Supplementary Material of "Using prior-data conflict to tune Bayesian regularized regression models"

Timofei Biziaev[1], Karen Kopciuk[1,3,4], Thierry Chekouo[1,2*]

[1*]Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, T2N 1N4, AB, Canada.
[2]Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, Minneapolis, 55455, MN, USA.
[3]Cancer Epidemiology and Prevention Research, Cancer Care Alberta, Alberta Health Services, 2210 2 street SW, Calgary, T2S 3C3, AB, Canada.
[4]Department of Oncology, Community Health Sciences, University of Calgary, 2500 University Drive NW, Calgary, T2N 1N4, AB, Canada.

*Corresponding author(s). E-mail(s): tchekouo@umn.edu;
Contributing authors: timofei.biziaev@ucalgary.ca;
kakopciu@ucalgary.ca;

1 Supplementary information

Additional plots and further simulation results can be found in this Supplementary Material document. Tables S1 and S3 are expanded versions of Tables 2 and 3 (in the main manuscript) that include the performance of penalized regression models like the LASSO (Tibshirani, 1996), Elastic-Net (Zou and Hastie, 2005), and the Spike-and-Slab LASSO (Ročková and George, 2018). For the LASSO and Elastic-Net, tuning parameters are selected via 10-fold cross validation, and a mixing weight of $\alpha = 0.6$ is used in the Elastic-Net, encouraging a light grouping effect. Cross-validation and estimation are done in the R package `glmnet`. For the Spike-and-Slab LASSO a sequence of spike penalty parameters $\lambda_0 \in \{5, 6, 7, ..., 100\}$ is used, which is similar to the sequence used in the simulation of the original paper (Ročková and George, 2018), and a slab penalty parameter fixed at the default $\lambda_1 = 1$. Estimation is done with the R package `SSLASSO`. Tables S2 and S4 are simulation study results for additional uncorrelated simulation scenarios where the active effect size is ± 1 as opposed to ± 0.5, including the performance of penalized regression models. Figures S1, S2, S3, S4 show the average prior-to-posterior divergences of λ^2 or τ^2 across replicates of a scenario, similar to Figure 6. Tables S7, S8 show simulation results with $n = 50$ and $n = 200$ observations and effects of size ± 1 for which the MB-BLASSO shows improved variable selection performance over the NI-BLASSO. Figure S5 shows the relationship between the marginal maximum-likelihood estimate (MMLE) of λ^2 and the prior-to-posterior divergence of λ^2 when it is given a Gamma prior. Figures S6 and S7 show divergences of λ^2 in the Bayesian LASSO for varying kernel density bandwidths. Figures S8 and S9 relate to the prior-to-posterior divergence of λ^2 for large prior means μ_{λ^2}. Included also is R code for performing linear spike-and-slab regression with $\tau^2 \sim$ Inverse-Gamma$(\mu_{\tau^2}, \sigma_{\tau^2})$ and Beta distributed prior inclusion probability, as well as implementing the Bayesian LASSO with $\lambda^2 \sim$ Gamma$(\mu_{\lambda^2}, \sigma_{\lambda^2})$, both using JAGS interfaced with the R package `runjags`.

1.1 Additional simulations

Table S1 Uncorrelated simulation study results for $p = 200$ with effect size of ± 0.5 using 90 to 100 replicates, including performance of penalized regression models. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-SS: Minimum-Divergence Spike-and-Slab, NI-SS: Noninformative Spike-and-Slab, MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior, SSLASSO: Spike-and-Slab LASSO, EN: Elastic Net.

		$C = 4$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.96 (0.013)	0.22 (0.024)	0.00 (0.000)	0.84 (0.019)
	10	0.96 (0.010)	0.22 (0.023)	0.00 (0.000)	0.84 (0.017)
NI-SS		0.90 (0.024)	0.21 (0.029)	0.00 (0.000)	0.81 (0.025)
MD-BLASSO	1	0.54 (0.034)	0.17 (0.035)	0.01 (0.001)	0.62 (0.032)
	10	0.53 (0.034)	0.20 (0.037)	0.01 (0.001)	0.60 (0.033)
NI-BLASSO		0.77 (0.022)	0.07 (0.012)	0.00 (0.000)	0.82 (0.016)
SSLASSO		0.65 (0.036)	0.12 (0.031)	0.01 (0.001)	0.72 (0.033)
LASSO		0.99 (0.004)	0.77 (0.014)	0.00 (0.000)	0.36 (0.016)
EN		1.00 (0.004)	0.80 (0.011)	0.00 (0.000)	0.31 (0.014)
		$C = 8$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.95 (0.009)	0.22 (0.018)	0.00 (0.000)	0.84 (0.013)
	10	0.94 (0.013)	0.23 (0.020)	0.00 (0.000)	0.83 (0.016)
NI-SS		0.88 (0.025)	0.24 (0.026)	0.00 (0.001)	0.79 (0.024)
MD-BLASSO	1	0.45 (0.026)	0.13 (0.029)	0.02 (0.001)	0.56 (0.027)
	10	0.41 (0.028)	0.19 (0.036)	0.02 (0.001)	0.52 (0.031)
NI-BLASSO		0.55 (0.017)	0.03 (0.007)	0.02 (0.001)	0.69 (0.015)
SSLASSO		0.54 (0.037)	0.16 (0.036)	0.02 (0.001)	0.61 (0.036)
LASSO		0.99 (0.004)	0.75 (0.009)	0.00 (0.000)	0.39 (0.011)
EN		0.99 (0.004)	0.77 (0.008)	0.00 (0.000)	0.37 (0.010)
		$C = 16$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.90 (0.11)	0.38 (0.023)	0.04 (0.017)	0.71 (0.019)
	10	0.90 (0.010)	0.38 (0.022)	0.02 (0.010)	0.70 (0.019)
NI-SS		0.80 (0.026)	0.32 (0.025)	0.02 (0.002)	0.70 (0.024)
MD-BLASSO	1	0.32 (0.018)	0.13 (0.029)	0.06 (0.001)	0.45 (0.022)
	10	0.24 (0.018)	0.29 (0.043)	0.06 (0.001)	0.35 (0.025)
NI-BLASSO		0.29 (0.011)	0.04 (0.013)	0.06 (0.001)	0.43 (0.014)
SSLASSO		0.53 (0.040)	0.29 (0.042)	0.04 (0.003)	0.57 (0.041)
LASSO		0.97 (0.007)	0.71 (0.007)	0.00 (0.001)	0.44 (0.008)
EN		0.97 (0.008)	0.72 (0.007)	0.00 (0.001)	0.42 (0.008)

Table S2 Uncorrelated simulation study results for $p = 200$ with effect size of ± 1 using 90 to 100 replicates, including performance of penalized regression models. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-SS: Minimum-Divergence Spike-and-Slab, NI-SS: Noninformative Spike-and-Slab, MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior, SSLASSO: Spike-and-Slab LASSO, EN: Elastic Net.

