JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 551, 2011–2014
https://doi.org/10.1080/01621459.2025.2540256

Taylor & Francis
Taylor & Francis Group

Check for updates

Correction: Bhattacharya, A., Pati, D., Pillai, N.S., and Dunson, D.B. (2015), "Dirichlet–Laplace Priors for Optimal Shrinkage," *Journal of the American Statistical Association*, 110, 1479–1490, DOI: 10.1080/01621459.2014.960967

A Note on Simulation Methods for the Dirichlet-Laplace Prior

Luis Gruber[a], Gregor Kastner[a], Anirban Bhattacharya[b], Debdeep Pati[c], Natesh Pillai[d], and David Dunson[e]

[a]Department of Statistics, University of Klagenfurt, Klagenfurt, Austria; [b]Department of Statistics, Texas A&M University, College Station, TX; [c]Department of Statistics, University of Wisconsin–Madison, Madison, WI; [d]Department of Statistics, Harvard University, Cambridge, MA; [e]Department of Statistical Science, Duke University, Durham, NC

ABSTRACT

Bhattacharya et al. introduce a novel prior, the Dirichlet-Laplace (DL) prior, and propose a Markov chain Monte Carlo (MCMC) method to simulate posterior draws under this prior in a conditionally Gaussian setting. The original algorithm samples from conditional distributions in the wrong order, that is, it does not correctly sample from the joint posterior distribution of all latent variables. This note details the issue and provides two simple solutions: A correction to the original algorithm and a new algorithm based on an alternative, yet equivalent, formulation of the prior. This corrigendum does not affect the theoretical results in Bhattacharya et al.

1. Model, Prior and Proposed MCMC Algorithm

Consider the standard normal means setting

$$y_i = \theta_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0,1), \quad 1 \le i \le n. \tag{1}$$

In Bhattacharya et al. (2015), a prior distribution for the vector of means, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$, is proposed. It can be represented as the following Gaussian scale-mixture:

$$\theta_j \sim \mathcal{N}(0, \psi_j \phi_j^2 \tau^2), \quad \psi_j \sim \mathcal{E}xp(1/2), \quad \boldsymbol{\phi} \sim \mathcal{D}ir(a, \ldots, a),$$
$$\tau \sim \mathcal{G}(na, 1/2). \tag{2}$$

This prior has attractive theoretical properties, was successfully applied to a range of practical and econometric problems (e.g., Zhang and Bondell 2018; Kastner and Huber 2020; Bhattacharyya et al. 2022; Huber, Kastner, and Pfarrhofer 2025), and consequently became a widely used continuous shrinkage prior.

To be of use for practical analysis, a posterior simulator via Markov chain Monte Carlo (MCMC) is developed in Bhattacharya et al. (2015). It aims at simulating draws from the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \tau, \boldsymbol{\psi} | \boldsymbol{y})$. The algorithm is intended to mix well, and thus relies on blocking and marginalization in order to reduce autocorrelation. The sampler cycles through the following steps:

1. $\boldsymbol{\theta} | \boldsymbol{\psi}, \boldsymbol{\phi}, \tau, \boldsymbol{y}$: $\theta_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ independently, with $\sigma_j^2 = \left(1 + \frac{1}{\psi_j \phi_j^2 \tau^2}\right)^{-1}$ and $\mu_j = \sigma_j^2 y_j, j = 1, \ldots, n$.
2. $\boldsymbol{\psi} | \boldsymbol{\theta}, \boldsymbol{\phi}, \tau$: Sample $\tilde{\psi}_j | \theta_j, \phi_j, \tau \sim iG(\frac{\phi_j \tau}{|\theta_j|}, 1)$ independently and set $\psi_j = \frac{1}{\tilde{\psi}_j}$.
3. $\tau | \boldsymbol{\phi}, \boldsymbol{\theta} \sim giG(na - n, 1, 2\sum_{j=1}^n \frac{|\theta_j|}{\phi_j})$.
4. $\boldsymbol{\phi} | \boldsymbol{\theta}$: Sample $T_j | \theta_j \sim giG(a-1, 1, 2|\theta_j|)$ independently and set $\phi_j = \frac{T_j}{\sum_{j=1}^n T_j}$.

