Supplemental material

One-step estimator paths for concave regularization

7 Implementation via coordinate descent

We use Coordinate descent (CD; e.g., Luenberger and Ye, 2008) to minimize (3) at each step along the path. CD is a local optimization algorithm that cycles through minimization of the conditional objective for individual parameters when the remaining parameters are fixed. Algorithms of this type have have become popular in L_1 penalized estimation since the work by Friedman et al. (2007) and Wu and Lange (2008).

Our CD routine, outlined in Algorithm 1, is a solver for penalized weighted-least squares problems as defined in equation (21) below. This applies directly in Gaussian regression, and for non-Gaussian models we follow Friedman et al. (2010) and apply CD inside an outer loop of iteratively re-weighted-least-squares (IRLS; e.g., Green, 1984). Given current parameter values $\hat{\boldsymbol{\beta}}$, the Newton-Raphson update for maximum likelihood estimation is $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} - \mathbf{H}^{-1}\mathrm{g}$, where $\mathbf{H}$ is the information matrix with elements $h_{jk} = \partial^2 l/\partial\beta_j\partial\beta_k|_{\hat{\boldsymbol{\beta}}}$ and $\mathbf{g}$ is coefficient gradient (see Appendix 8). For exponential family linear models we can write $\mathbf{H} = \mathbf{X}'\mathbf{V}\mathbf{X}$ and $\mathbf{g} = \mathbf{X}'\mathbf{V}(\mathbf{z} - \hat{\boldsymbol{\eta}})$, where $\mathbf{V} = \mathrm{diag}(\mathbf{v})$, $\mathbf{v} = [v_1 \ldots v_n]$ are 'weights', $\mathbf{z} = [z_1 \ldots z_n]$ are transformed 'response', and $\hat{\eta}_i = \hat{\alpha} + \mathbf{x}_i\hat{\boldsymbol{\beta}}$. In Gaussian regression, $v_i = 1$, $z_i = \hat{\eta}_i - y_i$, and the update is an exact solution. For binomial regression, $v_i = q_i(1 - q_i)$ and $z_i = \hat{\eta}_i - (y_i - q_i)/v_i$, where $q_i = (1 + \exp[-\hat{\eta}_i])^{-1}$ is the estimated probability of success.

This yields $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{z}$, such that the Newton update solves a weighted-least-squares problem. Adding L_1 costs, the minimization objective from (3) becomes

$$\underset{\alpha,\beta_1\ldots\beta_p\in\mathbb{R}}{\mathrm{argmin}} \sum_i \frac{v_i}{2}(\alpha + \mathbf{x}_i'\boldsymbol{\beta} - z_i)^2 + n\sum_j \omega_j\lambda|\beta_j|. \tag{21}$$

Our solver iterates between CD on (21) and, for non-Gaussian models, updates to $\mathbf{v}$ and $\mathbf{z}$. Each t^{th} segment IRLS routine initializes $[\hat{\alpha}, \hat{\boldsymbol{\beta}}]$ at solutions for λ^{t-1}, or at $[\hat{\alpha}, \mathbf{0}]$ for $t = 1$. In the `gamlr` implementation, a full pass update of all parameters is done only at the first CD iteration; otherwise coordinates with currently inactive (zero) $\hat{\beta}_j$ are not updated. Once the descent converges for this *active set*, IRLS $\mathbf{v}$ and $\mathbf{z}$ are updated and we begin a new CD loop with a full pass update. The routine stops when maximum squared change in β_j scaled by its information over one of these full pass updates is less than some tolerance threshold, `thresh`. The default in `gamlr` uses a relative tolerance of 10^{-7} times null model deviance.

Algorithm 1 Coordinate descent

Set $\mathtt{vh_j} = \sum_i v_i(x_{ij} - \bar{x}_j)^2$ and $\mathtt{vx_j} = \sum_i v_i x_{ij}$ for $j = 1 \dots p$.

while $\max\limits_{j=1\dots p} \mathtt{vh_j}\Delta_j^2 > \mathtt{thresh}$:

for j=1...p:

set $\mathtt{vg_j} = -\sum_i x_{ij}v_i(z_i - \hat{\eta}_i)$ and $\mathtt{ghb} = \mathtt{vg_j} - \mathtt{vh_j}\hat{\beta}_j$

if $|\mathtt{ghb}| < n\lambda^t\omega_j^t$: $\Delta_j = -\hat{\beta}_j$

else: $\Delta_j = -(\mathtt{vg_j} - \mathrm{sign}(\mathtt{ghb})n\lambda^t\omega_j^t)/\mathtt{vh_j}$.

update $\hat{\beta}_j \stackrel{+}{=} \Delta_j$, $\hat{\alpha} \stackrel{+}{=} -\mathtt{vx_j}\Delta_j$, and $\hat{\boldsymbol{\eta}} = \hat{\alpha} + \mathbf{X}'\hat{\boldsymbol{\beta}}$.

7.1 Descent convergence

Despite the non-differentiability of $|\beta_j|$ at zero, Tseng (2001) establishes local convergence for CD on (21) as a consequence of penalty separability: the non-differentiable part of our objective is a sum of functions on only a single coordinate. Thus CD solves each weighted-least squares problem, and the full algorithm converges if IRLS does. For non-Gaussian models, convergence of such nested L_1-penalized IRLS algorithms is shown in Lee et al. (2014).

7.2 Quasi-Newton acceleration

Under high collinearity and large γ, one may wish to accelerate convergence via a quasi-Newton step (e.g., Lange, 2010). Acceleration is applied to $\boldsymbol{\theta} = [\alpha, \boldsymbol{\beta}]$, and a move is accepted only if it leads to a decrease in the objective. Suppose that $\hat{\boldsymbol{\theta}}^{(0)}$, $\hat{\boldsymbol{\theta}}^{(-1)}$, and $\hat{\boldsymbol{\theta}}^{(-2)}$ are the current, previous, and previous-to-previous parameter estimates. Write $M(\hat{\boldsymbol{\theta}}^{(t)})$ as the implied CD update map $\hat{\boldsymbol{\theta}}^{(t)} \rightarrow \hat{\boldsymbol{\theta}}^{(t+1)}$, such that the algorithm converges at $\hat{\boldsymbol{\theta}} - M(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. With $\mathbf{u} = \hat{\boldsymbol{\theta}}^{(-1)} - \hat{\boldsymbol{\theta}}^{(-2)}$ and $\mathbf{v} = \hat{\boldsymbol{\theta}}^{(0)} - \hat{\boldsymbol{\theta}}^{(-1)}$, a secant approximation to the gradient of M is $\partial M/\partial\hat{\theta}_l \approx \mathrm{v}_l/\mathrm{u}_l$. An approximate Newton-Raphson step to solve for the root of $\hat{\boldsymbol{\theta}} - M(\hat{\boldsymbol{\theta}})$ updates each coordinate $\hat{\theta}_l \leftarrow \hat{\theta}_l^{(-1)} - (\hat{\theta}_l^{(-1)} - \hat{\theta}_l^{(0)})/(1 - \mathrm{v}_l/\mathrm{u}_l)$ which can be re-written as $\hat{\theta}_l = (1 - \mathrm{w}_l)\hat{\theta}_l^{(-1)} + \mathrm{w}_l\hat{\theta}_l^{(0)}$ where $\mathrm{w}_l = \mathrm{u}_l/(\mathrm{u}_l - \mathrm{v}_l)$.

8 Gradient, curvature, and path starts

The negative log likelihood objective in Gaussian regression is $l(\alpha, \boldsymbol{\beta}) = 0.5 \sum_i (y_i - \eta_i)^2$ with gradient $g_j(\boldsymbol{\beta}) = \partial l / \partial \beta_j = -\sum_i x_{ij}(y_i - \eta_i)$, and coordinate curvature $h_j(\boldsymbol{\beta}) = \partial^2 l / \partial \beta_j^2 = \sum_i x_{ij}^2$. In logistic regression, set $y_i = 1$ for 'success' and $y_i = 0$ for 'failure' and write $q_i = (1+\exp[-\eta_i])^{-1}$ as the probability of success. Then $l(\alpha, \boldsymbol{\beta}) = \sum_i -y_i\eta_i + \log(1+\exp[\eta_i])$, $g_j(\boldsymbol{\beta}) = \partial l / \partial \beta_j = -\sum_i x_{ij}(y_i - q_i)$, and $h_j(\boldsymbol{\beta}) = \partial^2 l / \partial \beta_j^2 = \sum_i x_{ij}^2 q_i(1 - q_i)$. In each case, it is implied that $\hat{\alpha}$ has been set to minimize $l(\alpha, \hat{\boldsymbol{\beta}})$.

For L_1 costs $c_j(|\beta_j|) = |\beta_j|$, the infimum λ such that $\hat{\boldsymbol{\beta}} = \mathbf{0}$ is available analytically as $\lambda^1 = n^{-1} \max\{|g_j(\mathbf{0})|,\ j = 1 \ldots p\}$, the maximum mean absolute gradient for the null model with $\boldsymbol{\beta} = \mathbf{0}$. This formula is used to obtain our starting values for the path algorithms.

9 False Discovery Control

A common goal in high-dimensional estimation is support recovery – having the set $\{j : \hat{\beta}_j \neq 0\} = \{j : \beta_j \neq 0\}$ for some 'true' $\boldsymbol{\beta}$. For standard lasso estimated $\hat{\boldsymbol{\beta}}$, many authors have shown (e.g., Buhlmann and van de Geer, 2011; Zou, 2006) that to get exact support recovery asymptotically or with high probability requires an *irrepresentability condition* which limits the size of least-squares projections from 'true support' onto spurious covariates.

DEFINITION 9.1. *The $(\theta, S, \mathbf{v})$-irrepresentable condition for $\theta \in [0, 1]$ and $\mathbf{v} \in \mathbb{R}^s$ holds that,*

$$|\boldsymbol{x}_j' \mathbf{X}_S (\mathbf{X}_S' \mathbf{X}_S)^{-1} \mathbf{v}| \leq \theta \ \ \forall\, j \notin S \tag{22}$$

This is often presented with $\mathbf{v} = \mathbf{1}$.[1] It can be a strict design restriction; for example, Buhlmann and van de Geer (2011) show a single variable that is highly correlated with many columns of $\mathbf{X}_S$ leading to failure. Much of the literature on concave penalization has focused on achieving support recovery *without* such conditions; see, e.g., Fan et al. (2014) for a recent overview. Our results will require irrepresentable conditions with $\mathbf{v} = \boldsymbol{\omega}_S$, which becomes less restrictive as one is able to shrink weights ω_j for $j \in S$. See the remarks for more discussion.

Our comparison of interest is between $\hat{S} = \{j : \hat{\beta}_j \neq 0\}$, for $\hat{\boldsymbol{\beta}}$ from weighted-L_1 penalized estimation, and $S = \{j : \beta_j^\nu \neq 0\}$ for $\boldsymbol{\beta}^\nu$ the L_0 penalized estimator from Theorem 3.1. Whether looking to an L_0 oracle or a sparse truth, our experience is that exact support recovery does not occur in practice (e.g., see the simulation in Section 5). Thus, we instead focus on ability of the weighted-lasso to minimize *false discoveries*: $\hat{\beta}_j \neq 0$ when $\beta_j^\nu = 0$.

[1] Wainwright (2009) shows that (22) with $\theta = 1$, $\mathbf{v} = \mathbf{1}$ is necessary for lasso sign recovery in the *noiseless* setting.

THEOREM 9.1. *Consider the setting of Theorem 3.1. If* $\omega_{S^c}^{\min} = 1$ *and* $\lambda > \sqrt{2\nu}$ *then*

$$\|\mathbf{X}'_{S^c}\mathbf{X}_S(\mathbf{X}'_S\mathbf{X}_S)^{-1}\boldsymbol{\omega}_S\|_\infty \leq 1 - \frac{\sqrt{2\nu}}{\lambda_t} \;\;\Rightarrow\;\; \hat{S} \cap S^c = \varnothing. \tag{23}$$

The result follows directly from the sign recovery lemma 9.1.

Remarks

- From Theorem 7.4 in Buhlmann and van de Geer (2011), the irrepresentability condition holds with $|\boldsymbol{x}'_j\mathbf{X}_S(\mathbf{X}'_S\mathbf{X}_S)^{-1}\boldsymbol{\omega}_S| \leq \frac{\|\boldsymbol{\omega}_S\|}{\sqrt{s}}\theta_{\text{adap}}(S)$ where $\theta_{\text{adap}}(S)$ is their 'adaptive restricted regression' coefficient. Of interest here, they show that $\theta_{\text{adap}}(S) \leq \sqrt{s}/\Lambda_{\min}(S)$ where $\Lambda_{\min}(S)$ is the minimum eigenvalue of $\mathbf{X}'_S\mathbf{X}_S/n$. Thus, (i) can be replaced by the restriction $\Lambda_{\min}(S) \geq \|\boldsymbol{\omega}_S\|(1 - \sqrt{2\nu}/(\omega_{S^c}^{\min}\lambda))^{-1} = \sqrt{s}L$, with L from Theorem 3.1, and small values for L appear key in both predictive performance and support recovery.

- Without irrepresentability, limits on false discovery are more pessimistic. Convergence conditions imply that for $j \in S^c \cap \hat{S}$ we have $n\lambda\omega_j = |\boldsymbol{x}'_j(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})| \leq |\boldsymbol{x}'_j\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\nu)| + |\boldsymbol{x}'_j\mathbf{e}^S| \leq n\left(2\|\boldsymbol{\omega}_S\|/\phi(L,S) + \sqrt{2\nu}/\lambda\right)\ \forall\, j$. Dividing by $n\lambda\omega_j$ and counting yields

$$|S^c \cap \hat{S}| \leq \left|\frac{1}{\boldsymbol{\omega}_{S^c\cap\hat{S}}}\right| \left(\frac{2\|\omega_S\|}{\phi(L,S)} + \frac{\sqrt{2\nu}}{\lambda}\right) \tag{24}$$

Without the ability to make ω_j very big for $j \in S^c$ (e.g., as in a thresholding procedure like that of Zhou 2009), the result in (24) has little to say about false discovery control.

