Supplemental Materials for "Sparsity Oriented Importance Learning for High-dimensional Linear Regression"

Chenglong Ye [*], Yi Yang [†], Yuhong Yang [‡]

August 27, 2017

Part A: Weighting using generalized fiducial inference.

Based on Fisher's controversial fiducial idea, Lai et al. (2015) proposed the generalized fiducial inference applied to "large p small n" problem. Their paper concerns the generalized fiducial inference for the linear regression case. For each candidate model $\mathcal{A}^k$, the fiducial probability for the model is

$$p(\mathcal{A}^k) \propto R(\mathcal{A}^k) \equiv \Gamma(\frac{n-|\mathcal{A}^k|}{2})(\pi RSS_{\mathcal{A}^k})^{-\frac{n-|\mathcal{A}^k|-1}{2}} n^{-\frac{|\mathcal{A}^k|+1}{2}} \begin{pmatrix} p \\ |\mathcal{A}^k| \end{pmatrix}^{-\gamma},$$

where $RSS_{\mathcal{A}^k}$ is the residual sum of squares of $\mathcal{A}^k$. For a practical reason, the authors approximate the above fiducial probability by

$$r(\mathcal{A}^k) \approx R(\mathcal{A}^k) / \sum_{l=1}^{K} R(\mathcal{A}^l).$$

We can use $r(\mathcal{A}^k)$ as the weight w_k for each candidate model. It is shown in their paper that the true model will have the highest fiducial probability among all the candidate models.

[*]School of Statistics, University of Minnesota (yexxx323@umn.edu)
[†]Corresponding author, Department of Mathematics and Statistics, McGill University (yi.yang6@mcgill.ca)
[‡]School of Statistics, University of Minnesota (yangx374@umn.edu)

Part B: Additional simulation results.

In this part, we provide the results of Example S1-S6, whose settings are described in Table 1 of the main body of the article. These results support our conclusions as discussed in Section 5.1.

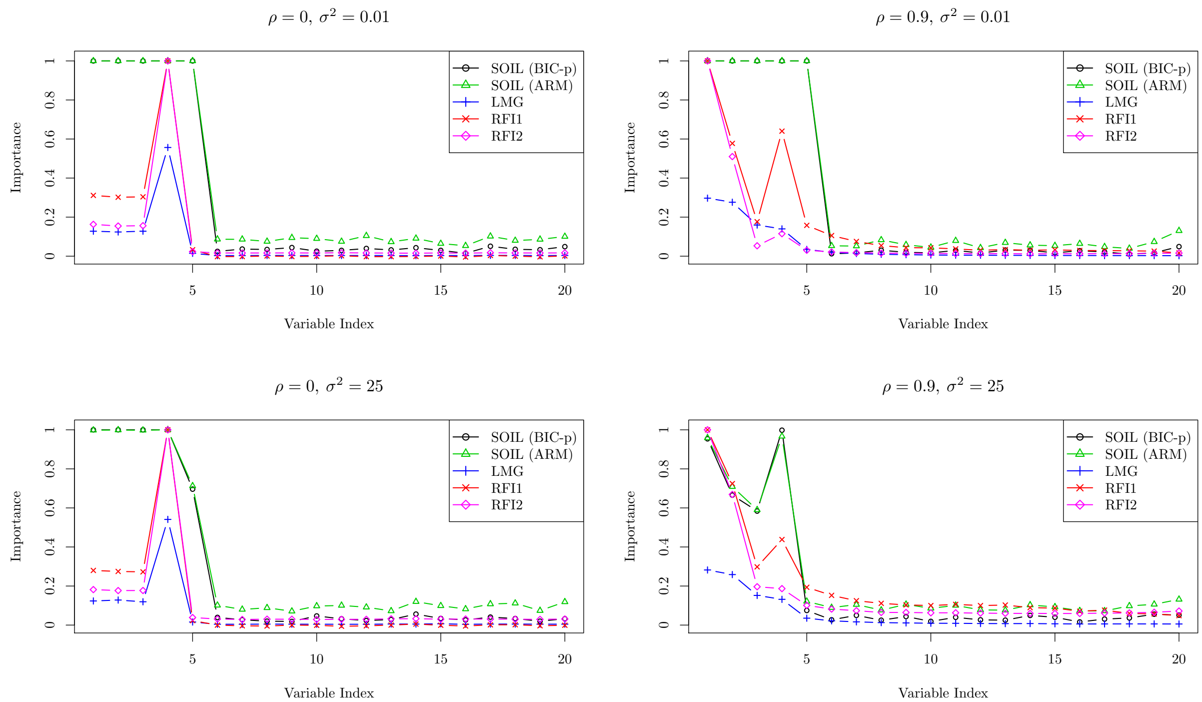

Figure S1: Simulation results for Example S1, where $n = 150$, $p = 20$. The true coefficients $\boldsymbol{\beta}^* = (4, 4, 4, -6\sqrt{2}, \frac{3}{4}, 0, ..., 0)$.

