Adaptive Shrinkage with a Nonparametric Bayesian Lasso

Santiago Marin, Bronwyn Loong and Anton H. Westveld

The 2026 ISBA World Meeting, Nagoya, Japan

✉ `santiago.marinardila@anu.edu.au`

Australian National University

Research School of Finance, Actuarial Studies & Statistics

Motivation and contribution

Modern datasets contain a large number of variables, so identifying which predictors are the most important ones—also known as **variable selection**—is a crucial and important task.

In a Bayesian framework, **shrinkage priors** provide a natural way to induce sparsity by shrinking the model parameters towards zero, pursuing a more parsimonious model.

- An ideal shrinkage prior should be **adaptive** to different signal levels, ensuring that small effects are ruled out, while keeping relatively intact the important ones.
- We develop the **nonparametric Bayesian Lasso**, an adaptive and flexible shrinkage prior for Bayesian regression and variable selection.
- We extend the spike-and-slab Lasso by placing a **Dirichlet process prior** on the shrinkage parameters. The result is a prior on the regression coefficients that can be seen as an infinite mixture of Laplace distributions, all offering different amounts of regularization, ensuring a more adaptive and flexible shrinkage.

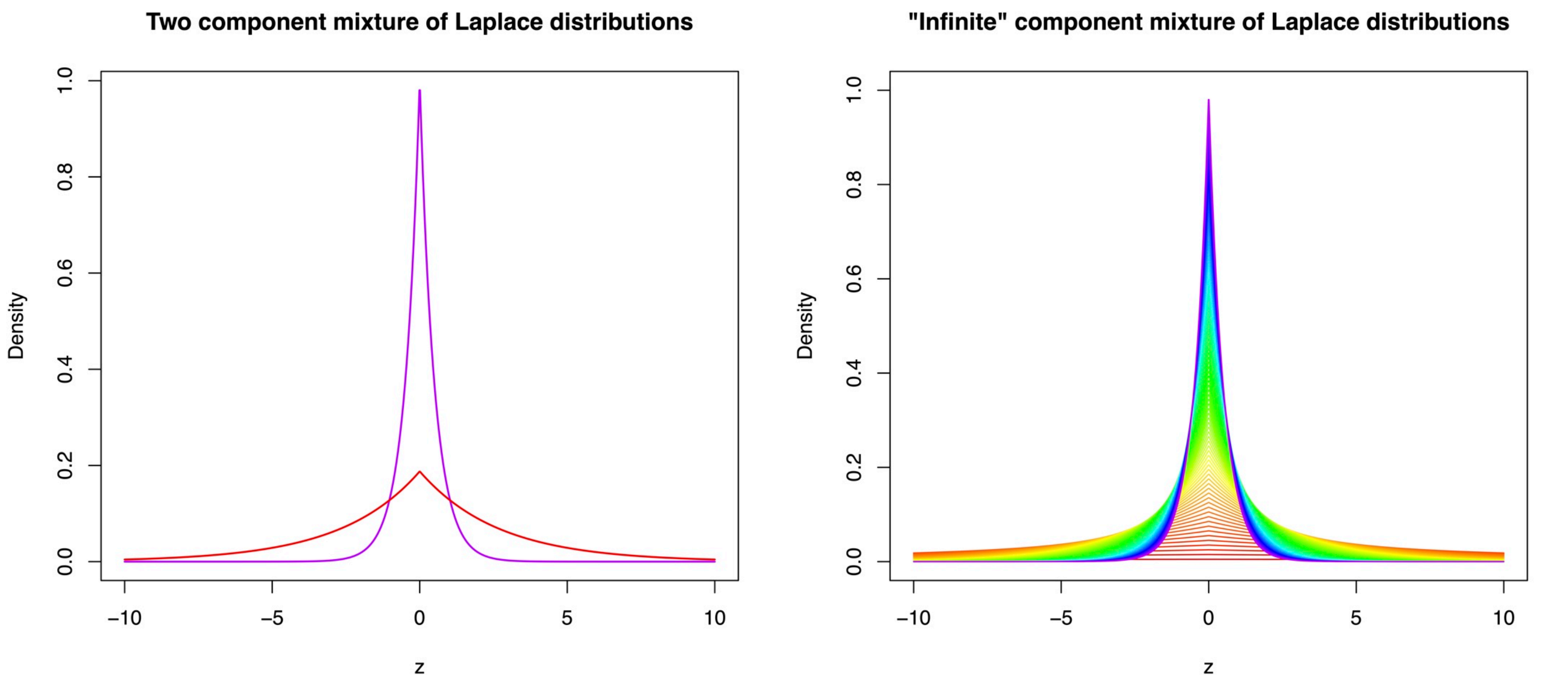


Figure 1: A two component mixture and an "infinite" component mixture of Laplace distributions.

Background

The spike-and-slab Lasso

Following [2], a spike-and-slab Lasso prior on each β_j, for $j \in \{1, \ldots, p\}$, is given by

$$p(\beta_j \,|\, \pi_1, \pi_2) = \sum_{k=1}^{2} \pi_k \, \psi(\beta_j \,|\, 0, \lambda_k),$$

where $\psi(\cdot \,|\, 0, \lambda)$ denotes the density of a Laplace-distributed random variable with location zero and rate parameter λ.

