# Web-based Supporting Information for: Incorporating Prior Information with Fused Sparse Group Lasso: Application to Prediction of Clinical Measures from Neuroimages

Joanne C. Beer, Howard J. Aizenstein,
Stewart J. Anderson, and Robert T. Krafty

## Contents



## List of Tables

# List of Figures

# 1 Web Appendix A: Fusion penalty D matrix example

For a $2 \times 2 \times 2$ cubic image,

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

# 2 Web Appendix B: Optimization Algoritm

## 2.1 Update step for $\boldsymbol{\theta}_j$

The subgradient with respect to $\boldsymbol{\theta}_j$ is

$$\frac{\partial \mathcal{L}_\rho}{\partial \boldsymbol{\theta}_j} = \lambda_j w_j \frac{\partial \|\boldsymbol{\theta}_j\|_2}{\partial \boldsymbol{\theta}_j} + \boldsymbol{\mu}_j + \rho \left(\boldsymbol{\theta}_j - \mathbf{K}_j \boldsymbol{\beta}\right).$$

In general, the subgradient of the $\ell_2$-norm $\|\mathbf{q}\|_2$ is $\mathbf{q}/\|\mathbf{q}\|_2$ if $\mathbf{q} \neq \mathbf{0}$ and $\{\mathbf{r} \mid \|\mathbf{r}\|_2 \leq 1\}$ if $\mathbf{q} = \mathbf{0}$. Therefore, if $\boldsymbol{\theta}_j \neq \mathbf{0}$, the condition $\partial \mathcal{L}_\rho^{\theta_j} / \partial \boldsymbol{\theta}_j = \mathbf{0}$ implies that

$$\boldsymbol{\theta}_j \left(1 + \frac{\lambda_j w_j}{\rho \|\boldsymbol{\theta}_j\|_2}\right) = \mathbf{K}_j \boldsymbol{\beta} - \frac{\boldsymbol{\mu}_j}{\rho}.$$

Let $\boldsymbol{\eta}_j = \mathbf{K}_j \boldsymbol{\beta} - \frac{\boldsymbol{\mu}_j}{\rho}$. The solution can be written in terms of the vector soft-thresholding operator $\mathcal{S}_\kappa(\mathbf{a}) = (1 - \kappa/\|\mathbf{a}\|_2)_+ \mathbf{a}$, where $\mathcal{S}_\kappa(\mathbf{0}) = \mathbf{0}$ and $(\cdot)_+ = \max(0, \cdot)$:

$$\begin{aligned} \boldsymbol{\theta}_j^{t+1} &= \mathcal{S}_{1/\rho}(\boldsymbol{\eta}_j) \\ &= \left(1 - \frac{\lambda_j w_j}{\rho \|\boldsymbol{\eta}_j\|_2}\right)_+ \boldsymbol{\eta}_j. \end{aligned}$$

## 2.2 Stopping Criteria

We use the stopping criteria described in Boyd et al. (2011). The algorithm terminates when the primal and dual residuals are small enough to achieve a linear combination of preselected levels of absolute ($\epsilon_{\text{abs}}$) and relative ($\epsilon_{\text{rel}}$) tolerance. Suitable values for $\epsilon_{\text{abs}}$ and $\epsilon_{\text{rel}}$ will depend on the specific application and scale of the data. Let the primal and dual residuals at iteration $t$ be denoted as $r^t = \boldsymbol{\theta}^t - \mathbf{K}\boldsymbol{\beta}^t$ and $s^t = \rho\mathbf{K}^T(\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t-1})$, respectively. The stopping criteria are $\|r^t\|_2 \leq \epsilon^t_{\text{pri}}$ and $\|s^t\|_2 \leq \epsilon^t_{\text{dual}}$, where

$$\epsilon^t_{\text{pri}} = \sqrt{p}\ \epsilon_{\text{abs}} + \epsilon_{\text{rel}}\ \max\left(\|\mathbf{K}\boldsymbol{\beta}^t\|_2, \|\boldsymbol{\theta}^t\|_2\right),$$

$$\epsilon^t_{\text{dual}} = \sqrt{|\boldsymbol{\theta}^t|}\ \epsilon_{\text{abs}} + \epsilon_{\text{rel}}\ \|\mathbf{K}^T\boldsymbol{\mu}^t\|_2,$$

and $|\boldsymbol{\theta}^t|$ represents the number of elements in $\boldsymbol{\theta}^t$.

## 2.3 Adaptive Step-size

To accelerate the convergence of the ADMM algorithm, we implement an adaptive step-size, $\rho$, following the procedure proposed by He et al. (2000) and implemented in Huo and Tseng (2017);

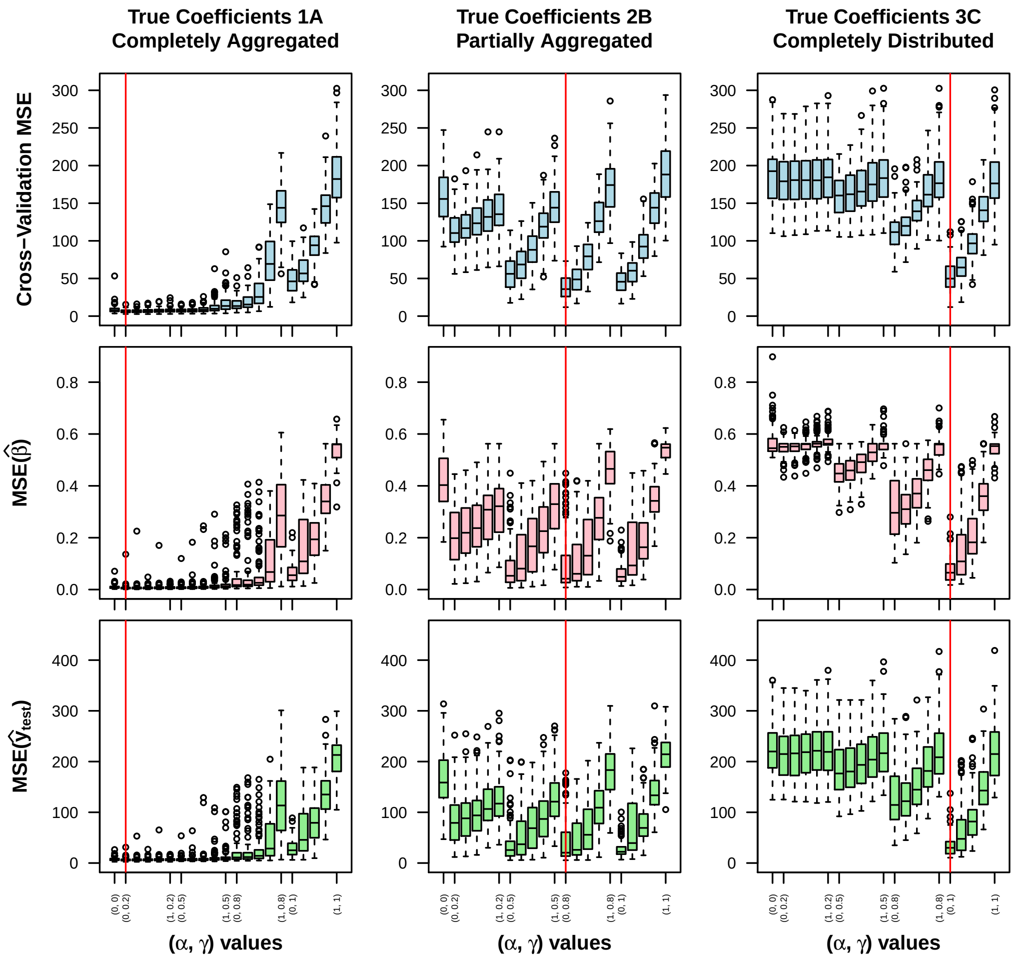
$$\rho^{t+1} = \begin{cases} \tau\rho^t & \text{if } \|r^t\|_2 > \eta\|s^t\|_2 \\ \rho^t/\tau & \text{if } \|r^t\|_2 < \eta\|s^t\|_2 \\ \rho^t & \text{otherwise} \end{cases},$$

where $r^t$ is the primal residual and $s^t$ is the dual residual at iteration $t$, defined above, and we set $\tau = 2$ and $\eta = 10$. This method helps to balance the primal and dual residuals so they converge to zero simultaneously.

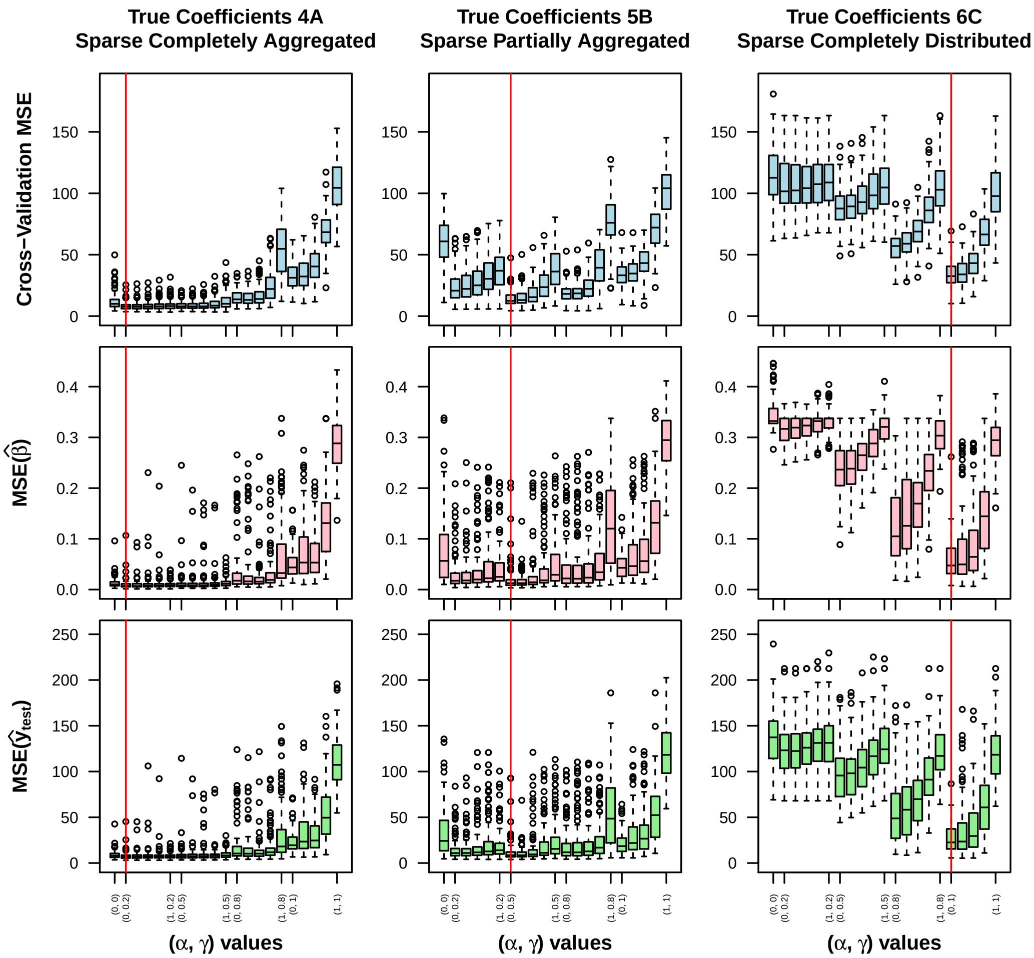
# 3 Web Appendix C: Simulation Study Results