9.1 Sign Recovery

LEMMA 9.1. *Under the setting of Theorem 3.1, with* $\hat{S} = \{j : \hat{\beta}_j \neq 0\}$*, if* $\omega_{S^c}^{\min}\lambda > \sqrt{2\nu}$ *then*

$$|\boldsymbol{x}'_j\mathbf{X}_S(\mathbf{X}'_S\mathbf{X}_S)^{-1}\boldsymbol{\omega}_S| \leq 1 - \frac{\sqrt{2\nu}}{\lambda\omega_j}\ \ \forall\, j \in S^c \Rightarrow \hat{S} \cap S^c = \varnothing. \tag{25}$$

If in addition $|(\mathbf{X}'_S\mathbf{X}_S)^{-1}\mathbf{X}'_S\mathbf{y}|_\infty > n\lambda\,|(\mathbf{X}'_S\mathbf{X}_S)^{-1}\boldsymbol{\omega}_S|_\infty$*, then* $\text{sgn}(\hat{\boldsymbol{\beta}}) = \text{sgn}(\boldsymbol{\beta}^\nu)$.

Proof. The Karush-Kuhn-Tucker (KKT) conditions at weighted-L_1 minimization convergence imply that

$$\mathbf{x}'_j\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\nu) + \mathbf{x}'_j\mathbf{e}^S = -n\lambda\zeta_j \;\text{ for }\; j = 1\ldots p \tag{26}$$

where $|\zeta_j| = \omega_j$ for $j \in \hat{S}$ and $|\zeta_j| \leq \omega_j$ for $j \in \hat{S}^c$. Following closely related proofs in Wainwright (2006, 2009); Zhou et al. (2009), $\hat{S} \cap S^c = \varnothing$ occurs if and only if these KKT conditions hold for projections restricted to S,

$$\mathbf{X}_S'\mathbf{X}_S(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^\nu) + \mathbf{X}_S'\mathbf{e}^S = -n\lambda\boldsymbol{\zeta}_S \;\;\Rightarrow\;\; \hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^\nu = -n\lambda(\mathbf{X}_S'\mathbf{X}_S)^{-1}\boldsymbol{\zeta}_S. \tag{27}$$

Thus all of the spurious regressors in S^c will have $\hat{\beta}_j = 0$ if and only if

$$\boldsymbol{x}_j'\mathbf{X}_S(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_S^\nu) - \boldsymbol{x}_j'\mathbf{e}^S \leq n\lambda\zeta_j \;\;\Leftarrow\;\; 1 - \frac{|x_j'\mathbf{e}^S|}{n} \geq 1 - \frac{\sqrt{2\nu}}{\lambda\omega_j} \geq |\boldsymbol{x}_j'\mathbf{X}_S(\mathbf{X}_S'\mathbf{X}_S)^{-1}\boldsymbol{\omega}_S|. \tag{28}$$

Finally, for sign recovery on $j \in S$ we need $|\beta_j^\nu| - |\beta_j^\nu - \hat{\beta}_j| > 0 \;\; \forall\, j \;\in\; S$, and our stated condition follows from $\boldsymbol{\beta^\nu}_S = (\mathbf{X}_S'\mathbf{X}_S)^{-1}\mathbf{X}_S'\mathbf{y}$ and $\boldsymbol{\beta^\nu}_S - \hat{\boldsymbol{\beta}}_S = n\lambda(\mathbf{X}_S'\mathbf{X}_S)^{-1}\boldsymbol{\zeta}_S$. □

10 Extra proofs

10.1 Stagewise Regression

Theorem 3.1 uses the following simple result for stagewise regression – iterative fitting of new covariates to the residuals of an existing linear model (as in, e.g., Goldberger 1961).

LEMMA 10.1. *Say* $\mathrm{MSE}_S = \|\mathbf{X}\boldsymbol{\beta}^S - \mathbf{y}\|^2/n$ *and* $\mathrm{cov}(\boldsymbol{x}_j, \mathbf{e}^S) = \boldsymbol{x}_j'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^S)/n$ *are sample variance and covariances. Then for any* $j \in 1 \ldots p$*,*

$$\mathrm{cov}^2(\boldsymbol{x}_j, \mathbf{e}^S) \leq \mathrm{MSE}_S - \mathrm{MSE}_{S\cup j}$$

Proof. From the well-known property on the correlation coefficient (R^2) for linear models, in-sample correlation and variances are such that

$$\frac{\mathrm{cov}^2(\boldsymbol{x}_j, \mathbf{e}^S)}{\mathrm{var}(\boldsymbol{x}_j)\mathrm{var}(\mathbf{e}^S)} = 1 - \frac{\mathrm{var}(\mathbf{e}^S - \tilde{\beta}_j\boldsymbol{x}_j)}{\mathrm{var}(\mathbf{e}^S)}$$

where $\tilde{\beta}_j = \boldsymbol{x}_j'\mathbf{e}^S/(\boldsymbol{x}_j'\boldsymbol{x}_j)$ is the stagewise coefficient estimate. Since $\mathrm{var}(\boldsymbol{x}_j) = 1$, multiplying everything by $\mathrm{var}(\mathbf{e}^S)$ yields $\mathrm{cov}^2(\boldsymbol{x}_j, \mathbf{e}^S) = \mathrm{var}(\mathbf{e}^S) - \mathrm{var}(\mathbf{e}^S - \tilde{\beta}_j\boldsymbol{x}_j) \leq \mathrm{var}(\mathbf{e}^S) - \mathrm{var}(\mathbf{e}^{S\cup j})$. The last inequality holds because $\mathbf{e}^{S\cup j}$, residuals from OLS on $\mathbf{X}_{S\cup j}$, have the smallest-possible sum of squares for that set of covariates. With $\mathrm{var}(\mathbf{e}^S) = \mathrm{MSE}_S$, etc, we are done. □

10.2 Bayesian MAP

PROPOSITION 10.1. $\hat{\boldsymbol{\beta}}$ *solves (14) if and only if it is also in the solution to (13).*

Proof. The conditional posterior mode for each τ_j given β_j is $\tau(\beta_j) = \gamma s/(1 + \gamma|\beta_j|)$. Any joint solution $[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}]$ for (13) thus consists of $\hat{\tau}_j = \tau(\hat{\beta}_j)$; otherwise, it is always possible to decrease the objective by replacing $\hat{\tau}_j$. Setting each $\tau_j = \tau(\beta_j)$ in (13) and removing constant terms yields (14). Moreover, the solution to (13) solves (14): otherwise, there would need to be a point on the profile slice of (13) defined by $\tau_j = \tau(\hat{\beta}_j)$ that is lower than its minimum. $\square$

For a Bayesian it is odd to be solving for $\boldsymbol{\tau}$ rather than marginalizing over its uncertainty. However, recognizing the form of a gamma density in (12), $\pi(\beta_j, \tau_j)$ integrates over τ_j to yield the marginal prior $\pi(\beta_j) = 0.5s\,(1 + \gamma|\beta_j|)^{-(s+1)}$. This is the generalized double Pareto density, as in Armagan et al. (2013). Since $-\log \pi(\beta_j) \propto (s + 1)\log(1 + \gamma|\beta_j|)$, the *profile* MAP solution to (13) is also the *marginal* MAP for $\boldsymbol{\beta}$ under $\mathrm{Ga}(s - 1, 1/\gamma)$ priors on each τ_j.

11 Stability

A strong form of stability comes from convexity of the penalized objective in (1). This requires that the minimum eigenvalue of $\mathbf{H}(\boldsymbol{\beta})$, the Hessian matrix of second derivatives of $l(\boldsymbol{\beta})$, is greater than $|c''(\beta_j)|\ \forall j$. For penalized least-squares under log costs, this amounts to requiring that the minimum eigenvalue of $\mathbf{H} = \mathbf{X}'\mathbf{X}$ is greater than $\lambda\gamma^2$.[2] In the simple *standardized orthogonal covariate* case, this has an easy interpretation in the context of our Bayesian model from Section 4.1: for Gaussian regression, $h_j = \sum_i x_{ij}^2 = n$ and the objective is convex if prior variance on each τ_j is less than the number of observations. For logistic regression you need $\mathrm{var}(\tau_j) < n/4$, since $\mathbf{H}$ now depends upon the coefficient values.

In real examples, however, we cannot rely upon objective convexity. A more useful definition of stability requires continuity of the implied coefficient function, $\hat{\beta}(\mathbf{y})$, in an imagined univariate regression problem (or for orthogonal covariates). This is one of the key requirements of concave penalties listed by Fan and Li (2001). Many popular concave cost functions, such as the SCAD and MCP, have been engineered to have this continuity property. Conveniently, Zou and Li (2008) show that OSE LLA solutions have this property even if the target objective does not. For example, even though the log penalty *does not* generally lead to continuous thresholding, their result implies that the GL solutions are continuous for $\gamma < \infty$.

A theoretically richer form of stability is Lipschitz continuity of the implied prediction function, $\hat{\boldsymbol{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}(\mathbf{y})$, which requires that $\|\hat{\mathbf{y}}(\mathbf{y}_1) - \hat{\mathbf{y}}(\mathbf{y}_2)| \leq L\|\mathbf{y}_1 - \mathbf{y}_2\|$ for some finite constant L on all possible $\mathbf{y}_1, \mathbf{y}_2$. Zou et al. (2007) establish Lipschitz continuity for L_1 estimated

[2] If ν is an eigenvalue of $\mathbf{H}$, then $(\mathbf{H} - \nu\mathbf{I})\mathbf{v} = 0$ for some nonzero $\mathbf{v}$; the negative log posterior Hessian at zero is $\mathbf{H} - \lambda\gamma^2\mathbf{I}$ and $(\mathbf{H} - \lambda\gamma^2\mathbf{I} + s\gamma^2\mathbf{I} - \nu\mathbf{I})\mathbf{v} = 0$ so that $\nu - s\gamma^2$ is an eigenvalue of the minimization objective.

predictors as part of their derivation of a degrees-of-freedom estimator. Thus, conditional upon values for the coefficient-specific weights, POSE and GL are trivially Lipschitz continuous. Unconditionally, we do not believe that the paths have this guarantee. However, we'll see in the next section that a heuristic degrees-of-freedom estimator that takes such stability for granted performs well as the basis for model selection.

Finally, the basic and most important type of stability is practical path continuity: by this, we mean that solutions change slowly enough along the path so that computational costs are kept within budget. A regularization path can be built from a continuous thresholding function, or perhaps even be Lipschitz stable, but none of that matters if it takes too long to fit. For example, Figure 4 shows timings growing rapidly with large γ for the hockey data of Section 6, even though all of these specifications are theoretically stable by some criteria.

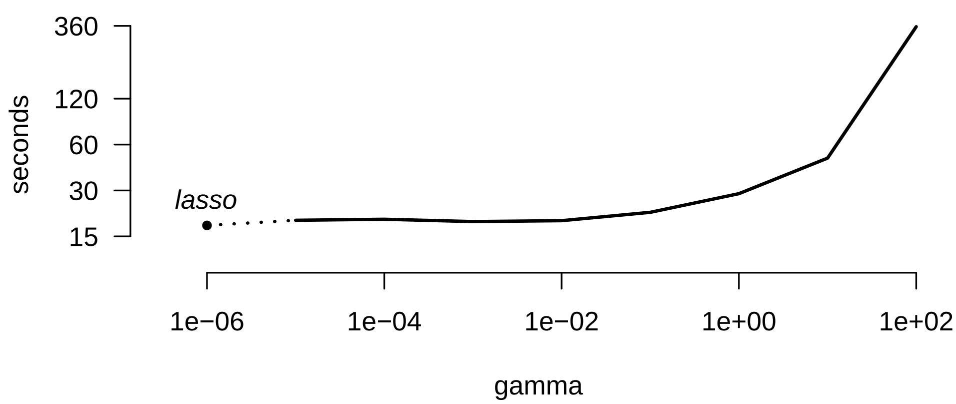

Figure 4: Timings for the hockey data path fits of Section 6 on a length-100 grid with $\lambda^{100} = 0.01\lambda^1$.

12 Information Criteria

We would like to choose a model that performs well in predicting new data. 'Good prediction' can be measured in a variety of ways. A common and coherent framework is to consider minimizing Kullback-Leibler (KL) divergence. Say $g(\mathbf{y})$ is the true data generating process, and $f(\mathbf{y}; \boldsymbol{\eta}, \phi)$ is the parametric density under study, which we suppose here is a natural exponential family with $\mathbb{E}[\mathbf{y}] = \boldsymbol{\eta}$ and dispersion ϕ. Then we wish to minimize

$$\mathrm{KL}(\boldsymbol{\eta}, \phi) = \mathbb{E}_g \log g(\mathbf{y}) - \mathbb{E}_g \log f(\mathbf{y}; \boldsymbol{\eta}, \phi), \tag{29}$$

the expected difference between log true density and our parametric approximation. Since $\mathbb{E}_g \log g(\mathbf{y})$ is constant, this leads one to minimize $Q(\boldsymbol{\eta}, \phi) = -\mathbb{E}_g \log f(\mathbf{y}; \boldsymbol{\eta}, \phi)$, the expected negative log likelihood. There is no requirement that g is a member of the family defined by f.

If parameters are to be estimated as $[\boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}]$, functions of random sample $\mathbf{y} \sim g$, then $Q(\boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}})$ is itself a random variable and one chooses estimators to minimize its expectation. *Crucially, we imagine a double-sample expectation*, where the minimization objective is

$$\mathbb{E}_{\mathbf{y}|g}\mathbb{E}_{\tilde{\mathbf{y}}|g} \log f(\tilde{\mathbf{y}}; \boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}). \tag{30}$$

The notation here indicates that inner and outer expectations are based on two *independent* random samples from g: $\mathbf{y}$ for training, upon which $\boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}$ are calculated, and $\tilde{\mathbf{y}}$ for validation.

Information criteria (IC) are analytic approximations to metrics like (30).[3] They take the form

$$-2\log f(\mathbf{y}; \boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}) + c(df) \tag{31}$$

where $c(df)$ is cost of the *degrees-of-freedom* used in $\boldsymbol{\eta}_{\mathbf{y}}$ – e.g., for $\mathbf{y} \sim (\boldsymbol{\eta}, \sigma^2\mathbf{I})$, Efron et al. (2004) defines $df = \sigma^{-2}\sum_i \mathrm{cov}(\eta_{\mathbf{y}i}, y_i)$.