It is clear that the spike-and-slab Lasso is a **finite mixture of two Laplace distributions**, where $\lambda_1 \gg \lambda_2$ such that $\psi(\cdot \,|\, 0, \lambda_1)$ (the spike) thresholds small coefficients, while $\psi(\cdot \,|\, 0, \lambda_2)$ (the slab) keeps large effects unaltered.

Dirichlet process mixtures

Consider a **Dirichlet process mixture** given by

$$z_i | \theta_i \overset{\text{ind}}{\sim} f(z|\theta_i), \qquad \theta_i | H \overset{\text{iid}}{\sim} H, \qquad H \sim \text{DP}(\alpha, H_0).$$

If we set $f(\cdot|\theta) = \psi(\cdot \,|\, 0, \theta_k)$ and let $\{\theta_k\}_{k\in\mathbb{N}} \overset{\text{iid}}{\sim} H_0$, we would have that

$$f(z|H) = \sum_{k=1}^{\infty} w_k \psi(z \,|\, 0, \theta_k).$$

Note the similarities between $f(z|H)$ and the spike-and-slab Lasso. The main difference is that $f(z|H)$ corresponds to an **infinite component mixture**, generalizing the spike-and-slab Lasso.

The nonparametric Bayesian Lasso

Our full model is

$$\begin{gathered}
\boldsymbol{y} \,|\, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \\
\beta_1, \ldots, \beta_p \,|\, \tau_1^2, \ldots, \tau_p^2 \overset{\text{ind}}{\sim} N(0, \tau_j^2), \\
p(\tau_1^2, \ldots, \tau_p^2 \,|\, \lambda_1^2, \ldots, \lambda_p^2) = \prod_{j=1}^{p} \frac{\lambda_j^2}{2} \exp\left\{ -\frac{\lambda_j^2}{2} \tau_j^2 \right\}, \\
\lambda_1^2, \ldots, \lambda_p^2 \,|\, G \overset{\text{iid}}{\sim} G, \\
G \sim \text{DP}(\alpha, G_0).
\end{gathered}$$

To complete our model specification, we set G_0 to be a gamma distribution with shape parameter a and rate parameter b, and set $p(\sigma^2) \propto 1/\sigma^2$.

Numerical simulations

The true coefficient vector, $\boldsymbol{\beta}_{\text{true}}$, is constructed as $\boldsymbol{\beta}_{\text{true}} = (\underbrace{10, \ldots, 10}_{5}, \underbrace{2, \ldots, 2}_{15}, \underbrace{0, \ldots, 0}_{180})' \in \mathbb{R}^{200}$.

