

Online Supplementary Material for “Sign-based shrinkage based on an asymmetric LASSO penalty”

S1 Proof of Lemma 2.1

Under the ordinary least squares model the ASYMLASSO solution can be written as follows:

$$\hat{\beta}(\tau) = \arg \min_{\beta} \left\{ \frac{1}{2} \|\mathbf{y} - X\beta\|_2^2 + \lambda \sum_{j=1}^p (|\beta_j| + (2\tau - 1)\beta_j) \right\}.$$

Define $\epsilon = \mathbf{y} - X\beta$ as the residual vector and let A be the event such that $\{\lambda(1 + 2|\tau - 1|) \geq \|\epsilon^T X\|_\infty\}$. Then, if A holds,

$$\begin{aligned} \frac{1}{2} \|X(\hat{\beta} - \beta)\|_2^2 &\leq \epsilon^T X(\hat{\beta} - \beta) + \lambda(\|\beta\|_1 - \|\hat{\beta}\|_1) + \lambda(2\tau - 1) \left(\sum_{j=1}^p \beta_j - \sum_{j=1}^p \hat{\beta}_j \right) \\ &\leq |\epsilon^T X|_\infty \|\hat{\beta} - \beta\|_1 + \lambda(\|\beta\|_1 - \|\hat{\beta}\|_1) + \lambda(2\tau - 1) \left(\sum_{j=1}^p \beta_j - \sum_{j=1}^p \hat{\beta}_j \right) \\ &\leq |\epsilon^T X|_\infty \|\hat{\beta} - \beta\|_1 + \lambda(\|\beta\|_1 - \|\hat{\beta}\|_1) + \lambda|2\tau - 1| (\|\beta\|_1 - \|\hat{\beta}\|_1) \\ &\leq \|\epsilon^T X\|_\infty (\|\hat{\beta}\|_1 + \|\beta\|_1) + \lambda(\|\beta\|_1 - \|\hat{\beta}\|_1) + \lambda|2\tau - 1| (\|\beta\|_1 - \|\hat{\beta}\|_1) \\ &= \|\hat{\beta}\|_1 \{\|\epsilon^T X\|_\infty - \lambda(1 + |2\tau - 1|)\} + \|\beta\|_1 \{\|\epsilon^T X\|_\infty + \lambda(1 + |2\tau - 1|)\} \\ &\leq 2\lambda(1 + |2\tau - 1|) \|\beta\|_1, \end{aligned} \tag{1}$$

where the second inequality is due to Holder’s inequality, the third and fourth inequalities are due to the triangle inequality, and the last inequality is due to our assumption on A .

S2 Additional Tables and Figures

Table S1: Asymmetric LASSO (ASYMLASSO) with varying values for τ where $n = 400$, $\Sigma = (0.5^{|i-j|})_{ij}$, $\mu = 0.10$, and $\beta^* = (-0.03, 0, 0, -0.03, -0.03, 0.03, 0.03, 0, 0, 0.03)$. The tuning parameter λ was selected using five-fold cross validation between an evenly-spaced grid $[0.05, 0.95]$. Results are averaged over 100 simulations. $\hat{\tau}_{CV}$ is the average value of τ selected via cross validation for each of the 100 simulations (P_j = proportion of simulations where β_j is correctly identified as non-zero). See Section 3.1 in the main text for more details.

σ	Method	P_1	P_4	P_5	P_6	P_7	P_{10}
0.5	$\hat{\tau}_{CV} = 0.41$	0.38	0.41	0.33	0.37	0.44	0.43
	$\tau = 0.05$	0.31	0.33	0.35	0.44	0.54	0.58
	$\tau = 0.25$	0.32	0.37	0.34	0.47	0.54	0.56
	$\tau = 0.50$	0.41	0.41	0.37	0.38	0.44	0.39
	$\tau = 0.75$	0.47	0.47	0.39	0.34	0.38	0.34
	$\tau = 0.95$	0.46	0.44	0.37	0.32	0.34	0.32
0.3	$\hat{\tau}_{CV} = 0.47$	0.72	0.84	0.70	0.81	0.87	0.86
	$\tau = 0.05$	0.68	0.76	0.67	0.87	0.90	0.92
	$\tau = 0.25$	0.75	0.84	0.66	0.86	0.90	0.92
	$\tau = 0.50$	0.86	0.93	0.75	0.80	0.89	0.88
	$\tau = 0.75$	0.90	0.93	0.83	0.81	0.85	0.84
	$\tau = 0.95$	0.89	0.92	0.81	0.78	0.80	0.81