Web Table C1: Optimal ($\alpha$, $\gamma$) for each scenario, based on the most frequent lowest error out of 100 simulation iterations

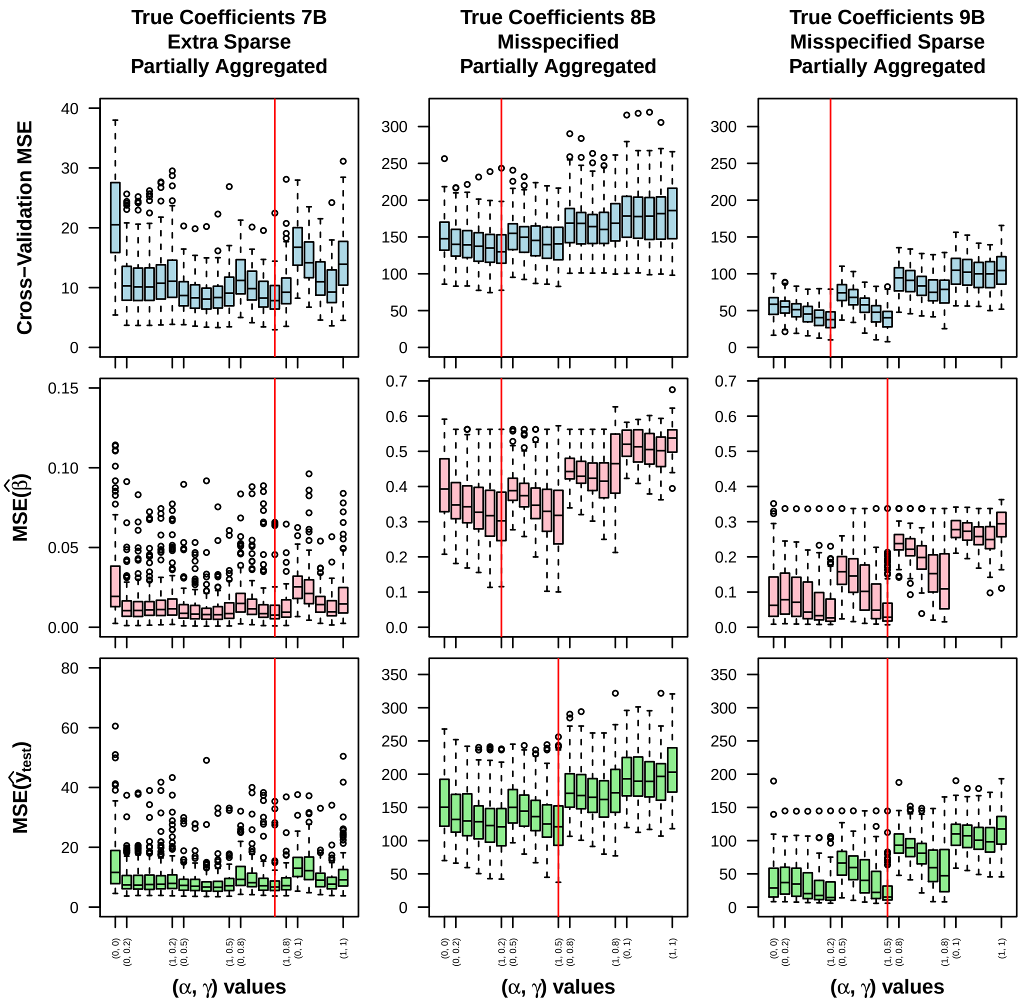
| True coefficient scenario | Mean CVE | MSE($\widehat{\boldsymbol{\beta}}$) | MSE($\widehat{\mathbf{y}}_{\text{test}}$) |
|---|---|---|---|
| 1A. Completely aggregated | (0.0, 0.2) | (0.0, 0.2) | (0.0, 0.2) |
| 2B. Partially aggregated | (0.0, 0.8) | (0.0, 0.8) | (0.0, 0.8) |
| 3C. Completely distributed | (0.0, 1.0) | (0.0, 1.0) | (0.0, 1.0) |
| 4A. Sparse completely aggregated | (0.0, 0.2) | (0.0, 0.2) | (0.0, 0.2) |
| 5B. Sparse partially aggregated | (0.0, 0.5) | (0.0, 0.5) | (0.0, 0.5) |
| 6C. Sparse completely distributed | (0.0, 1.0) | (0.0, 1.0) | (0.0, 1.0) |
| 7B. Extra sparse partially aggregated | (0.8, 0.8) | (0.8, 0.8) | (0.8, 0.8) |
| 8B. Misspecified partially aggregated | (1.0, 0.2) | (1.0, 0.2) | (1.0, 0.5) |
| 9B. Misspecified sparse partially aggregated | (1.0, 0.2) | (1.0, 0.5) | (1.0, 0.5) |




Web Figure C1: Simulation study results for true coefficients 1A, 2B, and 3C. Values of $\gamma \in \{0, 0.2, 0.5, 0.8, 1\}$ increase from left to right on the $x$-axis, corresponding to complete fusion penalty on the left ($\gamma = 0$) and complete sparsity penalties on the right ($\gamma = 1$). Intervals of increasing $\alpha \in \{0, 0.2, 0.5, 0.8, 1\}$ values correspond to complete group penalty on the left ($\alpha = 0$) and complete $\ell_1$ lasso penalty on the right ($\alpha = 1$). Red vertical lines indicate $(\alpha, \gamma)$ combination yielding most frequent lowest error over 100 simulations. MSE: mean squared error.




Web Figure C2: Simulation study results for true coefficients 4A, 5B, and 6C. Values of $\gamma \in \{0, 0.2, 0.5, 0.8, 1\}$ increase from left to right on the $x$-axis, corresponding to complete fusion penalty on the left ($\gamma = 0$) and complete sparsity penalties on the right ($\gamma = 1$). Intervals of increasing $\alpha \in \{0, 0.2, 0.5, 0.8, 1\}$ values correspond to complete group penalty on the left ($\alpha = 0$) and complete $\ell_1$ lasso penalty on the right ($\alpha = 1$). Red vertical lines indicate $(\alpha, \gamma)$ combination yielding most frequent lowest error over 100 simulations. MSE: mean squared error.




Web Figure C3: Simulation study results for true coefficients 7B, 8B, and 9B. Values of $\gamma \in \{0, 0.2, 0.5, 0.8, 1\}$ increase from left to right on the $x$-axis, corresponding to complete fusion penalty on the left ($\gamma = 0$) and complete sparsity penalties on the right ($\gamma = 1$). Intervals of increasing $\alpha \in \{0, 0.2, 0.5, 0.8, 1\}$ values correspond to complete group penalty on the left ($\alpha = 0$) and complete $\ell_1$ lasso penalty on the right ($\alpha = 1$). Red vertical lines indicate $(\alpha, \gamma)$ combination yielding most frequent lowest error over 100 simulations. MSE: mean squared error.

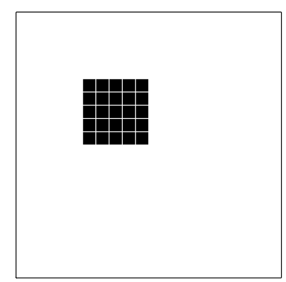
Web Table C2: Simulation results scenario 1A – Completely aggregated