Consider a Gaussian regression model where $\boldsymbol{\eta}_{\mathbf{y}}$ is an estimate for $\boldsymbol{\eta} = \mathbb{E}\mathbf{y}$ using df degrees of freedom, and set $\phi_{\mathbf{y}} = \sigma^2_{\mathbf{y}} = \sum_i (y_i - \eta_{\mathbf{y}i})^2/n$. We'll derive

$$df\frac{n}{n-df-1} \approx \mathbb{E}_{\mathbf{y}|g}\left[\log f(\mathbf{y}; \boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}) - \mathbb{E}_{\tilde{\mathbf{y}}|g}\log f(\tilde{\mathbf{y}}; \boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}})\right], \tag{32}$$

such that AICc's complexity penalty is the expected bias that results from taking the fitted log likelihood as an estimate for (30). First, by cancellation the inner term of (32) simplifies as

$$\log f(\mathbf{y}; \boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}) - \mathbb{E}_{\tilde{\mathbf{y}}|g}\log f(\tilde{\mathbf{y}}; \boldsymbol{\eta}_{\mathbf{y}}, \phi_{\mathbf{y}}) = \frac{\mathbb{E}_{\tilde{\mathbf{y}}|g}\sum_i(\tilde{y}_i - \eta_{\mathbf{y}i})^2}{2\sigma^2_{\mathbf{y}}} - \frac{n}{2}. \tag{33}$$

Now, assume that the *true* model is linear and that the data were generated from $\mathbf{y} \sim g(\boldsymbol{\eta}, \sigma^2\mathbf{I})$. The Mallows (1973) C_p formula holds that $n\sigma^2_{\mathbf{y}} + 2\sigma^2 df$ is an unbiased estimator for expected sum of square errors $\mathbb{E}_{\tilde{\mathbf{y}}|g}\sum_i(\tilde{y}_i - \eta_{\mathbf{y}i})^2/n$, such that

$$\frac{\mathbb{E}_{\tilde{\mathbf{y}}|g}\sum_i(\tilde{y}_i - \eta_{\mathbf{y}i})^2}{2\sigma^2_{\mathbf{y}}} - \frac{n}{2} \;\approx\; \frac{n\sigma^2_{\mathbf{y}} + 2\sigma^2 df}{2\sigma^2_{\mathbf{y}}} - \frac{n}{2} \;=\; df\frac{\sigma^2}{\sigma^2_{\mathbf{y}}}. \tag{34}$$

At this point, we see that the standard AIC approximation results from equating $\sigma^2 \approx \mathbb{E}_{\mathbf{y}|g}\sigma^2_{\mathbf{y}}$, so that $df\,\mathbb{E}_{\mathbf{y}|g}[\sigma^2/\sigma^2_{\mathbf{y}}] \approx df$. This will underpenalize complexity whenever residual variance

[3] Not all IC target (30). For example, the 'Bayesian' BIC, with $c(df) = \log(n)df$ (Schwarz, 1978), is derived (Kass and Raftery, 1995) as Laplace approximation to the negative log of the *marginal likelihood*. We include the BIC as a comparator to AIC and AICc in our examples.

$\sigma^2_{\mathbf{y}}$ tends to be smaller than the true variance σ^2 – that is, whenever the model is overfit. In contrast, AICc applies the chi-squared goodness of fit result $n\sigma^2_{\mathbf{y}}/\sigma^2 \sim \chi^2_{n-df-1}$ to obtain

$$\mathbb{E}_{\mathbf{y}|g}\left[\frac{\sigma^2}{\sigma^2_{\mathbf{y}}}df\right] = n\mathbb{E}_{\mathbf{y}|g}\left[\frac{1}{n\sigma^2_{\mathbf{y}}/\sigma^2}\right]df = \frac{n}{n-df-1}df. \tag{35}$$

Multiplying by -2 and subtracting from $-2\log f(\mathbf{y};\boldsymbol{\eta}_{\mathbf{y}},\sigma_{\mathbf{y}})$ yields the AICc.

13 Hockey players

Ten-fold CV results are shown in Figure 5 for γ of 0, 1, and 10. The OOS error minima are around the same in each case – average deviance slightly above 1.16 – but errors increase much faster away from optimality with larger γ. We also see that AICc selection is always between the CV.min and CV.1se selctions: at $\gamma = 0$ AICc matches the CV.1se choice, while at $\gamma = 10$ it has moved right to the CV.1se selection. Our heuristic might be over-estimating df for large-γ models (especially under this very collinear design), but one would also suspect that CV estimates of minimum deviance are biased downward more dramatically for larger γ than for low-variance small-γ estimators.

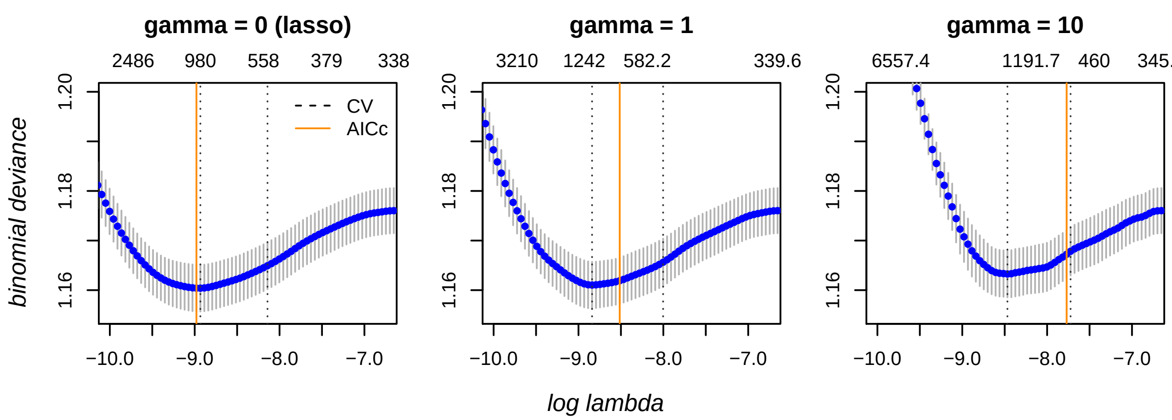

Figure 5: Hockey example 10-fold CV: mean OOS deviance $\pm$1se, with minimum-error and 1SE selection rules marked with black dotted lines, and solid orange line showing AICc selection.

14 Full simulation results

Continuous-response data are simulated from a $p = 1000$ dimensional linear model

$$y \sim \mathrm{N}\left(\mathbf{x}'\boldsymbol{\beta}, \sigma^2\right) \text{ where } \beta_j = (-1)^j \exp\left(-\frac{j}{\mathsf{d}}\right) \text{ for } j = 1 \dots p \text{ and, given } \mathbf{z} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right),$$

$$\text{either } \textit{dense design: } x_j = z_j \text{ or } \textit{sparse design: } x_j \overset{ind}{\sim} \mathrm{Bern}\left(1/(1+e^{-z_j})\right). \tag{36}$$

Each simulation draws $n = 1000$ means $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$, and two independent response samples $\mathbf{y}, \tilde{\mathbf{y}} \sim \mathrm{N}(\boldsymbol{\eta}, \sigma^2\mathbf{I})$. Residual variance σ^2 and covariate correlation $\boldsymbol{\Sigma}$ are adjusted across runs. In the first case, we define σ^2 through *signal-to-noise* ratios $\mathrm{sd}(\boldsymbol{\eta})/\sigma$ of $1/2$, 1, and 2. In the latter case, multicollinearity is parametrized via $\Sigma_{jk} = \rho^{|j-k|}$, and we consider $\rho = 0, 0.5,$ and 0.9. Finally, the coefficient decay rate d controls the effective sparsity: how much $\boldsymbol{\beta}$ is *measurably* different from zero. See Figure 6 for illustration; we consider d of 10, 50, 100, and 200.

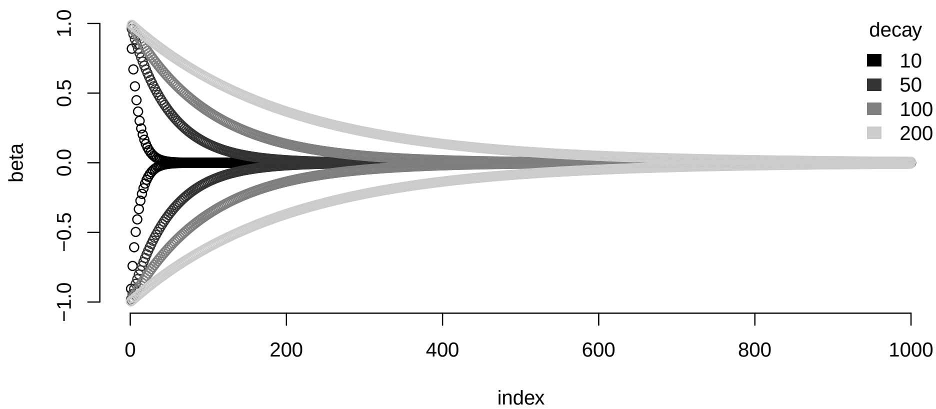

Figure 6: The linear model coefficients for our simulation in 36.

Results for both sparse and dense designs, over a set of 1000 datasets, are presented in the following tables. We first report out-of-sample $R^2 = 1 - \mathrm{var}(\tilde{\mathbf{y}} - \boldsymbol{\eta}_{\mathbf{y}})/\mathrm{var}(\tilde{\mathbf{y}})$, followed by tables showing false discovery and sensitivity with respect to the C_p oracle.

References

Armagan, A., D. B. Dunson, and J. Lee (2013). Generalized double pareto shrinkage. *Statistica Sinica 23*, 119.

Buhlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data.* Springer.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics 32*, 407–499.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

Fan, J., L. Xue, and H. Zou (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics 42*(3), 819–849.

Friedman, J., T. Hastie, H. Hofling, and R. Tibshirani (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics 1*, 302–332.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software 33*, 1–22.

Goldberger, A. S. (1961). Stepwise least squares: Residual analysis and specification error. *Journal of the American Statistical Association 56*(296), 998.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society, Series B 46*, 149–192.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–795.

Lange, K. (2010). *Numerical Analysis for Statisticians* (2nd ed.). Springer.

Lee, J., Y. Sun, and M. Saunders (2014). Proximal newton-type methods for minimizing convex objective functions in composite form. *SIAM Journal on Optimization 24*, 1420–1443.

Luenberger, D. G. and Y. Ye (2008). *Linear and Nonlinear Programming* (3rd ed.). Springer.

Mallows, C. L. (1973). Some comments on CP. *Technometrics 15*, 661–675.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications 109*, 475–494.

Wainwright, M. J. (2006). Sharp thresholds for high-dimensional and noisy recovery of sparsity. *UC Berkeley Technical Report*.

Wainwright, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using L1-constrained quadratic programming (lasso). *IEEE Transactions on Information Theory 55*(5), 2183–2202.

Wu, T. T. and K. Lange (2008). Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics 2*, 1–21.

Zhou, S. (2009). Thresholding procedures for high-dimensional variable selection and statistical estimation.pdf. *Advances in Neural Information Processing Systems 22*.

Zhou, S., S. van de Geer, and P. Bhlmann (2009). Adaptive lasso for high dimensional regression and gaussian graphical modeling. *arXiv preprint arXiv:0903.2515*.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.

Zou, H., T. Hastie, and R. Tibshirani (2007). On the degrees of freedom of the lasso. *The Annals of Statistics 35*, 2173–2192.

Zou, H. and R. Li (2008). One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics 36*(4), 1509–1533.

			% Worse than oracle C_p									
			lasso		GL $\gamma = 1$		GL $\gamma = 10$		marginal AL			
$\mathrm{sd}(\boldsymbol{\eta})/\sigma$	d	ρ	*AICc*	*CV*	*AICc*	*CV*	*AICc*	*CV*	*AICc*	*CV*	MCP	C_p R^2
		0	3	3	2	2	2	**1**	2	2	**1**	*0.79*
	10	*0.5*	5	5	4	4	3	**2**	8	8	**2**	*0.79*
		0.9	7	6	6	5	4	4	10	10	**3**	*0.79*
		0	6	5	5	**4**	5	5	5	5	**4**	*0.77*
2	*50*	*0.5*	11	8	9	7	**6**	**6**	14	14	**6**	*0.77*
		0.9	14	10	11	9	**7**	**7**	44	44	**7**	*0.77*
		0	9	6	7	**5**	7	7	8	7	**5**	*0.75*
	100	*0.5*	18	**9**	13	**9**	**9**	**9**	21	17	**9**	*0.75*
		0.9	23	11	16	10	10	**9**	56	56	11	*0.75*
		0	11	**-1**	6	0	3	7	6	2	**-1**	*0.67*
	200	*0.5*	47	**5**	20	7	6	74	30	21	**5**	*0.67*
		0.9	90	39	84	42	**7**	92	66	67	58	*0.68*
		0	9	9	7	7	8	6	9	10	**5**	*0.48*
	10	*0.5*	16	16	14	13	10	9	18	18	**7**	*0.48*
		0.9	23	22	20	19	16	14	**10**	**10**	11	*0.48*
		0	19	**17**	18	**17**	27	23	**17**	18	**17**	*0.44*
1	*50*	*0.5*	38	32	32	29	28	34	39	38	**25**	*0.44*
		0.9	70	68	70	54	66	56	46	46	**31**	*0.44*
		0	26	**20**	24	25	45	52	22	22	**20**	*0.40*
	100	*0.5*	64	57	56	61	51	95	51	**47**	57	*0.40*
		0.9	86	88	90	88	89	88	**65**	67	85	*0.40*
		0	35	**21**	27	47	66	94	23	**21**	**21**	*0.34*
	200	*0.5*	84	87	93	94	84	99	58	**52**	87	*0.34*
		0.9	91	93	95	93	96	94	**81**	83	93	*0.34*
		0	27	28	27	**24**	48	25	44	56	**24**	*0.17*
	10	*0.5*	50	52	46	45	63	41	71	77	**29**	*0.17*
		0.9	44	45	44	43	42	42	29	**28**	38	*0.18*
		0	**55**	57	74	71	101	94	72	90	58	*0.13*
0.5	*50*	*0.5*	**90**	95	98	96	101	100	108	117	95	*0.13*
		0.9	79	80	81	79	79	77	83	81	**76**	*0.13*
		0	**62**	69	96	93	102	101	84	112	68	*0.09*
	100	*0.5*	**94**	98	101	100	102	101	118	132	98	*0.09*
		0.9	**87**	89	92	88	95	89	95	92	**87**	*0.09*
		0	**54**	65	110	100	102	104	99	157	67	*0.04*
	200	*0.5*	**92**	98	102	102	102	103	142	169	98	*0.04*
		0.9	**91**	94	99	95	101	99	108	103	95	*0.05*

Table 3: The dense design version for Table 1 from the main paper. Average out-of-sample R^2, reported as % worse than the C_p oracle, across 1000 different samples from (36) under the dense design.