		$C = 4$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	1.00 (0.000)	0.05 (0.010)	0.00 (0.000)	0.97 (0.006)
	10	1.00 (0.000)	0.05 (0.010)	0.00 (0.000)	0.97 (0.006)
NI-SS		1.00 (0.000)	0.05 (0.009)	0.00 (0.000)	0.97 (0.005)
MD-BLASSO	1	0.99 (0.006)	0.06 (0.010)	0.00 (0.000)	0.96 (0.007)
	10	0.97 (0.013)	0.05 (0.010)	0.00 (0.000)	0.95 (0.010)
NI-BLASSO		1.00 (0.000)	0.02 (0.007)	0.00 (0.000)	0.99 (0.004)
SSLASSO		1.00 (0.000)	0.00 (0.003)	0.00 (0.00)	1.00 (0.002)
LASSO		1.00 (0.000)	0.78 (0.012)	0.00 (0.000)	0.35 (0.015)
EN		1.00 (0.000)	0.83 (0.008)	0.00 (0.000)	0.28 (0.001)
		$C = 8$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	1.00 (0.000)	0.04 (0.007)	0.00 (0.000)	0.98 (0.004)
	10	1.00 (0.000)	0.04 (0.007)	0.00 (0.000)	0.98 (0.004)
NI-SS		1.00 (0.000)	0.04 (0.007)	0.00 (0.000)	0.98 (0.004)
MD-BLASSO	1	0.90 (0.028)	0.09 (0.022)	0.00 (0.001)	0.89 (0.026)
	10	0.79 (0.038)	0.13 (0.028)	0.01 (0.002)	0.79 (0.035)
NI-BLASSO		0.98 (0.005)	0.00 (0.002)	0.00 (0.000)	0.99 (0.003)
SSLASSO		1.00 (0.000)	0.00 (0.002)	0.00 (0.000)	1.00 (0.001)
LASSO		1.00 (0.000)	0.76 (0.008)	0.00 (0.000)	0.38 (0.011)
EN		1.00 (0.000)	0.80 (0.006)	0.00 (0.000)	0.34 (0.008)
		$C = 16$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	1.00 (0.000)	0.06 (0.006)	0.00 (0.000)	0.97 (0.003)
	10	1.00 (0.000)	0.06 (0.006)	0.00 (0.000)	0.97 (0.003)
NI-SS		1.00 (0.000)	0.05 (0.006)	0.00 (0.000)	0.97 (0.003)
MD-BLASSO	1	0.59 (0.036)	0.14 (0.033)	0.03 (0.003)	0.65 (0.037)
	10	0.27 (0.037)	0.61 (0.049)	0.06 (0.003)	0.30 (0.041)
NI-BLASSO		0.54 (0.013)	0.00 (0.002)	0.04 (0.001)	0.69 (0.011)
SSLASSO		1.00 (0.000)	0.01 (0.003)	0.00 (0.000)	0.99 (0.002)
LASSO		1.00 (0.000)	0.73 (0.006)	0.00 (0.000)	0.42 (0.008)
EN		1.00 (0.000)	0.75 (0.005)	0.00 (0.000)	0.39 (0.006)

Table S3 Uncorrelated simulation study results for $p = 500$ with effect size of ± 0.5 using 90 to 100 replicates, including performance of penalized regression models. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-SS: Minimum-Divergence Spike-and-Slab, NI-SS: Noninformative Spike-and-Slab, MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior, SSLASSO: Spike-and-Slab LASSO, EN: Elastic Net.

		$C = 5$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.91 (0.017)	0.22 (0.019)	0.00 (0.000)	0.82 (0.016)
	10	0.91 (0.017)	0.22 (0.019)	0.00 (0.000)	0.82 (0.017)
NI-SS		0.88 (0.025)	0.20 (0.025)	0.00 (0.000)	0.82 (0.023)
MD-BLASSO	1	0.21 (0.018)	0.33 (0.047)	0.01 (0.000)	0.31 (0.024)
	10	0.18 (0.019)	0.41 (0.049)	0.01 (0.000)	0.27 (0.025)
NI-BLASSO		0.23 (0.018)	0.28 (0.045)	0.01 (0.000)	0.33 (0.024)
SSLASSO		0.59 (0.037)	0.21 (0.038)	0.00 (0.000)	0.63 (0.034)
LASSO		0.98 (0.006)	0.81 (0.009)	0.00 (0.000)	0.31 (0.012)
EN		0.98 (0.006)	0.84 (0.007)	0.00 (0.000)	0.27 (0.010)
		$C = 10$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.77 (0.023)	0.28 (0.022)	0.00 (0.000)	0.73 (0.020)
	10	0.76 (0.021)	0.27 (0.019)	0.00 (0.000)	0.73 (0.018)
NI-SS		0.69 (0.036)	0.30 (0.035)	0.01 (0.001)	0.66 (0.033)
MD-BLASSO	1	0.06 (0.008)	0.57 (0.050)	0.02 (0.000)	0.10 (0.013)
	10	0.05 (0.008)	0.63 (0.049)	0.02 (0.000)	0.09 (0.014)
NI-BLASSO		0.06 (0.008)	0.55 (0.050)	0.02 (0.000)	0.10 (0.013)
SSLASSO		0.46 (0.040)	0.40 (0.044)	0.01 (0.001)	0.47 (0.038)
LASSO		0.90 (0.017)	0.75 (0.012)	0.00 (0.000)	0.36 (0.012)
EN		0.92 (0.014)	0.79 (0.009)	0.00 (0.000)	0.33 (0.010)
		$C = 20$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.45 (0.015)	0.43 (0.019)	0.02 (0.001)	0.49 (0.014)
	10	0.46 (0.013)	0.39 (0.018)	0.02 (0.001)	0.51 (0.013)
NI-SS		0.39 (0.028)	0.53 (0.034)	0.05 (0.014)	0.39 (0.026)
MD-BLASSO	1	0.01 (0.003)	0.73 (0.046)	0.04 (0.000)	0.03 (0.005)
	10	0.01 (0.003)	0.77 (0.065)	0.04 (0.000)	0.02 (0.006)
NI-BLASSO		0.01 (0.002)	0.74 (0.044)	0.04 (0.000)	0.03 (0.005)
SSLASSO		0.11 (0.019)	0.80 (0.033)	0.04 (0.001)	0.12 (0.020)
LASSO		0.67 (0.020)	0.70 (0.013)	0.01 (0.001)	0.39 (0.010)
EN		0.70 (0.022)	0.71 (0.015)	0.01 (0.001)	0.36 (0.010)

Table S4 Uncorrelated simulation study results for $p = 500$ with effect size of ± 1 using 90 to 100 replicates, including performance of penalized regression models. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-SS: Minimum-Divergence Spike-and-Slab, NI-SS: Noninformative Spike-and-Slab, MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior, SSLASSO: Spike-and-Slab LASSO, EN: Elastic Net.

		$C = 5$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	1.00 (0.000)	0.04 (0.008)	0.00 (0.000)	0.98 (0.004)
	10	1.00 (0.000)	0.05 (0.008)	0.00 (0.000)	0.97 (0.005)
NI-SS		1.00 (0.000)	0.03 (0.008)	0.00 (0.000)	0.98 (0.004)
MD-BLASSO	1	0.73 (0.028)	0.03 (0.017)	0.00 (0.000)	0.80 (0.024)
	10	0.52 (0.035)	0.16 (0.037)	0.00 (0.000)	0.61 (0.035)
NI-BLASSO		0.80 (0.020)	0.00 (0.000)	0.00 (0.000)	0.87 (0.015)
SSLASSO		1.00 (0.000)	0.01 (0.004)	0.00 (0.000)	1.00 (0.002)
LASSO		1.00 (0.000)	0.82 (0.008)	0.00 (0.000)	0.30 (0.011)
EN		1.00 (0.000)	0.86 (0.005)	0.00 (0.000)	0.24 (0.007)
		$C = 10$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	1.00 (0.000)	0.04 (0.006)	0.00 (0.000)	0.98 (0.003)
	10	1.00 (0.000)	0.04 (0.006)	0.00 (0.000)	0.98 (0.003)
NI-SS		1.00 (0.001)	0.04 (0.007)	0.00 (0.000)	0.98 (0.004)
MD-BLASSO	1	0.14 (0.012)	0.28 (0.045)	0.02 (0.000)	0.22 (0.018)
	10	0.06 (0.010)	0.62 (0.049)	0.02 (0.000)	0.10 (0.015)
NI-BLASSO		0.16 (0.012)	0.20 (0.040)	0.02 (0.000)	0.15 (0.017)
SSLASSO		1.00 (0.000)	0.00 (0.000)	0.00 (0.000)	1.00 (0.000)
LASSO		1.00 (0.000)	0.80 (0.006)	0.00 (0.000)	0.33 (0.008)
EN		1.00 (0.000)	0.83 (0.005)	0.00 (0.000)	0.28 (0.007)
		$C = 20$			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.94 (0.017)	0.17 (0.028)	0.03 (0.017)	0.85 (0.027)
	10	0.96 (0.013)	0.17 (0.028)	0.05 (0.022)	0.86 (0.026)
NI-SS		0.95 (0.020)	0.08 (0.015)	0.00 (0.001)	0.93 (0.019)
MD-BLASSO	1	0.02 (0.003)	0.69 (0.046)	0.04 (0.000)	0.04 (0.006)
	10	0.01 (0.003)	0.83 (0.038)	0.04 (0.000)	0.02 (0.005)
NI-BLASSO		0.02 (0.003)	0.60 (0.049)	0.04 (0.000)	0.04 (0.006)
SSLASSO		0.92 (0.026)	0.06 (0.023)	0.00 (0.001)	0.92 (0.026)
LASSO		0.92 (0.014)	0.74 (0.006)	0.00 (0.001)	0.40 (0.001)
EN		0.90 (0.019)	0.77 (0.007)	0.00 (0.001)	0.36 (0.009)

