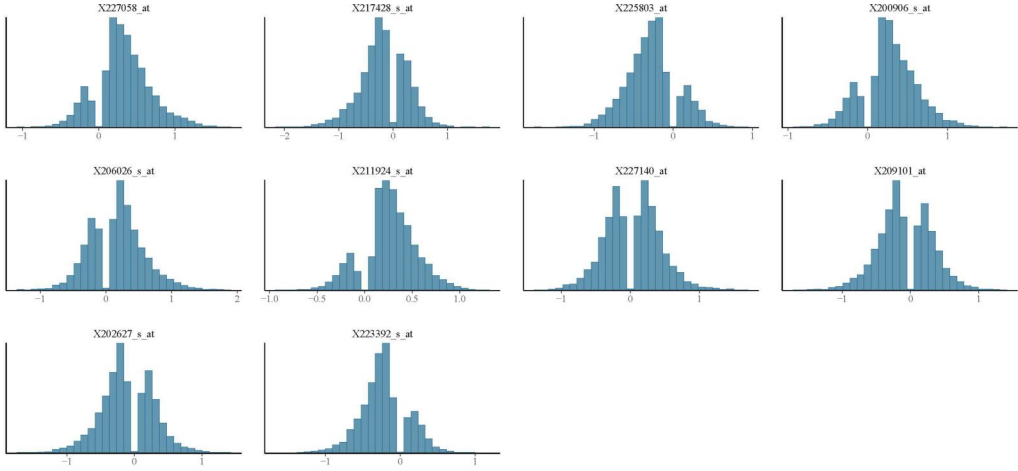**Figure 12.** Histograms using BRL method for CC data (top 10 predictors based on posterior median).

Figure 13. Histograms using XBRL method for CC data (top 10 predictors based on posterior median).
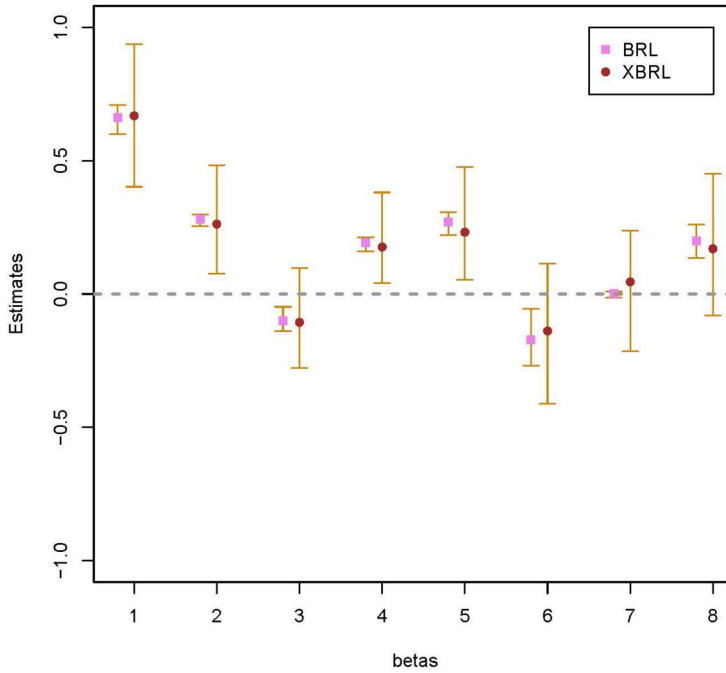

Figure 14. Credible interval plots for PC data.

genes ($p = 172$). We refer to this dataset as CC, which is available from Rossell and Telesca (2017).

We additionally consider the benchmark Boston housing (BH) dataset available from the R package `mlbench`, thus showcasing the applicability of our method in both small and large datasets.