It is important to note that a key feature is the joint update of $\boldsymbol{\psi}, \boldsymbol{\phi}, \tau | \boldsymbol{\theta}$.

2. On the Ordering of MCMC Updates

Joint updates, that is, updates where the order of updating single elements matters, are a common source of errors in MCMC algorithms. Geweke (2004) draws direct attention to this problem; nevertheless, it has since appeared undetected in other influential publications (e.g., Primiceri 2005; Del Negro and Primiceri 2015). Because joint updates appear to be such a common source for errors and to facilitate exposition, we first abstract the problem to a general level.

CONTACT David Dunson *dunson@duke.edu* Department of Statistical Science, Duke University, Box 90251, Durham, NC 27708-0251 USA.
A slightly modified version of this article is included in Luis Gruber's master thesis (Gruber 2025). In the process of preparing this corrigendum, it has come to our attention that the issue was independently uncovered by Onorati, Dunson, and Canale (2025).
© 2025 American Statistical Association

Assume the joint distribution $p(d,e,f)$, where the full conditional distributions $p(d|e,f)$, $p(e|d,f)$, and $p(f|e,d)$ are of well known forms. To generate M draws from $p(d,e,f)$, one can successively cycle through

1. $d^{(m)} \sim p(d|e^{(m-1)},f^{(m-1)})$,
2. $e^{(m)} \sim p(e|d^{(m)},f^{(m-1)})$,
3. $f^{(m)} \sim p(f|d^{(m)},e^{(m)})$,

where the parenthesized superscript denotes the respective MCMC iteration. The three updating steps are interchangeable, since they consist of sampling from full conditional distributions.

Things get slightly more involved by marginalizing. Assume that the conditional distribution $p(e|f)$—marginalized over d—is of well-known form. Then, the product $p(e|f)p(d|e,f)$ is the marginal-conditional decomposition of the joint conditional distribution $p(d,e|f)$. A correct sampler, exploiting the joint update of $p(d,e|f)$, cycles through

1. $f^{(m)} \sim p(f|d^{(m-1)},e^{(m-1)})$,
2. (a) $e^{(m)} \sim p(e|f^{(m)})$,
   (b) $d^{(m)} \sim p(d|e^{(m)},f^{(m)})$.

In that case, the steps 2(a) and 2(b) are *not* interchangeable. A correct algorithm samples e conditional on f—marginalized over d—and then samples d conditional on both e and f just drawn. In short, to sample from the joint distribution of two variables, it is possible to first draw from the marginal of one and then generate the other conditionally on that draw, thereby following the principle *marginal first—conditional later*.

Taking a closer look at the ingredients of the steps 2., 3., and 4., of the algorithm proposed in Bhattacharya et al. (2015), one observes that in the conditionals of both $\tau|\boldsymbol{\phi},\boldsymbol{\theta}$ and $\boldsymbol{\phi}|\boldsymbol{\theta}$, the auxiliary variable $\boldsymbol{\psi}$ is integrated out. In addition, the conditional $\boldsymbol{\phi}|\boldsymbol{\theta}$ is marginalized over τ. Together with the full conditional $\boldsymbol{\psi}|\boldsymbol{\theta},\boldsymbol{\phi},\tau$, those two conditionals form the marginal-conditional decomposition of the joint conditional $\boldsymbol{\phi},\tau,\boldsymbol{\psi}|\boldsymbol{\theta}$:

$$\begin{aligned} p(\boldsymbol{\phi},\tau,\boldsymbol{\psi}|\boldsymbol{\theta}) &= p(\boldsymbol{\phi},\tau|\boldsymbol{\theta})p(\boldsymbol{\psi}|\boldsymbol{\theta},\boldsymbol{\phi},\tau) \\ &= p(\boldsymbol{\phi}|\boldsymbol{\theta})p(\tau|\boldsymbol{\phi},\boldsymbol{\theta})p(\boldsymbol{\psi}|\boldsymbol{\theta},\boldsymbol{\phi},\tau). \end{aligned} \tag{3}$$