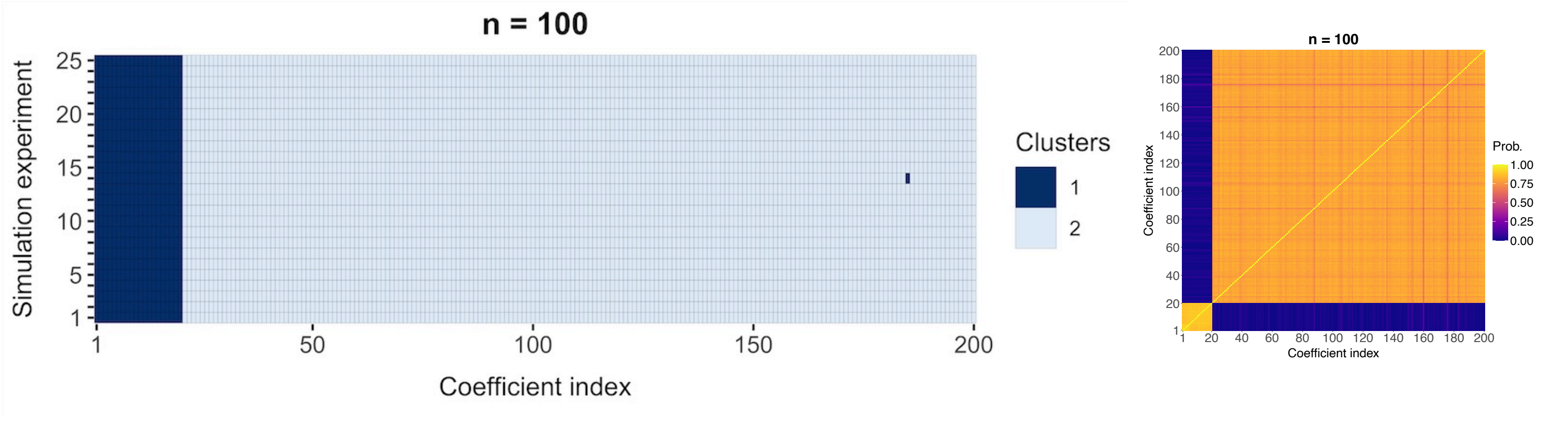


Figure 2: Clustering of the coefficients and posterior co-clustering probabilities recovered by the nonparametric Bayesian Lasso.

Asymptotic support recovery

Let's also study whether or not the posterior distribution from the nonparametric Bayesian Lasso asymptotically concentrates around the true parameter values. As such, let us compute

$$\text{concentration} = \frac{1}{p} \sum_{j=1}^{p} \mathbb{P}\left(\beta_j \in \{\beta_{\text{true},j} \pm \epsilon\} | \boldsymbol{y}, \mathbf{X}\right)$$

as a function of the sample size, where $\epsilon \to 0$.