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2$($\hat{y}^*_{\text{i test}}$) | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 6 | 5 | 13 | 8.94 (5.57) | 5.54 (8.29) | 0.010 (0.008) | 7.79 (3.42) | 5.95 (3.08) | 1.93 (2.35) | 2.09 (0.65) |
| 0.0 | 0.2 | 76 | 71 | 53 | 6.55 (2.03) | 3.45 (1.80) | 0.008 (0.013) | 6.72 (3.34) | 6.35 (2.97) | 1.65 (2.05) | 1.51 (0.84) |
| 0.2 | 0.2 | 4 | 1 | 2 | 6.68 (2.13) | 3.57 (1.94) | 0.009 (0.022) | 7.05 (5.19) | 6.29 (2.97) | 1.76 (2.19) | 1.93 (1.37) |
| 0.5 | 0.2 | 1 | 3 | 4 | 6.98 (2.37) | 3.85 (2.36) | 0.008 (0.006) | 6.93 (3.14) | 6.42 (2.79) | 1.74 (2.17) | 1.29 (0.42) |
| 0.8 | 0.2 | 1 | 2 | 5 | 7.44 (2.79) | 4.24 (3.12) | 0.009 (0.017) | 7.52 (6.36) | 6.42 (3.03) | 1.94 (2.44) | 1.75 (0.90) |
| 1.0 | 0.2 | 1 | 0 | 3 | 7.88 (3.25) | 4.66 (3.94) | 0.008 (0.005) | 7.20 (2.78) | 6.05 (3.14) | 1.87 (2.36) | 1.51 (0.41) |
| 0.0 | 0.5 | 9 | 16 | 16 | 7.66 (2.53) | 4.25 (2.62) | 0.009 (0.012) | 7.52 (5.40) | 5.68 (3.01) | 1.98 (2.47) | 1.64 (0.77) |
| 0.2 | 0.5 | 2 | 2 | 1 | 8.02 (2.81) | 4.60 (3.05) | 0.011 (0.019) | 8.06 (6.72) | 5.64 (3.02) | 2.24 (2.76) | 2.05 (1.06) |
| 0.5 | 0.5 | 0 | 0 | 1 | 9.15 (3.80) | 5.68 (4.57) | 0.014 (0.033) | 9.84 (16.66) | 5.87 (3.30) | 2.69 (3.37) | 3.10 (2.16) |
| 0.8 | 0.5 | 0 | 0 | 1 | 12.34 (8.02) | 9.19 (11.02) | 0.019 (0.038) | 11.41 (14.99) | 5.43 (3.43) | 3.68 (4.59) | 4.15 (2.50) |
| 1.0 | 0.5 | 0 | 0 | 0 | 17.86 (12.93) | 14.80 (15.66) | 0.023 (0.037) | 13.07 (16.54) | 5.54 (3.75) | 4.40 (5.59) | 4.73 (2.74) |
| 0.0 | 0.8 | 0 | 0 | 1 | 15.90 (7.80) | 10.99 (7.98) | 0.058 (0.088) | 26.30 (35.33) | 3.84 (3.34) | 10.67 (13.46) | 13.14 (10.68) |
| 0.2 | 0.8 | 0 | 0 | 0 | 19.26 (10.35) | 14.21 (10.66) | 0.057 (0.093) | 25.98 (36.81) | 3.82 (3.45) | 10.36 (13.04) | 13.24 (10.78) |
| 0.5 | 0.8 | 0 | 0 | 0 | 31.00 (17.92) | 24.45 (16.92) | 0.064 (0.090) | 28.27 (33.80) | 3.55 (3.69) | 12.68 (16.19) | 13.81 (10.41) |
| 0.8 | 0.8 | 0 | 0 | 0 | 72.25 (31.30) | 50.31 (24.60) | 0.112 (0.106) | 48.57 (44.95) | 6.24 (9.94) | 26.05 (33.08) | 20.48 (12.58) |
| 1.0 | 0.8 | 0 | 0 | 0 | 141.90 (32.96) | 73.90 (32.59) | 0.295 (0.172) | 118.56 (72.32) | 69.10 (162.39) | 82.91 (106.96) | 37.95 (22.28) |
| 0.0 | 1.0 | 0 | 0 | 0 | 47.35 (17.60) | 30.46 (14.63) | 0.067 (0.042) | 30.44 (17.95) | 4.74 (5.46) | 15.61 (20.22) | 12.51 (4.46) |
| 0.2 | 1.0 | 0 | 0 | 0 | 60.57 (19.61) | 37.54 (16.93) | 0.154 (0.114) | 63.14 (45.42) | 5.56 (8.35) | 40.35 (52.79) | 23.55 (13.62) |
| 0.5 | 1.0 | 0 | 0 | 0 | 93.20 (21.19) | 52.57 (22.68) | 0.196 (0.088) | 80.99 (40.17) | 12.36 (15.73) | 54.97 (71.33) | 26.35 (8.74) |
| 0.8 | 1.0 | 0 | 0 | 0 | 145.06 (29.35) | 68.60 (29.24) | 0.346 (0.083) | 137.98 (42.60) | 39.12 (47.25) | 106.62 (139.48) | 37.42 (9.39) |
| 1.0 | 1.0 | 0 | 0 | 0 | 186.05 (40.96) | 83.08 (38.29) | 0.538 (0.045) | 210.16 (38.85) | 376.83 (409.69) | 195.01 (256.55) | 29.93 (7.02) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2$($\hat{y}^*_{\text{i test}}$) and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 1A: Complete Group**



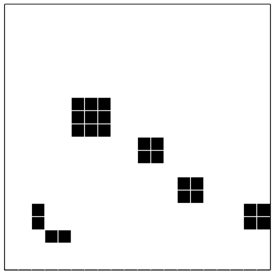
Web Table C3: Simulation results scenario 2B – Partially aggregated

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2(\hat{y}^*_{\text{i test}})$ | Var$(\hat{y}^*_{\text{i test}})$ |
| 0.0 | 0.0 | 0 | 0 | 0 | 161.36 (38.23) | 69.57 (25.79) | 0.412 (0.100) | 167.54 (57.01) | 78.79 (182.06) | 131.13 (202.89) | 44.98 (15.58) |
| 0.0 | 0.2 | 0 | 0 | 0 | 113.64 (24.05) | 59.66 (26.82) | 0.209 (0.115) | 85.08 (49.99) | 6.57 (8.37) | 58.40 (88.40) | 28.47 (12.76) |
| 0.2 | 0.2 | 0 | 0 | 0 | 118.22 (24.93) | 60.80 (26.99) | 0.228 (0.115) | 92.77 (51.14) | 7.31 (9.69) | 64.68 (97.75) | 29.98 (12.59) |
| 0.5 | 0.2 | 0 | 0 | 0 | 125.96 (27.07) | 62.55 (27.38) | 0.244 (0.110) | 98.68 (48.34) | 9.23 (10.92) | 70.24 (106.28) | 31.01 (11.59) |
| 0.8 | 0.2 | 0 | 0 | 0 | 134.38 (29.66) | 64.90 (27.22) | 0.290 (0.112) | 116.32 (49.83) | 15.43 (28.79) | 86.54 (130.15) | 35.32 (12.20) |
| 1.0 | 0.2 | 0 | 0 | 0 | 140.55 (30.60) | 66.63 (27.36) | 0.314 (0.114) | 124.66 (51.72) | 21.21 (42.36) | 94.66 (141.73) | 37.67 (13.44) |
| 0.0 | 0.5 | 0 | 23 | 22 | 56.99 (22.74) | 39.61 (21.30) | 0.088 (0.088) | 38.10 (38.52) | 2.91 (4.28) | 21.33 (31.40) | 14.98 (10.04) |
| 0.2 | 0.5 | 0 | 0 | 1 | 67.52 (23.86) | 45.00 (22.91) | 0.129 (0.116) | 53.70 (47.89) | 3.22 (5.26) | 32.74 (47.75) | 21.92 (14.92) |
| 0.5 | 0.5 | 0 | 1 | 1 | 88.87 (24.52) | 53.73 (26.47) | 0.177 (0.120) | 73.06 (48.94) | 4.42 (6.96) | 48.77 (71.19) | 27.08 (15.57) |
| 0.8 | 0.5 | 0 | 0 | 0 | 119.12 (25.94) | 61.10 (28.37) | 0.228 (0.120) | 92.88 (51.75) | 9.06 (12.88) | 65.96 (96.20) | 31.53 (13.15) |
| 1.0 | 0.5 | 0 | 0 | 0 | 146.06 (31.46) | 68.59 (31.76) | 0.324 (0.118) | 129.73 (53.16) | 28.45 (101.18) | 99.57 (145.03) | 39.56 (14.04) |
| 0.0 | 0.8 | 93 | 63 | 64 | 38.41 (15.34) | 26.35 (15.05) | 0.101 (0.118) | 42.27 (44.70) | 3.03 (3.95) | 23.61 (33.58) | 20.20 (16.04) |
| 0.2 | 0.8 | 0 | 2 | 2 | 49.51 (18.34) | 33.16 (17.59) | 0.115 (0.115) | 48.01 (44.68) | 3.01 (4.66) | 28.82 (41.05) | 21.28 (14.90) |
| 0.5 | 0.8 | 0 | 0 | 0 | 78.65 (21.80) | 47.11 (21.52) | 0.163 (0.118) | 67.12 (46.56) | 4.45 (7.00) | 45.50 (64.89) | 26.37 (15.66) |
| 0.8 | 0.8 | 0 | 0 | 0 | 129.37 (28.36) | 62.76 (28.92) | 0.275 (0.115) | 110.72 (47.70) | 17.12 (32.92) | 85.88 (123.17) | 35.10 (13.10) |
| 1.0 | 0.8 | 0 | 0 | 0 | 174.06 (39.29) | 78.30 (35.14) | 0.457 (0.089) | 183.83 (50.87) | 97.17 (162.28) | 159.57 (230.09) | 41.68 (12.49) |
| 0.0 | 1.0 | 7 | 11 | 10 | 45.76 (15.16) | 28.73 (14.82) | 0.063 (0.041) | 27.66 (16.91) | 4.22 (4.79) | 15.26 (21.43) | 12.20 (4.40) |
| 0.2 | 1.0 | 0 | 0 | 0 | 59.46 (17.51) | 35.71 (17.10) | 0.156 (0.129) | 65.20 (53.71) | 4.44 (5.57) | 41.57 (58.72) | 26.68 (17.32) |
| 0.5 | 1.0 | 0 | 0 | 0 | 92.27 (22.19) | 48.47 (21.58) | 0.190 (0.098) | 77.20 (40.61) | 9.89 (11.65) | 55.19 (78.11) | 28.15 (10.45) |
| 0.8 | 1.0 | 0 | 0 | 0 | 143.86 (30.55) | 67.49 (29.89) | 0.350 (0.082) | 140.95 (40.99) | 55.55 (145.28) | 113.13 (160.89) | 39.75 (10.90) |
| 1.0 | 1.0 | 0 | 0 | 0 | 189.43 (41.04) | 84.10 (35.06) | 0.536 (0.040) | 214.65 (41.14) | 327.31 (377.91) | 206.33 (297.76) | 30.69 (7.50) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2(\hat{y}^*_{\text{i test}})$ and Var$(\hat{y}^*_{\text{i test}})$ were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 2B: Complete Group**



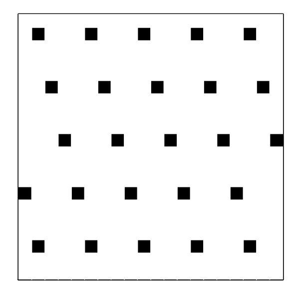
Web Table C4: Simulation results scenario 3C – Completely distributed