Table S5 Correlated simulation study results with effect sizes ±0.5 and ±1 using 90 to 100 replicates, including performance of penalized regression models. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-SS: Minimum-Divergence Spike-and-Slab, NI-SS: Noninformative Spike-and-Slab, MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior, SSLASSO: Spike-and-Slab LASSO, EN: Elastic Net.

		$P = 517$, Correlated Semi-Synthetic Design			
		$C = 5, \pm 1$ EFFECT			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.93 (0.012)	0.08 (0.013)	0.00 (0.000)	0.92 (0.012)
	10	0.92 (0.012)	0.08 (0.012)	0.00 (0.000)	0.92 (0.012)
NI-SS		0.91 (0.013)	0.09 (0.014)	0.00 (0.000)	0.91 (0.013)
MD-BLASSO	1	0.12 (0.012)	0.49 (0.049)	0.01 (0.000)	0.19 (0.018)
	10	0.07 (0.010)	0.70 (0.046)	0.01 (0.000)	0.11 (0.016)
NI-BLASSO		0.13 (0.012)	0.42 (0.049)	0.01 (0.000)	0.21 (0.018)
SSLASSO		0.87 (0.015)	0.14 (0.017)	0.00 (0.000)	0.86 (0.016)
LASSO		0.96 (0.009)	0.85 (0.006)	0.00 (0.000)	0.26 (0.009)
EN		0.99 (0.005)	0.88 (0.003)	0.00 (0.000)	0.21 (0.005)
		$C = 5, \pm 0.5$ EFFECT			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.31 (0.017)	0.42 (0.033)	0.01 (0.000)	0.42 (0.019)
	10	0.31 (0.019)	0.41 (0.034)	0.01 (0.000)	0.43 (0.020)
NI-SS		0.41 (0.024)	0.46 (0.034)	0.01 (0.000)	0.40 (0.022)
MD-BLASSO	1	0.00 (0.000)	1.00 (0.000)	0.01 (0.000)	0.00 (0.000)
	10	0.00 (0.000)	1.00 (0.000)	0.01 (0.000)	0.00 (0.000)
NI-BLASSO		0.01 (0.003)	0.97 (0.017)	0.01 (0.000)	0.01 (0.001)
SSLASSO		0.29 (0.020)	0.61 (0.03)	0.01 (0.000)	0.30 (0.02)
LASSO		0.66 (0.020)	0.87 (0.007)	0.00 (0.000)	0.21 (0.01)
EN		0.76 (0.018)	0.89 (0.004)	0.00 (0.000)	0.17 (0.01)
		$C = 10, \pm 1$ EFFECT			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.90 (0.010)	0.09 (0.011)	0.00 (0.000)	0.90 (0.010)
	10	0.90 (0.011)	0.10 (0.010)	0.00 (0.000)	0.90 (0.010)
NI-SS		0.90 (0.011)	0.09 (0.010)	0.00 (0.000)	0.90 (0.010)
MD-BLASSO	1	0.09 (0.008)	0.33 (0.047)	0.02 (0.000)	0.16 (0.013)
NI-BLASSO		0.10 (0.008)	0.29 (0.048)	0.02 (0.000)	0.17 (0.013)
SSLASSO		0.87 (0.011)	0.11 (0.011)	0.00 (0.000)	0.88 (0.010)
LASSO		0.88 (0.010)	0.85 (0.004)	0.00 (0.000)	0.26 (0.01)
EN		0.90 (0.009)	0.86 (0.004)	0.00 (0.000)	0.20 (0.003)
		$C = 10, \pm 0.5$ EFFECT			
Model	$Var(\tau^2)$	TPR	FDR	FNDR	F1
MD-SS	1	0.36 (0.015)	0.47 (0.025)	0.01 (0.000)	0.42 (0.015)
	10	0.36 (0.015)	0.47 (0.026)	0.01 (0.000)	0.42 (0.015)
NI-SS		0.44 (0.018)	0.50 (0.027)	0.02 (0.011)	0.41 (0.017)
MD-BLASSO	1	0.01 (0.002)	0.95 (0.022)	0.02 (0.000)	0.01 (0.004)
	10	0.00 (0.002)	0.97 (0.017)	0.02 (0.000)	0.01 (0.003)
NI-BLASSO		0.01 (0.003)	0.89 (0.032)	0.02 (0.000)	0.02 (0.006)
SSLASSO		0.32 90.016)	0.48 (0.027)	0.01 (0.000)	0.37 (0.017)
LASSO		0.59 (0.013)	0.85 (0.005)	0.01 (0.000)	0.23 (0.007)
EN		0.64 90.013)	0.86 (0.004)	0.01 (0.000)	0.22 (0.007)

Fig. S1 Average simulated prior-to-posterior KL Divergences of τ^2 in the spike-and-slab model based on 90-100 replicates. Vertical dashed lines denote the first local minimum. For uncorrelated $p = 200$ scenario with $var(\tau^2) = 1$ (red) and $var(\tau^2) = 10$ (black)

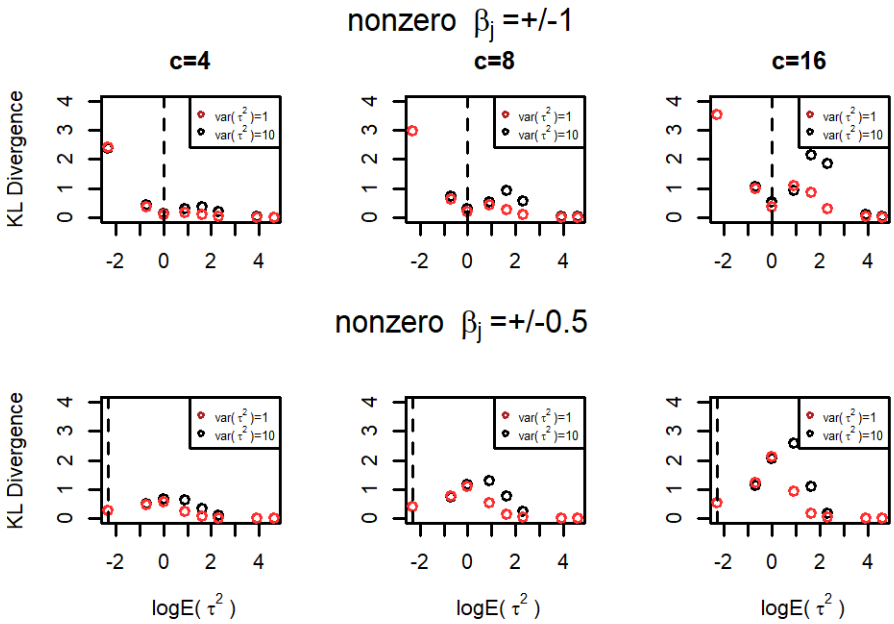


Fig. S2 Average simulated inverse prior-to-posterior KL Divergences of λ^2 in the Bayesian LASSO model based on 90-100 replicates. Vertical dashed lines denote the first local maximum. For uncorrelated $p = 200$ scenario with $var(\lambda^2) = 1$ (red) and $var(\lambda^2) = 10$ (black)