Clearly, the proposed algorithm cycles through the conditionals in the wrong order, because within the joint update of $\boldsymbol{\phi},\tau,\boldsymbol{\psi}|\boldsymbol{\theta}$ it does not adhere to the principle *marginal first—conditional later*.

3. Two Solutions

In this section, we present two solutions how to correctly sample from the posterior distribution under the prior in (2). The first one is simply a correction to the original one. The second builds on another—yet equivalent—formulation of the prior.

The corrected original algorithm cycles through the following steps:

1. $\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{\phi},\tau,\boldsymbol{y}$: $\theta_j \sim \mathcal{N}(\mu_j,\sigma_j^2)$ independently, with $\sigma_j^2 = \left(1+\frac{1}{\psi_j\phi_j^2\tau^2}\right)^{-1}$ and $\mu_j = \sigma_j^2 y_j, j = 1,\ldots,n$.
2. For the joint update of $p(\boldsymbol{\phi},\tau,\boldsymbol{\psi}|\boldsymbol{\theta})$ sample
   (a) $T_j|\theta_j \sim giG(a-1,1,2|\theta_j|)$ independently and set $\phi_j = \frac{T_j}{\sum_{j=1}^n T_j}$,
   (b) $\tau|\boldsymbol{\phi},\boldsymbol{\theta} \sim giG(na-n,1,2\sum_{j=1}^n \frac{|\theta_j|}{\phi_j})$,
   (c) $\tilde{\psi}_j|\theta_j,\phi_j,\tau \sim iG(\frac{\phi_j\tau}{|\theta_j|},1)$ independently and set $\psi_j = \frac{1}{\tilde{\psi}_j}$.

Note that the ordering within the joint update of $p(\boldsymbol{\phi},\tau,\boldsymbol{\psi}|\boldsymbol{\theta})$ is not interchangeable.

The second solution builds on an alternative formulation of the prior, which is mentioned in sec. 3 (eq. 10) in Bhattacharya et al. (2015). It uses the fact that $\lambda_j = \phi_j\tau \sim \mathcal{G}(a,1/2)$ independently for $j = 1,\ldots,n$:

$$\theta_j \sim \mathcal{N}(0,\psi_j\lambda_j^2), \quad \psi_j \sim \mathcal{E}xp(1/2), \quad \lambda_j \sim \mathcal{G}(a,1/2). \tag{4}$$

The conditional posterior distributions of the alternative formulation are standard and hence not derived. The algorithm associated with the alternative formulation cycles through the following steps:

1. $\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{\lambda},\boldsymbol{y}$: $\theta_j \sim \mathcal{N}(\mu_j,\sigma_j^2)$ independently, with $\sigma_j^2 = \left(1+\frac{1}{\psi_j\lambda_j^2}\right)^{-1}$ and $\mu_j = \sigma_j^2 y_j, j = 1,\ldots,n$.
2. For the joint update of $p(\boldsymbol{\lambda},\boldsymbol{\psi}|\boldsymbol{\theta})$ sample
   (a) $\lambda_j|\theta_j \sim giG(a-1,1,2|\theta_j|)$ independently,
   (b) $\tilde{\psi}_j|\theta_j,\lambda_j \sim iG(\frac{\lambda_j}{|\theta_j|},1)$ independently and set $\psi_j = \frac{1}{\tilde{\psi}_j}$.