3.2.2. Results
It is highly satisfactory to observe that the proposed XBRL method exhibits competitive performance on these datasets, especially with respect to better or comparable prediction accuracy than
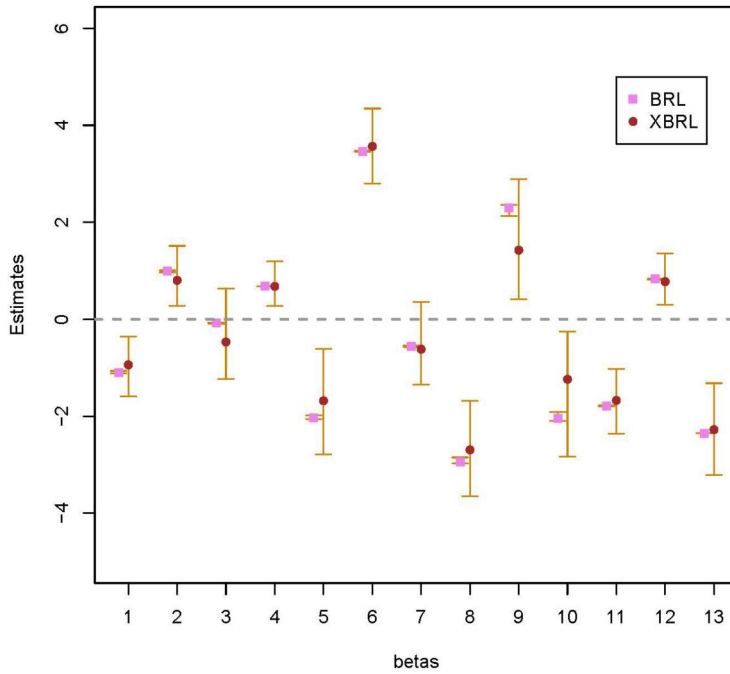
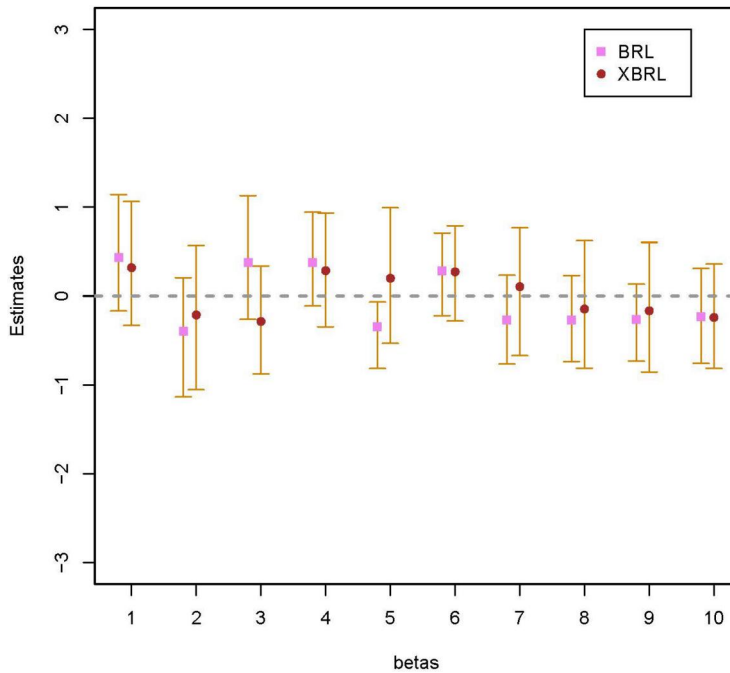**Figure 15.** Credible interval plots for BH data.

Figure 16. Credible interval plots for CC data (top 10 predictors based on posterior median).