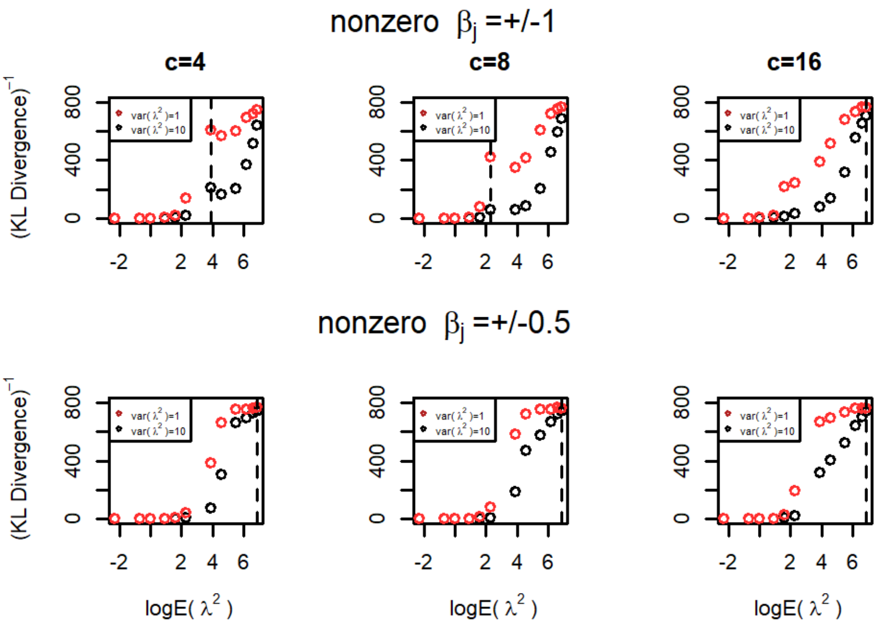


Fig. S3 Average simulated prior-to-posterior KL Divergences of τ^2 in the spike-and-slab model based on 90-100 replicates. Vertical dashed lines denote the first local minimum. For uncorrelated $p = 500$ scenario with $var(\tau^2) = 1$ (red) and $var(\tau^2) = 10$ (black)

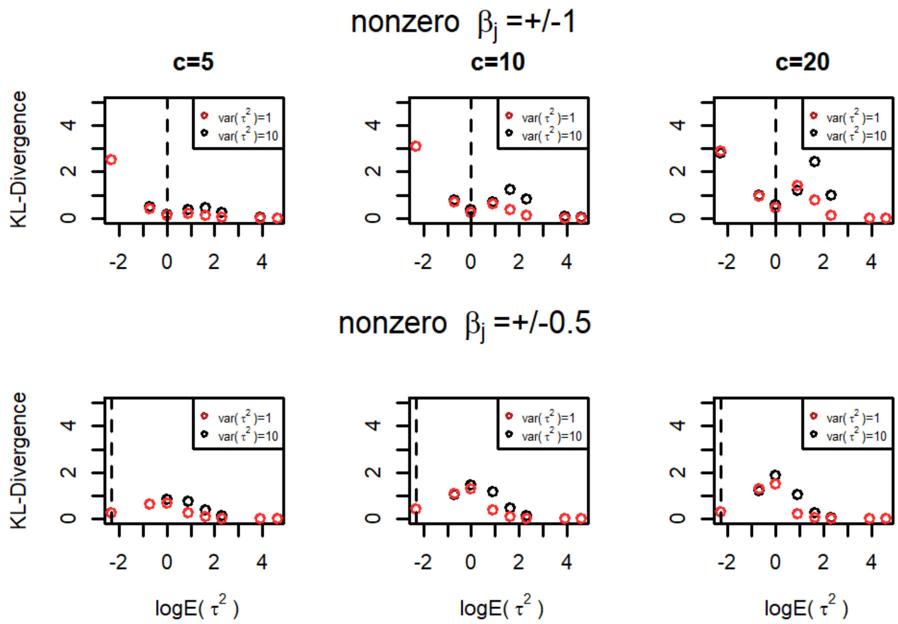
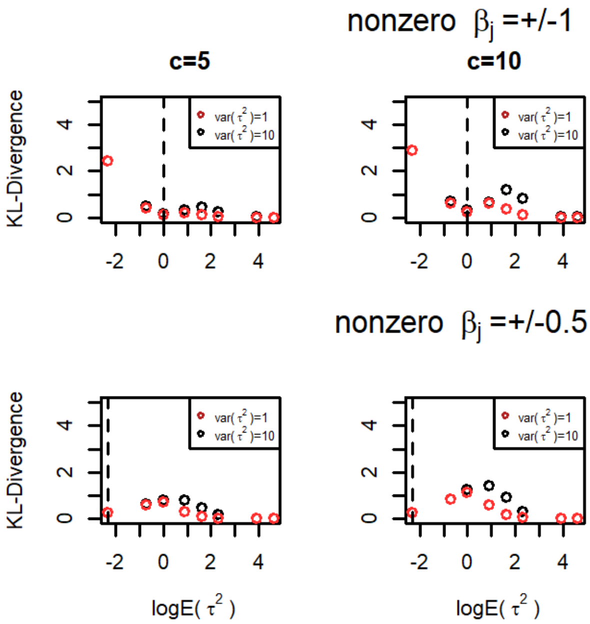


Fig. S4 Average simulated prior-to-posterior KL Divergences of τ^2 in the spike-and-slab model based on 90-100 replicates. Vertical dashed lines denote the first local minimum. For correlated $p = 517$ scenario with $var(\tau^2) = 1$ (red) and $var(\tau^2) = 10$ (black)



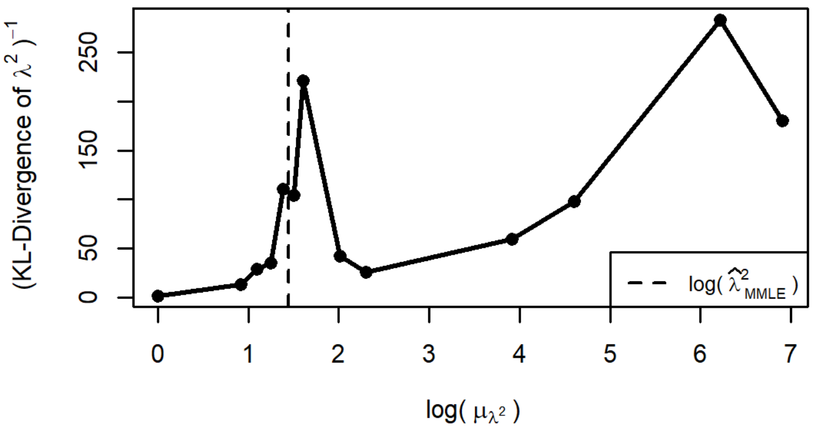
1.2 Relationship between prior-to-posterior Kullback-Leibler divergence and the Marginal Maximum-Likelihood Estimate (MMLE)

In this subsection we provide some intuition and simulation results that relate the notion of minimum-divergence to the marginal maximum-likelihood estimate of a parameter.