Table 4: Predictive R^2, for **dense** design and $\mathbf{d = 10}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.76	0.76	0.77	0.77	0.77	
CV.min	0.77	0.77	**0.78**	0.77	**0.78**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.77	0.77	**0.78**	0.77		$\rho = 0$
AIC	0.70	0.70	0.68	0.77		C_p R^2 = 0.79
BIC	0.75	0.76	0.77	0.77		
CV.1se	0.74	0.74	0.76	0.71	0.77	
CV.min	0.75	0.76	0.77	0.73	**0.78**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.75	0.76	0.77	0.73		$\rho = 0.5$
AIC	0.71	0.71	0.70	0.73		C_p R^2 = 0.79
BIC	0.70	0.72	0.75	0.72		
CV.1se	0.73	0.73	0.75	0.70	0.76	
CV.min	0.74	0.75	0.76	0.71	**0.77**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.74	0.75	0.76	0.71		$\rho = 0.9$
AIC	0.74	0.75	0.76	0.71		C_p R^2 = 0.79
BIC	0.65	0.67	0.70	0.71		
CV.1se	0.40	0.42	0.43	**0.45**	0.44	
CV.min	0.44	**0.45**	**0.45**	0.43	**0.45**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.44	**0.45**	0.44	0.44		$\rho = 0$
AIC	0.18	0.17	0.11	0.42		C_p R^2 = 0.48
BIC	0.40	0.41	0.41	0.44		
CV.1se	0.36	0.38	0.41	0.39	0.43	
CV.min	0.40	0.42	0.44	0.39	**0.45**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.40	0.42	0.43	0.39		$\rho = 0.5$
AIC	0.21	0.20	0.15	0.39		C_p R^2 = 0.48
BIC	0.31	0.34	0.38	0.35		
CV.1se	0.31	0.34	0.39	0.41	0.42	
CV.min	0.37	0.39	0.41	**0.44**	0.43	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.37	0.38	0.40	**0.44**		$\rho = 0.9$
AIC	0.35	0.34	0.30	**0.44**		C_p R^2 = 0.48
BIC	0.26	0.26	0.27	0.42		
CV.1se	0.08	0.09	0.09	**0.13**	0.10	
CV.min	**0.13**	**0.13**	**0.13**	0.08	**0.13**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.13**	**0.13**	0.09	0.10		$\rho = 0$
AIC	-0.40	-0.43	-0.51	-0.07		C_p R^2 = 0.17
BIC	0.09	0.09	0.04	0.12		
CV.1se	0.03	0.04	0.06	0.07	0.09	
CV.min	0.08	0.10	0.10	0.04	**0.12**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.09	0.09	0.07	0.05		$\rho = 0.5$
AIC	-0.37	-0.39	-0.48	-0.04		C_p R^2 = 0.17
BIC	0.04	0.04	0.04	0.05		
CV.1se	0.06	0.06	0.09	0.12	0.07	
CV.min	0.10	0.10	0.10	**0.13**	0.11	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.10	0.10	0.10	**0.13**		$\rho = 0.9$
AIC	-0.15	-0.18	-0.32	0.12		C_p R^2 = 0.18
BIC	0.10	0.10	0.10	0.11		

Table 5: Predictive R^2, for **dense** design and $\mathbf{d} = \mathbf{50}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.71	0.72	0.72	0.72	0.72	
CV.min	0.73	0.73	0.73	0.73	**0.74**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.72	0.73	0.73	0.73		$\rho = 0$
AIC	0.67	0.67	0.64	0.73		C_p R^2 = 0.77
BIC	0.63	0.65	0.67	0.68		
CV.1se	0.68	0.69	0.70	0.64	0.71	
CV.min	0.70	0.71	**0.72**	0.66	**0.72**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.68	0.70	**0.72**	0.66		$\rho = 0.5$
AIC	0.67	0.67	0.65	0.67		C_p R^2 = 0.77
BIC	0.27	0.49	0.64	0.47		
CV.1se	0.67	0.68	0.70	0.39	0.71	
CV.min	0.69	0.70	0.71	0.43	**0.72**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.66	0.68	0.71	0.43		$\rho = 0.9$
AIC	0.69	0.71	0.71	0.43		C_p R^2 = 0.77
BIC	0.13	0.13	0.14	0.35		
CV.1se	0.31	0.31	0.29	**0.36**	0.33	
CV.min	**0.36**	**0.36**	0.34	**0.36**	**0.36**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.35	**0.36**	0.32	**0.36**		$\rho = 0$
AIC	0.13	0.11	0.05	0.31		C_p R^2 = 0.44
BIC	0.14	0.17	0.02	0.27		
CV.1se	0.20	0.23	0.22	0.25	0.30	
CV.min	0.30	0.31	0.29	0.27	**0.33**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.27	0.30	0.31	0.27		$\rho = 0.5$
AIC	0.14	0.12	0.07	0.26		C_p R^2 = 0.44
BIC	0.01	0.01	0.00	0.04		
CV.1se	0.05	0.09	0.12	0.19	0.26	
CV.min	0.14	0.20	0.19	0.24	**0.3**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.13	0.13	0.15	0.24		$\rho = 0.9$
AIC	0.25	0.25	0.18	0.24		C_p R^2 = 0.44
BIC	0.08	0.08	0.08	0.09		
CV.1se	0.01	0.00	-0.00	0.05	0.01	
CV.min	**0.06**	0.04	0.01	0.01	0.05	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.06**	0.03	-0.00	0.04		$\rho = 0$
AIC	-0.44	-0.49	-0.56	-0.19		C_p R^2 = 0.13
BIC	0.00	-0.00	-0.00	0.01		
CV.1se	-0.00	-0.00	-0.00	**0.01**	-0.00	
CV.min	**0.01**	0.00	0.00	-0.02	**0.01**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.01**	0.00	-0.00	-0.01		$\rho = 0.5$
AIC	-0.43	-0.47	-0.54	-0.21		C_p R^2 = 0.13
BIC	-0.00	-0.00	-0.00	0.00		
CV.1se	0.00	0.00	0.00	0.01	0.00	
CV.min	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.03**	0.02	**0.03**	0.02		$\rho = 0.9$
AIC	-0.27	-0.31	-0.45	0.01		C_p R^2 = 0.13
BIC	0.02	0.02	0.02	**0.03**		

Table 6: Predictive R^2, for **dense** design and $\mathbf{d = 100}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.68	0.68	0.67	0.68	0.69	
CV.min	**0.71**	**0.71**	0.69	0.70	**0.71**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.68	0.69	0.70	0.69		$\rho = 0$
AIC	0.66	0.65	0.63	0.70		C_p R^2 = 0.75
BIC	0.25	0.46	0.60	0.56		
CV.1se	0.64	0.65	0.66	0.56	0.66	
CV.min	**0.68**	**0.68**	**0.68**	0.62	**0.68**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.61	0.65	**0.68**	0.59		$\rho = 0.5$
AIC	0.66	0.65	0.64	0.64		C_p R^2 = 0.75
BIC	0.01	0.01	0.16	0.03		
CV.1se	0.63	0.64	0.66	0.24	0.65	
CV.min	0.67	**0.68**	**0.68**	0.33	0.67	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.58	0.63	**0.68**	0.33		$\rho = 0.9$
AIC	0.67	**0.68**	**0.68**	0.34		C_p R^2 = 0.75
BIC	0.07	0.07	0.07	0.08		
CV.1se	0.24	0.21	0.10	0.30	0.26	
CV.min	**0.32**	0.30	0.19	0.31	**0.32**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.29	0.30	0.22	0.31		$\rho = 0$
AIC	0.12	0.09	0.04	0.26		C_p R^2 = 0.40
BIC	0.00	0.00	-0.00	0.08		
CV.1se	0.03	0.04	0.00	0.17	0.06	
CV.min	0.17	0.15	0.02	**0.21**	0.17	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.14	0.18	0.20	0.19		$\rho = 0.5$
AIC	0.12	0.10	0.05	**0.21**		C_p R^2 = 0.40
BIC	0.00	0.00	-0.00	0.01		
CV.1se	0.01	0.01	0.01	0.06	0.01	
CV.min	0.05	0.05	0.05	0.13	0.06	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.05	0.04	0.04	0.14		$\rho = 0.9$
AIC	**0.21**	0.20	0.13	0.17		C_p R^2 = 0.40
BIC	0.04	0.04	0.04	0.04		
CV.1se	0.00	-0.00	-0.00	**0.03**	-0.00	
CV.min	**0.03**	0.01	-0.00	-0.01	**0.03**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.03**	0.00	-0.00	0.01		$\rho = 0$
AIC	-0.45	-0.50	-0.56	-0.23		C_p R^2 = 0.09
BIC	-0.00	-0.00	-0.05	0.00		
CV.1se	-0.00	-0.00	-0.00	0.00	-0.00	
CV.min	0.00	-0.00	-0.00	-0.03	0.00	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.01**	-0.00	-0.00	-0.02		$\rho = 0.5$
AIC	-0.44	-0.49	-0.55	-0.23		C_p R^2 = 0.09
BIC	-0.00	-0.00	-0.00	-0.00		
CV.1se	-0.00	-0.00	-0.00	0.00	-0.00	
CV.min	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.01**	**0.01**	0.00	0.00		$\rho = 0.9$
AIC	-0.32	-0.37	-0.49	-0.01		C_p R^2 = 0.09
BIC	**0.01**	0.00	0.00	**0.01**		

Table 7: Predictive R^2, for **dense** design and $\mathbf{d = 200}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.64	0.63	0.54	0.63	0.66	
CV.min	**0.68**	0.67	0.63	0.66	**0.68**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.60	0.63	0.65	0.63		$\rho = 0$
AIC	0.65	0.64	0.62	0.67		C_p R^2 = 0.67
BIC	0.00	0.00	0.42	0.08		
CV.1se	0.52	0.49	0.06	0.44	0.54	
CV.min	0.64	0.63	0.18	0.53	0.64	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.36	0.54	0.63	0.47		$\rho = 0.5$
AIC	**0.65**	0.64	0.63	0.61		C_p R^2 = 0.67
BIC	0.00	0.00	0.02	0.01		
CV.1se	0.18	0.18	0.01	0.08	0.11	
CV.min	0.42	0.39	0.05	0.22	0.28	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.07	0.11	0.63	0.23		$\rho = 0.9$
AIC	**0.65**	**0.65**	**0.65**	0.34		C_p R^2 = 0.68
BIC	0.03	0.03	0.03	0.04		
CV.1se	0.15	0.06	0.00	0.25	0.17	
CV.min	**0.27**	0.18	0.02	**0.27**	**0.27**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.22	0.25	0.12	0.26		$\rho = 0$
AIC	0.11	0.08	0.03	0.22		C_p R^2 = 0.34
BIC	0.00	-0.00	0.00	0.01		
CV.1se	0.00	0.00	-0.00	0.11	0.00	
CV.min	0.04	0.02	0.00	0.16	0.04	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.06	0.03	0.06	0.14		$\rho = 0.5$
AIC	0.11	0.08	0.04	**0.18**		C_p R^2 = 0.34
BIC	0.00	-0.00	-0.00	0.00		
CV.1se	0.00	0.00	0.00	0.01	-0.00	
CV.min	0.02	0.02	0.02	0.06	0.02	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.03	0.02	0.01	0.07		$\rho = 0.9$
AIC	**0.18**	0.15	0.09	0.14		C_p R^2 = 0.34
BIC	0.01	0.01	0.01	0.02		
CV.1se	-0.00	-0.00	-0.00	**0.02**	-0.00	
CV.min	**0.02**	-0.00	-0.00	-0.03	0.01	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.02**	-0.00	-0.00	0.00		$\rho = 0$
AIC	-0.45	-0.52	-0.57	-0.25		C_p R^2 = 0.04
BIC	-0.00	-0.00	-0.41	-0.00		
CV.1se	**0**	**0**	**0**	**0**	**0**	
CV.min	**0**	**0**	**0**	-0.03	**0**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0**	**0**	**0**	-0.02		$\rho = 0.5$
AIC	-0.44	-0.50	-0.56	-0.25		C_p R^2 = 0.04
BIC	**0**	**0**	-0.01	**0**		
CV.1se	**0**	**0**	**0**	**0**	**0**	
CV.min	**0**	**0**	**0**	**0**	**0**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0**	**0**	**0**	**0**		$\rho = 0.9$
AIC	-0.34	-0.41	-0.52	-0.03		C_p R^2 = 0.05
BIC	**0**	**0**	**0**	**0**		

Table 8: Predictive R^2, for **sparse** design and $\mathbf{d = 10}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.76	0.76	0.77	0.77	0.77	
CV.min	0.77	0.77	**0.78**	**0.78**	**0.78**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.77	0.77	**0.78**	0.77		$\rho = 0$
AIC	0.70	0.70	0.68	**0.78**		C_p R^2 = 0.79
BIC	0.75	0.76	0.77	0.77		
CV.1se	0.75	0.76	0.77	0.77	0.77	
CV.min	0.77	0.77	**0.78**	0.77	**0.78**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.77	0.77	**0.78**	0.77		$\rho = 0.5$
AIC	0.70	0.70	0.68	0.77		C_p R^2 = 0.79
BIC	0.75	0.76	0.77	0.76		
CV.1se	0.75	0.76	0.77	0.76	0.77	
CV.min	0.77	0.77	**0.78**	0.77	**0.78**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.77	0.77	**0.78**	0.77		$\rho = 0.9$
AIC	0.70	0.70	0.69	0.77		C_p R^2 = 0.79
BIC	0.74	0.75	0.77	0.76		
CV.1se	0.41	0.42	0.43	**0.45**	0.44	
CV.min	0.44	**0.45**	**0.45**	0.43	**0.45**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.44	**0.45**	0.44	0.44		$\rho = 0$
AIC	0.18	0.17	0.11	0.42		C_p R^2 = 0.48
BIC	0.40	0.42	0.43	0.44		
CV.1se	0.40	0.41	0.43	0.44	0.44	
CV.min	0.44	0.44	**0.45**	0.43	**0.45**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.44	0.44	0.44	0.44		$\rho = 0.5$
AIC	0.18	0.17	0.11	0.41		C_p R^2 = 0.48
BIC	0.39	0.41	0.43	0.43		
CV.1se	0.40	0.41	0.43	0.44	0.44	
CV.min	0.43	0.44	**0.45**	0.43	**0.45**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.43	0.44	0.44	0.43		$\rho = 0.9$
AIC	0.18	0.17	0.12	0.41		C_p R^2 = 0.48
BIC	0.39	0.41	0.43	0.43		
CV.1se	0.08	0.09	0.10	**0.13**	0.10	
CV.min	**0.13**	**0.13**	**0.13**	0.08	**0.13**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.13**	**0.13**	0.08	0.10		$\rho = 0$
AIC	-0.40	-0.43	-0.52	-0.08		C_p R^2 = 0.17
BIC	0.09	0.10	0.08	0.12		
CV.1se	0.07	0.08	0.09	**0.13**	0.10	
CV.min	0.12	**0.13**	**0.13**	0.07	**0.13**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.12	**0.13**	0.08	0.10		$\rho = 0.5$
AIC	-0.40	-0.43	-0.51	-0.08		C_p R^2 = 0.17
BIC	0.08	0.09	0.08	0.11		
CV.1se	0.07	0.08	0.09	0.12	0.10	
CV.min	0.12	**0.13**	**0.13**	0.08	**0.13**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.12	**0.13**	0.08	0.10		$\rho = 0.9$
AIC	-0.40	-0.42	-0.51	-0.07		C_p R^2 = 0.17
BIC	0.08	0.09	0.08	0.11		