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2$($\hat{y}^*_{\text{i test}}$) | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 0 | 0 | 0 | 186.18 (38.07) | 81.51 (30.54) | 0.571 (0.060) | 226.26 (53.30) | 257.34 (330.13) | 202.26 (339.84) | 23.99 (9.36) |
| 0.0 | 0.2 | 0 | 0 | 0 | 179.75 (36.07) | 78.18 (31.74) | 0.546 (0.030) | 218.21 (54.85) | 86.61 (174.00) | 193.44 (325.28) | 25.60 (6.94) |
| 0.2 | 0.2 | 0 | 0 | 0 | 180.28 (36.15) | 78.04 (31.97) | 0.548 (0.028) | 218.14 (53.68) | 145.35 (269.96) | 193.26 (324.96) | 25.84 (6.91) |
| 0.5 | 0.2 | 0 | 0 | 0 | 181.73 (36.47) | 78.74 (32.52) | 0.555 (0.029) | 221.58 (55.77) | 139.96 (243.00) | 196.26 (329.17) | 26.18 (6.87) |
| 0.8 | 0.2 | 0 | 0 | 0 | 183.22 (36.54) | 79.77 (32.22) | 0.564 (0.031) | 225.68 (57.14) | 159.41 (251.41) | 199.34 (331.90) | 27.71 (7.17) |
| 1.0 | 0.2 | 0 | 0 | 0 | 185.10 (37.10) | 80.98 (31.04) | 0.571 (0.033) | 228.76 (59.70) | 162.93 (262.06) | 202.02 (336.57) | 27.99 (7.42) |
| 0.0 | 0.5 | 0 | 0 | 0 | 160.39 (26.65) | 70.79 (29.82) | 0.451 (0.058) | 183.95 (50.80) | 40.76 (142.82) | 155.86 (261.08) | 25.36 (5.65) |
| 0.2 | 0.5 | 0 | 0 | 0 | 163.30 (28.21) | 71.44 (29.99) | 0.461 (0.056) | 188.14 (51.68) | 65.74 (185.52) | 159.73 (266.73) | 25.92 (6.00) |
| 0.5 | 0.5 | 0 | 0 | 0 | 170.64 (32.13) | 73.87 (30.19) | 0.488 (0.049) | 197.49 (52.89) | 48.66 (111.00) | 169.75 (283.79) | 26.10 (5.70) |
| 0.8 | 0.5 | 0 | 0 | 0 | 179.83 (36.16) | 77.06 (31.98) | 0.527 (0.041) | 211.97 (55.64) | 94.74 (176.03) | 187.21 (311.47) | 24.64 (5.42) |
| 1.0 | 0.5 | 0 | 0 | 0 | 186.83 (37.65) | 80.09 (31.39) | 0.559 (0.033) | 223.33 (57.10) | 202.68 (316.35) | 204.64 (341.07) | 22.36 (5.31) |
| 0.0 | 0.8 | 0 | 0 | 0 | 111.58 (23.99) | 57.98 (24.72) | 0.313 (0.113) | 129.59 (56.85) | 19.17 (100.26) | 98.94 (165.22) | 26.51 (9.93) |
| 0.2 | 0.8 | 0 | 0 | 0 | 121.00 (23.23) | 60.04 (25.52) | 0.320 (0.092) | 132.44 (51.23) | 21.33 (100.50) | 101.30 (169.70) | 27.07 (8.34) |
| 0.5 | 0.8 | 0 | 0 | 0 | 140.00 (23.37) | 64.36 (27.52) | 0.374 (0.085) | 152.60 (48.86) | 33.04 (105.68) | 120.12 (201.28) | 30.05 (8.78) |
| 0.8 | 0.8 | 0 | 0 | 0 | 165.60 (29.98) | 71.43 (30.36) | 0.461 (0.064) | 186.33 (51.75) | 62.33 (109.93) | 153.97 (255.86) | 30.33 (7.86) |
| 1.0 | 0.8 | 0 | 0 | 0 | 183.26 (37.49) | 78.29 (31.71) | 0.547 (0.040) | 220.87 (55.85) | 143.25 (151.26) | 194.99 (325.58) | 26.05 (6.40) |
| 0.0 | 1.0 | 100 | 97 | 97 | 52.43 (20.49) | 34.19 (17.19) | 0.075 (0.049) | 34.30 (21.80) | 5.08 (6.53) | 17.11 (28.67) | 13.01 (5.45) |
| 0.2 | 1.0 | 0 | 3 | 3 | 65.85 (22.59) | 40.02 (18.90) | 0.156 (0.130) | 64.80 (51.88) | 6.30 (8.77) | 37.80 (62.75) | 23.92 (16.18) |
| 0.5 | 1.0 | 0 | 0 | 0 | 96.66 (23.95) | 50.94 (21.39) | 0.209 (0.106) | 86.93 (46.70) | 11.53 (15.86) | 55.23 (92.19) | 28.21 (12.10) |
| 0.8 | 1.0 | 0 | 0 | 0 | 142.42 (24.75) | 65.24 (27.57) | 0.365 (0.082) | 149.83 (44.95) | 54.64 (110.55) | 112.07 (186.71) | 34.32 (10.64) |
| 1.0 | 1.0 | 0 | 0 | 0 | 181.07 (37.81) | 79.12 (32.55) | 0.543 (0.038) | 219.54 (55.27) | 274.52 (333.78) | 190.52 (317.56) | 30.71 (7.40) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2$($\hat{y}^*_{\text{i test}}$) and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 3C: Complete Group**



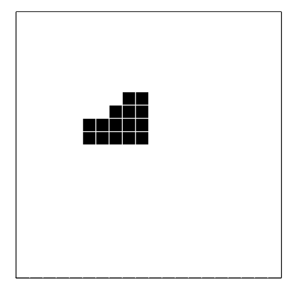
## Web Table C5: Simulation results scenario 4A – Sparse completely aggregated

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2$($\hat{y}^*_{\text{i test}}$) | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 0 | 2 | 7 | 12.23 (7.20) | 7.87 (7.47) | 0.013 (0.011) | 9.03 (4.82) | 5.33 (2.67) | 2.14 (3.68) | 2.52 (0.88) |
| 0.0 | 0.2 | 38 | 33 | 23 | 8.32 (3.45) | 5.04 (3.97) | 0.010 (0.012) | 7.81 (4.91) | 5.67 (2.77) | 1.88 (3.18) | 1.77 (0.79) |
| 0.2 | 0.2 | 12 | 11 | 9 | 8.37 (3.53) | 5.05 (4.02) | 0.010 (0.012) | 7.95 (5.37) | 5.62 (2.80) | 1.89 (3.20) | 1.83 (0.80) |
| 0.5 | 0.2 | 8 | 14 | 9 | 8.52 (3.67) | 5.11 (4.18) | 0.012 (0.026) | 8.94 (11.15) | 5.75 (2.82) | 2.11 (3.56) | 2.37 (1.56) |
| 0.8 | 0.2 | 8 | 8 | 5 | 8.84 (3.98) | 5.29 (4.38) | 0.011 (0.021) | 8.47 (9.11) | 5.84 (2.97) | 2.03 (3.39) | 2.15 (1.18) |
| 1.0 | 0.2 | 2 | 3 | 11 | 9.12 (4.25) | 5.49 (4.79) | 0.010 (0.007) | 7.77 (3.13) | 5.58 (3.05) | 1.90 (3.14) | 1.77 (0.57) |
| 0.0 | 0.5 | 9 | 8 | 11 | 8.79 (3.22) | 5.38 (3.94) | 0.013 (0.026) | 9.15 (11.37) | 5.20 (2.97) | 2.37 (3.87) | 2.49 (1.64) |
| 0.2 | 0.5 | 9 | 8 | 10 | 8.73 (3.31) | 5.39 (4.07) | 0.013 (0.024) | 9.30 (11.08) | 5.33 (3.01) | 2.38 (3.88) | 2.51 (1.67) |
| 0.5 | 0.5 | 9 | 7 | 7 | 9.02 (3.75) | 5.67 (4.81) | 0.016 (0.028) | 10.23 (11.72) | 5.47 (2.88) | 2.76 (4.51) | 3.21 (2.58) |
| 0.8 | 0.5 | 1 | 4 | 6 | 10.09 (4.81) | 6.61 (6.05) | 0.014 (0.023) | 9.57 (10.40) | 5.80 (3.12) | 2.59 (4.17) | 2.65 (1.54) |
| 1.0 | 0.5 | 4 | 2 | 2 | 12.34 (7.18) | 8.79 (9.16) | 0.014 (0.010) | 9.39 (4.47) | 5.82 (3.87) | 2.80 (4.47) | 2.51 (0.80) |
| 0.0 | 0.8 | 0 | 0 | 0 | 15.21 (6.11) | 9.74 (6.95) | 0.038 (0.054) | 18.73 (21.73) | 4.53 (3.52) | 6.64 (10.63) | 7.56 (6.41) |
| 0.2 | 0.8 | 0 | 0 | 0 | 15.02 (6.42) | 9.78 (7.49) | 0.036 (0.054) | 17.78 (21.21) | 4.70 (3.67) | 6.16 (9.83) | 7.26 (6.10) |
| 0.5 | 0.8 | 0 | 0 | 0 | 16.42 (8.14) | 11.26 (9.63) | 0.025 (0.036) | 14.05 (15.74) | 5.32 (4.22) | 4.63 (7.35) | 4.83 (2.97) |
| 0.8 | 0.8 | 0 | 0 | 0 | 25.02 (14.00) | 18.42 (14.67) | 0.034 (0.044) | 17.36 (17.36) | 6.09 (5.18) | 6.54 (10.30) | 6.49 (4.51) |
| 1.0 | 0.8 | 0 | 0 | 0 | 54.74 (23.08) | 38.92 (23.38) | 0.070 (0.073) | 31.33 (31.13) | 19.64 (99.65) | 13.73 (22.20) | 12.18 (8.35) |
| 0.0 | 1.0 | 0 | 0 | 0 | 32.87 (10.61) | 19.86 (11.04) | 0.049 (0.028) | 23.43 (12.33) | 5.06 (5.72) | 10.08 (15.68) | 8.82 (2.91) |
| 0.2 | 1.0 | 0 | 0 | 0 | 34.33 (11.52) | 21.08 (11.64) | 0.078 (0.064) | 34.09 (25.99) | 4.80 (5.85) | 16.70 (26.39) | 12.39 (7.50) |
| 0.5 | 1.0 | 0 | 0 | 0 | 42.24 (13.71) | 25.65 (13.13) | 0.070 (0.049) | 30.99 (19.56) | 7.06 (8.01) | 14.75 (23.58) | 11.47 (5.42) |
| 0.8 | 1.0 | 0 | 0 | 0 | 69.08 (15.29) | 37.38 (16.70) | 0.127 (0.068) | 54.56 (30.75) | 27.79 (100.97) | 29.18 (47.24) | 17.96 (7.64) |
| 1.0 | 1.0 | 0 | 0 | 0 | 106.45 (20.56) | 48.07 (21.44) | 0.283 (0.048) | 112.35 (28.24) | 174.63 (293.11) | 79.35 (128.34) | 25.01 (6.42) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2$($\hat{y}^*_{\text{i test}}$) and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 4A: Sparse Group**



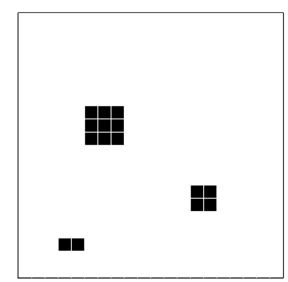
Web Table C6: Simulation results scenario 5B – Sparse partially aggregated