Suppose we have the model/likelihood $f(\boldsymbol{y}|\theta)$ and prior $\theta \sim \pi(\theta)$. Prior-data conflict checks are used to check if $\pi(\theta)$ lies in a region where the likelihood for θ is very low. As such, our idea is that a prior on θ that conflicts very little with the data (as measured, in our case, by the prior-to-posterior Kullback-Leibler divergence) should place much of its mass in a region where the likelihood for θ is high. Our motivation then is that a locally minimal divergence for λ^2 or τ^2 would occur somewhere near the marginal maximum-likelihood estimate (MMLE) of λ^2 or τ^2. MMLE's of λ^2 or τ^2 are not fast to obtain (especially in high-dimensions) and our methodology aimed to elicit a prior that placed much of its mass near the MMLE without actually estimating it.

We illustrate this with a simple example using the Bayesian LASSO. We generate $n = 100$ observations from a linear regression model with $p = 10$, $\boldsymbol{\beta} = (-1, 1, 1, -1, 0, 0, 0, 0, 0, 0)$, $\beta_0 = -1$, and predictors $\boldsymbol{X} = (X_1, X_2, ..., X_{10})$ drawn independently from a standard normal distribution. Using the Monte-Carlo EM algorithm proposed in Casella (2001) with initial value 3 and 50 runs of the Gibbs sampler, we obtain an estimate of λ to be $\hat{\lambda}_{MMLE} = 2.06$ by taking the average of the last 10 updates of λ. We then compute the prior-to-posterior divergence of λ^2 when it is assumes the prior Gamma$(\mu_{\lambda^2}, \sigma^2_{\lambda} = 1)$ at each $\mu_{\lambda^2} \in \{1, 2, 2.5, 3, 3.5, 4, 4.5, 5, 10, 50, 100, 500, 1000\}$. Figure S5 shows that the MMLE of λ^2 is in a region where the inverse divergence of λ^2 is quite large.

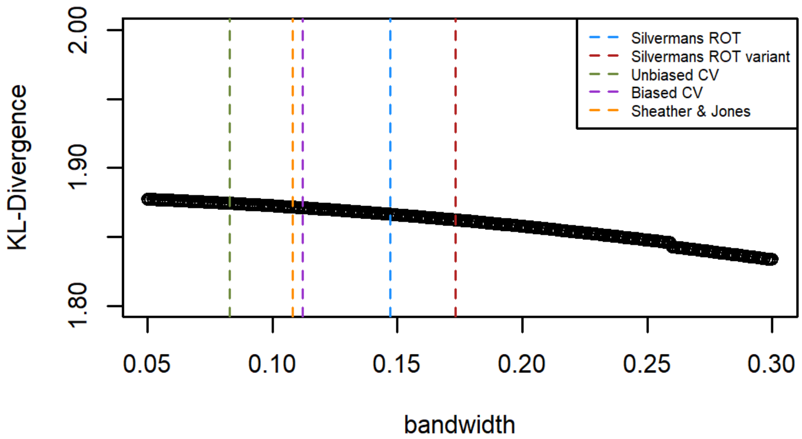
Fig. S5 Inverse prior-to-posterior KL divergences of λ^2 in the Bayesian LASSO with $\lambda^2 \sim$ Gamma$(\mu_{\lambda^2}, \sigma^2_{\lambda^2} = 1)$ for varying prior mean μ_{λ^2}. For $n = 100$ observations simulated from a linear regression model with $\boldsymbol{\beta} = (-1, 1, 1, -1, 0, 0, 0, 0, 0, 0)^T$, $\beta_0 = -1$, and predictors $\boldsymbol{X} = (X_1, X_2, ..., X_{10})$ drawn independently from a standard normal distribution. $\hat{\lambda}_{MMLE}$ is estimated using the MC-EM algorithm with initial value 3 and 50 runs of the Gibbs sampler, taking the average of the last 10 updates of λ as the MMLE.



1.3 Bandwidth sensitivity analysis

Bandwidth selection is critical for kernel density estimation, with smaller bandwidths corresponding to coarser density estimates and larger ones correspond to smoother density estimates. In our methodology the R function `density()` is used to compute kernal density estimates. This function allows the user to specify a custom bandwidth directly or to use one of five bandwidth selectors. The selectors are Silverman's rule of thumb given as $h = 0.9 * \min\{\hat{\sigma}, \frac{\text{IQR}}{1.34}\}n^{-1/5}$ where $\hat{\sigma}$ is the sample standard deviation and IQR the interquartile range, a scaled version of Silverman's rule of thumb given as $h = 1.06 * \min\{\hat{\sigma}, \frac{\text{IQR}}{1.34}\}n^{-1/5}$, and cross-validation methods referred to as

Fig. S6 Prior-to-posterior KL divergences of λ^2 in the Bayesian LASSO for varying kernel density bandwidth using $E(\lambda^2) = 1$, $var(\lambda^2) = 1$. For a realization of data from uncorrelated $n = 100$, $p = 200$, $c = 8$, $+/-0.5$ effects scenario



unbiased and biased cross-validation that minimize an integrated squared-error and asymptotic mean integrated squared-error, respectively. Expressions for these squared errors can be found in Sheather (2004). Lastly, the Sheather & Jones method proposed in Sheather and Jones (1991) can be used. Our methodology uses Silverman's rule of thumb. To investigate sensitivity of divergence to bandwidth we compute prior-to-posterior divergences of λ^2 in the Bayesian LASSO for varying bandwidths, including those given by the bandwidth selectors just described. We do this for a realization of data from the simulation scenario with $n = 100$ observations, $p = 200$ predictors of whom $c = 8$ are significant, with effects of $+/-0.5$. Here λ^2 is given a Gamma($\mu_{\lambda^2} = 1, \sigma^2_{\lambda^2} = 1$) prior and 20,000 samples are drawn from it's posterior distribution. We compute $\mathrm{KL}(\pi(\lambda^2|\boldsymbol{y}_0) \,||\, \pi(\lambda^2))$ for bandwidths ranging from 0.05 to 0.3. We see from Figure S6 that larger bandwidths correspond to smaller divergences.

To investigate whether our methodology is sensitive to bandwidth we compute prior-to-posterior divergences of λ^2 for the above scenario at all prior λ^2 means considered in the manuscript, for each bandwidth selector. Prior means considered are $\mu_{\lambda^2} \in \{0.1, 0.5, 1, 2.5, 5, 10, 50, 100, 250, 500, 750, 1000\}$ with fixed $var(\lambda^2) = 1$.

Fig. S7 Inverse prior-to-posterior KL divergences of λ^2 in the Bayesian LASSO for varying kernel density bandwidth and prior λ^2 mean. For uncorrelated $n = 100$, $p = 200$, $c = 8$, $+/-0.5$ effects scenario with $var(\lambda^2) = 1$

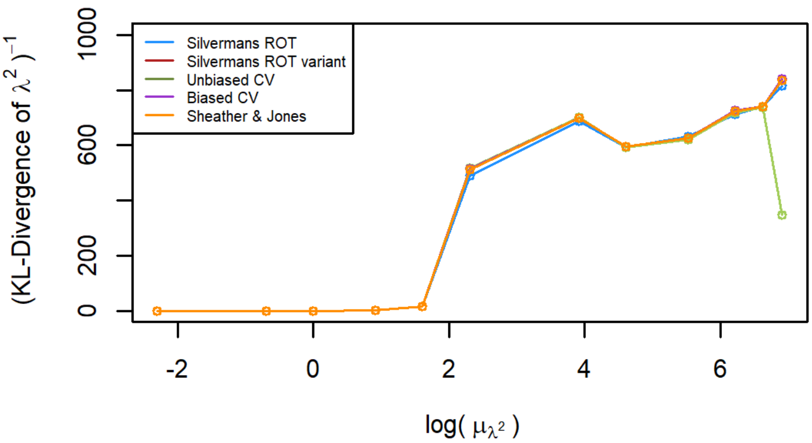


Figure S7 shows there is little difference in divergence between bandwidths considered, with the exception of the divergence using the unbiased cross validation bandwidth at $E(\lambda^2) = 1000$. The first local maximum of $1/\mathrm{KL}(\pi(\lambda^2|\boldsymbol{y}_0) \mid\mid \pi(\lambda^2))$ occurs at $E(\lambda^2) = 50$ for all bandwidths, meaning the proposed MD-BLASSO model specification is the same regardless of bandwidth for this realization of data at this simulation configuration.