Table 9: Predictive R^2, for **sparse** design and $\mathbf{d} = \mathbf{50}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.71	0.71	0.72	0.72	0.72	
CV.min	0.73	0.73	0.73	0.73	**0.74**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.72	0.73	0.73	0.73		$\rho = 0$
AIC	0.67	0.67	0.64	0.73		C_p R^2 = 0.77
BIC	0.63	0.66	0.69	0.68		
CV.1se	0.71	0.71	0.71	0.71	0.72	
CV.min	0.73	0.73	0.73	0.72	**0.74**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.72	0.73	0.73	0.72		$\rho = 0.5$
AIC	0.67	0.66	0.64	0.72		C_p R^2 = 0.77
BIC	0.61	0.65	0.69	0.67		
CV.1se	0.71	0.71	0.71	0.71	0.72	
CV.min	0.73	0.73	0.73	0.72	**0.74**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.72	0.73	0.73	0.72		$\rho = 0.9$
AIC	0.67	0.67	0.64	0.72		C_p R^2 = 0.77
BIC	0.61	0.65	0.69	0.66		
CV.1se	0.31	0.31	0.29	0.36	0.33	
CV.min	0.36	0.36	0.33	0.36	0.36	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.35	**0.37**	0.30	0.36		$\rho = 0$
AIC	0.13	0.11	0.05	0.30		C_p R^2 = 0.44
BIC	0.14	0.21	0.13	0.27		
CV.1se	0.30	0.30	0.29	0.35	0.32	
CV.min	**0.36**	**0.36**	0.33	0.35	**0.36**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.34	**0.36**	0.30	0.35		$\rho = 0.5$
AIC	0.13	0.11	0.05	0.29		C_p R^2 = 0.44
BIC	0.10	0.18	0.12	0.25		
CV.1se	0.30	0.30	0.28	0.35	0.32	
CV.min	0.35	**0.36**	0.33	0.35	**0.36**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.34	**0.36**	0.31	0.35		$\rho = 0.9$
AIC	0.13	0.11	0.05	0.30		C_p R^2 = 0.44
BIC	0.09	0.17	0.11	0.24		
CV.1se	0.01	0.00	0.00	0.05	0.01	
CV.min	**0.06**	0.04	0.01	0.01	0.05	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.06**	0.05	-0.06	0.04		$\rho = 0$
AIC	-0.45	-0.49	-0.56	-0.20		C_p R^2 = 0.13
BIC	0.00	0.00	-0.00	0.01		
CV.1se	0.00	0.00	-0.00	**0.05**	0.00	
CV.min	**0.05**	0.03	0.00	0.00	**0.05**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.05**	**0.05**	-0.07	0.03		$\rho = 0.5$
AIC	-0.45	-0.49	-0.56	-0.21		C_p R^2 = 0.13
BIC	0.00	0.00	-0.00	0.01		
CV.1se	0.00	0.00	-0.00	0.04	0.00	
CV.min	0.04	0.03	0.00	0.01	0.04	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.05**	**0.05**	-0.06	0.03		$\rho = 0.9$
AIC	-0.44	-0.49	-0.56	-0.20		C_p R^2 = 0.13
BIC	0.00	0.00	-0.00	0.01		

Table 10: Predictive R^2, for **sparse** design and $\mathbf{d = 100}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.68	0.68	0.67	0.68	0.69	
CV.min	0.70	**0.71**	0.69	0.70	**0.71**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.68	0.69	0.70	0.69		$\rho = 0$
AIC	0.66	0.65	0.63	0.70		C_p R^2 = 0.75
BIC	0.24	0.50	0.63	0.56		
CV.1se	0.68	0.68	0.66	0.67	0.69	
CV.min	**0.7**	**0.7**	0.69	0.69	**0.7**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.67	0.69	**0.7**	0.68		$\rho = 0.5$
AIC	0.66	0.65	0.63	0.69		C_p R^2 = 0.75
BIC	0.12	0.46	0.63	0.53		
CV.1se	0.68	0.68	0.67	0.67	0.69	
CV.min	**0.7**	**0.7**	0.69	0.69	**0.7**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.67	0.69	**0.7**	0.68		$\rho = 0.9$
AIC	0.66	0.65	0.63	0.69		C_p R^2 = 0.75
BIC	0.10	0.45	0.63	0.51		
CV.1se	0.24	0.21	0.09	0.30	0.26	
CV.min	**0.32**	0.30	0.19	0.31	**0.32**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.29	0.31	0.24	0.31		$\rho = 0$
AIC	0.11	0.09	0.04	0.26		C_p R^2 = 0.40
BIC	0.00	0.02	0.00	0.07		
CV.1se	0.22	0.20	0.08	0.29	0.24	
CV.min	**0.31**	0.29	0.17	0.30	**0.31**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.28	0.30	0.24	0.30		$\rho = 0.5$
AIC	0.11	0.09	0.04	0.25		C_p R^2 = 0.40
BIC	0.00	0.01	-0.00	0.05		
CV.1se	0.22	0.20	0.07	0.29	0.24	
CV.min	**0.31**	0.29	0.17	0.30	**0.31**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.28	0.30	0.23	0.30		$\rho = 0.9$
AIC	0.11	0.09	0.04	0.25		C_p R^2 = 0.40
BIC	0.00	0.01	0.00	0.05		
CV.1se	0.00	-0.00	-0.00	**0.03**	-0.00	
CV.min	**0.03**	0.01	-0.00	-0.01	**0.03**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.03**	0.01	-0.04	0.01		$\rho = 0$
AIC	-0.45	-0.51	-0.57	-0.23		C_p R^2 = 0.09
BIC	-0.00	-0.00	-0.00	0.00		
CV.1se	0.00	-0.00	-0.00	0.02	-0.00	
CV.min	0.02	0.00	-0.00	-0.02	0.02	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.03**	0.01	-0.05	0.01		$\rho = 0.5$
AIC	-0.45	-0.51	-0.57	-0.24		C_p R^2 = 0.09
BIC	-0.00	-0.00	-0.00	0.00		
CV.1se	-0.00	-0.00	-0.00	0.02	-0.00	
CV.min	0.02	0.00	-0.00	-0.02	0.02	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.03**	0.01	-0.05	0.01		$\rho = 0.9$
AIC	-0.45	-0.50	-0.57	-0.23		C_p R^2 = 0.09
BIC	-0.00	-0.00	-0.00	0.00		

Table 11: Predictive R^2, for **sparse** design and $\mathbf{d = 200}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.64	0.62	0.55	0.63	0.66	
CV.min	**0.68**	0.67	0.63	0.66	**0.68**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.60	0.64	0.65	0.63		$\rho = 0$
AIC	0.65	0.64	0.62	0.67		C_p R^2 = 0.67
BIC	0.00	0.01	0.27	0.08		
CV.1se	0.63	0.62	0.53	0.62	0.65	
CV.min	0.67	0.67	0.62	0.65	**0.68**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.59	0.63	0.65	0.62		$\rho = 0.5$
AIC	0.65	0.64	0.62	0.66		C_p R^2 = 0.67
BIC	0.00	0.01	0.22	0.05		
CV.1se	0.63	0.62	0.51	0.61	0.65	
CV.min	0.67	0.67	0.61	0.64	**0.68**	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.58	0.63	0.65	0.61		$\rho = 0.9$
AIC	0.65	0.64	0.62	0.66		C_p R^2 = 0.67
BIC	0.00	0.01	0.21	0.04		
CV.1se	0.15	0.06	0.00	0.25	0.17	
CV.min	**0.27**	0.18	0.02	**0.27**	**0.27**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.22	0.26	0.18	0.26		$\rho = 0$
AIC	0.11	0.08	0.03	0.22		C_p R^2 = 0.34
BIC	0.00	0.00	-0.00	0.01		
CV.1se	0.12	0.04	0.00	0.24	0.14	
CV.min	0.25	0.15	0.01	**0.26**	0.25	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.20	0.25	0.18	0.25		$\rho = 0.5$
AIC	0.10	0.08	0.03	0.21		C_p R^2 = 0.34
BIC	0.00	0.00	-0.00	0.01		
CV.1se	0.12	0.04	0.00	0.23	0.13	
CV.min	**0.25**	0.16	0.01	**0.25**	**0.25**	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.20	**0.25**	0.18	0.24		$\rho = 0.9$
AIC	0.11	0.08	0.03	0.22		C_p R^2 = 0.34
BIC	0.00	0.00	-0.00	0.01		
CV.1se	-0.00	-0.00	-0.00	**0.02**	-0.00	
CV.min	0.01	-0.00	-0.00	-0.03	0.01	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.02**	-0.03	-0.02	0.00		$\rho = 0$
AIC	-0.46	-0.52	-0.57	-0.25		C_p R^2 = 0.04
BIC	-0.00	-0.00	-0.02	-0.00		
CV.1se	-0.00	-0.00	-0.00	0.01	-0.00	
CV.min	0.01	-0.00	-0.00	-0.03	0.01	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.02**	-0.02	-0.03	-0.00		$\rho = 0.5$
AIC	-0.46	-0.52	-0.57	-0.26		C_p R^2 = 0.04
BIC	-0.00	-0.00	-0.00	-0.00		
CV.1se	-0.00	-0.00	-0.00	**0.01**	-0.00	
CV.min	**0.01**	-0.00	-0.00	-0.03	**0.01**	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	**0.01**	-0.02	-0.03	-0.01		$\rho = 0.9$
AIC	-0.45	-0.52	-0.57	-0.25		C_p R^2 = 0.04
BIC	-0.00	-0.00	-0.01	-0.00		

Table 12: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **dense** design and $\mathbf{d = 10}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.36 | 0.75	0.21 | 0.71	0.02 | 0.60	0.24 | 0.65	0.00 | 0.57	
CV.min	0.72 | 0.85	0.64 | 0.83	0.22 | 0.72	0.55 | 0.73	0.20 | 0.71	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.70 | 0.84	0.60 | 0.82	0.10 | 0.67	0.52 | 0.72		$\rho = 0$
AIC	0.95 | 0.95	0.95 | 0.95	0.95 | 0.94	0.53 | 0.72		
BIC	0.22 | 0.71	0.12 | 0.68	0.00 | 0.57	0.18 | 0.63		$\bar{s}_{C_p}$ = 33.2
CV.1se	0.72 | 0.70	0.61 | 0.66	0.17 | 0.56	0.51 | 0.51	0.04 | 0.55	
CV.min	0.85 | 0.80	0.81 | 0.77	0.53 | 0.67	0.60 | 0.55	0.25 | 0.63	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.82 | 0.78	0.78 | 0.75	0.44 | 0.65	0.60 | 0.55		$\rho = 0.5$
AIC	0.94 | 0.95	0.94 | 0.95	0.94 | 0.94	0.60 | 0.55		
BIC	0.49 | 0.59	0.36 | 0.57	0.05 | 0.49	0.49 | 0.51		$\bar{s}_{C_p}$ = 32.9
CV.1se	0.87 | 0.65	0.85 | 0.60	0.73 | 0.50	0.04 | 0.37	0.48 | 0.50	
CV.min	0.90 | 0.76	0.89 | 0.73	0.81 | 0.60	0.06 | 0.40	0.66 | 0.57	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.89 | 0.72	0.87 | 0.68	0.78 | 0.57	0.06 | 0.40		$\rho = 0.9$
AIC	0.90 | 0.79	0.90 | 0.79	0.88 | 0.78	0.06 | 0.40		
BIC	0.80 | 0.43	0.78 | 0.40	0.49 | 0.31	0.06 | 0.39		$\bar{s}_{C_p}$ = 30.4
CV.1se	0.27 | 0.64	0.13 | 0.59	0.01 | 0.48	0.46 | 0.68	0.01 | 0.47	
CV.min	0.72 | 0.79	0.61 | 0.76	0.15 | 0.62	0.81 | 0.81	0.25 | 0.65	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.72 | 0.79	0.57 | 0.75	0.03 | 0.52	0.77 | 0.79		$\rho = 0$
AIC	0.97 | 0.95	0.97 | 0.95	0.97 | 0.93	0.85 | 0.84		
BIC	0.19 | 0.61	0.07 | 0.57	0.00 | 0.40	0.16 | 0.58		$\bar{s}_{C_p}$ = 26.2
CV.1se	0.66 | 0.55	0.53 | 0.51	0.12 | 0.42	0.76 | 0.57	0.07 | 0.44	
CV.min	0.85 | 0.72	0.81 | 0.68	0.43 | 0.54	0.86 | 0.67	0.33 | 0.55	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.84 | 0.70	0.79 | 0.66	0.31 | 0.51	0.85 | 0.65		$\rho = 0.5$
AIC	0.96 | 0.96	0.96 | 0.95	0.96 | 0.94	0.87 | 0.67		
BIC	0.41 | 0.43	0.29 | 0.41	0.02 | 0.32	0.51 | 0.43		$\bar{s}_{C_p}$ = 25.8
CV.1se	0.68 | 0.31	0.80 | 0.34	0.77 | 0.33	0.32 | 0.44	0.61 | 0.35	
CV.min	0.90 | 0.59	0.90 | 0.56	0.84 | 0.45	0.53 | 0.54	0.76 | 0.45	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.91 | 0.56	0.89 | 0.51	0.77 | 0.39	0.53 | 0.54		$\rho = 0.9$
AIC	0.95 | 0.86	0.95 | 0.88	0.94 | 0.88	0.53 | 0.54		
BIC	0.08 | 0.05	0.01 | 0.04	0.01 | 0.05	0.34 | 0.45		$\bar{s}_{C_p}$ = 23.3
CV.1se	0.12 | 0.39	0.06 | 0.34	0.01 | 0.27	0.63 | 0.64	0.01 | 0.27	
CV.min	0.70 | 0.68	0.54 | 0.61	0.14 | 0.43	0.89 | 0.81	0.36 | 0.54	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.73 | 0.69	0.43 | 0.57	0.01 | 0.22	0.85 | 0.77		$\rho = 0$
AIC	0.98 | 0.95	0.98 | 0.94	0.98 | 0.91	0.94 | 0.88		
BIC	0.13 | 0.41	0.03 | 0.33	0.00 | 0.09	0.17 | 0.44		$\bar{s}_{C_p}$ = 19.2
CV.1se	0.16 | 0.11	0.19 | 0.15	0.07 | 0.14	0.80 | 0.42	0.10 | 0.22	
CV.min	0.81 | 0.46	0.78 | 0.44	0.35 | 0.30	0.92 | 0.61	0.44 | 0.37	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.84 | 0.50	0.71 | 0.41	0.04 | 0.13	0.90 | 0.57		$\rho = 0.5$
AIC	0.98 | 0.95	0.98 | 0.95	0.98 | 0.93	0.94 | 0.74		
BIC	0.10 | 0.11	0.04 | 0.09	0.00 | 0.05	0.17 | 0.12		$\bar{s}_{C_p}$ = 19.0
CV.1se	0.00 | 0.07	0.00 | 0.07	0.00 | 0.07	0.44 | 0.28	0.01 | 0.07	
CV.min	0.57 | 0.09	0.45 | 0.07	0.15 | 0.07	0.82 | 0.64	0.35 | 0.12	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.70 | 0.10	0.20 | 0.07	0.00 | 0.07	0.81 | 0.64		$\rho = 0.9$
AIC	0.96 | 0.86	0.97 | 0.90	0.97 | 0.88	0.83 | 0.68		
BIC	0.09 | 0.07	0.01 | 0.07	0.00 | 0.07	0.13 | 0.12		$\bar{s}_{C_p}$ = 16.4