Both algorithms are intrinsically connected: The conditional posterior distribution of $T_j|\theta_j$ in the corrected algorithm, which is based on Theorem 2.1 in Bhattacharya et al. (2015), is of the same distribution as the conditional posterior distribution of $\lambda_j|\theta_j$ in the alternative algorithm. The explanation is as follows: Since a Dirichlet-distributed vector always sums up to unity, the following holds: $\sum_{j=1}^n \lambda_j = \sum_{j=1}^n \phi_j\tau = \tau$. That is, ϕ_j for $j = 1,\ldots,n$ can be recovered with $\phi_j = \frac{\lambda_j}{\sum_{j=1}^n \lambda_j}$. Hence, it must be that $T_j \propto \lambda_j$.

We conclude this section by noting that a similar Gibbs sampler based on (4) was employed by the authors to produce some of the results in the article during the revision process, and also in Pati et al. (2014)—the only difference being that λ_j was sampled given θ_j as well as ψ_j (i.e., no blocking was performed). Doing so avoided the need to sample from a generalized inverse Gaussian distribution, whose MATLAB implementation at the time was slow.

4. "Getting It Right"

In the following, we demonstrate that the proposed algorithm does not generate draws from the distributions it aims at sampling from. Here, we are applying a slight variation of "Getting it right" test proposed in Geweke (2004). In a nutshell: Let $\boldsymbol{\kappa} = (\boldsymbol{\psi},\boldsymbol{\phi},\tau)'$ denote the vector of hyperparameters. Then, one possibility to sample from the joint distribution of $p(\boldsymbol{\theta},\boldsymbol{\kappa})$ is the *marginal-conditional simulator* (mcs): It first samples $\boldsymbol{\kappa}^{(m)} \sim p(\boldsymbol{\kappa})$, and then $\boldsymbol{\theta}^{(m)} \sim p(\boldsymbol{\theta}|\boldsymbol{\kappa}^{(m)})$. The resulting sequence $\{\boldsymbol{\kappa}^{(m)},\boldsymbol{\theta}^{(m)}\}$ is iid. Another way of sampling from the joint distribution of $p(\boldsymbol{\theta},\boldsymbol{\kappa})$ is the *successive-conditional simulator* (scs). It is initialized with one single draw $\boldsymbol{\kappa}^{(0)} \sim p(\boldsymbol{\kappa})$. Then, it successively iterates over $\boldsymbol{\theta}^{(m)} \sim p(\boldsymbol{\theta}|\boldsymbol{\kappa}^{(m-1)})$ and $\boldsymbol{\kappa}^{(m)} \sim q(\boldsymbol{\kappa}|\boldsymbol{\kappa}^{(m-1)},\boldsymbol{\theta}^{(m)})$, where $q(\cdot)$ is the transition kernel of the posterior MCMC algorithm with respect to the joint conditional

