FARMER SCHOOL OF BUSINESS

Comparing Supersaturated Screening Designs using Exact Screening Probabilities

Maria Weese
Spring Research Conference 2023
Banff, Alberta Canada

Joint work with:

Kade Young, NC State

Jon Stallrich, NC State

Byran Smucker, Miami University

David Edwards, VCU

Screening Experiments

Small n to determine which of p factors drive response

Factors take on two levels: ± 1

Focus on analysis based on main effect model

$$y = \beta_0 + X\beta + e \qquad e \sim^{iid} N(0, \sigma^2)$$

Assumptions:

- Factor sparsity ($k < p$ factors important)
- For active effects, $\beta_j \gg \sigma$
- Negligible interactions

Screening Design and Analysis Approach

$n \geq p + 1$ (Least Squares ☺):

- Orthogonal designs ($X^T X = nI$) minimize variance
- Least-squares backwards elimination possible

$n < p + 1$ (No Least Squares ☹):

- Supersaturated case
- Find design with $X^T X \approx nI$
- Often a regularization method

Heuristic Supersaturated Optimality Measures

Define: $S = X^T X = (s_{ij})$ (sometimes includes intercept)

1. $\mathrm{E}(s^2)$ or $\mathrm{UE}(s^2)$: minimize the average of s_{ij}^2

2. $\mathrm{Var}(s+)$: minimize $\mathrm{Var}(s_{ij})$ average $s_{ij} > 0$. These designs perform well when effect directions are known.

3. Pareto Efficient Designs (PED): choose designs on pareto front of two Gauss-Dantzig heuristic criteria. This is the first design construction method to connect to penalized estimation (Singh and Stufken (2022)).

Gauss-Dantzig is a Typical Analysis Method for SSD

Dantzig estimates are the solution to:

$$\min_{\hat{\boldsymbol{\beta}} \in \mathbb{R}^k} \|\hat{\boldsymbol{\beta}}\|_1 \quad \text{subject to } \|\mathbf{X}^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\|_\infty \leq \delta$$

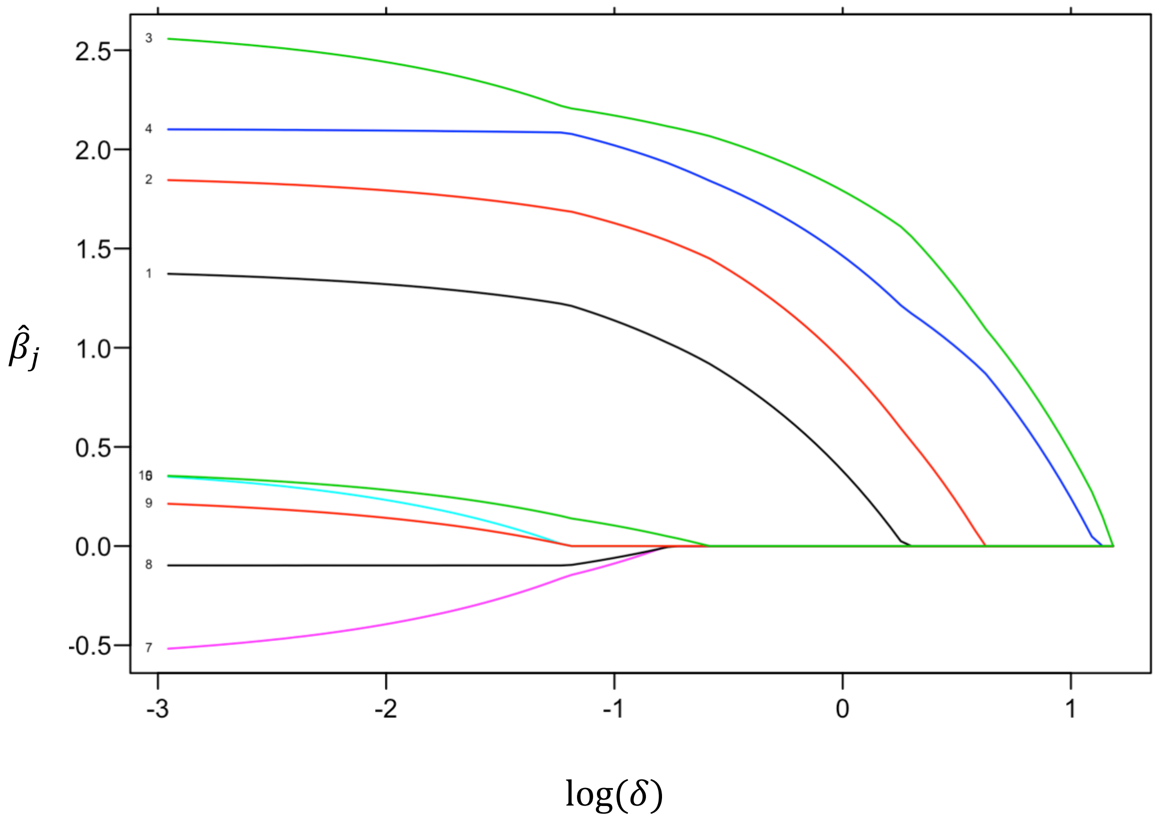
Candes and Tao (2007) developed a two-stage approach which Phoa et al. (2009) used to analyze SSDs.