BRL, while remaining up to 20 times faster than published methods (Table 3). Overall, these findings suggest that the proposed XBRL method is a useful alternative to existing Bayesian nonlocal regularization methods with computational demands that are competitive with the current fastest

alternatives. Gelman-Rubin and Geweke diagnostics for all three datasets are represented in Tables 4–6 along with trace plots (Figures 2–7) and histograms (Figures 8–13), confirming good mixing for both BRL and XBRL. They also reveal that BRL and XBRL can differ in detecting the bimodality of posterior marginals and in indicating uncertainty about which covariates belong in the model. In terms of feature selection based on 95% credible intervals (Figures 14–16), there is a general consensus on most features across datasets. However, occasional disagreements do arise. This does not imply that one conclusion is definitively correct and the other is unequivocally wrong in a specific setting. Rather, it highlights the substantial divergence between these two conclusions and underscores the value for practitioners in having both options available, which we have provided as part of our open-source implementation freely available at: https://github.com/himelmallick/BayesRecipe.

4. Conclusions

We have proposed an accelerated Bayesian reciprocal LASSO (XBRL) method for simultaneous coefficient estimation and variable selection in linear regression. Our approach is based on the observation that the reciprocal LASSO estimate can be interpreted as a Bayesian posterior mode estimate when the regression parameters are assigned independent inverse Laplace priors. On simulated and real data, we have demonstrated that XBRL is amenable to faster posterior estimation while performing as well as or better than custom BRL implementations that rely on auxiliary latent variables. Computationally efficient extension of this approach to other models, such as GLMs, survival, count, and zero-inflated regression models, may yield further advantages, potentially leading to a multi-model framework for accelerated reciprocal regularization under a single estimation umbrella.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Jingyu He 🔟 http://orcid.org/0000-0003-0370-8867

References

Bai, R., V. Ročková, and E. I. George. 2021. Spike-and-slab meets LASSO: A review of the spike-and-slab lasso. In *Handbook of Bayesian Variable Selection*, 81–108. New York, Imprint: Chapman and Hall/CRC.

Bhadra, A., J. Datta, N. G. Polson, and B. Willard. 2019. LASSO meets horseshoe: A survey. *Statistical Science* 34 (3):405–27. doi:10.1214/19-STS700.

Bon, J. 2019. Stochastic nets. PhD thesis, University of Western Australia.

Calon, A., E. Espinet, S. Palomo-Ponce, D. V. Tauriello, M. Iglesias, M. V. Céspedes, M. Sevillano, C. Nadal, P. Jung, X. H.-F. Zhang, et al. 2012. Dependency of colorectal cancer on a TGF-β-driven program in stromal cells for metastasis initiation. *Cancer Cell* 22 (5):571–84. doi:10.1016/j.ccr.2012.08.013.

Cao, X., K. Khare, and M. Ghosh. 2020. High-dimensional posterior consistency for hierarchical non-local priors in regression. *Bayesian Analysis* 15 (1). doi:10.1214/19-BA1154.

Cao, X., and K. Lee. 2022. Bayesian inference on hierarchical nonlocal priors in generalized linear models. *Bayesian Analysis* 1 (1): 1–24.

Carvalho, C. M., N. G. Polson, and J. G. Scott. 2010. The horseshoe estimator for sparse signals. *Biometrika* 97 (2): 465–80. doi:10.1093/biomet/asq017.

Hahn, P. R., J. He, and H. F. Lopes. 2019. Efficient sampling for Gaussian linear regression with arbitrary priors. *Journal of Computational and Graphical Statistics* 28 (1):142–54. doi:10.1080/10618600.2018.1482762.

Johndrow, J., P. Orenstein, and A. Bhattacharya. 2020. Scalable approximate MCMC algorithms for the horseshoe prior. *Journal of Machine Learning Research* 21 (73):1–61.

Johnson, V. E., and D. Rossell. 2010. On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 72 (2):143–70. doi:10.1111/j.1467-9868.2009.00730.x.

Johnson, V. E., and D. Rossell. 2012. Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association* 107 (498):649–60. doi:10.1080/01621459.2012.682536.

Kim, H.-J. 2007. Moments of a class of internally truncated normal distributions. *Communications for Statistical Applications and Methods* 14 (3):679–86. doi:10.5351/CKSS.2007.14.3.679.