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2(\hat{y}^*_{\text{i test}})$ | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 0 | 0 | 0 | 60.74 (18.60) | 38.84 (16.15) | 0.081 (0.072) | 34.43 (28.12) | 11.66 (44.69) | 17.92 (23.46) | 14.90 (7.58) |
| 0.0 | 0.2 | 0 | 7 | 6 | 24.71 (13.21) | 18.45 (12.05) | 0.031 (0.037) | 15.36 (13.71) | 3.85 (3.29) | 7.13 (9.11) | 5.51 (2.68) |
| 0.2 | 0.2 | 1 | 0 | 2 | 25.92 (13.67) | 19.27 (12.27) | 0.030 (0.038) | 15.41 (15.01) | 3.73 (3.36) | 6.97 (8.91) | 5.66 (3.09) |
| 0.5 | 0.2 | 0 | 0 | 0 | 28.54 (14.53) | 21.27 (12.75) | 0.038 (0.047) | 18.73 (20.58) | 3.57 (4.02) | 8.35 (10.72) | 7.58 (4.50) |
| 0.8 | 0.2 | 0 | 0 | 1 | 32.64 (15.26) | 24.24 (13.61) | 0.050 (0.058) | 23.24 (24.56) | 4.09 (5.82) | 10.96 (13.93) | 9.98 (5.95) |
| 1.0 | 0.2 | 0 | 0 | 0 | 36.81 (15.72) | 27.36 (14.23) | 0.042 (0.043) | 19.95 (17.93) | 4.26 (5.15) | 9.40 (12.06) | 7.98 (4.04) |
| 0.0 | 0.5 | 87 | 74 | 61 | 14.56 (6.62) | 9.83 (6.12) | 0.021 (0.032) | 12.36 (14.26) | 3.89 (2.96) | 5.12 (6.57) | 3.94 (2.39) |
| 0.2 | 0.5 | 8 | 8 | 11 | 15.38 (7.25) | 10.67 (6.86) | 0.017 (0.016) | 10.45 (7.47) | 3.91 (3.00) | 4.51 (5.79) | 2.88 (1.30) |
| 0.5 | 0.5 | 1 | 6 | 7 | 18.32 (9.19) | 13.45 (9.18) | 0.033 (0.051) | 16.23 (20.11) | 4.04 (3.06) | 7.29 (9.46) | 7.05 (5.14) |
| 0.8 | 0.5 | 0 | 1 | 2 | 25.99 (13.22) | 19.58 (13.43) | 0.043 (0.052) | 20.52 (21.99) | 3.89 (3.89) | 9.89 (12.72) | 8.56 (5.60) |
| 1.0 | 0.5 | 0 | 0 | 0 | 38.40 (16.84) | 28.04 (15.05) | 0.054 (0.058) | 25.11 (24.34) | 4.70 (6.05) | 12.71 (16.37) | 10.56 (5.90) |
| 0.0 | 0.8 | 2 | 4 | 5 | 18.94 (7.27) | 12.35 (6.54) | 0.046 (0.059) | 21.12 (22.36) | 3.31 (3.29) | 11.08 (14.21) | 9.64 (7.29) |
| 0.2 | 0.8 | 1 | 0 | 4 | 19.60 (7.81) | 13.11 (7.20) | 0.049 (0.065) | 22.70 (25.69) | 3.32 (3.47) | 11.59 (14.84) | 10.21 (8.00) |
| 0.5 | 0.8 | 0 | 0 | 1 | 23.72 (10.21) | 16.48 (9.79) | 0.045 (0.057) | 20.91 (21.97) | 3.81 (4.15) | 10.87 (13.89) | 9.13 (6.20) |
| 0.8 | 0.8 | 0 | 0 | 0 | 40.48 (16.25) | 27.24 (14.98) | 0.054 (0.052) | 25.03 (22.14) | 5.83 (6.82) | 13.90 (17.85) | 9.86 (5.74) |
| 1.0 | 0.8 | 0 | 0 | 0 | 77.15 (20.95) | 40.52 (18.02) | 0.134 (0.094) | 56.78 (39.95) | 25.28 (45.77) | 37.83 (49.00) | 20.71 (12.13) |
| 0.0 | 1.0 | 0 | 0 | 0 | 33.75 (9.65) | 20.27 (9.30) | 0.047 (0.027) | 21.80 (12.18) | 5.00 (5.13) | 12.38 (15.65) | 8.36 (3.00) |
| 0.2 | 1.0 | 0 | 0 | 0 | 35.27 (10.14) | 21.65 (10.06) | 0.070 (0.059) | 30.93 (25.13) | 4.39 (5.31) | 18.67 (23.87) | 12.18 (6.50) |
| 0.5 | 1.0 | 0 | 0 | 0 | 43.42 (12.12) | 26.68 (12.58) | 0.077 (0.063) | 33.82 (26.18) | 6.12 (7.23) | 21.20 (26.92) | 13.29 (7.28) |
| 0.8 | 1.0 | 0 | 0 | 0 | 70.41 (16.46) | 36.26 (15.87) | 0.129 (0.068) | 54.71 (31.25) | 27.25 (99.98) | 38.48 (49.36) | 18.39 (7.99) |
| 1.0 | 1.0 | 0 | 0 | 0 | 102.05 (19.98) | 43.72 (19.03) | 0.290 (0.051) | 118.21 (29.39) | 144.70 (261.19) | 101.53 (130.82) | 27.52 (8.09) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2(\hat{y}^*_{\text{i test}})$ and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 5B: Sparse Group**



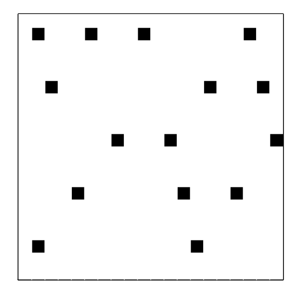
Web Table C7: Simulation results scenario 6C – Sparse completely distributed

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2(\hat{y}^*_{\text{i test}})$ | Var$(\hat{y}^*_{\text{i test}})$ |
| 0.0 | 0.0 | 0 | 0 | 0 | 114.11 (22.85) | 48.67 (19.49) | 0.346 (0.030) | 137.31 (30.12) | 227.33 (327.26) | 133.51 (202.74) | 15.02 (5.65) |
| 0.0 | 0.2 | 0 | 0 | 0 | 106.42 (20.47) | 45.42 (18.89) | 0.314 (0.026) | 125.47 (27.25) | 48.10 (106.77) | 117.40 (180.09) | 16.73 (4.82) |
| 0.2 | 0.2 | 0 | 0 | 0 | 106.70 (20.55) | 45.49 (18.96) | 0.316 (0.026) | 125.97 (26.69) | 119.30 (264.32) | 118.29 (181.42) | 16.78 (4.76) |
| 0.5 | 0.2 | 0 | 0 | 0 | 107.73 (20.83) | 45.43 (18.96) | 0.320 (0.024) | 127.27 (26.33) | 100.34 (204.82) | 119.21 (183.36) | 17.14 (4.82) |
| 0.8 | 0.2 | 0 | 0 | 0 | 109.24 (21.23) | 46.65 (20.24) | 0.326 (0.022) | 130.24 (26.47) | 114.59 (220.97) | 121.97 (187.39) | 17.05 (4.86) |
| 1.0 | 0.2 | 0 | 0 | 0 | 110.24 (21.33) | 46.98 (20.48) | 0.331 (0.022) | 132.43 (27.80) | 142.40 (246.78) | 124.97 (192.42) | 16.31 (4.76) |
| 0.0 | 0.5 | 0 | 0 | 0 | 88.49 (16.21) | 41.65 (16.50) | 0.237 (0.049) | 97.52 (29.58) | 20.62 (101.95) | 83.78 (129.96) | 17.29 (4.77) |
| 0.2 | 0.5 | 0 | 0 | 0 | 90.00 (16.73) | 41.72 (16.47) | 0.240 (0.045) | 99.07 (29.60) | 22.52 (84.51) | 84.68 (131.69) | 17.67 (4.92) |
| 0.5 | 0.5 | 0 | 0 | 0 | 94.13 (18.25) | 42.84 (17.81) | 0.262 (0.040) | 106.71 (28.62) | 30.94 (106.01) | 93.03 (144.94) | 18.00 (4.74) |
| 0.8 | 0.5 | 0 | 0 | 0 | 101.05 (20.87) | 44.43 (19.01) | 0.288 (0.033) | 117.69 (29.07) | 56.95 (144.44) | 103.71 (160.88) | 18.82 (5.06) |
| 1.0 | 0.5 | 0 | 0 | 0 | 105.72 (22.16) | 45.74 (19.36) | 0.316 (0.028) | 127.86 (29.32) | 135.74 (278.01) | 113.37 (175.64) | 20.15 (5.62) |
| 0.0 | 0.8 | 0 | 1 | 1 | 55.42 (12.68) | 32.90 (13.89) | 0.128 (0.076) | 54.93 (34.12) | 5.08 (6.90) | 38.94 (60.79) | 15.38 (8.56) |
| 0.2 | 0.8 | 0 | 0 | 0 | 59.46 (12.62) | 34.04 (14.98) | 0.146 (0.083) | 61.82 (35.45) | 7.28 (12.95) | 45.02 (69.95) | 16.99 (9.68) |
| 0.5 | 0.8 | 0 | 0 | 0 | 69.60 (13.26) | 36.45 (16.20) | 0.166 (0.065) | 70.64 (31.68) | 13.11 (33.77) | 51.99 (81.43) | 17.36 (6.58) |
| 0.8 | 0.8 | 0 | 0 | 0 | 87.08 (17.96) | 40.70 (17.92) | 0.229 (0.052) | 95.89 (31.57) | 26.82 (30.56) | 76.15 (118.67) | 19.39 (5.77) |
| 1.0 | 0.8 | 0 | 0 | 0 | 103.03 (22.38) | 46.77 (18.43) | 0.301 (0.034) | 122.19 (29.10) | 134.00 (235.90) | 107.22 (167.16) | 19.25 (5.42) |
| 0.0 | 1.0 | 85 | 46 | 46 | 33.97 (10.56) | 21.11 (9.57) | 0.058 (0.037) | 27.31 (14.85) | 5.84 (5.54) | 14.33 (22.74) | 9.60 (4.35) |
| 0.2 | 1.0 | 15 | 43 | 42 | 35.44 (11.07) | 22.21 (9.97) | 0.081 (0.075) | 36.55 (32.01) | 5.83 (6.00) | 20.11 (31.56) | 13.97 (10.13) |
| 0.5 | 1.0 | 0 | 10 | 11 | 43.39 (12.19) | 26.67 (11.73) | 0.088 (0.070) | 39.44 (29.69) | 6.97 (7.63) | 22.39 (35.05) | 14.11 (8.47) |
| 0.8 | 1.0 | 0 | 0 | 0 | 67.77 (14.00) | 35.34 (16.81) | 0.143 (0.068) | 62.04 (31.64) | 19.32 (25.39) | 40.18 (62.61) | 18.71 (8.20) |
| 1.0 | 1.0 | 0 | 0 | 0 | 100.76 (21.93) | 45.50 (17.48) | 0.290 (0.040) | 119.29 (30.49) | 177.75 (284.93) | 98.77 (153.94) | 21.48 (6.27) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2(\hat{y}^*_{\text{i test}})$ and Var$(\hat{y}^*_{\text{i test}})$ were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 6C: Sparse Group**