1. Apply a threshold value to the Dantzig estimates at each value of δ.
2. Calculates the value of some information criterion using least squares for each set of active factors.

We will refer to this two-stage procedure as the Gauss-Dantzig.

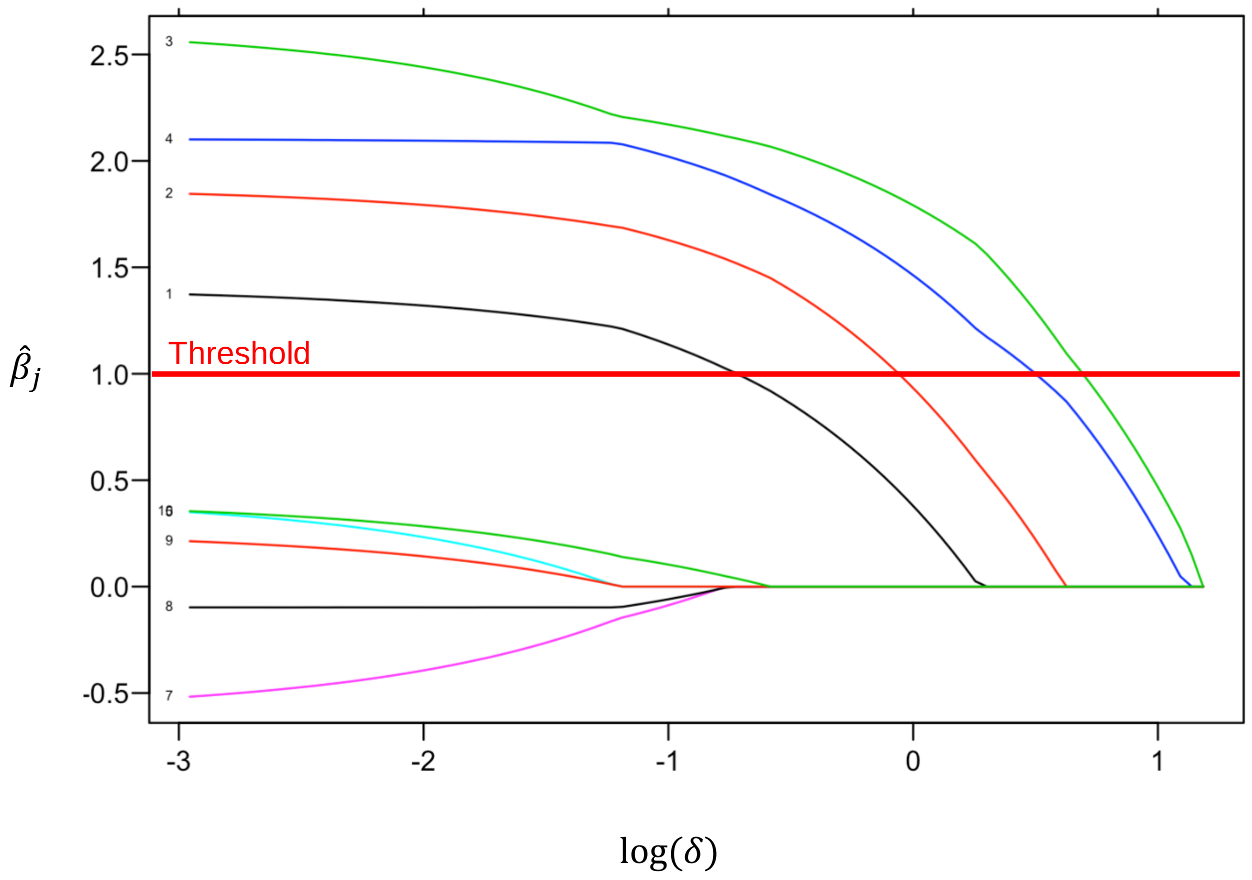
Gauss-Dantzig Selector Step 1

Solve the Dantzig selector along grid of δ (solution path).



