Simultaneous Confidence Intervals Using Entire Solution Paths

Xiaorui Zhu*; Peng Wang; Yichen Qin; University of Cincinnati

SCAN ME

Introduction

An ideal simultaneous confidence intervals (SCI) for sparse linear model

1) should be as tight as possible and achieve the nominal confidence level simultaneously (coverage probability, the width of intervals of nonzero and zero coefficients);
2) **should be able to imply the variable selection results** in a way that the truly relevant and irrelevant coefficients could have nonzero and zero width intervals, respectively.

We propose a general approach to construct simultaneous confidence intervals based on entire solution paths and residual bootstraps.

General Approach for Constructing SCI

We define an **outlyingness score** for each bootstrap estimator to measure the relative location of a bootstrap estimator among all B bootstrap estimators as follow:

$$O^{(b)} = g(\hat{\boldsymbol{\beta}}) = (o_1^{(b)}, \dots, o_d^{(b)}) \in \mathbb{R}^{+d},\ b \in 1, \dots, B.$$

Then, we can rule out α percent of outlying bootstrap estimators among all to construct the simultaneous confidence intervals with confidence level 1−α.

Two special instances of outlyingness score:

1. $O^{\mathrm{F},(b)} = (o^{\mathrm{F},(b)}) = g^{\mathrm{F}}(\hat{\boldsymbol{\beta}}) = \hat{F}(\gamma_b, \gamma_f) = \frac{(\mathrm{RSS}_{\gamma_b} - \mathrm{RSS}_{\gamma_f})/(df_{\gamma_b} - df_{\gamma_f})}{\mathrm{RSS}_{\gamma_f}/df_{\gamma_f}}$

2. $O^{\mathrm{MaxMin},(b)} = (o_{\max}^{(b)}, o_{\min}^{(b)}) = g^{\mathrm{MaxMin}}(\hat{\boldsymbol{\beta}})$
$$= \left(\max_{j\in\{1,\dots,p\}}\left(\frac{\hat{\beta}_j^{(b)} - \bar{\hat{\beta}}_j}{s.e._{\hat{\beta}_j}}\right), \min_{j\in\{1,\dots,p\}}\left(\frac{\hat{\beta}_j^{(b)} - \bar{\hat{\beta}}_j}{s.e._{\hat{\beta}_j}}\right)\right)$$

Procedure: Simultaneous Confidence Intervals

Step 1 : Apply residual bootstrap for variable selection to obtain: $\{\hat{\boldsymbol{\beta}}^{(b)}\}_{b=1}^{B}$;

Step 2 : Construct outlyingness score: $O^{(b)} = (o_1, o_2, \dots, o_d) = g(\hat{\boldsymbol{\beta}}) \in \mathbb{R}^{+d}$;

Step 3 : Construct a set $\mathcal{A}_\alpha \subset \{1, \dots, B\}$:

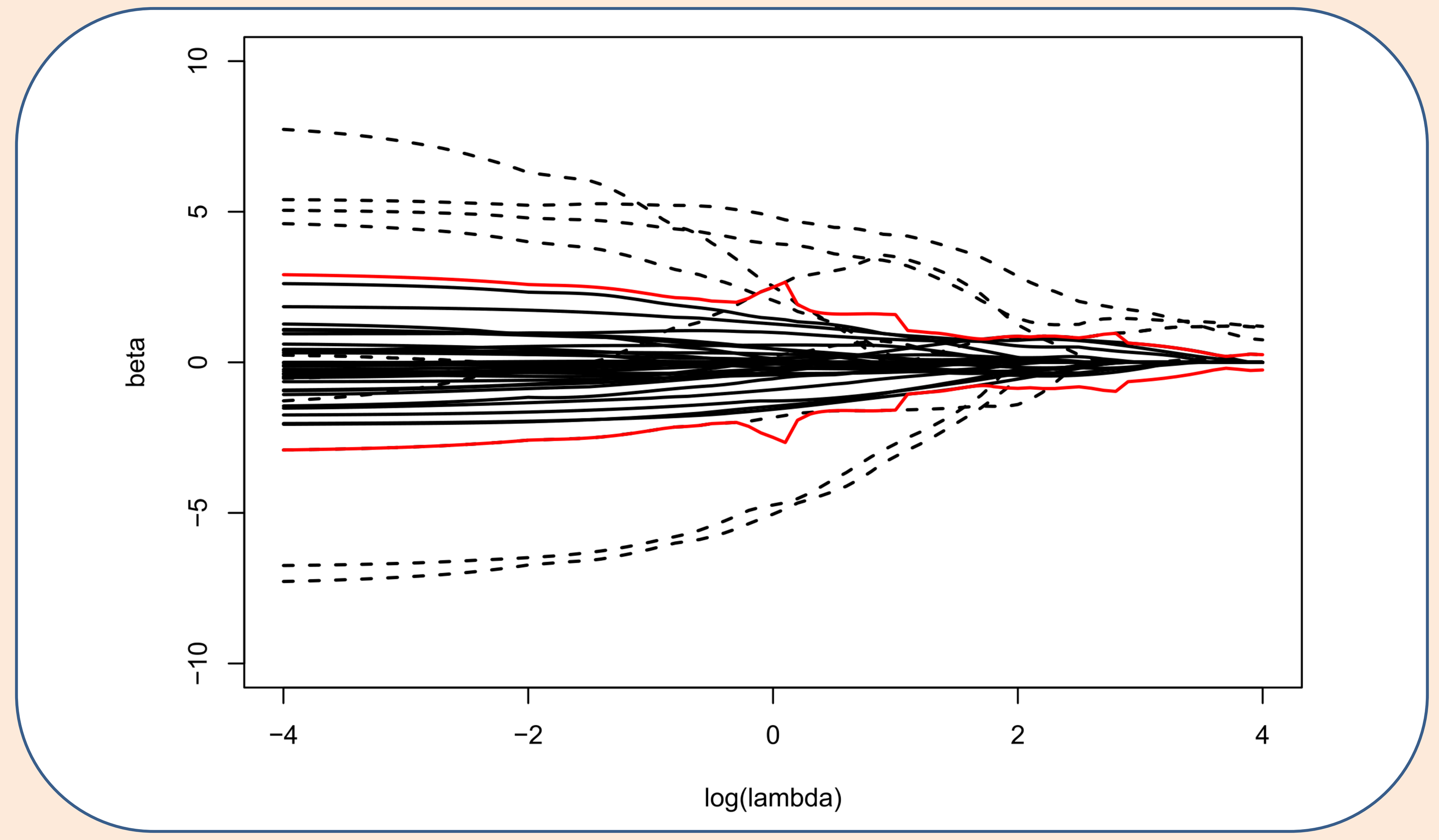
$$\mathcal{A}_\alpha = \{b \in (1, \dots, B);\ o_i^{(b)} \le q_i(1 - \tfrac{\alpha}{d}), i = 1, \dots, d\},$$

where $q_i(1-\frac{\alpha}{d})$ is $(1-\frac{\alpha}{d})$ quintile of o_i;

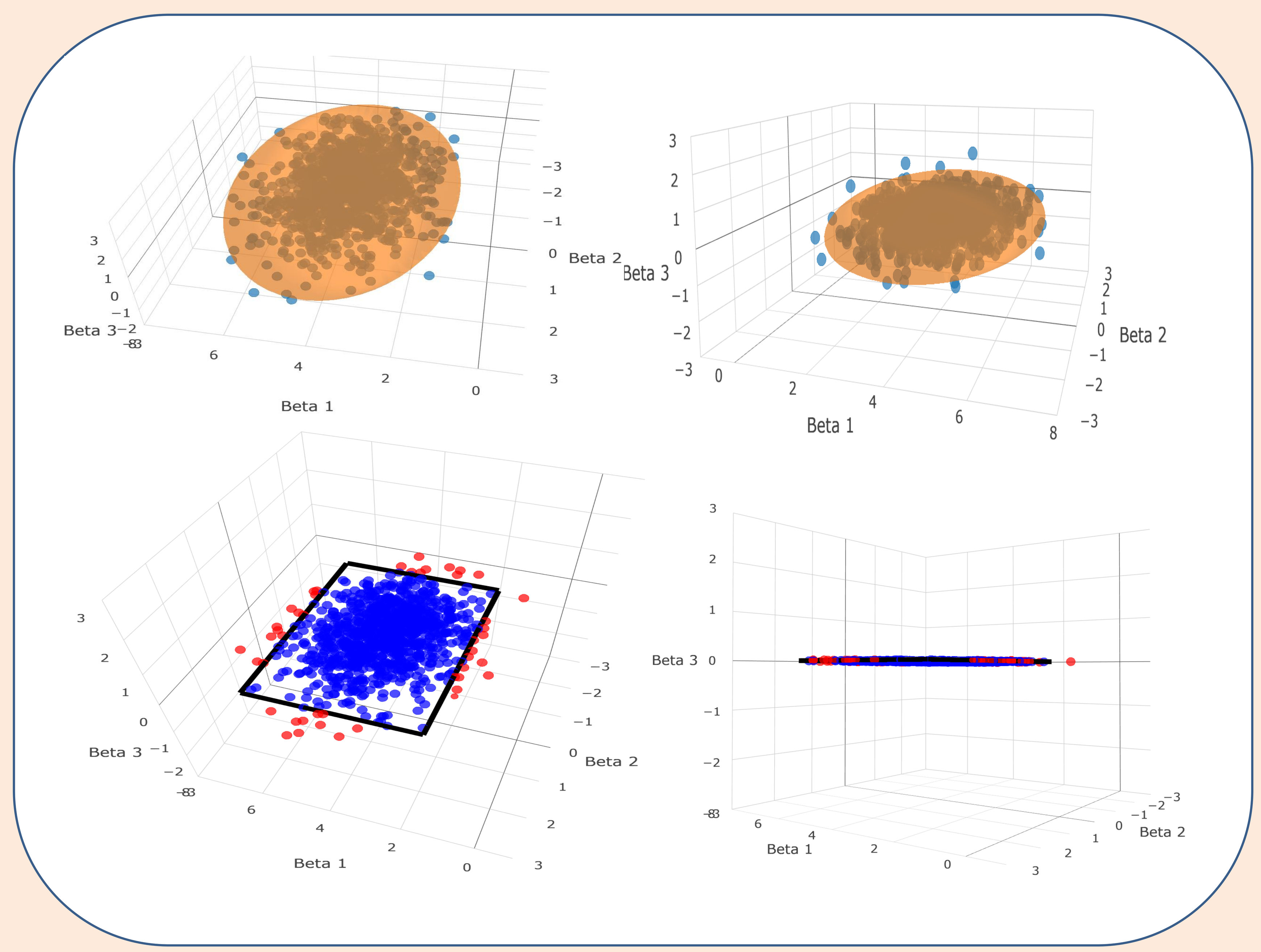
Step 4 : Construct the simultaneous confidence intervals (SCI) as:

$$\mathrm{SCI}(1-\alpha) = \left\{\boldsymbol{\beta} \in \mathbb{R}^p;\ \min_{b\in\mathcal{A}_\alpha} \beta_j^{(b)} \le \beta_j \le \max_{b\in\mathcal{A}_\alpha} \beta_j^{(b)}, j = 1, \dots, p\right\}.$$

Selection by Partitioning the Solution Paths (SPSP)



Geometrical Differences: Proposed vs. Debiased



Theoretical Results

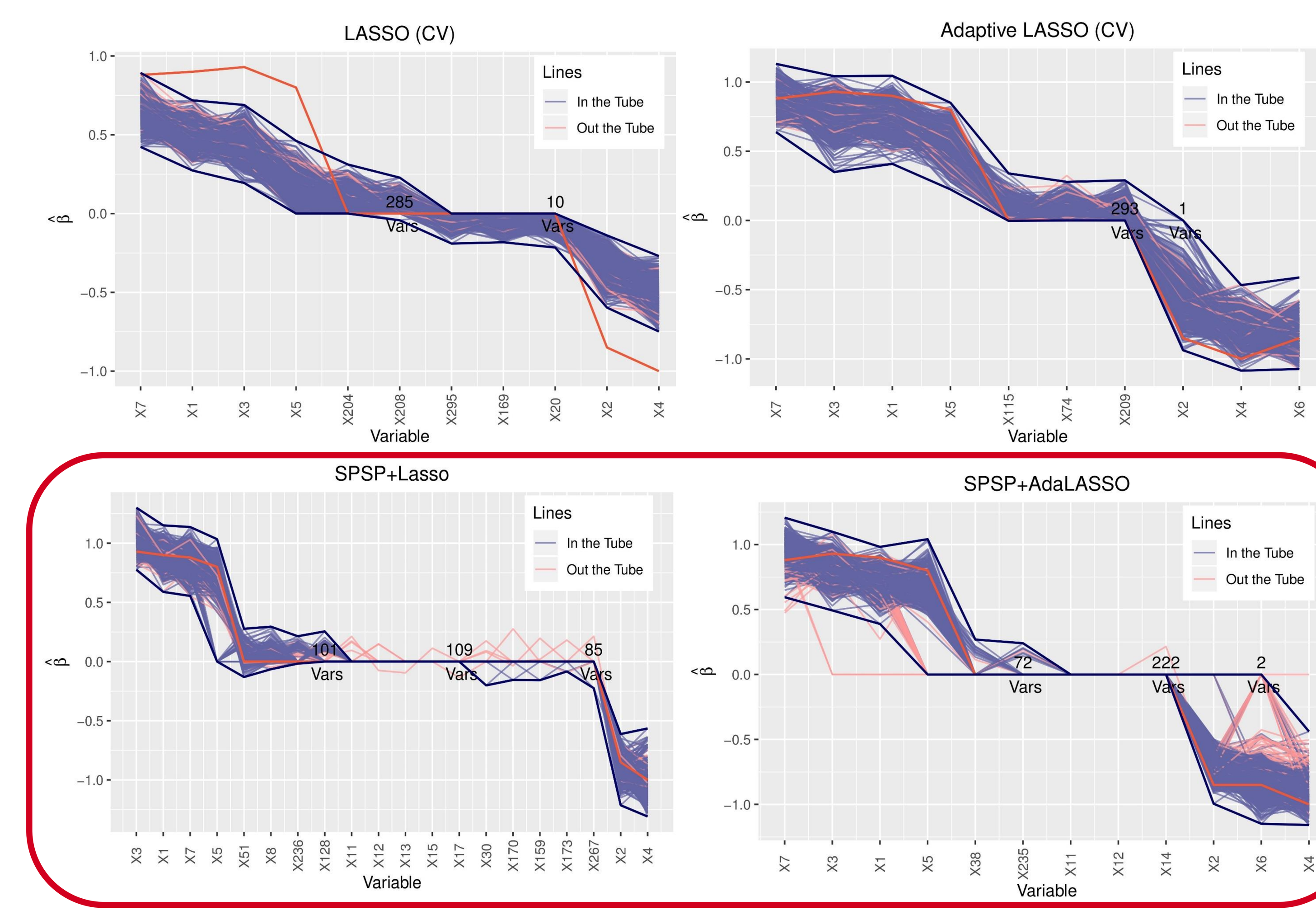
Theorem: Under mild assumptions, for $\alpha \in (0, 1)$ and all $\boldsymbol{\beta} \in \mathbb{R}^p$, we have