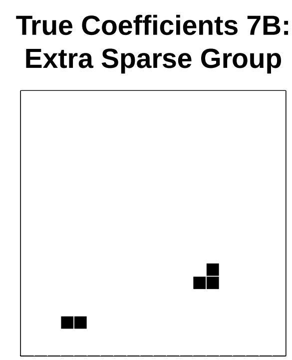


Web Table C8: Simulation results scenario 7B – Extra sparse partially aggregated

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2$($\hat{y}^*_{\text{i test}}$) | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 0 | 0 | 0 | 21.39 (7.72) | 14.64 (7.10) | 0.031 (0.027) | 15.50 (11.15) | 15.59 (100.02) | 4.22 (5.27) | 5.49 (2.16) |
| 0.0 | 0.2 | 0 | 1 | 0 | 11.32 (4.64) | 7.02 (4.27) | 0.014 (0.013) | 9.15 (5.33) | 4.54 (2.69) | 2.20 (2.71) | 2.14 (0.79) |
| 0.2 | 0.2 | 1 | 0 | 0 | 11.25 (4.64) | 6.93 (4.19) | 0.014 (0.014) | 9.25 (5.78) | 4.46 (2.72) | 2.19 (2.71) | 2.25 (0.82) |
| 0.5 | 0.2 | 1 | 1 | 2 | 11.34 (4.78) | 6.98 (4.26) | 0.015 (0.014) | 9.60 (6.17) | 4.65 (2.94) | 2.26 (2.78) | 2.46 (0.88) |
| 0.8 | 0.2 | 1 | 0 | 0 | 11.70 (5.04) | 7.16 (4.52) | 0.016 (0.015) | 9.86 (6.72) | 4.42 (3.18) | 2.27 (2.79) | 2.75 (0.99) |
| 1.0 | 0.2 | 0 | 0 | 0 | 12.16 (5.30) | 7.31 (4.63) | 0.017 (0.016) | 10.31 (7.24) | 4.56 (3.40) | 2.36 (2.91) | 2.97 (1.05) |
| 0.0 | 0.5 | 0 | 5 | 4 | 9.21 (3.09) | 5.47 (3.30) | 0.012 (0.012) | 8.45 (4.74) | 4.73 (3.00) | 1.97 (2.37) | 1.72 (0.72) |
| 0.2 | 0.5 | 3 | 8 | 8 | 8.76 (2.94) | 5.11 (3.07) | 0.012 (0.010) | 8.32 (4.62) | 4.93 (3.04) | 1.87 (2.26) | 1.68 (0.68) |
| 0.5 | 0.5 | 22 | 19 | 12 | 8.45 (2.88) | 4.80 (2.91) | 0.011 (0.011) | 7.89 (5.35) | 5.27 (3.08) | 1.64 (1.97) | 1.68 (0.72) |
| 0.8 | 0.5 | 15 | 18 | 18 | 8.69 (3.13) | 4.99 (2.97) | 0.011 (0.010) | 7.92 (4.58) | 6.25 (3.49) | 1.65 (1.96) | 1.77 (0.70) |
| 1.0 | 0.5 | 5 | 7 | 9 | 9.59 (3.74) | 5.60 (3.61) | 0.012 (0.010) | 8.31 (4.68) | 6.39 (3.81) | 1.72 (2.05) | 2.08 (0.71) |
| 0.0 | 0.8 | 1 | 0 | 0 | 11.58 (3.43) | 6.92 (3.90) | 0.019 (0.015) | 11.11 (5.83) | 4.41 (3.54) | 3.07 (3.61) | 2.67 (0.96) |
| 0.2 | 0.8 | 0 | 0 | 3 | 10.28 (3.18) | 6.14 (3.49) | 0.016 (0.015) | 10.06 (6.21) | 5.04 (3.82) | 2.55 (2.99) | 2.38 (0.89) |
| 0.5 | 0.8 | 9 | 9 | 8 | 8.87 (2.92) | 5.22 (3.04) | 0.015 (0.017) | 9.36 (6.85) | 6.07 (4.16) | 2.14 (2.53) | 2.29 (1.02) |
| 0.8 | 0.8 | 34 | 31 | 32 | 8.46 (3.02) | 4.98 (3.32) | 0.011 (0.011) | 8.21 (5.33) | 7.68 (4.65) | 1.67 (1.98) | 1.87 (0.74) |
| 1.0 | 0.8 | 4 | 0 | 2 | 10.04 (4.01) | 5.96 (3.80) | 0.013 (0.010) | 8.54 (4.59) | 8.80 (5.41) | 1.76 (2.11) | 2.28 (0.75) |
| 0.0 | 1.0 | 0 | 0 | 0 | 16.85 (4.01) | 9.27 (4.88) | 0.027 (0.011) | 14.07 (5.50) | 6.22 (5.43) | 4.37 (5.13) | 4.01 (1.23) |
| 0.2 | 1.0 | 0 | 0 | 0 | 14.63 (3.80) | 8.29 (4.41) | 0.026 (0.017) | 13.58 (6.35) | 6.05 (5.20) | 4.23 (4.94) | 3.69 (1.16) |
| 0.5 | 1.0 | 0 | 0 | 2 | 11.53 (3.51) | 6.69 (3.62) | 0.016 (0.009) | 9.80 (4.24) | 6.59 (5.15) | 2.45 (2.87) | 2.46 (0.80) |
| 0.8 | 1.0 | 4 | 1 | 0 | 10.01 (3.56) | 6.00 (3.47) | 0.013 (0.009) | 8.66 (4.16) | 9.40 (6.85) | 1.93 (2.25) | 2.14 (0.78) |
| 1.0 | 1.0 | 0 | 0 | 0 | 14.43 (5.59) | 9.23 (5.71) | 0.020 (0.016) | 11.51 (7.50) | 14.31 (12.13) | 2.80 (3.28) | 3.60 (1.40) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2$($\hat{y}^*_{\text{i test}}$) and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

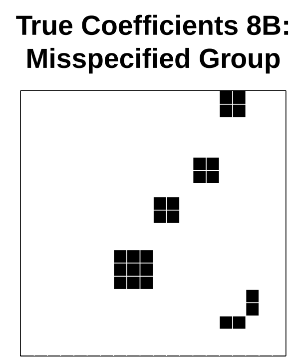


Web Table C9: Simulation results scenario 8B – Misspecified partially aggregated

| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2$($\hat{y}^*_{\text{i test}}$) | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 10 | 1 | 2 | 151.50 (34.20) | 65.97 (26.18) | 0.398 (0.096) | 156.39 (46.87) | 49.86 (118.57) | 126.56 (177.64) | 38.71 (13.11) |
| 0.0 | 0.2 | 4 | 0 | 3 | 142.43 (28.98) | 63.77 (25.98) | 0.357 (0.082) | 140.30 (41.22) | 15.97 (30.86) | 113.70 (167.41) | 32.32 (9.21) |
| 0.2 | 0.2 | 4 | 2 | 1 | 140.45 (28.87) | 63.39 (25.84) | 0.348 (0.087) | 137.27 (41.70) | 29.16 (101.83) | 110.26 (162.61) | 32.15 (9.51) |
| 0.5 | 0.2 | 4 | 4 | 6 | 138.00 (28.85) | 63.85 (26.03) | 0.333 (0.092) | 131.93 (42.26) | 25.00 (76.71) | 104.08 (154.64) | 32.38 (9.60) |
| 0.8 | 0.2 | 4 | 13 | 3 | 135.65 (29.47) | 63.57 (26.44) | 0.324 (0.102) | 128.45 (45.06) | 42.49 (144.88) | 100.22 (149.36) | 32.60 (10.29) |
| 1.0 | 0.2 | 34 | 34 | 33 | 134.09 (29.94) | 64.33 (27.88) | 0.313 (0.103) | 123.58 (44.67) | 27.30 (76.44) | 95.64 (142.71) | 32.65 (10.18) |
| 0.0 | 0.5 | 2 | 2 | 2 | 152.96 (28.96) | 66.43 (28.43) | 0.394 (0.056) | 153.40 (34.31) | 9.00 (19.43) | 128.86 (195.84) | 28.78 (5.87) |
| 0.2 | 0.5 | 1 | 1 | 1 | 150.01 (28.38) | 66.00 (28.78) | 0.381 (0.064) | 148.79 (35.11) | 15.31 (54.18) | 123.31 (186.70) | 29.12 (6.57) |
| 0.5 | 0.5 | 0 | 2 | 3 | 145.38 (28.18) | 65.84 (29.93) | 0.359 (0.074) | 140.95 (36.72) | 12.78 (25.78) | 113.80 (173.23) | 29.90 (6.99) |
| 0.8 | 0.5 | 9 | 7 | 8 | 141.57 (28.70) | 65.28 (30.36) | 0.330 (0.094) | 129.48 (42.97) | 15.98 (28.68) | 99.90 (152.01) | 32.08 (9.23) |
| 1.0 | 0.5 | 13 | 33 | 34 | 141.65 (30.88) | 66.16 (29.11) | 0.321 (0.109) | 125.84 (46.55) | 23.17 (34.29) | 95.93 (146.64) | 33.59 (10.50) |
| 0.0 | 0.8 | 1 | 1 | 1 | 169.75 (35.38) | 71.04 (27.84) | 0.455 (0.048) | 175.19 (38.10) | 31.86 (111.09) | 153.02 (234.84) | 26.48 (5.53) |
| 0.2 | 0.8 | 1 | 0 | 1 | 167.49 (34.50) | 70.38 (28.86) | 0.445 (0.051) | 172.08 (38.26) | 18.44 (35.66) | 147.54 (226.18) | 27.36 (5.95) |
| 0.5 | 0.8 | 0 | 0 | 1 | 164.91 (33.28) | 70.34 (31.17) | 0.432 (0.060) | 166.88 (36.26) | 20.55 (37.99) | 139.73 (213.46) | 30.46 (6.64) |
| 0.8 | 0.8 | 2 | 0 | 0 | 164.13 (34.34) | 72.87 (31.52) | 0.422 (0.073) | 163.97 (39.91) | 32.03 (48.05) | 132.80 (202.47) | 33.80 (8.08) |
| 1.0 | 0.8 | 3 | 0 | 0 | 170.82 (35.91) | 74.67 (32.96) | 0.454 (0.098) | 176.08 (49.83) | 93.30 (137.74) | 139.90 (214.04) | 38.19 (11.20) |
| 0.0 | 1.0 | 2 | 0 | 0 | 181.79 (40.05) | 75.84 (30.05) | 0.517 (0.040) | 198.43 (42.48) | 206.69 (375.26) | 177.13 (270.24) | 25.31 (5.95) |
| 0.2 | 1.0 | 1 | 0 | 0 | 180.80 (39.62) | 76.67 (29.93) | 0.510 (0.046) | 197.24 (42.05) | 83.45 (124.26) | 176.52 (270.56) | 24.01 (5.80) |
| 0.5 | 1.0 | 1 | 0 | 1 | 179.73 (39.58) | 76.43 (30.09) | 0.502 (0.050) | 193.35 (39.75) | 86.22 (177.72) | 169.61 (258.94) | 27.29 (6.54) |
| 0.8 | 1.0 | 1 | 0 | 0 | 180.22 (39.41) | 76.71 (30.79) | 0.498 (0.050) | 192.15 (41.71) | 116.20 (217.00) | 165.13 (250.72) | 29.82 (7.26) |
| 1.0 | 1.0 | 3 | 0 | 0 | 184.40 (40.95) | 80.37 (32.84) | 0.532 (0.046) | 205.88 (43.85) | 320.65 (391.41) | 175.88 (266.54) | 32.77 (8.43) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2$($\hat{y}^*_{\text{i test}}$) and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.