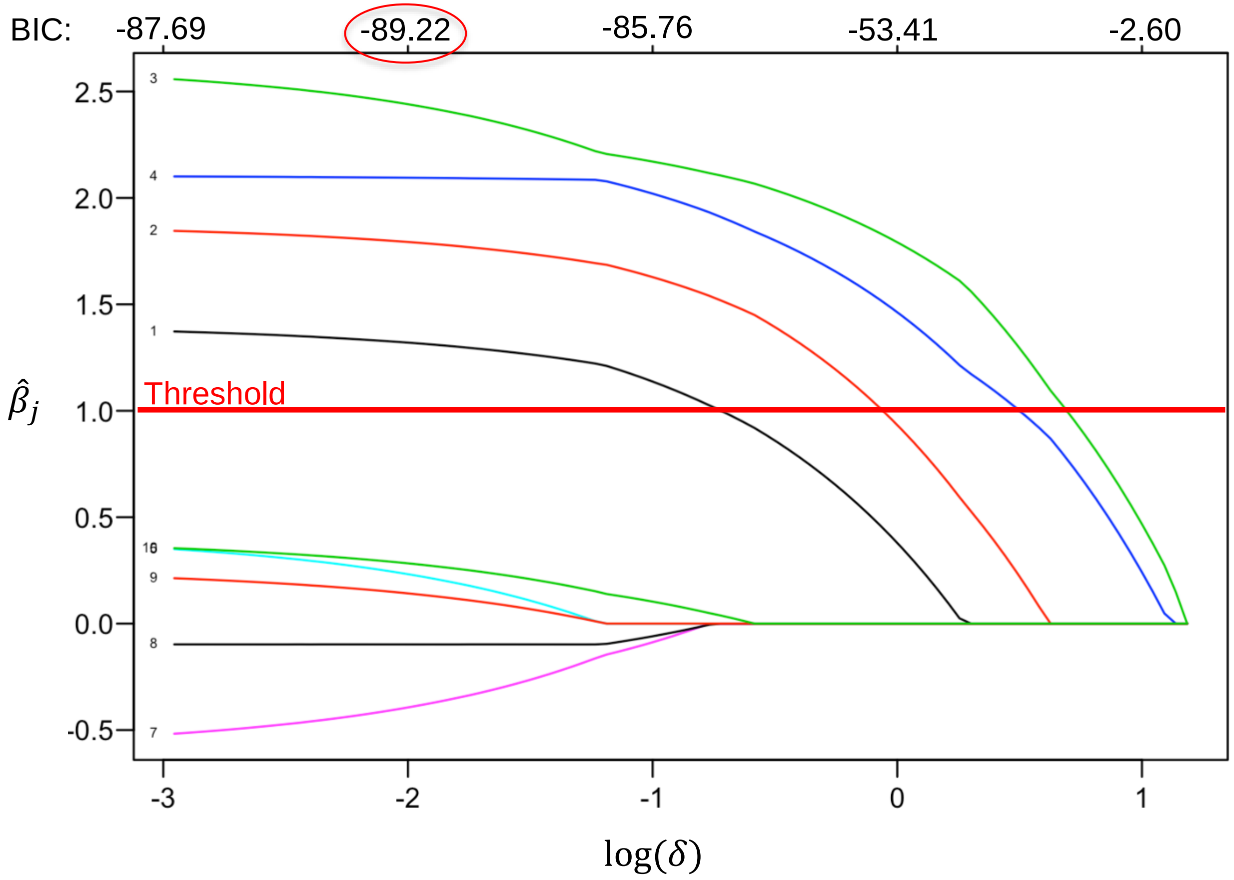
Gauss-Dantzig Selector Step 2

For each δ, apply a threshold the Dantzig estimates to find active factors.



Gauss-Dantzig Selector step 3

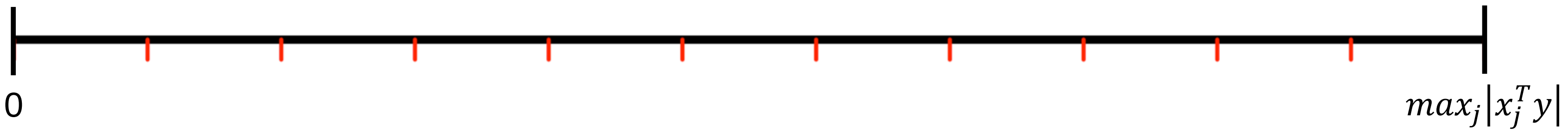
Fit OLS on active factors for each δ and apply information criteria on OLS fits to select δ.



Potential Variations: Tuning Grid Density

The density of grid of δ exploring the tuning parameter space impacts selection.