1.4 Prior-to-posterior Kullback-Leibler divergence for large $\boldsymbol{\lambda^2}$ in the Bayesian LASSO

We note that for large values of μ_{λ^2} or μ_{τ^2} the prior-to-posterior divergence of λ^2 or τ^2, respectively, becomes very small and stabilizes. We restrict our explanation of this to the Bayesian LASSO as the intuition is the same for spike-and-slab regression. When μ_{λ^2} becomes large, the prior variance on regression effects $\beta_1, ..., \beta_p$ becomes small, inducing substantial shrinkage of effects towards 0. Such a prior places into the regression model the assumption that all effects are basically 0. Because of this assumption of 0 effects, new observations are viewed by the model as having occurred due to chance or error, resulting in a large posterior error variance σ^2. Consequently, the prior distributions of $\boldsymbol{\beta}$ and λ^2 are left largely unchanged by the observed data, resulting in small prior-to-posterior divergence at those parameters.
As a simple illustration we fit the Bayesian LASSO model with $\lambda^2 \sim \mathrm{Gamma}(\mu_{\lambda^2}, \sigma^2_{\lambda^2} = 1)$ for $\mu_{\lambda^2} \in \{1, 10, 100, 500, 1000\}$ to a simulated dataset. 100 observations are simulated from a linear regression model with true $\boldsymbol{\beta} = (1, -1, -1, -1, 0, 0, ..., 0)^T$ of size $p = 20$ with intercept $\beta_0 = -1$, standard normal error variances and standard normal independent predictors $X_1, X_2, ..., X_{20}$. We see

in Figure S8 that the inverse divergences of λ^2 become large at $\log(\mu_{\lambda^2}) = 6.2$ and $\log(\mu_{\lambda^2}) = 6.9$ corresponding to prior means $\mu_{\lambda^2} = 500$ or 1000. Figure S9 illustrates how with increasing shrinkage the posterior error variance $\pi(\sigma^2|\boldsymbol{y}_0)$ takes on larger and larger values.

Fig. S8 Inverse prior-to-posterior KL divergences of λ^2 in the Bayesian LASSO for varying prior λ^2 mean and $var(\lambda^2) = \sigma^2_{\lambda^2} = 1$. For $n = 100$ observations simulated from a linear regression model with true $\boldsymbol{\beta} = (1, -1, -1, -1, 0, 0, ..., 0)^T$ of size $p = 20$.

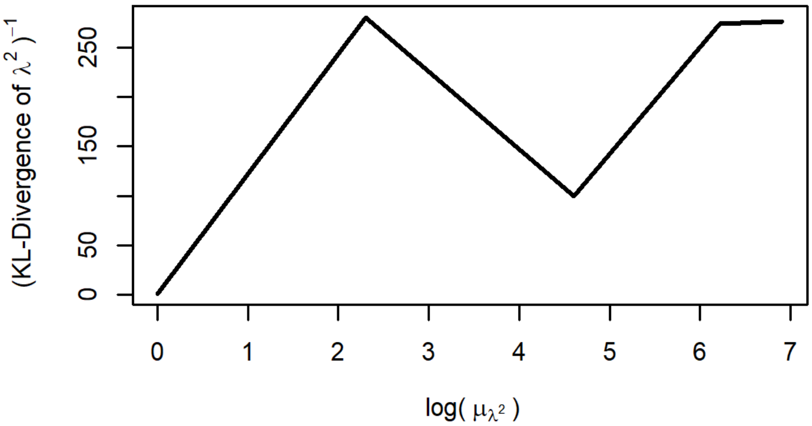
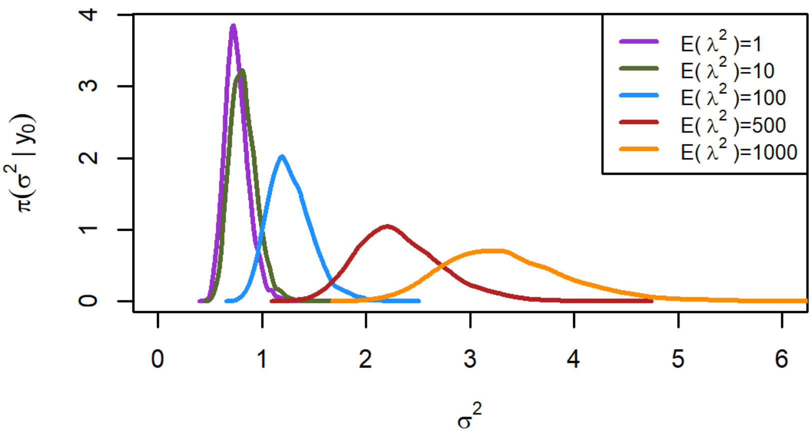


Fig. S9 Posterior distributions of error variance σ^2 in the Bayesian LASSO for varying prior λ^2 mean and $var(\lambda^2) = 1$. For $n = 100$ observations simulated from a linear regression model with true $\boldsymbol{\beta} = (1, -1, -1, -1, 0, 0, ..., 0)^T$ of size $p = 20$.



1.5 Performance of the MD-BLASSO for sample sizes $n = 50$ and $n = 200$

In most simulations the MD-BLASSO tends to perform slightly worse than the NI-BLASSO, due to over-shrinkage of effects. We provide here a scenario where the MD-BLASSO clearly outperforms the NI-BLASSO. We consider an uncorrelated scenario with $n = 50$, $p = 200$, with $c = 4$ nonzero effects of size ± 1. For this configuration we find that the MD-BLASSO with small $var(\lambda^2) = \sigma^2_{\lambda^2}$ of 0.005 provided the best performance. We use the same sequence of prior means as in the main simulation study, namely, $\mu_{\lambda^2} \in \{0.1, 0.5, 1, 2.5, 5, 10, 50, 100, 250, 500, 750, 1000\}$. We fit also the non-informative configuration of the Bayesian LASSO which uses $\lambda^2 \sim \text{Gamma}(\alpha = 10, \beta = 0.1)$ corresponding to $E(\lambda^2) = 100$ and $Var(\lambda^2) = 10^3$ as in previous simulation scenarios. Table S6 gives the 95% posterior credible intervals from both models for the first realization of data. We find that the non-informative configuration induces more shrinkage than the minimum-divergence configuration, and as a result fails to select 2 of the 4 significant covariates. The minimum-divergence configuration selects all 4 relevant covariates and no others. Table S7 shows the variable selection results over 100 replications of this data set. We see that though true positive rates are not high for either models, the MD-BLASSO obtains more true positives, lower false-discoveries, and a better balance between true positive and false discovery as seen in a larger F_1 score.

Table S6 Posterior credible intervals for the truly nonzero regression effects using MD-BLASSO and the NI-BLASSO fit to $n = 50$ responses from the regression model with $\boldsymbol{\beta} = (-1, 1, 1, -1, 0, 0, ..., 0)^T$ of length $p = 200$, intercept $\beta_0 = -1$, independent standard normal predictors $X_1, X_2, ..., X_{200}$, and standard normal error variance σ^2.