Table 13: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **dense** design and $\mathbf{d} = \mathbf{50}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.50 | 0.76	0.36 | 0.71	0.07 | 0.56	0.44 | 0.67	0.15 | 0.62	
CV.min	0.65 | 0.84	0.56 | 0.80	0.19 | 0.65	0.58 | 0.74	0.39 | 0.74	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.56 | 0.79	0.47 | 0.77	0.29 | 0.69	0.53 | 0.72		$\rho = 0$
AIC	0.84 | 0.94	0.84 | 0.93	0.84 | 0.91	0.63 | 0.77		$\bar{s}_{C_p}$ = 122.8
BIC	0.21 | 0.59	0.11 | 0.55	0.01 | 0.43	0.22 | 0.54		
CV.1se	0.68 | 0.74	0.62 | 0.70	0.33 | 0.56	0.65 | 0.58	0.24 | 0.55	
CV.min	0.74 | 0.84	0.70 | 0.80	0.47 | 0.66	0.69 | 0.65	0.42 | 0.64	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.68 | 0.74	0.63 | 0.72	0.48 | 0.68	0.67 | 0.62		$\rho = 0.5$
AIC	0.84 | 0.95	0.84 | 0.94	0.83 | 0.93	0.70 | 0.67		$\bar{s}_{C_p}$ = 122.4
BIC	0.25 | 0.24	0.35 | 0.39	0.15 | 0.41	0.46 | 0.35		
CV.1se	0.74 | 0.74	0.71 | 0.70	0.60 | 0.58	0.56 | 0.17	0.54 | 0.54	
CV.min	0.77 | 0.83	0.75 | 0.80	0.64 | 0.67	0.61 | 0.21	0.59 | 0.61	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.73 | 0.72	0.71 | 0.70	0.62 | 0.65	0.61 | 0.20		$\rho = 0.9$
AIC	0.77 | 0.83	0.77 | 0.84	0.74 | 0.84	0.61 | 0.21		$\bar{s}_{C_p}$ = 119.4
BIC	0.05 | 0.01	0.01 | 0.01	0.00 | 0.01	0.47 | 0.14		
CV.1se	0.43 | 0.58	0.27 | 0.50	0.07 | 0.33	0.52 | 0.62	0.26 | 0.49	
CV.min	0.66 | 0.73	0.54 | 0.66	0.19 | 0.45	0.68 | 0.73	0.57 | 0.68	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.59 | 0.69	0.50 | 0.65	0.32 | 0.51	0.62 | 0.69		$\rho = 0$
AIC	0.90 | 0.93	0.90 | 0.92	0.90 | 0.90	0.79 | 0.82		$\bar{s}_{C_p}$ = 90.1
BIC	0.07 | 0.25	0.03 | 0.23	0.00 | 0.02	0.17 | 0.38		
CV.1se	0.62 | 0.42	0.60 | 0.42	0.33 | 0.28	0.71 | 0.48	0.38 | 0.37	
CV.min	0.76 | 0.66	0.72 | 0.62	0.48 | 0.42	0.77 | 0.59	0.54 | 0.49	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.72 | 0.57	0.69 | 0.57	0.59 | 0.56	0.75 | 0.55		$\rho = 0.5$
AIC	0.90 | 0.95	0.90 | 0.94	0.90 | 0.92	0.82 | 0.75		$\bar{s}_{C_p}$ = 89.7
BIC	0.01 | 0.01	0.00 | 0.01	0.00 | 0.00	0.09 | 0.04		
CV.1se	0.06 | 0.04	0.18 | 0.10	0.17 | 0.10	0.65 | 0.18	0.61 | 0.30	
CV.min	0.59 | 0.20	0.63 | 0.35	0.43 | 0.24	0.73 | 0.29	0.72 | 0.42	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.67 | 0.16	0.31 | 0.14	0.21 | 0.16	0.73 | 0.28		$\rho = 0.9$
AIC	0.87 | 0.88	0.86 | 0.88	0.86 | 0.85	0.73 | 0.29		$\bar{s}_{C_p}$ = 86.2
BIC	0.06 | 0.01	0.01 | 0.01	0.00 | 0.01	0.11 | 0.03		
CV.1se	0.09 | 0.06	0.03 | 0.02	0.00 | 0.00	0.66 | 0.45	0.07 | 0.05	
CV.min	0.61 | 0.41	0.36 | 0.23	0.07 | 0.05	0.81 | 0.64	0.60 | 0.40	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.64 | 0.43	0.30 | 0.21	0.00 | 0.00	0.76 | 0.57		$\rho = 0$
AIC	0.94 | 0.92	0.94 | 0.90	0.94 | 0.90	0.90 | 0.81		$\bar{s}_{C_p}$ = 56.0
BIC	0.01 | 0.01	0.00 | 0.00	0.00 | 0.00	0.06 | 0.04		
CV.1se	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.69 | 0.16	0.16 | 0.01	
CV.min	0.44 | 0.07	0.31 | 0.04	0.10 | 0.01	0.85 | 0.38	0.48 | 0.07	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.63 | 0.13	0.11 | 0.02	0.00 | 0.00	0.84 | 0.32		$\rho = 0.5$
AIC	0.94 | 0.94	0.94 | 0.93	0.94 | 0.91	0.91 | 0.74		$\bar{s}_{C_p}$ = 55.6
BIC	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.09 | 0.01		
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.01	0.14 | 0.03	0.00 | 0.01	
CV.min	0.54 | 0.04	0.43 | 0.03	0.10 | 0.02	0.82 | 0.17	0.17 | 0.02	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.66 | 0.05	0.12 | 0.02	0.00 | 0.02	0.82 | 0.19		$\rho = 0.9$
AIC	0.93 | 0.88	0.93 | 0.87	0.94 | 0.83	0.84 | 0.29		$\bar{s}_{C_p}$ = 52.2
BIC	0.06 | 0.02	0.01 | 0.02	0.00 | 0.01	0.09 | 0.03		

Table 14: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **dense** design and $\mathbf{d} = \mathbf{100}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.49 | 0.78	0.38 | 0.72	0.16 | 0.55	0.43 | 0.67	0.40 | 0.73	
CV.min	0.60 | 0.85	0.52 | 0.81	0.27 | 0.65	0.52 | 0.74	0.56 | 0.83	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.48 | 0.77	0.42 | 0.74	0.36 | 0.71	0.47 | 0.70		$\rho = 0$
AIC	0.77 | 0.94	0.77 | 0.93	0.77 | 0.90	0.59 | 0.79		
BIC	0.07 | 0.22	0.08 | 0.36	0.09 | 0.44	0.20 | 0.45		$\bar{s}_{C_p}$ = 191.6
CV.1se	0.62 | 0.75	0.57 | 0.70	0.39 | 0.58	0.59 | 0.55	0.39 | 0.58	
CV.min	0.68 | 0.85	0.63 | 0.81	0.47 | 0.69	0.64 | 0.65	0.45 | 0.66	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.60 | 0.69	0.56 | 0.69	0.46 | 0.68	0.60 | 0.58		$\rho = 0.5$
AIC	0.77 | 0.95	0.77 | 0.95	0.76 | 0.93	0.67 | 0.71		
BIC	0.01 | 0.00	0.00 | 0.00	0.08 | 0.12	0.05 | 0.02		$\bar{s}_{C_p}$ = 191.6
CV.1se	0.67 | 0.74	0.64 | 0.71	0.54 | 0.60	0.58 | 0.13	0.51 | 0.54	
CV.min	0.70 | 0.85	0.67 | 0.82	0.57 | 0.70	0.62 | 0.22	0.54 | 0.62	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.65 | 0.64	0.62 | 0.67	0.55 | 0.68	0.62 | 0.21		$\rho = 0.9$
AIC	0.70 | 0.86	0.69 | 0.86	0.66 | 0.85	0.62 | 0.23		
BIC	0.05 | 0.01	0.01 | 0.01	0.00 | 0.01	0.06 | 0.01		$\bar{s}_{C_p}$ = 190.4
CV.1se	0.41 | 0.47	0.25 | 0.33	0.06 | 0.11	0.51 | 0.56	0.40 | 0.47	
CV.min	0.61 | 0.67	0.47 | 0.54	0.18 | 0.25	0.62 | 0.67	0.61 | 0.67	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.53 | 0.58	0.48 | 0.55	0.48 | 0.50	0.56 | 0.61		$\rho = 0$
AIC	0.84 | 0.92	0.84 | 0.90	0.85 | 0.89	0.74 | 0.80		
BIC	0.00 | 0.01	0.00 | 0.00	0.00 | 0.00	0.06 | 0.09		$\bar{s}_{C_p}$ = 143.2
CV.1se	0.17 | 0.07	0.16 | 0.07	0.01 | 0.00	0.66 | 0.34	0.23 | 0.09	
CV.min	0.61 | 0.39	0.51 | 0.31	0.11 | 0.03	0.71 | 0.50	0.55 | 0.33	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.61 | 0.31	0.52 | 0.34	0.53 | 0.45	0.69 | 0.42		$\rho = 0.5$
AIC	0.84 | 0.94	0.84 | 0.93	0.84 | 0.90	0.78 | 0.74		
BIC	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.06 | 0.01		$\bar{s}_{C_p}$ = 143.0
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.38 | 0.05	0.01 | 0.01	
CV.min	0.48 | 0.02	0.40 | 0.02	0.08 | 0.01	0.71 | 0.18	0.20 | 0.03	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.58 | 0.03	0.12 | 0.01	0.00 | 0.01	0.71 | 0.20		$\rho = 0.9$
AIC	0.81 | 0.87	0.81 | 0.86	0.81 | 0.83	0.73 | 0.29		
BIC	0.06 | 0.01	0.01 | 0.01	0.00 | 0.01	0.07 | 0.01		$\bar{s}_{C_p}$ = 139.3
CV.1se	0.04 | 0.01	0.01 | 0.00	0.00 | 0.00	0.67 | 0.34	0.09 | 0.01	
CV.min	0.53 | 0.23	0.20 | 0.05	0.06 | 0.01	0.79 | 0.53	0.54 | 0.23	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.60 | 0.27	0.14 | 0.06	0.00 | 0.00	0.75 | 0.45		$\rho = 0$
AIC	0.92 | 0.91	0.92 | 0.89	0.92 | 0.90	0.88 | 0.77		
BIC	0.01 | 0.00	0.00 | 0.00	0.08 | 0.07	0.05 | 0.01		$\bar{s}_{C_p}$ = 77.6
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.64 | 0.09	0.25 | 0.00	
CV.min	0.38 | 0.03	0.21 | 0.01	0.08 | 0.00	0.84 | 0.28	0.48 | 0.03	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.57 | 0.06	0.04 | 0.00	0.00 | 0.00	0.83 | 0.23		$\rho = 0.5$
AIC	0.92 | 0.92	0.92 | 0.90	0.92 | 0.90	0.89 | 0.70		
BIC	0.02 | 0.00	0.00 | 0.00	0.00 | 0.00	0.09 | 0.00		$\bar{s}_{C_p}$ = 77.5
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.07 | 0.01	0.05 | 0.01	
CV.min	0.49 | 0.03	0.39 | 0.02	0.13 | 0.01	0.81 | 0.07	0.32 | 0.02	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.64 | 0.04	0.11 | 0.01	0.01 | 0.01	0.82 | 0.08		$\rho = 0.9$
AIC	0.92 | 0.87	0.92 | 0.84	0.92 | 0.82	0.85 | 0.18		
BIC	0.04 | 0.01	0.01 | 0.01	0.00 | 0.00	0.10 | 0.01		$\bar{s}_{C_p}$ = 73.6