Coarse Grid: 10 evenly spaced values excluding endpoints

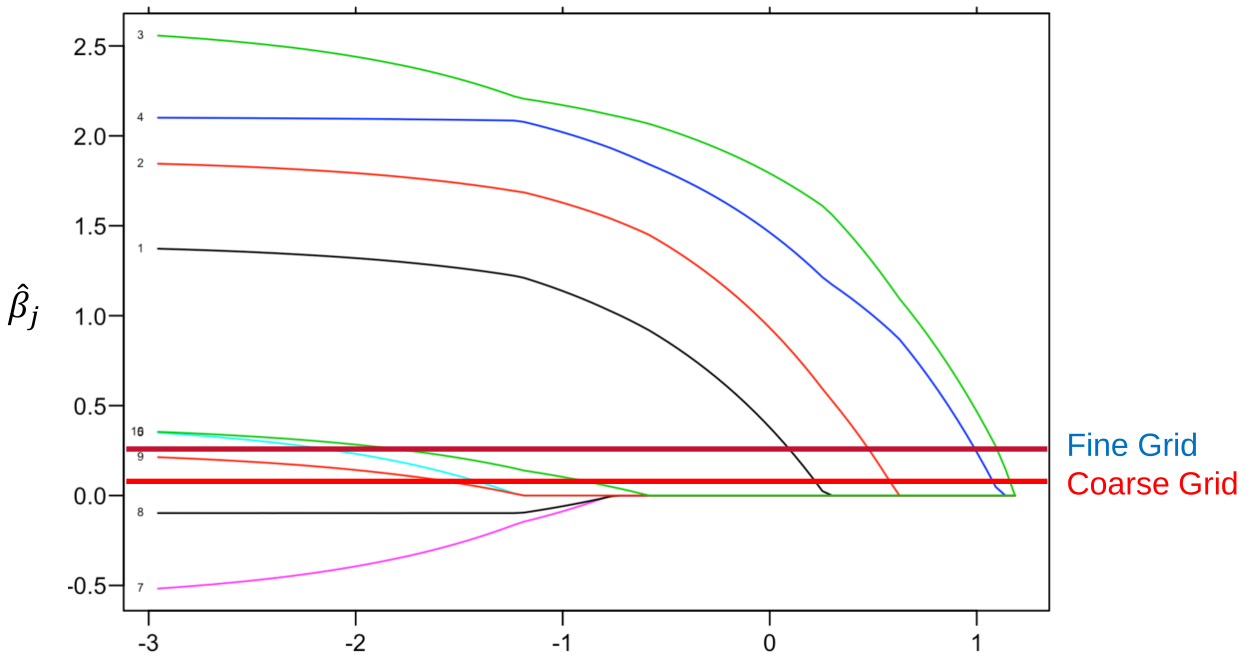


Fine Grid: 100 evenly spaced values including endpoints

Potential Variations: Threshold Value

Without knowledge of σ, data driven thresholds are used:

- Threshold = $0.1 \times max_j|\hat{\beta}_j|$ (the value of 0.1 can be changed)
- This maximum occurs at the smallest δ in the solution path



Demonstration of Impact of Potential Variations

We compare PED and $Var(s+)$ designs with the Gauss-Dantzig Selector via simulation.

- Use the data driven threshold and standard BIC statistic to select δ
- Comparison over 6 scenarios of differing signal-to-noise ratios (SN) and different numbers of active effects

We compare coarse vs. fine δ grids.

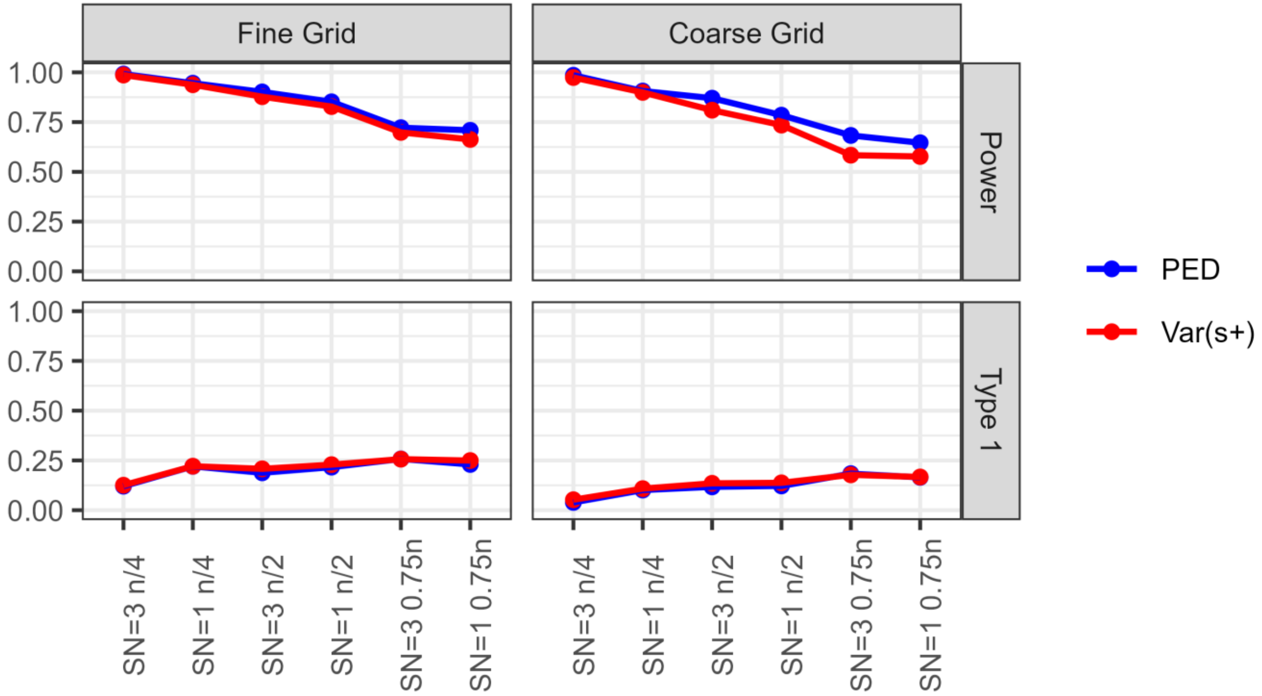
Power: Proportion of truly active factors selected as potentially active.

Type 1 Error: Proportion of truly inactive effects selected as potentially active.

Simulation Comparison

Scenario: n=14, p=24 with effect signs known

- Coarse δ grid shows an increase in power for the PED over Var(s+).
- Fine δ grid does not show the same increase.