Effect	True value	MD-BLASSO posterior mean (95% CI)	NI-BLASSO posterior mean (95% CI)
β_1	-1	-0.64 (-1.13, -0.99)	-0.48 (-0.99, -0.85)
β_2	1	0.44 (0.01, 0.09)	0.32 (-0.03, 0.02)
β_3	1	0.51 (0.06, 0.17)	0.39 (-0.01, 0.05)
β_4	-1	-0.48 (-0.97, -0.83)	-0.38 (-0.87, -0.74)

Table S7 Uncorrelated simulation study results for $n = 50$, $p = 200$, $c = 4$ significant effects of size ± 1 using 100 replicates. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior

Model	$Var(\lambda^2)$	TPR	FDR	FNDR	F1
MD-BLASSO	0.005	0.39 (0.029)	0.22 (0.042)	0.01 (0.001)	0.49 (0.032)
NI-BLASSO	10^3	0.24 (0.023)	0.39 (0.049)	0.02 (0.000)	0.33 (0.029)

We find additionally that with a sample size of $n = 200$, $p = 200$ predictors of which $c = 16$ are significant with effects of ± 1, the MD-BLASSO with hypervariance $\sigma^2_{\lambda^2} = 0.005$ on average outperforms the NI-BLASSO in variable selection. The MD-BLASSO has less false-discoveries and a higher F_1 score (Table S8).

Table S8 Uncorrelated simulation study results for $n = 200$, $p = 200$, $c = 16$ significant effects of size ± 1 using 100 replicates. Each entry refers to the average value of the metric and simulation standard error (in parentheses). MD-BLASSO: Minimum-Divergence Bayesian LASSO, NI-BLASSO: Bayesian LASSO with non-informative prior

Model	$Var(\lambda^2)$	TPR	FDR	FNDR	F1
MD-BLASSO	0.005	1.00 (0.000)	0.02 (0.004)	0.00 (0.000)	0.99 (0.002)
NI-BLASSO	10^3	1.00 (0.000)	0.07 (0.007)	0.00 (0.000)	0.96 (0.004)

Further investigation is required to understand the relationship between the hypervariance $\sigma^2_{\lambda^2}$ and performance of the MD-BLASSO, and under what circumstances a smaller hypervariance like 0.005 improves selection as compared to a larger hypervariance like 1 or 10.

2 Code

2.1 Bayesian LASSO

Below is code for running the Bayesian LASSO model with $\lambda^2 \sim \text{Gamma}(\mu_{\lambda^2}, \sigma^2_{\lambda^2})$

```
#BAYESIAN LASSO
#Loading the JAGS package
library(runjags)

############################################
#FUNCTIONS

#1. Function that computes the prior-to-posterior divergence of lambda^2,
#using kernel density estimate of the posterior

KLD_G=function(post_sample,post_density,prior_shape,prior_rate){

  #A kernel density estimate may be non-zero for negative values
  #which is not possible for a posterior sample of lambda^2.
  #So we first compute the area of the kernel density estimate
  #that is nonzero, in order to scale the posterior probabilities
  #when computing the KL-Divergence
  diff=post_density$x[2]-post_density$x[1]
  if (length(which(post_density$x<0))>0){
    Area=diff*sum(post_density$y[which(post_density$x>=0)])
  } else {Area=1}


  #NOW COMPUTE KLD FOR EACH PARAMETER
  post_kerndens=post_density
  goodind=which(post_kerndens$x>=0 & post_kerndens$y>0)
  DIFF=(max(post_kerndens$x[goodind])-min(post_kerndens$x[goodind]))/length(goodind)
  post_kerndens_x=post_kerndens$x[goodind]
  post_kerndens_y=post_kerndens$y[goodind]/Area
  KLD_tau2=DIFF*sum(post_kerndens_y*(log(post_kerndens_y)-
        log(dgamma(x=post_kerndens_x,
        shape = prior_shape,rate=prior_rate))))

  l=list("Divergence"=KLD_tau2,"Area"=Area,"Differential"=DIFF)
  return(l)
}

#2. Bayesian LASSO model with lambda^2 ~ Inverse-Gamma(shape,rate),
#and sigma^2 ~ Inverse-Gamma(0.1,0.1)

BLASSO_LS <- "model {
  ########### BLASSO priors ##############
  lambda2 ~ dgamma(shape,rate)
  lambda2over2 <- lambda2/2
```

```
  sigma2_inv ~ dgamma(0.1,0.1)
  sigma2 <- pow(sigma2_inv,-1)

  # Sample beta-coefficients for each predictor
  for ( j in 1:P ) {
    tau2[j] ~ dexp(lambda2over2)
    beta[j] ~ dnorm(0,sigma2_inv/tau2[j])
  }
  #### Likelihood ####
  for ( i in 1:N ) {
    y[i] ~ dnorm(mu[i],sigma2_inv)
    mu[i] <- x[i,1:P]%*%beta[]
  }
}"

#3. Function that generates posterior samples of all unknown quantities
#in the Bayesian LASSO for a specified prior mean and variance on
#lambda^2, and computes an estimate of the prior-to-posterior divergence
#of lambda^2

MD_BLASSO = function(Y_tilde, X, mu_lambda2, sigma2_lambda2,
                     burnin_length, sampling_length){
  N=length(Y_tilde);P=dim(X)[2]
  #Compute the shape and rate parameters for lambda2 prior based on mu_lambda2 and
  #sigma2_lambda2
  shape_rate=c(mu_lambda2^2/sigma2_lambda2,mu_lambda2/sigma2_lambda2)
  shape=shape_rate[1];rate=shape_rate[2]
  #Setting initial values for MCMC sampler
  beta_inits1=rnorm(n=P,0,1);beta_inits2=rnorm(n=P,0,1);
  lambda2_inits1=rgamma(n=1,shape=shape_rate[1],rate=shape_rate[2])
  lambda2_inits2=rgamma(n=1,shape=shape_rate[1],rate=shape_rate[2])
  tau2_inits1=rexp(n=P,1/2);tau2_inits2=rexp(n=P,1/2)
  sigma2inv_inits1=rgamma(n=1,shape=0.1,rate=0.1)
  sigma2inv_inits2=rgamma(n=1,shape=0.1,rate=0.1)
  line_data <- list("x" = X, "y" = Y_tilde, "P" = P,
                    "N" = N,"shape"=shape_rate[1],"rate"=shape_rate[2])
  line_inits <- list(list("beta"=beta_inits1,"lambda2"=lambda2_inits1,
                          "tau2"=tau2_inits1,"sigma2_inv"=sigma2inv_inits1,
                          ".RNG.name" = "base::Wichmann-Hill",".RNG.seed" = 314159),
                     list("beta"=beta_inits2,"lambda2"=lambda2_inits2,
                          "tau2"=tau2_inits2,"sigma2_inv"=sigma2inv_inits2,
                          ".RNG.name" = "base::Super-Duper",".RNG.seed" = 3149))
  #Running the MCMC sampler
  blfit <- run.jags(model=BLASSO_LS, monitor=c("beta","tau2","lambda2","sigma2"),
                    data=line_data, n.chains=2, inits=line_inits,
```

```
                    burnin=burnin_length, adapt=1000,sample=sampling_length,
                    method="parallel")
  summary <- summary(blfit)
  #Computing the prior-to-posterior divergence of lambda2
  lambda2_samp = c(blfit$mcmc[[1]][,2*P+1],blfit$mcmc[[2]][,2*P+1])
  KLD = KLD_G(post_sample = lambda2_samp,post_density = density(lambda2_samp),
              prior_shape = shape_rate[1],prior_rate = shape_rate[2])
  lambda2_div <- KLD$Divergence

  return(list('SUMMARY'=summary, 'LAMBDA2 DIVERGENCE'=lambda2_div))

}

#################################################

#Generating N =100 observations from a linear model with
#P = 20 predictors, of whom C = 4 are significant.
#Effects are +/- 0.5, intercept is -1.

N=100;P=20;C=4;eff_size=0.5

set.seed(845)
X=NULL
for (k in 1:P){
  X=cbind(X,rnorm(N,0,1))
}
alpha=-1;betas=c(sample(c(-eff_size,eff_size),size=C,replace=T),rep(0,P-C))
X=scale(X);errors=rnorm(N,0,1)
Y=alpha+X%*%betas+errors
Y_tilde=c(Y-mean(Y))

#Setting prior mean mu_lambda2 = 100, and prior variance sigma2_lambda2 = 1,
#and fitting the Bayesian LASSO
MD = MD_BLASSO(Y_tilde,X,mu_lambda2=100,sigma2_lambda2=1,
               burnin_length=1000,sampling_length=2000)
MD$SUMMARY
MD$`LAMBDA2 DIVERGENCE`
```