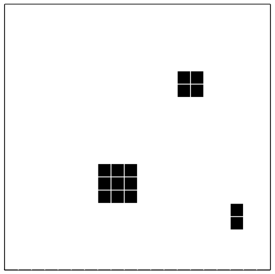
Web Table C10: Simulation results scenario 9B – Misspecified sparse partially distributed

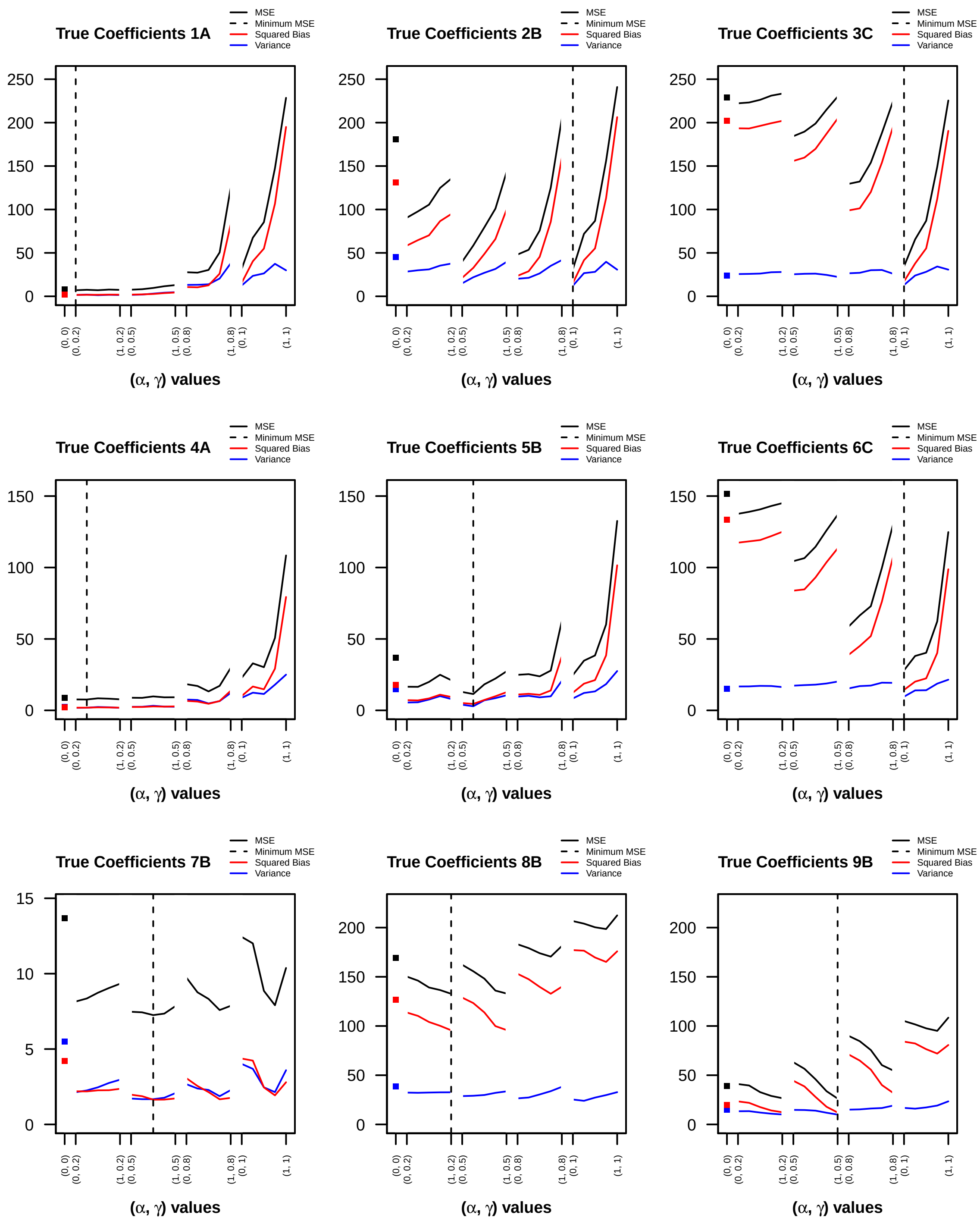
| | | Frequency of minimum | | | Mean (SD) over 100 simulation iterations | | | | | Mean (SD) over 100 test obs. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\gamma$ | Mean CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Mean CVE | SD CVE | MSE($\hat{\boldsymbol{\beta}}$) | MSE($\hat{\mathbf{y}}_{\text{test}}$) | Optimal $\lambda$ | Bias$^2$($\hat{y}^*_{\text{i test}}$) | Var($\hat{y}^*_{\text{i test}}$) |
| 0.0 | 0.0 | 4 | 5 | 5 | 57.63 (18.10) | 38.07 (18.88) | 0.094 (0.079) | 40.23 (32.95) | 17.78 (100.05) | 19.86 (37.68) | 15.08 (11.39) |
| 0.0 | 0.2 | 0 | 0 | 0 | 54.48 (14.43) | 34.60 (15.68) | 0.098 (0.068) | 41.60 (26.84) | 5.18 (11.44) | 23.39 (45.40) | 13.39 (9.13) |
| 0.2 | 0.2 | 0 | 0 | 0 | 50.72 (14.36) | 33.37 (15.50) | 0.094 (0.072) | 39.75 (28.07) | 4.99 (9.34) | 21.96 (42.93) | 13.49 (10.12) |
| 0.5 | 0.2 | 1 | 1 | 1 | 45.31 (14.56) | 31.60 (15.50) | 0.078 (0.069) | 33.59 (27.01) | 5.59 (9.42) | 17.59 (34.53) | 12.07 (10.10) |
| 0.8 | 0.2 | 1 | 5 | 2 | 40.78 (15.06) | 29.43 (15.19) | 0.066 (0.064) | 29.58 (26.39) | 6.11 (11.53) | 14.19 (28.09) | 10.92 (9.65) |
| 1.0 | 0.2 | 59 | 43 | 42 | 38.11 (15.14) | 28.23 (15.38) | 0.059 (0.062) | 26.82 (25.76) | 7.38 (18.76) | 12.39 (24.60) | 10.11 (9.29) |
| 0.0 | 0.5 | 0 | 0 | 0 | 73.91 (15.43) | 38.16 (17.32) | 0.162 (0.056) | 67.10 (23.91) | 12.96 (99.87) | 44.36 (86.98) | 14.78 (5.67) |
| 0.2 | 0.5 | 0 | 0 | 0 | 68.12 (14.82) | 36.94 (16.79) | 0.145 (0.063) | 59.69 (24.95) | 3.85 (9.75) | 38.63 (76.24) | 14.62 (6.70) |
| 0.5 | 0.5 | 1 | 0 | 0 | 58.08 (14.67) | 35.49 (15.60) | 0.114 (0.073) | 48.08 (30.29) | 4.32 (10.00) | 27.95 (54.86) | 14.01 (9.34) |
| 0.8 | 0.5 | 2 | 2 | 2 | 46.48 (15.79) | 32.89 (15.22) | 0.080 (0.070) | 35.77 (29.84) | 6.58 (14.87) | 17.91 (35.22) | 11.87 (9.44) |
| 1.0 | 0.5 | 31 | 44 | 47 | 40.06 (16.45) | 29.80 (14.84) | 0.058 (0.064) | 26.31 (25.13) | 7.95 (12.39) | 12.09 (23.73) | 10.01 (9.11) |
| 0.0 | 0.8 | 0 | 0 | 0 | 93.14 (19.77) | 42.68 (18.60) | 0.240 (0.036) | 96.69 (23.70) | 21.56 (103.07) | 71.04 (136.17) | 15.02 (3.10) |
| 0.2 | 0.8 | 0 | 0 | 0 | 89.82 (18.83) | 42.20 (18.52) | 0.223 (0.041) | 90.85 (23.12) | 11.24 (27.28) | 64.88 (125.21) | 15.28 (3.70) |
| 0.5 | 0.8 | 0 | 0 | 0 | 83.54 (17.86) | 40.94 (18.32) | 0.199 (0.052) | 81.61 (24.11) | 12.38 (32.14) | 55.69 (107.58) | 16.16 (5.00) |
| 0.8 | 0.8 | 1 | 0 | 0 | 76.45 (18.25) | 39.84 (18.52) | 0.154 (0.072) | 63.92 (31.37) | 15.04 (30.60) | 39.99 (78.20) | 16.66 (8.22) |
| 1.0 | 0.8 | 0 | 0 | 1 | 76.64 (20.42) | 42.13 (20.47) | 0.138 (0.098) | 58.10 (41.60) | 41.36 (115.35) | 32.13 (62.27) | 19.10 (14.95) |
| 0.0 | 1.0 | 0 | 0 | 0 | 104.42 (22.92) | 46.08 (19.97) | 0.281 (0.034) | 111.59 (25.27) | 108.80 (282.70) | 84.11 (158.93) | 16.77 (3.90) |
| 0.2 | 1.0 | 0 | 0 | 0 | 102.75 (22.40) | 46.01 (19.58) | 0.274 (0.035) | 109.20 (25.26) | 37.29 (71.43) | 82.28 (155.68) | 16.01 (3.63) |
| 0.5 | 1.0 | 0 | 0 | 0 | 99.39 (21.04) | 45.59 (19.68) | 0.260 (0.037) | 104.89 (25.43) | 26.53 (34.41) | 76.44 (144.45) | 17.27 (3.90) |
| 0.8 | 1.0 | 0 | 0 | 0 | 97.82 (21.17) | 44.51 (18.96) | 0.253 (0.047) | 102.92 (27.24) | 42.83 (51.51) | 71.98 (135.72) | 19.16 (5.69) |
| 1.0 | 1.0 | 0 | 0 | 0 | 103.87 (23.26) | 47.21 (18.98) | 0.286 (0.050) | 116.19 (30.34) | 183.24 (302.75) | 80.68 (152.09) | 23.43 (6.85) |