Simulation to Study Inconsistencies

Used designs from Marley and Woods (2010)

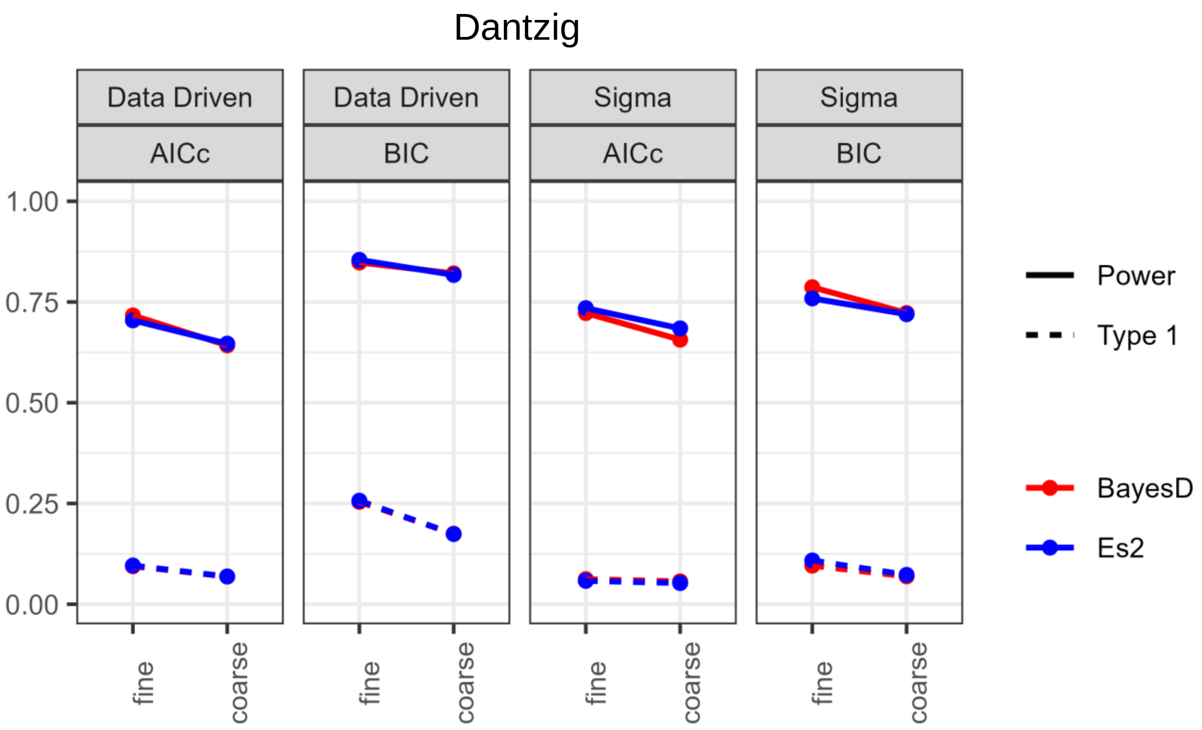
- Compare $E(s^2)$- and Bayes D- optimal designs
- Active effects were generated from $N(\mu, 0.2)$

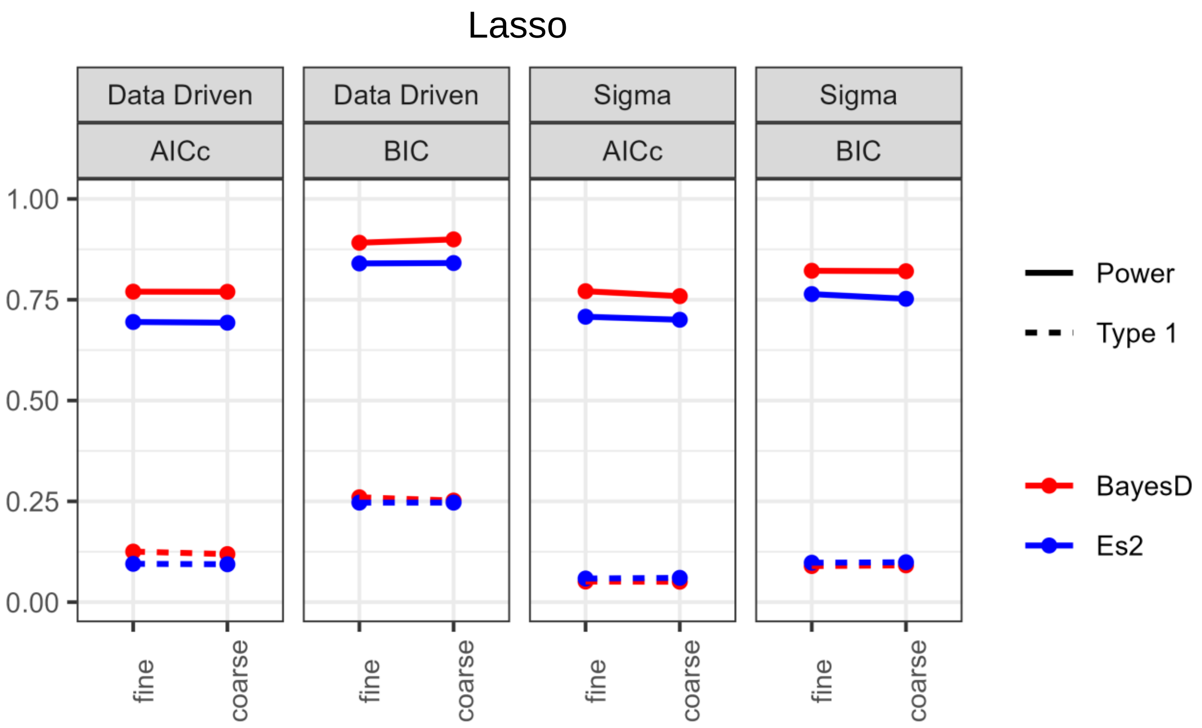
Studied the simulation results comparing following settings:

- Threshold: data driven vs. σ
- Information Criteria: BIC vs. AICc
- Tuning Grid: Fine Grid vs. Coarse Grid
- Dantzig vs. lasso

Simulation to study Inconsistencies

Scenario: n=14, k=24, μ=3 and c=6 with mixed signs





Type 1 Error and Power are sensitive to the analysis method settings.

A Different Approach

Needed: A way to evaluate and compare designs based on exact screening probabilities that is not reliant on simulation decisions.

Under the lasso, exact sign recovery probabilities can be simulated or calculated via a closed form expression developed in Stallrich et al. 2023.

arXiv > math > arXiv:2303.16843

Mathematics > Statistics Theory

[Submitted on 29 Mar 2023]

Optimal Supersaturated Designs for Lasso Sign Recovery

Jonathan W. Stallrich, Kade Young, Maria L. Weese, Byran J. Smucker, David J. Edwards

Supersaturated designs, in which the number of factors exceeds the number of runs, are often constructed under a heuristic criterion that measures a design's proximity to an unattainable orthogonal design. Such a criterion does not directly measure a design's quality in terms of screening. To address this disconnect, we develop optimality criteria to maximize the lasso's sign recovery probability. The criteria have varying amounts of prior knowledge about the model's parameters. We show that an orthogonal design is an ideal structure when the signs of the active factors are unknown. When the signs are assumed known, we show that a design whose columns exhibit small, positive correlations are ideal. Such designs are sought after by the Var(s+)-criterion. These conclusions are based on a continuous optimization framework, which rigorously justifies the use of established heuristic criteria. From this justification, we propose a computationally-efficient design search algorithm that filters through optimal designs under different heuristic criteria to select the one that maximizes the sign recovery probability under the lasso.

A Different Approach

Support Recovery:

- All truly active effects are estimated as non-zero
- All truly inactive effects are estimated as zero

Sign Recovery:

- All truly active effects are estimated as non-zero with the correct sign
- All truly inactive effects are estimated as zero

Lasso sign recovery probabilities are mathematically tractable.

Lasso Transformations

Data generated under model $y = \beta_0 + X\beta + e$

- $A =$ true active set of β $I = A^c$ = inactive set of β

Analysis of transformed model $y^* = X^*\beta^* + e^*$

- y^* =center y
- Center columns of X (remove intercept)
- $V = D\left(n^{-1}\left|\left|x_j - \bar{x}_j\right|\right|_2^2\right) \leq I$
- Use V to scale columns so $Diag\left(\frac{1}{n}X^{*T}X^*\right) = 1$
- $\beta^* = V^{1/2}\beta$ is design dependent

A closed-form expression for Lasso Solution

Lasso Estimate: $\hat{\beta}^* = \text{argmin}_\beta \frac{1}{2n}(y^* - X^*\beta)^T(y^* - X^*\beta) + \lambda||\beta||_1$

More notation

- X_A^* is X^* but only columns in A
- $C_{AA} = \frac{1}{n} {X^*}_A^T X_A^*$ and likewise C_{AI}
- $z_A =$ sign vector of β_A

If you have a known lasso support $\hat{A}$ and sign $z_{\hat{A}}$:

$$\hat{\beta}_{\hat{A}}^* = \frac{1}{n} C_{\hat{A}\hat{A}}^{-1} X_{\hat{A}}^{*T} y - \lambda C_{\hat{A}\hat{A}}^{-1} z_{\hat{A}} \qquad \hat{\beta}_{\hat{I}}^* = 0$$

Lasso Sign Recovery

Sign recovery occurs when two events of normal random variables hold:

$$I_\lambda | X, z_A = \{ |\boldsymbol{v}| < \lambda\sqrt{n} \}$$

$$\mu_v = \lambda\sqrt{n} C_{IA} C_{AA}^{-1} z_A \qquad \Sigma_v = \sigma^2 (C_{II} - C_{IA} C_{AA}^{-1} C_{AI})$$

$$S_\lambda | X, \beta_A = \left\{ \boldsymbol{u} < \sqrt{n}\, V_A^{\frac{1}{2}} |\beta_A| - \lambda\sqrt{n}\, D(C_{AA}^{-1} z_A z_A^T) \right\}$$

$$\mu_u = 0 \qquad \Sigma_u = \sigma^2 D(z_A) C_{AA}^{-1} D(z_A)$$

- $\phi_\lambda(X|\beta) = P(\hat{z}_A = z_A \,|X, \beta_A) = P[I_\lambda \cap S_\lambda \,|X, \beta_A]$
  $= P[I_\lambda | X, z_A] \times P[S_\lambda | X, \beta_A]$