2.2 Spike-and-Slab

Below is code for running the spike-and-slab model with $\tau^2 \sim \text{Inverse-Gamma}(\mu_{\tau^2}, \sigma^2_{\tau^2})$

```
#SPIKE-AND-SLAB
#Loading the JAGS package and Inverse-Gamma distribution package
library(runjags)
library(invgamma)

######################################################
#FUNCTIONS

#1. Function that computes the shape and scale of the inverse-gamma
#distribution for a given mean and variance
shape_scale_sqnser=function(museq,sigma2){
  seq=rbind((museq^2/sigma2)+2,museq*(museq^2+sigma2)/sigma2)
  rownames(seq)=c("shape","scale")
  return(seq)
}

#2. Function that computes the prior-to-posterior divergence of tau^2,
#using kernel density estimate of the posterior
KLD_IG=function(post_sample,post_density,prior_shape,prior_scale){

  #A kernel density estimate may be non-zero for negative values
  #which is not possible for a posterior sample of tau^2.
  #So we first compute the area of the kernel density estimate
  #that is nonzero, in order to scale the posterior probabilities
  #when computing the KL-Divergence
  diff=post_density$x[2]-post_density$x[1]
  if (length(which(post_density$x<0))>0){
    Area=diff*sum(post_density$y[which(post_density$x>=0)])
  } else {Area=1}


  #NOW COMPUTE KLD FOR EACH PARAMETER
  post_kerndens=post_density
  goodind=which(post_kerndens$x>=0 & post_kerndens$y>0)
  DIFF=(max(post_kerndens$x[goodind])-min(post_kerndens$x[goodind]))/length(goodind)
  post_kerndens_x=post_kerndens$x[goodind]
  post_kerndens_y=post_kerndens$y[goodind]/Area
  KLD_tau2=DIFF*sum(post_kerndens_y*(log(post_kerndens_y)-
        log(dinvgamma(x=post_kerndens_x,
        shape = prior_shape,rate=prior_scale))))

  l=list("Divergence"=KLD_tau2,"Area"=Area,"Differential"=DIFF)
  return(l)
}
```

```
#3. Spike-and-Slab with tau^2 ~ Inverse-Gamma(shape,rate),
#and sigma^2~ Inverse-Gamma(0.1,0.1)
NormalSpikeSlab_TS <- "model {
  ########### KM priors ##############
  tau2_inv ~ dgamma(shape,rate)
  tau2 <- pow(tau2_inv,-1)
  sigma2_inv ~ dgamma(0.1,0.1)
  sigma2 <- pow(sigma2_inv,-1)
  p_incl ~ dbeta(a,b)

  # Sample indicators and beta-coefficients for each predictor
  for ( j in 1:P ) {
    slab[j] ~ dnorm( 0 , tau2_inv)
    ind[j] ~ dbern(p_incl)
    beta[j] <- ind[j]*slab[j]
  }
  #### Likelihood ####
  for ( i in 1:N ) {
    y[i] ~ dnorm(mu[i],sigma2_inv)
    mu[i] <- x[i,1:P]%*%beta[]
  }
}"

#4. Function that generates posterior samples of all unknown quantities
#in spike-and-slab regression for a specified prior mean and variance on
#Inverse-Gamma distributed tau^2, and computes an estimate of the prior-to-posterior
#divergence of tau^2

MD_SS = function(Y_tilde, X, mu_tau2, sigma2_tau2,
                     burnin_length, sampling_length){
  N=length(Y_tilde);P=dim(X)[2]
  #Compute the shape and rate parameters for lambda2 prior
  #based on mu_lambda2, sigma2_lambda2
  shape_scale=shape_scale_sqnser(museq = mu_tau2,sigma2 = sigma2_tau2)
  #Setting initial values for MCMC sampler
  beta_inits1=rnorm(n=P,0,1);beta_inits2=rnorm(n=P,0,1);
  tau2inv_inits1=rgamma(n=1,shape=shape_scale[1],rate=shape_scale[2]);
  tau2inv_inits2=rgamma(n=1,shape=shape_scale[1],rate=shape_scale[2])
  sigma2inv_inits1=rgamma(n=1,shape=1,rate=1)
  sigma2inv_inits2=rgamma(n=1,shape=1,rate=1)
  line_data <- list("x" = X, "y" = Y_tilde, "P" = P, "N" = N,
                    "a" = 1, "b" = 1,"shape"=shape_scale[1],"rate"=shape_scale[2])
  line_inits <- list(list("ind" = rbinom(n=P,size = 1,prob = 0.5),
                          "slab" = beta_inits1,"tau2_inv"=tau2inv_inits1,
                          "sigma2_inv"=sigma2inv_inits1,
```

```
                         "p_incl" = 0.25,".RNG.name" = "base::Wichmann-Hill",
                         ".RNG.seed" = 314159),
                    list("ind" = rbinom(n=P,size = 1,prob = 0.5),
                         "slab" = beta_inits2, "tau2_inv"=tau2inv_inits2,
                         "sigma2_inv"=sigma2inv_inits2,
                         "p_incl" = 0.75,".RNG.name" = "base::Super-Duper",
                         ".RNG.seed" = 3149))
  #Sampling
  nssfit <- run.jags(model=NormalSpikeSlab_TS,
                    monitor=c("ind","slab","beta","tau2","sigma2","p_incl"),
                    data=line_data, n.chains=2, inits=line_inits,
                    burnin=BURNIN,adapt=1000,
                    sample=SAMPLING,method="parallel")
  summary <- summary(nssfit)

  #Computing the prior-to-posterior divergence of tau^2
  tau2_samp = c(nssfit$mcmc[[1]][,3*P+1],nssfit$mcmc[[2]][,3*P+1])
  KLD = KLD_IG(post_sample = tau2_samp[j],post_density = density(tau2_samp),
               prior_shape = shape_scale[1],prior_scale = shape_scale[2])
  tau2_div <- KLD$Divergence

  return(list('SUMMARY'=summary, 'TAU2 DIVERGENCE'=tau2_div))

}

#######################################################

#Generating N =100 observations from a linear model with
#P = 20 predictors, of whom C = 4 are significant.
#Effects are +/- 0.5, intercept is -1.
N=100;P=20;C=4;eff_size=0.5

set.seed(845)
X=NULL
for (k in 1:P){
  X=cbind(X,rnorm(N,0,1))
}
alpha=-1;betas=c(sample(c(-eff_size,eff_size),size=C,replace=T),rep(0,P-C))
X=scale(X);errors=rnorm(N,0,1)
Y=alpha+X%*%betas+errors
Y_tilde=c(Y-mean(Y))


#Setting prior mean mu_tau2 = 1, and prior variance sigma2_tau2 = 1,
#and fitting the spike-and-slab model
```

```
MD = MD_SS(Y_tilde,X,mu_tau2=1,sigma2_tau2=1,
               burnin_length=1000,sampling_length=2000)
MD$SUMMARY
MD$`TAU2 DIVERGENCE`
```

References

Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics 2*(4), 485–500.

Ročková, V. and E. I. George (2018). The spike-and-slab lasso. *Journal of the American Statistical Association 113*(521), 431–444.

Sheather, S. J. (2004). Density Estimation. *Statistical Science 19*(4), 588 – 597.

Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological) 53*(3), 683–690.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.

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), 301–320.