Notes:

- $\alpha$ controls balance between group lasso ($\alpha = 0$) and $\ell_1$ lasso ($\alpha = 1$) penalty terms.
  $\gamma$ controls balance between fusion ($\gamma = 0$) and sparsity ($\gamma = 1$) penalty terms.
- 'Frequency of minimum' columns give frequency of instances out of 100 of simulations yielding the minimum indicated error measure for each $(\alpha, \gamma)$ pair.
- Mean CVE (cross-validation error) is the minimum mean CVE over 5 cross-validation folds, which was used to select the optimal $\lambda$.
- MSE($\hat{\boldsymbol{\beta}}$) is the mean squared error of the estimated coefficients for the model fitted to the entire training sample of $n = 50$.
- MSE($\hat{\mathbf{y}}_{\text{test}}$) is the mean squared error of the estimated response for a test sample of $n = 50$.
- Bias$^2$($\hat{y}^*_{\text{i test}}$) and Var($\hat{y}^*_{\text{i test}}$) were calculated as follows: For each $(\alpha, \gamma)$ pair, the response for each observation in a test sample of $n = 100$ was calculated using each set of estimated coefficients from the 100 simulation iterations. The mean squared bias and variance of the predicted responses were calculated for each observation individually, and then the mean and standard deviations of these were calculated over the 100 observations.

**True Coefficients 9B:**
**Misspecified Sparse Group**






Web Figure C4: Mean over $n = 100$ test observations of the squared bias, variance, and mean squared error of predicted responses from models estimated in 100 simulation iterations. Values of $\gamma$ increase from left to right on the $x$-axis, corresponding to complete fusion penalty on the left ($\gamma = 0$) and complete sparsity penalties on the right ($\gamma = 1$). Intervals of increasing $\alpha$ values correspond to complete group penalty on the left ($\alpha = 0$) and complete $\ell_1$ lasso penalty on the right ($\alpha = 1$).

# 4 Web Appendix D: ABIDE Application Results

Web Table D1: ABIDE dataset descriptive summary

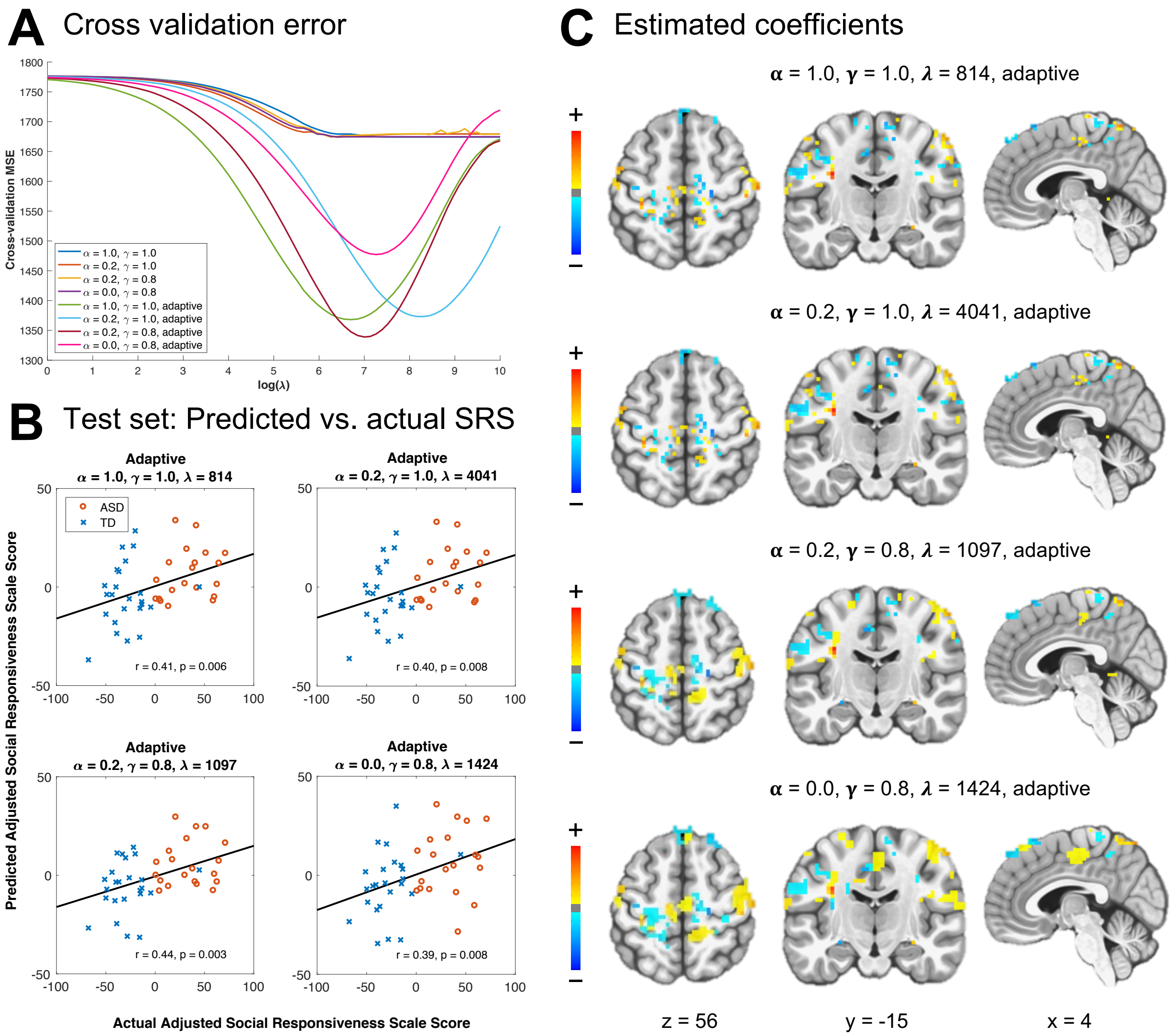
| | Overall ($n = 219$) | Training Set ($n = 175$) | Test Set ($n = 44$) | |
|---|---|---|---|---|
| | Mean (SD) | Mean (SD) | Mean (SD) | $p$ * |
| Age at scan | 17.4 (7.5) | 17.4 (7.3) | 17.4 (8.4) | 0.97 |
| Full-scale IQ | 108.6 (14.3) | 108.2 (13.8) | 110 (16.5) | 0.50 |
| Mean framewise displacement | 0.09 (0.06) | 0.09 (0.07) | 0.08 (0.05) | 0.71 |
| Social Responsiveness Scale score | 56.3 (42.7) | 56.6 (43.1) | 55.1 (41.8) | 0.84 |
| | $n$ (%) | $n$ (%) | $n$ (%) | $p$ * |
| Diagnosis group | | | | 0.50 |
| Autism spectrum disorder | 111 (50.7) | 91 (52.0) | 20 (45.5) | |
| Typically developing | 108 (49.3) | 84 (48.0) | 24 (54.5) | |
| Site of acquisition | | | | 0.95 |
| Leuven_1 | 22 (10.0) | 17 (9.7) | 5 (11.4) | |
| NYU | 101 (46.1) | 82 (46.9) | 19 (43.2) | |
| USM | 72 (32.9) | 57 (32.6) | 15 (34.1) | |
| Yale | 24 (11.0) | 19 (10.9) | 5 (11.4) | |
| Eye status at scan | | | | 0.09 |
| Open | 204 (93.2) | 166 (94.9) | 38 (86.4) | |
| Closed | 15 (6.8) | 9 (5.1) | 6 (13.6) | |

ABIDE: Autism Brain Imaging Data Exchange
* Welch's t-test for continuous variables, Fisher's exact test for categorical variables
Note: One training set participant was missing mean framewise displacement data.

Web Table D2: Linear regression modeling Social Responsiveness Scale score as outcome, fit to training set ($n = 175$)

| | Estimate | Std. Error | $t$ | $p$ |
|---|---|---|---|---|
| Intercept | 41.63 | 32.50 | 1.28 | 0.20 |
| Age at scan | 1.26 | 0.56 | 2.25 | 0.03 |
| Full-scale IQ | -0.18 | 0.24 | -0.77 | 0.45 |
| Mean framewise displacement | 130.09 | 50.93 | 2.55 | 0.01 |
| Site of acquisition | | | | |
| Leuven_1 | reference | | | |
| NYU | 9.74 | 12.42 | 0.78 | 0.43 |
| USM | -5.88 | 11.78 | -0.50 | 0.62 |
| Yale | -12.09 | 15.15 | -0.80 | 0.43 |
| Eye status at scan | | | | |
| Open | reference | | | |
| Closed | -1.65 | 15.04 | -0.11 | 0.91 |

Residual standard error: 42.18 on 167 degrees of freedom
Multiple R-squared: 0.08095
Adjusted R-squared: 0.04243
F-statistic: 2.101 on 7 and 167 degrees of freedom, $p = 0.04599$
Note: One training set participant was missing mean framewise displacement data. In order to fit the linear regression model to compute adjusted SRS scores, this missing value was imputed by a linear regression fit to the remaining $n = 174$ training participants, modeling mean framewise displacement as outcome with predictors age, IQ, social responsiveness scale score, site of acquisition, and eye status at scan.




Web Figure D1: (A) Five-fold cross-validation was carried out over a range of $\lambda$ values for several sets of $\alpha$ and $\gamma$ values. (B) Correlation of predicted and actual adjusted SRS scores for selected $\alpha$ and $\gamma$ values. Points are distinguished by autism spectrum disorder (ASD, red circle) and typically developing (TD, blue cross) diagnosis groups. (C) Estimated coefficients at the optimal $\lambda$ for selected $\alpha$ and $\gamma$ values. Higher coefficient values contribute to higher predicted Social Responsiveness Scale (SRS) scores, which indicate greater autistic social impairment.

# References

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3,** 1–122.

He, B., Yang, H., and Wang, S. (2000). Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications* **106,** 337–356.

Huo, Z. and Tseng, G. (2017). Integrative sparse k-means with overlapping group lasso in genomic applications for disease subtype discovery. *The Annals of Applied Statistics* **11,** 1011.