A Relaxation of Local Assumptions

- Calculation of $\phi_\lambda(X|\beta)$ requires known A, and β

- When A is unknown but z_A is known:
  - For a given number of active effects k, let $\Phi_\lambda(X|k,\beta)$ be the average $\phi_\lambda(X|\beta)$ over all possible A

- When both A and z_A are unknown:
  - For a given number of active effects k, let $\Phi_\lambda^{\pm}(X|k,\beta)$ be the average $\phi_\lambda(X|\beta)$ over all possible A and z_A

A Different Approach

Given number of active effects k and tuning parameter λ :

- $\Phi_\lambda(X|k,\beta)$ is the average sign recovery probability over all possible supports.
- $\Phi_\lambda^\pm(X|k,\beta)$ is the average sign recovery probability over all possible supports and effect directions.

Plot $\Phi_\lambda(X|k,\beta)$ or $\Phi_\lambda^\pm(X|k,\beta)$ over a range of $\log(\lambda)$

- Looking for design with larger Φ_λ over larger $\log(\lambda)$ range

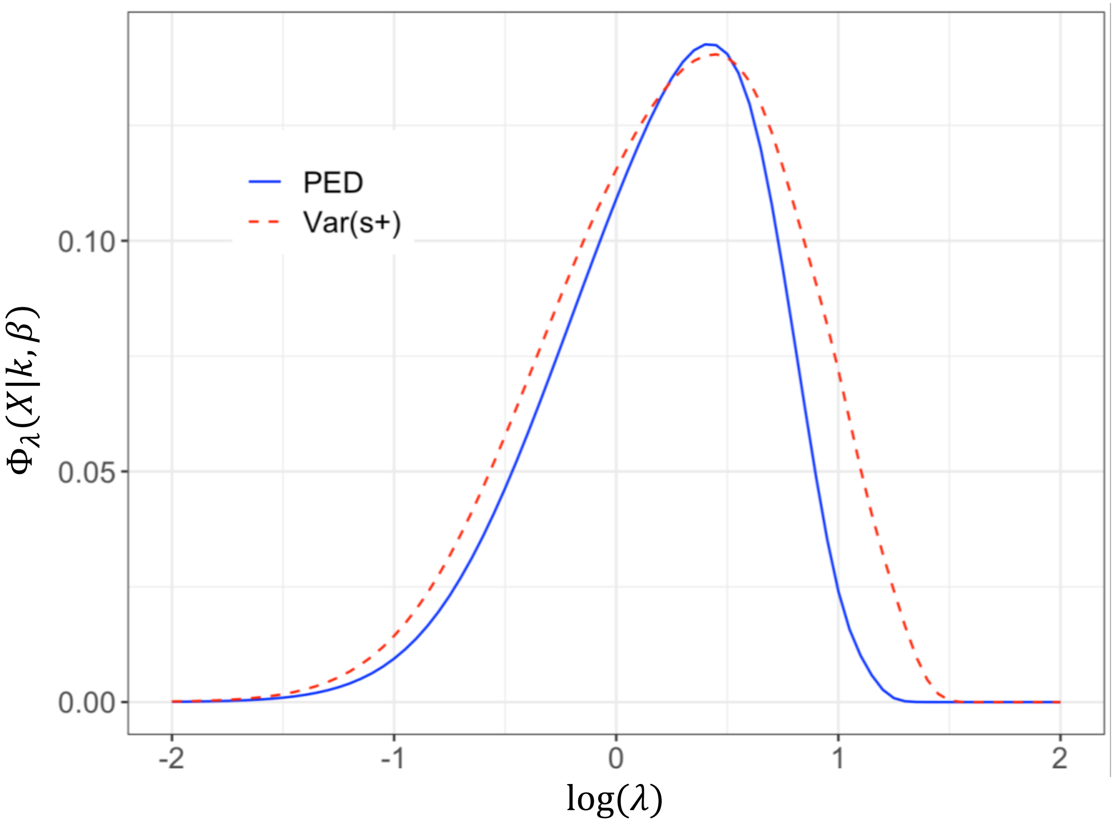
Allows for a more complete comparison of designs robust to tuning parameter value

If scalar measures are desired we recommend either:

- Integrating Φ_λ over the range of $\log(\lambda)$
- Taking the maximum Φ_λ over $\log(\lambda)$