Kyung, M., J. Gill, M. Ghosh, and G. Casella. 2010. Penalized regression, standard errors, and Bayesian LASSOs. *Bayesian Analysis* 5 (2):369–412.

Li, Q., and N. Lin. 2010. The Bayesian elastic net. *Bayesian Analysis* 5 (1):151–70. doi:10.1214/10-BA506.

Mallick, H. 2015. *Some contributions to Bayesian regularization methods with applications to genetics and clinical trials*. PhD thesis, The University of Alabama at Birmingham.

Mallick, H., R. Alhamzawi, E. Paul, and V. Svetnik. 2021. The reciprocal Bayesian LASSO. *Statistics in Medicine* 40 (22):4830–49. doi:10.1002/sim.9098.

Mallick, H., and N. Yi. 2013. Bayesian methods for high dimensional linear models. *Journal of Biometrics & Biostatistics* 1:005.

Mallick, H., and N. Yi. 2014. A new Bayesian LASSO. *Statistics and Its Interface* 7 (4):571–82. doi:10.4310/SII.2014.v7.n4.a12.

Mallick, H., and N. Yi. 2018. Bayesian bridge regression. *Journal of Applied Statistics* 45 (6):988–1008. doi:10.1080/02664763.2017.1324565.

Murray, I., R. Adams, and D. MacKay. 2010. Elliptical slice sampling. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 541–8. JMLR Workshop and Conference Proceedings.

Nikooienejad, A., W. Wang, and V. E. Johnson. 2016. Bayesian variable selection for binary outcomes in high-dimensional genomic studies using non-local priors. *Bioinformatics* 32 (9):1338–45. doi:10.1093/bioinformatics/btv764.

Nikooienejad, A., W. Wang, and V. E. Johnson. 2020. Bayesian variable selection for survival data using inverse moment priors. *The Annals of Applied Statistics* 14 (2): 809–28.

Park, T., and G. Casella. 2008. The Bayesian LASSO. *Journal of the American Statistical Association* 103 (482):681–6. doi:10.1198/016214508000000337.

Polson, N. G., J. G. Scott, and J. Windle. 2014. The Bayesian bridge. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76 (4):713–33. doi:10.1111/rssb.12042.

Rossell, D., and D. Telesca. 2017. Nonlocal priors for high-dimensional estimation. *Journal of the American Statistical Association* 112 (517):254–65. doi:10.1080/01621459.2015.1130634.

Sanyal, N., M.-T. Lo, K. Kauppi, S. Djurovic, O. A. Andreassen, V. E. Johnson, and C.-H. Chen. 2019. Gwasinlps: Non-local prior based iterative snp selection tool for genome-wide association studies. *Bioinformatics (Oxford, England)* 35 (1):1–11. doi:10.1093/bioinformatics/bty472.

Shin, M., A. Bhattacharya, and V. E. Johnson. 2018. Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statistica Sinica* 28 (2):1053.

Song, Q. 2018. An overview of reciprocal L1-regularization for high dimensional regression data. *Wiley Interdisciplinary Reviews: Computational Statistics* 10 (1):e1416.

Song, Q., and F. Liang. 2015. High-dimensional variable selection with reciprocal L1-regularization. *Journal of the American Statistical Association* 110 (512):1607–20. doi:10.1080/01621459.2014.984812.

Stamey, T. A., J. N. Kabalin, J. E. McNeal, I. M. Johnstone, F. Freiha, E. A. Redwine, and N. Yang. 1989. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. *The Journal of Urology* 141 (5):1076–83. doi:10.1016/s0022-5347(17)41175-x.

Tibshirani, R. 1996. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B* 58 (1):267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

Van Erp, S., D. L. Oberski, and J. Mulder. 2019. Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology* 89:31–50. doi:10.1016/j.jmp.2018.12.004.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67 (2):301–20. doi:10.1111/j.1467-9868.2005.00503.x.