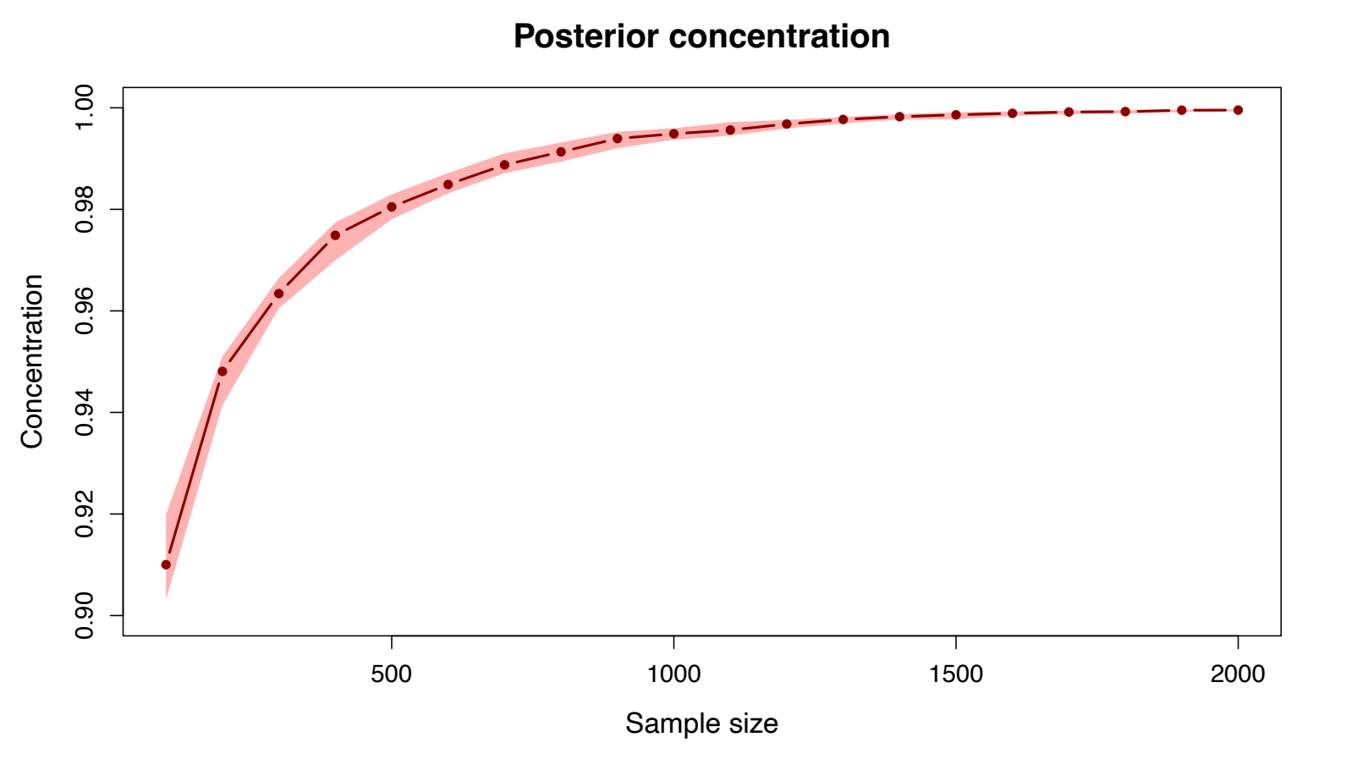


Figure 3: Posterior concentration as a function of the sample size.

Protein activity data

- To demonstrate the practical utility of our proposed method, we apply it to the protein activity data from [1].
- As comparison metrics, we use the cross-validated mean squared prediction error (CV MSPE) and the cross-validated expected log pointwise predictive density (CV ELPPD).
- As competitors, we consider the Bayesian bootstrap spike-and-slab Lasso (BB-SSL1—BB-SSL5), the horseshoe prior (H-P), the Bayesian Lasso (B-L), the Bayesian adaptive Lasso (B-AL), and the nonparametric Bayesian Lasso (BNP-L).