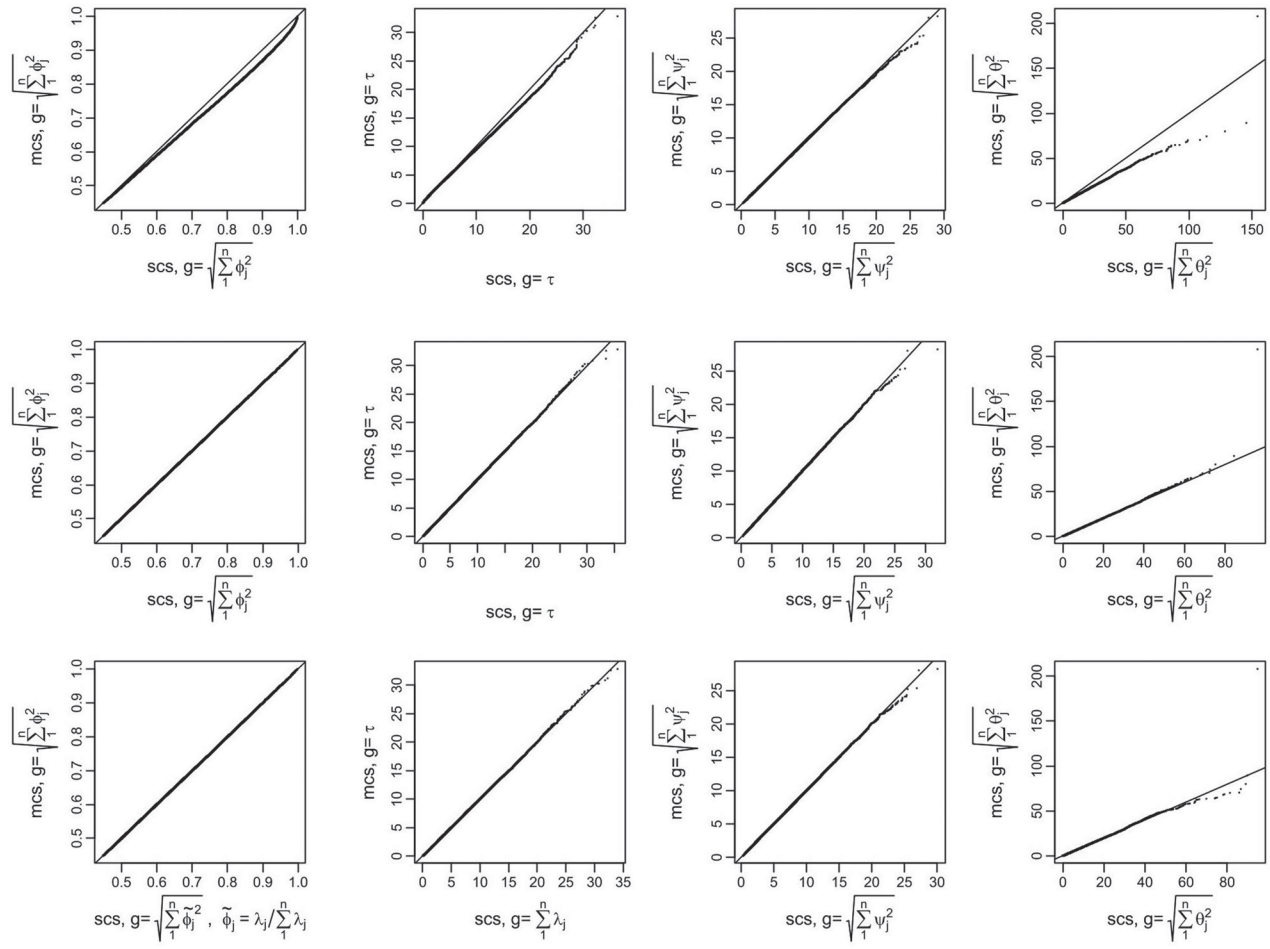

Figure 1. QQ-plots comparing samples from the *marginal-conditional sampler* (mcs) and the *successive-conditional sampler* (scs) with the original (top row), the corrected (middle row), and the alternative (bottom row) MCMC transition kernel. Plots are based on 250,000 iterations of each simulator. The concentration parameter is $a = 0.5$ and the dimensionality is $n = 5$.

posterior $p(\boldsymbol{\kappa}|\boldsymbol{\theta})$. Comparing the samples from both simulators allows for checking the correctness of the joint update.

The top row of Figure 1 depicts QQ-plots comparing samples from the marginal-conditional simulator to samples from the successive-conditional sampler with the transition kernel of the original algorithm for several parameters of the model.[1] For almost all considered test functions $g(\boldsymbol{\kappa})$, the quantiles do not align. The middle row of Figure 1 depicts QQ-plots for the corrected algorithm. The quantiles of the successive-conditional sampler with the corrected transition kernel align with the quantiles of the marginal-conditional sampler. The bottom row of Figure 1 depicts QQ-plots for the alternative algorithm. It demonstrates that the transformations $g^{(m)} = \frac{\lambda_j^{(m)}}{\sum_{j=1}^n \lambda_j^{(m)}}$ and $g^{(m)} = \sum_{j=1}^n \lambda_j^{(m)}$ of the successive-conditional simulator correspond to the sequences of $\phi_j^{(m)}$ and $\tau^{(m)}$ of the marginal-conditional simulator, respectively.