Table 15: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **dense** design and $\mathbf{d = 200}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.56 | 0.84	0.46 | 0.75	0.30 | 0.49	0.50 | 0.74	0.58 | 0.86	
CV.min	0.65 | 0.91	0.58 | 0.86	0.43 | 0.67	0.57 | 0.81	0.66 | 0.92	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.50 | 0.77	0.47 | 0.76	0.52 | 0.77	0.50 | 0.74		$\rho = 0$
AIC	0.77 | 0.96	0.76 | 0.95	0.77 | 0.92	0.63 | 0.86		$\bar{s}_{C_p}$ = 199.4
BIC	0.00 | 0.00	0.00 | 0.00	0.35 | 0.50	0.04 | 0.06		
CV.1se	0.61 | 0.68	0.53 | 0.57	0.06 | 0.06	0.60 | 0.47	0.58 | 0.67	
CV.min	0.68 | 0.87	0.63 | 0.82	0.20 | 0.19	0.64 | 0.62	0.69 | 0.88	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.56 | 0.44	0.57 | 0.63	0.56 | 0.76	0.61 | 0.51		$\rho = 0.5$
AIC	0.77 | 0.97	0.76 | 0.97	0.76 | 0.95	0.69 | 0.78		$\bar{s}_{C_p}$ = 199.4
BIC	0.00 | 0.00	0.00 | 0.00	0.02 | 0.02	0.05 | 0.01		
CV.1se	0.23 | 0.23	0.20 | 0.20	0.00 | 0.01	0.34 | 0.05	0.12 | 0.11	
CV.min	0.62 | 0.56	0.54 | 0.50	0.10 | 0.02	0.65 | 0.19	0.40 | 0.34	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.55 | 0.04	0.19 | 0.10	0.60 | 0.74	0.65 | 0.19		$\rho = 0.9$
AIC	0.72 | 0.92	0.71 | 0.92	0.69 | 0.91	0.66 | 0.35		$\bar{s}_{C_p}$ = 199.3
BIC	0.06 | 0.01	0.01 | 0.01	0.00 | 0.00	0.07 | 0.01		
CV.1se	0.38 | 0.30	0.13 | 0.08	0.00 | 0.00	0.52 | 0.48	0.39 | 0.31	
CV.min	0.58 | 0.57	0.37 | 0.29	0.05 | 0.02	0.60 | 0.60	0.58 | 0.58	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.49 | 0.43	0.50 | 0.45	0.39 | 0.33	0.54 | 0.51		$\rho = 0$
AIC	0.80 | 0.91	0.80 | 0.89	0.81 | 0.88	0.72 | 0.77		$\bar{s}_{C_p}$ = 187.3
BIC	0.00 | 0.00	0.00 | 0.00	0.06 | 0.06	0.03 | 0.01		
CV.1se	0.02 | 0.00	0.01 | 0.00	0.00 | 0.00	0.61 | 0.20	0.15 | 0.00	
CV.min	0.41 | 0.09	0.23 | 0.03	0.07 | 0.00	0.68 | 0.38	0.43 | 0.09	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.51 | 0.11	0.12 | 0.04	0.22 | 0.17	0.66 | 0.29		$\rho = 0.5$
AIC	0.80 | 0.93	0.80 | 0.91	0.80 | 0.89	0.75 | 0.72		$\bar{s}_{C_p}$ = 187.3
BIC	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.06 | 0.00		
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.11 | 0.01	0.01 | 0.00	
CV.min	0.46 | 0.02	0.38 | 0.01	0.11 | 0.01	0.69 | 0.08	0.26 | 0.01	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.56 | 0.03	0.12 | 0.01	0.01 | 0.00	0.70 | 0.09		$\rho = 0.9$
AIC	0.78 | 0.87	0.77 | 0.85	0.78 | 0.81	0.72 | 0.34		$\bar{s}_{C_p}$ = 186.1
BIC	0.05 | 0.01	0.01 | 0.00	0.00 | 0.00	0.10 | 0.01		
CV.1se	0.02 | 0.00	0.00 | 0.00	0.00 | 0.00	0.72 | 0.25	0.16 | 0.00	
CV.min	0.50 | 0.12	0.14 | 0.01	0.07 | 0.00	0.81 | 0.43	0.52 | 0.12	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.61 | 0.15	0.08 | 0.03	0.00 | 0.00	0.77 | 0.35		$\rho = 0$
AIC	0.91 | 0.90	0.91 | 0.89	0.91 | 0.91	0.88 | 0.72		$\bar{s}_{C_p}$ = 91.0
BIC	0.01 | 0.00	0.00 | 0.00	0.65 | 0.64	0.08 | 0.00		
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.64 | 0.06	0.28 | 0.00	
CV.min	0.38 | 0.02	0.20 | 0.00	0.08 | 0.00	0.84 | 0.22	0.50 | 0.02	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.58 | 0.04	0.03 | 0.00	0.00 | 0.00	0.83 | 0.18		$\rho = 0.5$
AIC	0.91 | 0.91	0.91 | 0.89	0.91 | 0.90	0.89 | 0.66		$\bar{s}_{C_p}$ = 90.7
BIC	0.02 | 0.00	0.00 | 0.00	0.02 | 0.02	0.11 | 0.00		
CV.1se	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.07 | 0.00	0.19 | 0.00	
CV.min	0.42 | 0.02	0.31 | 0.01	0.13 | 0.01	0.82 | 0.05	0.44 | 0.02	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.59 | 0.03	0.07 | 0.00	0.00 | 0.00	0.83 | 0.06		$\rho = 0.9$
AIC	0.91 | 0.85	0.91 | 0.83	0.91 | 0.82	0.85 | 0.15		$\bar{s}_{C_p}$ = 86.6
BIC	0.02 | 0.00	0.01 | 0.00	0.00 | 0.00	0.10 | 0.01		

Table 16: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **sparse** design and $\mathbf{d = 10}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.36 | 0.74	0.20 | 0.70	0.01 | 0.60	0.23 | 0.65	0.01 | 0.57	
CV.min	0.72 | 0.85	0.64 | 0.82	0.21 | 0.72	0.54 | 0.72	0.20 | 0.71	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.70 | 0.84	0.63 | 0.82	0.36 | 0.76	0.52 | 0.72		$\rho = 0$
AIC	0.95 | 0.95	0.95 | 0.95	0.95 | 0.94	0.53 | 0.72		$\bar{s}_{C_p}$ = 33.4
BIC	0.22 | 0.70	0.13 | 0.69	0.01 | 0.60	0.17 | 0.63		
CV.1se	0.41 | 0.74	0.25 | 0.71	0.02 | 0.60	0.28 | 0.65	0.01 | 0.57	
CV.min	0.74 | 0.84	0.66 | 0.82	0.24 | 0.72	0.56 | 0.72	0.21 | 0.71	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.72 | 0.84	0.65 | 0.82	0.39 | 0.76	0.54 | 0.71		$\rho = 0.5$
AIC	0.95 | 0.96	0.95 | 0.95	0.95 | 0.94	0.55 | 0.71		$\bar{s}_{C_p}$ = 33.3
BIC	0.24 | 0.70	0.15 | 0.68	0.01 | 0.60	0.20 | 0.62		
CV.1se	0.45 | 0.74	0.28 | 0.71	0.03 | 0.60	0.30 | 0.64	0.01 | 0.58	
CV.min	0.75 | 0.84	0.67 | 0.82	0.27 | 0.72	0.57 | 0.71	0.23 | 0.71	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.73 | 0.83	0.66 | 0.81	0.40 | 0.76	0.55 | 0.70		$\rho = 0.9$
AIC	0.95 | 0.96	0.95 | 0.96	0.95 | 0.94	0.56 | 0.71		$\bar{s}_{C_p}$ = 33.1
BIC	0.27 | 0.69	0.17 | 0.68	0.02 | 0.60	0.22 | 0.62		
CV.1se	0.27 | 0.64	0.13 | 0.59	0.01 | 0.47	0.45 | 0.68	0.01 | 0.46	
CV.min	0.72 | 0.79	0.61 | 0.76	0.14 | 0.61	0.81 | 0.81	0.26 | 0.65	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.72 | 0.79	0.63 | 0.77	0.35 | 0.67	0.77 | 0.79		$\rho = 0$
AIC	0.97 | 0.95	0.97 | 0.95	0.97 | 0.93	0.85 | 0.83		$\bar{s}_{C_p}$ = 26.4
BIC	0.19 | 0.61	0.10 | 0.58	0.00 | 0.45	0.16 | 0.57		
CV.1se	0.31 | 0.62	0.16 | 0.58	0.01 | 0.47	0.50 | 0.67	0.01 | 0.46	
CV.min	0.74 | 0.79	0.64 | 0.75	0.15 | 0.60	0.82 | 0.80	0.24 | 0.62	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.73 | 0.78	0.65 | 0.76	0.37 | 0.67	0.78 | 0.78		$\rho = 0.5$
AIC	0.97 | 0.95	0.97 | 0.95	0.97 | 0.93	0.86 | 0.83		$\bar{s}_{C_p}$ = 26.4
BIC	0.21 | 0.59	0.12 | 0.56	0.01 | 0.45	0.19 | 0.56		
CV.1se	0.35 | 0.63	0.19 | 0.58	0.02 | 0.47	0.51 | 0.66	0.01 | 0.46	
CV.min	0.75 | 0.79	0.66 | 0.75	0.18 | 0.60	0.81 | 0.79	0.27 | 0.63	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.74 | 0.79	0.67 | 0.76	0.39 | 0.67	0.79 | 0.77		$\rho = 0.9$
AIC	0.97 | 0.95	0.97 | 0.95	0.96 | 0.93	0.85 | 0.82		$\bar{s}_{C_p}$ = 26.2
BIC	0.23 | 0.59	0.13 | 0.56	0.01 | 0.45	0.21 | 0.56		
CV.1se	0.12 | 0.38	0.06 | 0.33	0.01 | 0.26	0.61 | 0.62	0.01 | 0.27	
CV.min	0.70 | 0.66	0.55 | 0.59	0.14 | 0.43	0.88 | 0.79	0.38 | 0.53	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.73 | 0.69	0.60 | 0.63	0.21 | 0.42	0.85 | 0.76		$\rho = 0$
AIC	0.98 | 0.94	0.98 | 0.94	0.98 | 0.91	0.94 | 0.88		$\bar{s}_{C_p}$ = 19.7
BIC	0.13 | 0.40	0.06 | 0.36	0.00 | 0.18	0.17 | 0.43		
CV.1se	0.14 | 0.36	0.07 | 0.33	0.01 | 0.26	0.66 | 0.62	0.01 | 0.26	
CV.min	0.72 | 0.66	0.59 | 0.60	0.14 | 0.42	0.89 | 0.79	0.36 | 0.52	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.75 | 0.68	0.64 | 0.64	0.25 | 0.45	0.86 | 0.75		$\rho = 0.5$
AIC	0.98 | 0.95	0.98 | 0.94	0.98 | 0.92	0.95 | 0.88		$\bar{s}_{C_p}$ = 19.4
BIC	0.13 | 0.37	0.07 | 0.34	0.00 | 0.18	0.19 | 0.41		
CV.1se	0.16 | 0.35	0.09 | 0.32	0.02 | 0.24	0.65 | 0.59	0.02 | 0.25	
CV.min	0.74 | 0.65	0.61 | 0.58	0.16 | 0.40	0.89 | 0.76	0.37 | 0.50	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.76 | 0.66	0.66 | 0.61	0.28 | 0.44	0.86 | 0.73		$\rho = 0.9$
AIC	0.98 | 0.95	0.98 | 0.94	0.98 | 0.92	0.94 | 0.86		$\bar{s}_{C_p}$ = 19.7
BIC	0.14 | 0.35	0.07 | 0.32	0.00 | 0.17	0.21 | 0.40		

Table 17: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **sparse** design and $\mathbf{d = 50}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.49 | 0.76	0.36 | 0.71	0.07 | 0.56	0.43 | 0.66	0.15 | 0.62	
CV.min	0.64 | 0.83	0.56 | 0.80	0.19 | 0.65	0.57 | 0.74	0.39 | 0.74	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.55 | 0.79	0.48 | 0.77	0.43 | 0.75	0.53 | 0.71		$\rho = 0$
AIC	0.84 | 0.94	0.84 | 0.94	0.84 | 0.91	0.63 | 0.77		$\bar{s}_{C_p}$ = 124.2
BIC	0.22 | 0.59	0.13 | 0.56	0.02 | 0.47	0.22 | 0.54		
CV.1se	0.52 | 0.76	0.39 | 0.71	0.08 | 0.56	0.47 | 0.66	0.13 | 0.60	
CV.min	0.66 | 0.84	0.58 | 0.80	0.21 | 0.65	0.59 | 0.73	0.37 | 0.72	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.57 | 0.78	0.50 | 0.77	0.44 | 0.75	0.55 | 0.71		$\rho = 0.5$
AIC	0.84 | 0.94	0.84 | 0.94	0.84 | 0.92	0.65 | 0.77		$\bar{s}_{C_p}$ = 123.8
BIC	0.23 | 0.57	0.15 | 0.55	0.02 | 0.47	0.25 | 0.53		
CV.1se	0.54 | 0.76	0.41 | 0.71	0.10 | 0.56	0.48 | 0.65	0.16 | 0.60	
CV.min	0.67 | 0.84	0.59 | 0.80	0.24 | 0.65	0.60 | 0.73	0.39 | 0.73	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.58 | 0.78	0.51 | 0.77	0.45 | 0.75	0.56 | 0.70		$\rho = 0.9$
AIC	0.84 | 0.94	0.84 | 0.94	0.84 | 0.91	0.65 | 0.76		$\bar{s}_{C_p}$ = 123.5
BIC	0.25 | 0.56	0.16 | 0.55	0.03 | 0.47	0.27 | 0.52		
CV.1se	0.43 | 0.58	0.27 | 0.50	0.07 | 0.34	0.52 | 0.62	0.26 | 0.50	
CV.min	0.66 | 0.73	0.54 | 0.66	0.20 | 0.46	0.68 | 0.73	0.57 | 0.68	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.60 | 0.69	0.55 | 0.67	0.62 | 0.69	0.62 | 0.69		$\rho = 0$
AIC	0.90 | 0.93	0.90 | 0.92	0.90 | 0.90	0.79 | 0.82		$\bar{s}_{C_p}$ = 90.2
BIC	0.08 | 0.24	0.06 | 0.29	0.01 | 0.12	0.17 | 0.38		
CV.1se	0.46 | 0.57	0.31 | 0.49	0.08 | 0.34	0.55 | 0.61	0.23 | 0.46	
CV.min	0.67 | 0.73	0.56 | 0.66	0.21 | 0.46	0.69 | 0.72	0.55 | 0.66	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.61 | 0.68	0.56 | 0.66	0.62 | 0.69	0.64 | 0.68		$\rho = 0.5$
AIC	0.90 | 0.94	0.90 | 0.93	0.90 | 0.90	0.80 | 0.82		$\bar{s}_{C_p}$ = 90.2
BIC	0.06 | 0.17	0.07 | 0.26	0.01 | 0.12	0.18 | 0.35		
CV.1se	0.48 | 0.57	0.34 | 0.50	0.10 | 0.33	0.56 | 0.61	0.26 | 0.47	
CV.min	0.68 | 0.73	0.58 | 0.67	0.24 | 0.46	0.69 | 0.71	0.56 | 0.66	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.62 | 0.69	0.57 | 0.67	0.63 | 0.69	0.65 | 0.68		$\rho = 0.9$
AIC	0.90 | 0.94	0.90 | 0.93	0.90 | 0.90	0.80 | 0.81		$\bar{s}_{C_p}$ = 89.2
BIC	0.06 | 0.16	0.08 | 0.24	0.01 | 0.10	0.20 | 0.34		
CV.1se	0.08 | 0.06	0.03 | 0.02	0.00 | 0.01	0.65 | 0.44	0.08 | 0.05	
CV.min	0.61 | 0.41	0.34 | 0.22	0.07 | 0.05	0.80 | 0.63	0.59 | 0.40	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.64 | 0.43	0.62 | 0.43	0.22 | 0.16	0.76 | 0.57		$\rho = 0$
AIC	0.94 | 0.92	0.94 | 0.90	0.94 | 0.90	0.90 | 0.81		$\bar{s}_{C_p}$ = 56.2
BIC	0.01 | 0.01	0.00 | 0.01	0.00 | 0.00	0.06 | 0.04		
CV.1se	0.07 | 0.04	0.03 | 0.02	0.00 | 0.00	0.67 | 0.43	0.09 | 0.03	
CV.min	0.60 | 0.36	0.34 | 0.19	0.08 | 0.04	0.81 | 0.61	0.58 | 0.35	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.64 | 0.39	0.62 | 0.41	0.29 | 0.20	0.77 | 0.55		$\rho = 0.5$
AIC	0.94 | 0.92	0.94 | 0.91	0.94 | 0.90	0.90 | 0.81		$\bar{s}_{C_p}$ = 56.2
BIC	0.01 | 0.01	0.01 | 0.01	0.00 | 0.00	0.06 | 0.03		
CV.1se	0.07 | 0.03	0.03 | 0.01	0.00 | 0.00	0.67 | 0.41	0.07 | 0.03	
CV.min	0.61 | 0.36	0.37 | 0.18	0.09 | 0.04	0.81 | 0.60	0.58 | 0.34	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.66 | 0.39	0.64 | 0.40	0.26 | 0.18	0.77 | 0.54		$\rho = 0.9$
AIC	0.94 | 0.92	0.94 | 0.91	0.94 | 0.90	0.90 | 0.80		$\bar{s}_{C_p}$ = 56.5
BIC	0.01 | 0.01	0.01 | 0.01	0.00 | 0.00	0.07 | 0.03		