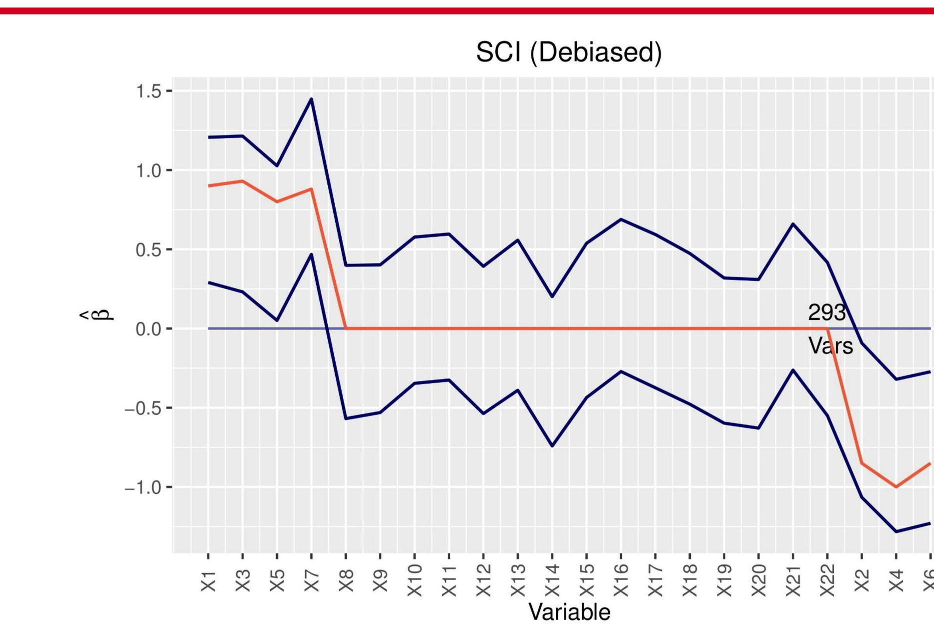
$$\mathbf{P}(\boldsymbol{\beta} \in \mathrm{SCI}_{n,(1-\alpha)}) \to 1 - \alpha \text{ as } n \to \infty.$$

Simulation Example

- p=300, n=200
- $\beta^* = (0.9, -0.85, 0.93, -1, 0.8, -0.85, 0.88)$
- Remaining coefficients equal zero
- Correlation between X_i and X_j is $0.5^{|i-j|}$

SPSP vs. Cross-Validation vs. Debiased





SCI	W.Nzero	W.Zero	Cover Pr	Avg Card	Med Card	Std Card
SPSP+AdaLasso(MaxMin)	0.60	0.04	96.50	68.31	59.00	51.66
SPSP+AdaLasso(F)	0.61	0.06	98.50			
SPSP+Lasso(MaxMin)	0.92	0.19	96.50	734.19	770.50	150.75
SPSP+Lasso(F)	0.92	0.19	96.50			
AdaLasso(MaxMin)	0.64	0.21	66.00	949.24	950.00	1.56
AdaLasso(F)	0.64	0.21	65.50			
Lasso(MaxMin)	0.54	0.25	0.00	950.00	950.00	0.00
Lasso(F)	0.54	0.25	0.00			
True model(MaxMin)	0.45	0.00	92.50	1.00	1.00	0.00
True model(F)	0.46	0.00	99.50			
SCI(Debiased)	0.97	0.97	98.00			