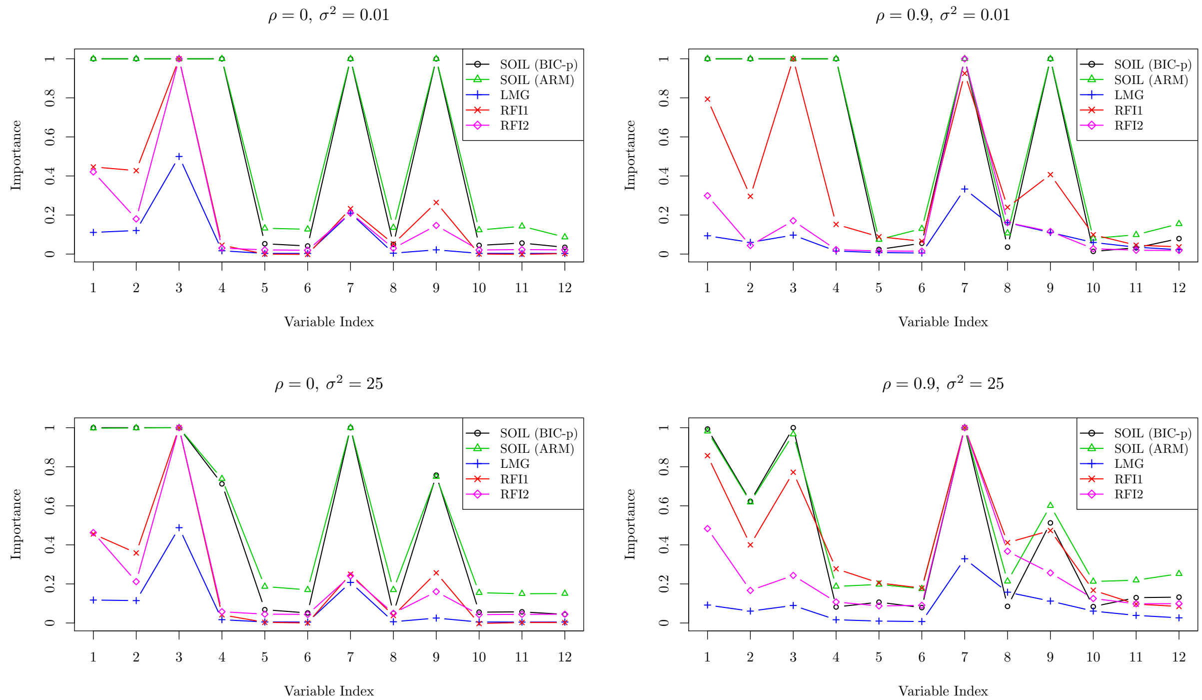

Figure S2: Simulation results for Example S2, where $n = 150$, $p = 6$. The true coefficients $\boldsymbol{\beta}^* = (4, 4, -6\sqrt{2}, \frac{3}{4}, 0, 0)^\intercal$. Add $(X_1^2, X_2^2, X_3^2, X_4^2, X_5^2, X_6^2)$ and corresponding coefficients $(\beta_7^*, \beta_8^*, \ldots, \beta_{12}^*)^\intercal = (4, 0, 1, 0, 0, 0)^\intercal$.