Table S2: Asymmetric LASSO (ASYMLASSO) with varying values for τ where $n = 400$, $\Sigma = (0.8^{1(i \neq j)})_{ij}$, $\mu = 0.10$, and $\beta^* = (-0.03, 0, 0, -0.03, -0.03, 0.03, 0.03, 0, 0, 0.03)$. The tuning parameter λ was selected using five-fold cross validation between an evenly-spaced grid $[0.05, 0.95]$. Results are averaged over 100 simulations. $\hat{\tau}_{CV}$ is the average value of τ selected via cross validation for each of the 100 simulations (P_j = proportion of simulations where β_j is correctly identified as non-zero). See Section 3.1 in the main text for more details.

σ	Method	P_1	P_4	P_5	P_6	P_7	P_{10}
0.5	$\hat{\tau}_{CV} = 0.42$	0.14	0.09	0.16	0.18	0.16	0.13
	$\tau = 0.05$	0.14	0.13	0.13	0.21	0.17	0.16
	$\tau = 0.25$	0.13	0.13	0.15	0.19	0.17	0.16
	$\tau = 0.50$	0.15	0.11	0.19	0.18	0.13	0.13
	$\tau = 0.75$	0.17	0.13	0.20	0.18	0.14	0.13
	$\tau = 0.95$	0.16	0.13	0.19	0.14	0.15	0.12
0.3	$\hat{\tau}_{CV} = 0.40$	0.21	0.18	0.20	0.24	0.23	0.21
	$\tau = 0.05$	0.17	0.17	0.18	0.29	0.24	0.21
	$\tau = 0.25$	0.19	0.15	0.19	0.28	0.23	0.23
	$\tau = 0.50$	0.22	0.19	0.23	0.24	0.24	0.23
	$\tau = 0.75$	0.25	0.22	0.21	0.23	0.23	0.18
	$\tau = 0.95$	0.27	0.22	0.25	0.24	0.24	0.19

Table S3: Comparison of ASYMLASSO to LASSO and the non-negative LASSO (nLASSO) based on 100 Monte Carlo replicates. (MSB = mean squared bias; FP = mean number of false positives; FN = mean number of false negatives (out of 5); P_j = proportion of simulations where β_j is correctly identified as non-zero; PMSE = Averaged predicted mean squared error.) See Section 3.2 in the main text for more details.

p	Method	MSB	FP	FN	P_1	P_4	P_6	P_7	P_{10}	PMSE
$\Sigma = I$ 50	Oracle	0.06	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.25
	LASSO	0.09	5.14	2.43	0.32	0.28	0.57	0.57	0.83	0.26
	nLASSO	0.09	4.23	2.02	0.37	0.38	0.72	0.64	0.87	0.26
	ASYMLASSO($\hat{\tau} = 0.23$)	0.09	5.40	2.06	0.35	0.39	0.69	0.64	0.87	0.26
200	Oracle	0.05	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.25
	LASSO	0.10	5.41	3.43	0.09	0.24	0.32	0.29	0.63	0.26
	nLASSO	0.10	5.64	3.11	0.17	0.27	0.39	0.34	0.72	0.26
	ASYMLASSO($\hat{\tau} = 0.27$)	0.10	7.59	3.09	0.18	0.28	0.38	0.35	0.72	0.26
$\Sigma = (0.80)_{ij}$ 50	Oracle	0.11	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.25
	LASSO	0.13	6.46	3.02	0.26	0.29	0.46	0.37	0.60	0.26
	nLASSO	0.12	5.93	3.03	0.28	0.27	0.44	0.37	0.61	0.26
	ASYMLASSO($\hat{\tau} = 0.26$)	0.13	6.80	2.98	0.28	0.28	0.46	0.38	0.62	0.26
200	Oracle	0.11	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.25
	LASSO	0.14	9.78	3.87	0.11	0.20	0.18	0.20	0.44	0.26
	nLASSO	0.13	9.32	3.86	0.12	0.19	0.18	0.20	0.45	0.26
	ASYMLASSO($\hat{\tau} = 0.25$)	0.14	10.22	3.85	0.11	0.20	0.19	0.20	0.45	0.26