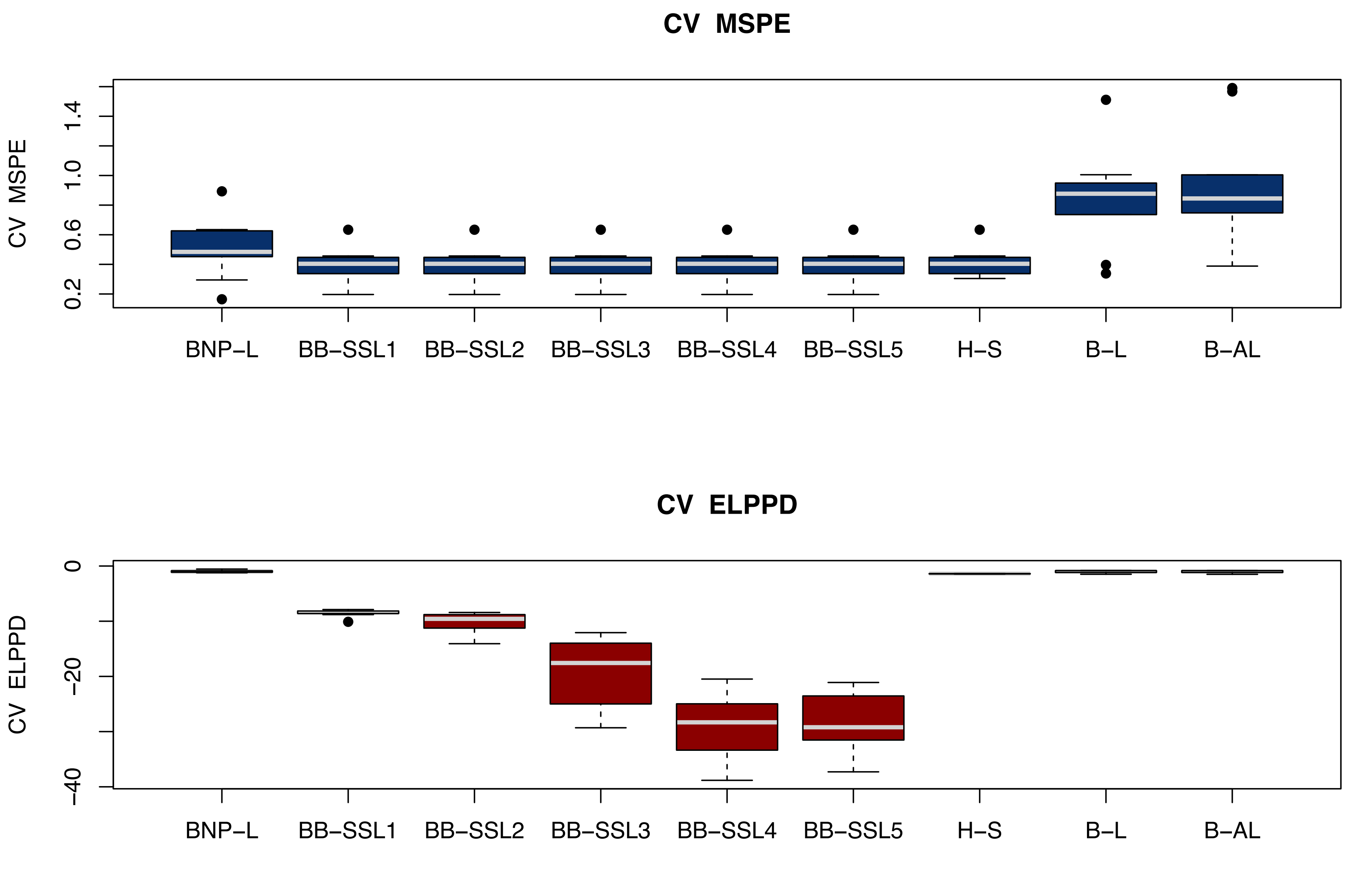


Figure 4: Ten-fold CV experiment on the Protein Activity Data.

Discussion

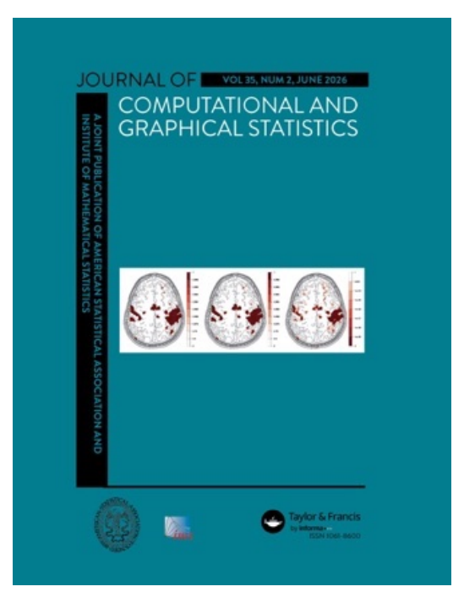
Our results suggest that the nonparametric Bayesian Lasso leads to coefficient recovery, variable selection accuracy, and out of sample predictions that are comparable to or better than those from state-of-the-art shrinkage priors, highlight the benefits of our proposed method.

On the whole, our nonparametric Bayesian Lasso expands and enriches the already vibrant world of Bayesian shrinkage priors!

References

[1] M. A. Clyde, J. Ghosh, and M. L. Littman. Bayesian adaptive sampling for variable selection and model averaging. *JCGS*, 20(1):80–101, 2011.

[2] V. Ročková and E. I. George. The spike-and-slab lasso. *JASA*, 113(521):431–444, 2018.



Adaptive Shrinkage with a Nonparametric Bayesian Lasso.
Marin, Loong and Westveld. (2026).
Journal of Computational and Graphical Statistics, **35**(2):854–864 🔓

"bnplasso"
R package. 🔓
`https://github.com/marinsantiago/bnplasso/`