Example 1

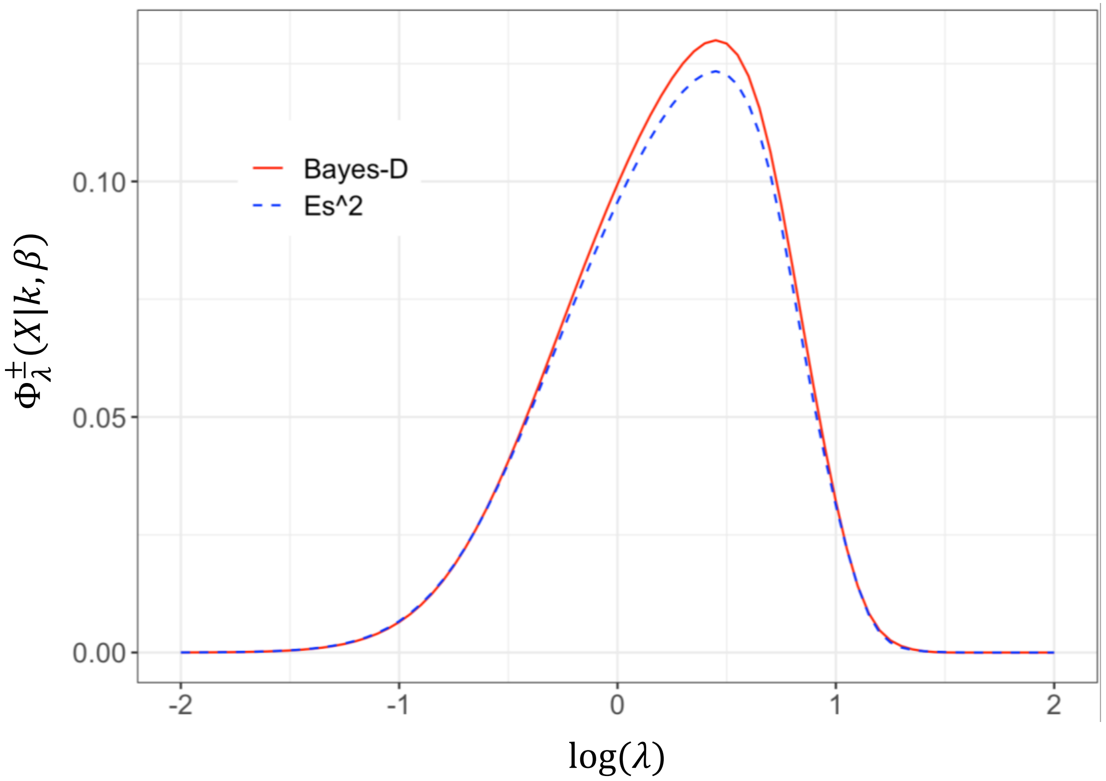
- Scenario: n=14, p=24, k=7 with SN=3 and effect signs known
- PED and $Var(s+)$
- Gauss Dantzig selector shows PED is better than $Var(s+)$ in simulation



- $Var(s+)$ is higher for almost all $\log(\lambda)$
- There are $\log(\lambda)$ where PED has a larger Φ_λ
- Advantages to either design

Example 2

- Scenario: n=14, p=24, k=6 with SN=3 and effect signs unknown
- $E(s^2)$- and Bayes-D- optimal designs from Marley and Woods, 2010
- Simulation: Dantzig-No difference, Lasso – Bayes-D is optimal



- Bayes-D dominates $E(s^2)$ over all $\log(\lambda)$
- Agrees with lasso simulation results

Summary

- Comparing/evaluating SSDs using simulations of the GDS can be sensitive to the simulation settings

- Our approach utilizes exact lasso sign recovery probabilities
  - Reasonably robust to tuning parameter selection
  - Unambiguous interpretation of results

- Future work:
  - Extend the lasso screening measure to the case where the set of factors selected by the lasso contains the true support and possibly a few extra factors.

References

Candes, Emmanuel, and Terence Tao. "The Dantzig selector: Statistical estimation when p is much larger than n." *The annals of Statistics* 35.6 (2007): 2313-2351.

Hastie, T., Tibshirani, R., and Wainwright, M. 2019. “Statistical learning with sparsity, the lasso and generalizations.” Chapman and Hall/CRC

Marley, C. J. and Woods, D. C. (2010), “A Comparison of design and model selection methods for supersaturated experiments,” *Computational Statistics and Data Analysis*, 54, 3158-3167.

Phoa, Frederick KH, Yu-Hui Pan, and Hongquan Xu. "Analysis of supersaturated designs via the Dantzig selector." *Journal of Statistical Planning and Inference* 139.7 (2009): 2362-2372.

Singh, R., & Stufken, J. (2023). Selection of two-level supersaturated designs for main effects models. *Technometrics*, *65*(1), 96-104..

Stallrich, J. W., Young, K., Weese, M. L., Smucker, B. J., and Edwards, D. J. (2023), “Optimal Supersaturated Designs for Lasso Sign Recovery,”. https://arxiv.org/abs/2303.16843

Weese, Maria L., David J. Edwards, and Byran J. Smucker. "A criterion for constructing powerful supersaturated designs when effect directions are known." *Journal of Quality Technology* 49.3 (2017): 265-277.

Weese, Maria L., et al. "Strategies for Supersaturated Screening: Group Orthogonal and Constrained Var (s) Designs." Technometrics 63.4 (2021): 443-455.

Thank you!

FARMER SCHOOL OF BUSINESS

Backup Slides

FARMER SCHOOL OF BUSINESS

Simulation to study Inconsistencies

Scenario: n=12, k=26, μ=3 and c=6 with effect directions unknown