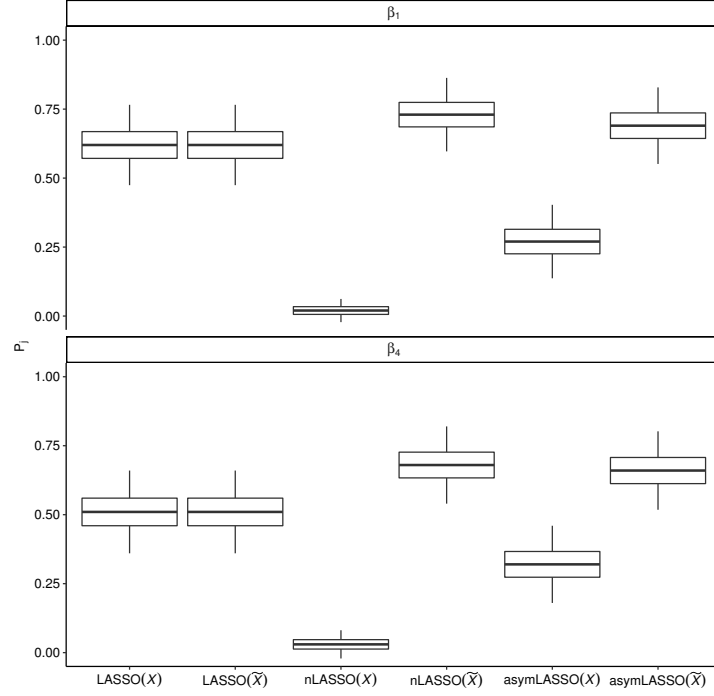


Figure S1: Box plot (mean \pm 3 standard deviations) of the inclusion probability (P_j) for the first and fourth nonzero coefficients ($\beta_{01} = \beta_{04} = -0.03$) when $n = 400$, $p = 50$, and $\Sigma = (0.5^{|i-j|})_{ij}$. A grid search between $[0.05, 0.95]$ is used to select τ for ASYMLASSO. Five-fold cross validation is used to select the final model for LASSO, nLASSO, and ASYMLASSO. Results are averaged over 100 Monte Carlo simulations. (X): Uses the design matrix X in the model fit; (\tilde{X}): Uses the design matrix \tilde{X} in the model fit, where \tilde{X} and X are identical except that the first and fourth columns are negated. See Section 3.3 in the main text for more details.

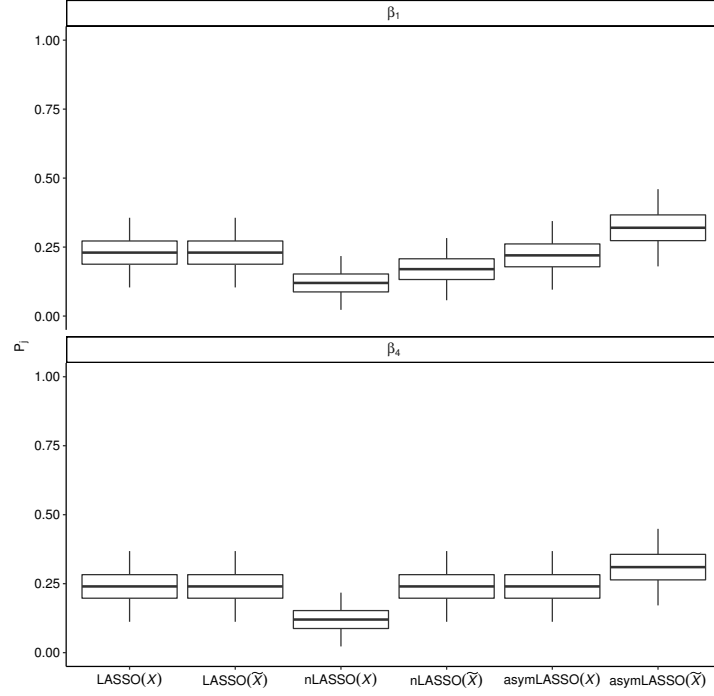


Figure S2: Box plot (mean \pm 3 standard deviations) of the inclusion probability (P_j) for the first and fourth nonzero coefficients ($\beta_{01} = \beta_{04} = -0.03$) when $n = 400$, $p = 50$, and $\Sigma = (0.8^{1(i \neq j)})_{ij}$. A grid search between $[0.05, 0.95]$ is used to select τ for ASYMLASSO. Five-fold cross validation is used to select the final model for LASSO, nLASSO, and ASYMLASSO. Results are averaged over 100 Monte Carlo simulations. (X): Uses the design matrix X in the model fit; (\tilde{X}): Uses the design matrix \tilde{X} in the model fit, where \tilde{X} and X are identical except that the first and fourth columns are negated. See Section 3.3 in the main text for more details.