5. Empirical Results

In this section we provide results of the simulation study in sec. 4 in Bhattacharya et al. (2015) based on both correct algorithms. Tables 1 and 2 in this article refer to Tables 1 and 2 in the original manuscript. Differences between the results in the original manuscript and the results presented here using the original algorithm are probably due to the use of different software packages. In general, the scores with respect to the original algorithm align with those presented in Bhattacharya et al. (2015) except for $\mathrm{DL}_{1/n}$ in the scenarios where $A \in \{5, 6\}$, that is in the scenarios with relatively weak signals. For $\mathrm{DL}_{1/n}$, the scores based on the correct algorithms are better for all considered dimensionalities, sparsity and signal levels compared to the scores based on the original algorithm. For $\mathrm{DL}_{1/2}$, the scores based on the correct algorithms are a little worse than the scores based on the original algorithm when the model size q_n is small or when the signal level A is high. Overall, the differences are greater for $\mathrm{DL}_{1/n}$.

Table 3 depicts the results for the high-dimensional simulation setting and refers to Table 3 in the original manuscript. Similarly, for $\mathrm{DL}_{1/2}$, the differences between the results based on the original algorithm and the correct algorithms are relatively small, whereas for $\mathrm{DL}_{1/n}$ the differences are considerable. For most considered scenarios, the scores based on the corrected algorithm and the scores based on the alternative algorithm align almost perfectly. We observe differences in the scenarios

[1] Computer code for all simulators can be found at *https://github.com/gregorkastner/dl*.

Table 1. Squared error comparison over 100 replicates.

	n	100						200					
	$\frac{q_n}{n}$%	5		10		20		5		10		20	
	A	5	6	5	6	5	6	5	6	5	6	5	6
$DL_{1/n}$	original	42.47	17.51	82.95	34.45	143.55	65.15	108.32	44.32	181.16	79.22	316.33	129.44
$DL_{1/n}$	corrected	16.37	7.59	35.37	14.87	66.81	32.42	38.19	15.40	73.99	33.43	140.61	66.51
$DL_{1/n}$	alternative	14.22	6.96	28.24	14.27	61.40	29.99	30.56	14.53	64.57	31.98	126.77	63.30
$DL_{1/2}$	original	12.47	11.19	20.31	17.73	36.83	31.62	25.22	22.69	41.81	36.60	74.11	63.31
$DL_{1/2}$	corrected	13.98	13.18	20.42	18.86	34.10	30.90	28.15	26.59	41.98	38.64	68.56	61.70
$DL_{1/2}$	alternative	13.98	13.25	20.44	18.84	34.17	30.84	28.18	26.49	41.89	38.51	68.67	61.75

NOTE: Average squared error across replicates reported for DL with $a = 1/n$ and $a = 1/2$ based on the original, the corrected, and the alternative algorithm. Signal strength $A = 5$ and $A = 6$.

Table 2. Squared error comparison over 100 replicates.

	n	100						200					
	$\frac{q_n}{n}$%	5		10		20		5		10		20	
	A	7	8	7	8	7	8	7	8	7	8	7	8
$DL_{1/n}$	original	7.43	5.96	14.98	12.86	32.47	25.90	18.51	12.65	35.69	26.90	66.90	52.19
$DL_{1/n}$	corrected	5.72	5.40	11.75	11.22	25.67	23.58	11.90	11.41	26.25	24.03	51.09	46.98
$DL_{1/n}$	alternative	5.73	5.42	11.73	11.16	25.15	23.54	11.93	11.38	25.88	23.85	50.02	46.87
$DL_{1/2}$	original	10.69	10.41	16.74	16.19	29.34	28.22	21.66	21.10	34.34	33.20	58.73	56.40
$DL_{1/2}$	corrected	12.83	12.63	18.25	17.84	29.41	28.67	25.98	25.56	37.23	36.35	58.83	57.16
$DL_{1/2}$	alternative	12.90	12.66	18.25	17.88	29.39	28.64	25.97	25.60	37.17	36.37	58.76	57.13

NOTE: Average squared error across replicates reported for DL with $a = 1/n$ and $a = 1/2$ based on the original, the corrected and the alternative algorithm. Signal strength $A = 7$ and $A = 8$.