Table 18: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **sparse** design and $\mathbf{d = 100}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.49 | 0.77	0.38 | 0.72	0.16 | 0.56	0.43 | 0.67	0.40 | 0.74	
CV.min	0.60 | 0.85	0.52 | 0.81	0.27 | 0.66	0.52 | 0.74	0.56 | 0.83	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.48 | 0.77	0.43 | 0.75	0.42 | 0.74	0.47 | 0.70		$\rho = 0$
AIC	0.77 | 0.94	0.77 | 0.93	0.77 | 0.90	0.59 | 0.79		$\bar{s}_{C_p}$ = 191.7
BIC	0.07 | 0.21	0.10 | 0.41	0.09 | 0.46	0.19 | 0.45		
CV.1se	0.51 | 0.77	0.41 | 0.72	0.17 | 0.56	0.45 | 0.67	0.38 | 0.71	
CV.min	0.61 | 0.85	0.54 | 0.81	0.29 | 0.66	0.54 | 0.73	0.54 | 0.81	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.49 | 0.76	0.44 | 0.74	0.42 | 0.74	0.49 | 0.69		$\rho = 0.5$
AIC	0.77 | 0.94	0.77 | 0.94	0.77 | 0.91	0.60 | 0.79		$\bar{s}_{C_p}$ = 192.1
BIC	0.04 | 0.10	0.11 | 0.37	0.10 | 0.46	0.21 | 0.42		
CV.1se	0.52 | 0.77	0.42 | 0.72	0.19 | 0.56	0.46 | 0.66	0.40 | 0.72	
CV.min	0.61 | 0.85	0.54 | 0.81	0.31 | 0.66	0.54 | 0.73	0.55 | 0.82	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.50 | 0.76	0.45 | 0.74	0.43 | 0.74	0.49 | 0.68		$\rho = 0.9$
AIC	0.77 | 0.94	0.77 | 0.94	0.77 | 0.91	0.61 | 0.78		$\bar{s}_{C_p}$ = 191.7
BIC	0.03 | 0.08	0.12 | 0.35	0.12 | 0.47	0.22 | 0.40		
CV.1se	0.41 | 0.47	0.25 | 0.33	0.06 | 0.10	0.51 | 0.56	0.40 | 0.47	
CV.min	0.60 | 0.67	0.47 | 0.54	0.18 | 0.24	0.62 | 0.67	0.60 | 0.67	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.53 | 0.59	0.51 | 0.58	0.61 | 0.61	0.56 | 0.61		$\rho = 0$
AIC	0.84 | 0.92	0.84 | 0.91	0.85 | 0.89	0.74 | 0.80		$\bar{s}_{C_p}$ = 143.0
BIC	0.00 | 0.01	0.01 | 0.02	0.00 | 0.00	0.05 | 0.09		
CV.1se	0.42 | 0.45	0.27 | 0.32	0.05 | 0.09	0.53 | 0.54	0.41 | 0.44	
CV.min	0.62 | 0.66	0.49 | 0.54	0.17 | 0.22	0.63 | 0.66	0.61 | 0.66	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.54 | 0.57	0.52 | 0.57	0.61 | 0.61	0.58 | 0.60		$\rho = 0.5$
AIC	0.84 | 0.93	0.84 | 0.91	0.85 | 0.89	0.75 | 0.79		$\bar{s}_{C_p}$ = 143.7
BIC	0.00 | 0.01	0.01 | 0.01	0.00 | 0.00	0.05 | 0.06		
CV.1se	0.44 | 0.45	0.30 | 0.32	0.06 | 0.08	0.54 | 0.53	0.42 | 0.43	
CV.min	0.62 | 0.66	0.51 | 0.54	0.19 | 0.22	0.63 | 0.65	0.62 | 0.66	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.55 | 0.56	0.53 | 0.57	0.61 | 0.61	0.59 | 0.59		$\rho = 0.9$
AIC	0.84 | 0.93	0.84 | 0.91	0.85 | 0.89	0.75 | 0.79		$\bar{s}_{C_p}$ = 143.5
BIC	0.00 | 0.00	0.00 | 0.01	0.00 | 0.00	0.05 | 0.06		
CV.1se	0.04 | 0.01	0.01 | 0.00	0.00 | 0.00	0.67 | 0.33	0.13 | 0.01	
CV.min	0.54 | 0.23	0.20 | 0.05	0.06 | 0.01	0.79 | 0.53	0.54 | 0.23	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.60 | 0.27	0.64 | 0.33	0.14 | 0.09	0.75 | 0.44		$\rho = 0$
AIC	0.92 | 0.91	0.92 | 0.89	0.92 | 0.90	0.88 | 0.76		$\bar{s}_{C_p}$ = 77.9
BIC	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.05 | 0.01		
CV.1se	0.03 | 0.01	0.00 | 0.00	0.00 | 0.00	0.68 | 0.31	0.14 | 0.01	
CV.min	0.51 | 0.18	0.18 | 0.04	0.06 | 0.01	0.80 | 0.51	0.50 | 0.17	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.59 | 0.23	0.61 | 0.29	0.17 | 0.10	0.76 | 0.43		$\rho = 0.5$
AIC	0.92 | 0.92	0.92 | 0.90	0.92 | 0.90	0.89 | 0.76		$\bar{s}_{C_p}$ = 78.2
BIC	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.06 | 0.01		
CV.1se	0.02 | 0.01	0.01 | 0.00	0.00 | 0.00	0.69 | 0.29	0.15 | 0.01	
CV.min	0.51 | 0.17	0.19 | 0.04	0.07 | 0.01	0.80 | 0.50	0.52 | 0.17	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.60 | 0.23	0.62 | 0.29	0.16 | 0.09	0.76 | 0.41		$\rho = 0.9$
AIC	0.92 | 0.91	0.92 | 0.89	0.92 | 0.90	0.89 | 0.76		$\bar{s}_{C_p}$ = 77.5
BIC	0.01 | 0.00	0.01 | 0.00	0.00 | 0.00	0.07 | 0.01		

Table 19: *False Discovery Rate* | *Sensitivity*, relative to C_p oracle, for **sparse** design and $\mathbf{d = 200}$.

	lasso	GL $\gamma = 1$	GL $\gamma = 10$	marginal AL	sparsenet MCP	
CV.1se	0.56 | 0.84	0.46 | 0.75	0.31 | 0.50	0.50 | 0.74	0.57 | 0.86	
CV.min	0.64 | 0.91	0.58 | 0.86	0.43 | 0.68	0.57 | 0.81	0.66 | 0.92	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.50 | 0.77	0.48 | 0.77	0.53 | 0.78	0.50 | 0.74		$\rho = 0$
AIC	0.77 | 0.96	0.76 | 0.95	0.77 | 0.92	0.63 | 0.86		$\bar{s}_{C_p}$ = 199.4
BIC	0.00 | 0.00	0.00 | 0.01	0.17 | 0.28	0.04 | 0.06		
CV.1se	0.57 | 0.83	0.48 | 0.74	0.31 | 0.49	0.52 | 0.72	0.58 | 0.85	
CV.min	0.65 | 0.91	0.58 | 0.86	0.43 | 0.67	0.58 | 0.80	0.67 | 0.93	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.51 | 0.75	0.49 | 0.77	0.53 | 0.79	0.52 | 0.73		$\rho = 0.5$
AIC	0.77 | 0.97	0.77 | 0.95	0.77 | 0.93	0.64 | 0.86		$\bar{s}_{C_p}$ = 199.4
BIC	0.00 | 0.00	0.00 | 0.00	0.14 | 0.23	0.04 | 0.04		
CV.1se	0.57 | 0.83	0.48 | 0.74	0.31 | 0.47	0.52 | 0.71	0.58 | 0.85	
CV.min	0.65 | 0.91	0.59 | 0.86	0.44 | 0.66	0.58 | 0.79	0.67 | 0.92	$\mathrm{sd}(\mu)/\sigma = 2$
AICc	0.51 | 0.75	0.49 | 0.76	0.53 | 0.78	0.52 | 0.71		$\rho = 0.9$
AIC	0.77 | 0.96	0.76 | 0.95	0.77 | 0.93	0.64 | 0.85		$\bar{s}_{C_p}$ = 199.4
BIC	0.00 | 0.00	0.00 | 0.00	0.14 | 0.22	0.04 | 0.03		
CV.1se	0.38 | 0.30	0.14 | 0.09	0.00 | 0.00	0.52 | 0.47	0.39 | 0.32	
CV.min	0.58 | 0.57	0.38 | 0.29	0.06 | 0.02	0.60 | 0.60	0.58 | 0.58	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.49 | 0.44	0.53 | 0.48	0.63 | 0.53	0.54 | 0.51		$\rho = 0$
AIC	0.80 | 0.91	0.80 | 0.89	0.81 | 0.88	0.72 | 0.77		$\bar{s}_{C_p}$ = 187.6
BIC	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.03 | 0.01		
CV.1se	0.36 | 0.24	0.10 | 0.05	0.00 | 0.00	0.54 | 0.46	0.37 | 0.25	
CV.min	0.58 | 0.54	0.36 | 0.25	0.05 | 0.02	0.61 | 0.59	0.58 | 0.55	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.50 | 0.41	0.54 | 0.47	0.63 | 0.54	0.56 | 0.49		$\rho = 0.5$
AIC	0.80 | 0.92	0.80 | 0.90	0.81 | 0.88	0.72 | 0.77		$\bar{s}_{C_p}$ = 187.8
BIC	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.03 | 0.01		
CV.1se	0.37 | 0.24	0.11 | 0.06	0.00 | 0.00	0.54 | 0.44	0.37 | 0.24	
CV.min	0.58 | 0.54	0.38 | 0.26	0.04 | 0.01	0.61 | 0.58	0.59 | 0.54	$\mathrm{sd}(\mu)/\sigma = 1$
AICc	0.51 | 0.40	0.54 | 0.46	0.64 | 0.53	0.56 | 0.48		$\rho = 0.9$
AIC	0.80 | 0.91	0.80 | 0.89	0.81 | 0.88	0.72 | 0.76		$\bar{s}_{C_p}$ = 187.5
BIC	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00	0.03 | 0.01		
CV.1se	0.02 | 0.00	0.00 | 0.00	0.00 | 0.00	0.72 | 0.24	0.22 | 0.01	
CV.min	0.51 | 0.12	0.16 | 0.01	0.07 | 0.00	0.81 | 0.43	0.53 | 0.12	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.60 | 0.15	0.68 | 0.27	0.07 | 0.03	0.78 | 0.35		$\rho = 0$
AIC	0.91 | 0.91	0.91 | 0.89	0.91 | 0.91	0.88 | 0.71		$\bar{s}_{C_p}$ = 90.6
BIC	0.01 | 0.00	0.00 | 0.00	0.02 | 0.02	0.09 | 0.00		
CV.1se	0.02 | 0.00	0.00 | 0.00	0.00 | 0.00	0.73 | 0.23	0.25 | 0.00	
CV.min	0.48 | 0.09	0.16 | 0.01	0.06 | 0.00	0.81 | 0.41	0.52 | 0.09	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.59 | 0.12	0.65 | 0.24	0.07 | 0.04	0.78 | 0.33		$\rho = 0.5$
AIC	0.91 | 0.91	0.91 | 0.89	0.91 | 0.91	0.88 | 0.71		$\bar{s}_{C_p}$ = 91.1
BIC	0.01 | 0.00	0.00 | 0.00	0.01 | 0.01	0.07 | 0.00		
CV.1se	0.01 | 0.00	0.00 | 0.00	0.00 | 0.00	0.72 | 0.21	0.23 | 0.00	
CV.min	0.47 | 0.09	0.16 | 0.01	0.06 | 0.00	0.81 | 0.40	0.51 | 0.08	$\mathrm{sd}(\mu)/\sigma = 0.5$
AICc	0.58 | 0.12	0.65 | 0.23	0.07 | 0.04	0.79 | 0.32		$\rho = 0.9$
AIC	0.91 | 0.90	0.91 | 0.89	0.91 | 0.91	0.88 | 0.71		$\bar{s}_{C_p}$ = 89.9
BIC	0.01 | 0.00	0.00 | 0.00	0.01 | 0.01	0.09 | 0.00		