Table 3. Squared error comparison over 100 replicates.

	n	1000					
	A	2	3	4	5	6	7
$DL_{1/n}$	original	384.14	795.03	1106.57	845.25	387.41	202.93
$DL_{1/n}$	corrected	365.64	667.63	671.93	381.47	182.34	132.72
$DL_{1/n}$	alternative	358.21	614.98	572.78	316.10	165.99	129.65
$DL_{1/2}$	original	268.00	314.29	266.72	215.07	189.34	178.42
$DL_{1/2}$	corrected	277.08	304.44	256.85	217.63	200.76	193.79
$DL_{1/2}$	alternative	277.23	304.57	257.24	217.54	200.77	193.86

NOTE: Average squared error for the posterior median reported for DL with $a = 1/n$ and $a = 1/2$ based on the original, the corrected, and the alternative algorithm.

where, for $DL_{1/n}$, the squared error loss is strikingly high, though. In these cases, the squared error loss is always the smallest for the alternative algorithm, which seems to be more robust for very small values of the parameter a. We hypothesize that this is to do with numerical and mixing issues, and leave a more detailed investigation for further research.

ORCID

Luis Gruber http://orcid.org/0000-0002-2399-738X
Gregor Kastner http://orcid.org/0000-0002-8237-8271
Debdeep Pati http://orcid.org/0000-0002-5345-8635
David Dunson http://orcid.org/0000-0003-4942-1597

References

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015), "Dirichlet–Laplace Priors for Optimal Shrinkage," *Journal of the American Statistical Association*, 110, 1479–1490. [2011,2012,2013]

Bhattacharyya, A., Pal, S., Mitra, R., and Rai, S. (2022), "Applications of Bayesian Shrinkage Prior Models in Clinical Research with Categorical Responses," *BMC Medical Research Methodology*, 22, 1–19. [2011]

Del Negro, M., and Primiceri, G. E. (2015), "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum," *The Review of Economic Studies*, 82, 1342–1345. [2011]

Geweke, J. (2004), "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association*, 99, 799–804. [2011,2012]

Gruber, L. (2025), "Shrinkage Priors for Bayesian Vector Autoregressions Featuring Stochastic Volatility Using the R Package bayesianVARs," Master's thesis, University of Klagenfurt. [2011]

Huber, F., Kastner, G., and Pfarrhofer, M. (2025), "Introducing Shrinkage in Heavy-Tailed State Space Models to Predict Equity Excess Returns," *Empirical Economics*, 68, 535–553. [2011]

Kastner, G., and Huber, F. (2020), "Sparse Bayesian Vector Autoregressions in Huge Dimensions," *Journal of Forecasting*, 39, 1142–1165. [2011]

Onorati, P., Dunson, D. B., and Canale, A. (2025), "On the Posterior Computation Under the Dirichlet-Laplace Prior," arXiv pre-print 2507.05214. [2011]

Pati, D., Bhattacharya, A., Pillai, N. S., and Dunson, D. (2014), "Posterior Contraction in Sparse Bayesian Factor Models for Massive Covariance Matrices," *The Annals of Statistics*, 42, 1102–1130. [2012]

Primiceri, G. E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821–852. [2011]

Zhang, Y., and Bondell, H. D. (2018), "Variable Selection via Penalized Credible Regions with Dirichlet–Laplace Global-Local Shrinkage Priors," *Bayesian Analysis*, 13, 823–844. [2011]