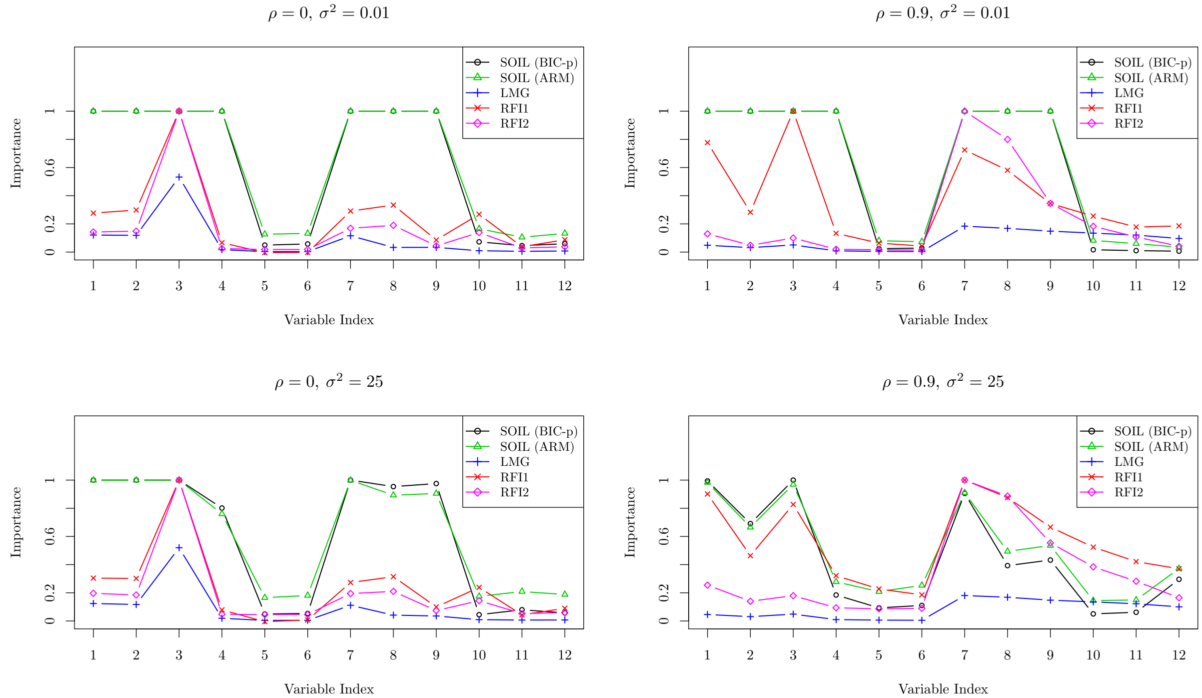

Figure S3: Simulation results for Example S3, where $n = 150$, $p = 6$. The true coefficient $\boldsymbol{\beta}^* = (4, 4, -6\sqrt{2}, \frac{3}{4}, 0, 0)^\intercal$. Add $(X_1X_2, X_1X_3, X_1X_4, X_2X_3, X_2X_4, X_3X_4)$ and corresponding coefficients $(\beta_7^*, \beta_8^*, \ldots, \beta_{12}^*)^\intercal = (4, 2, 2, 0, 0, 0)^\intercal$.

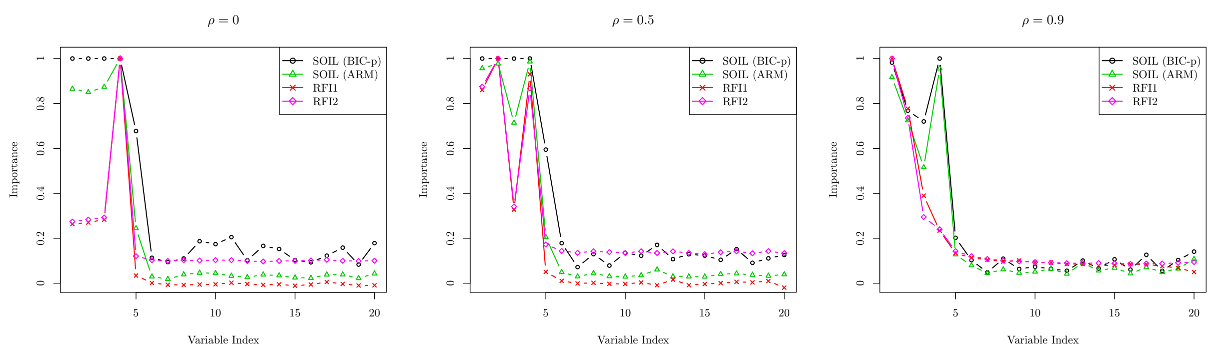

Figure S4: Simulation results for Example S4, where $n = 150$, $p = 20$. The true coefficients $\boldsymbol{\beta}^* = (4, 4, 4, -6\sqrt{2}, \frac{3}{4}, 0, ..., 0)$.

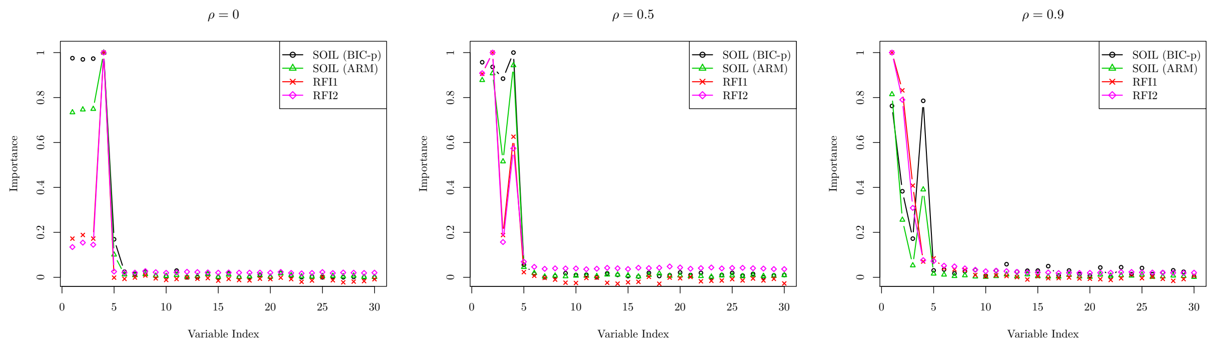

Figure S5: Simulation results for Example S5, where $n = 100$, $p = 200$. The true coefficients $\boldsymbol{\beta}^* = (4, 4, 4, -6\sqrt{2}, \frac{3}{4}, 0, ..., 0)$.

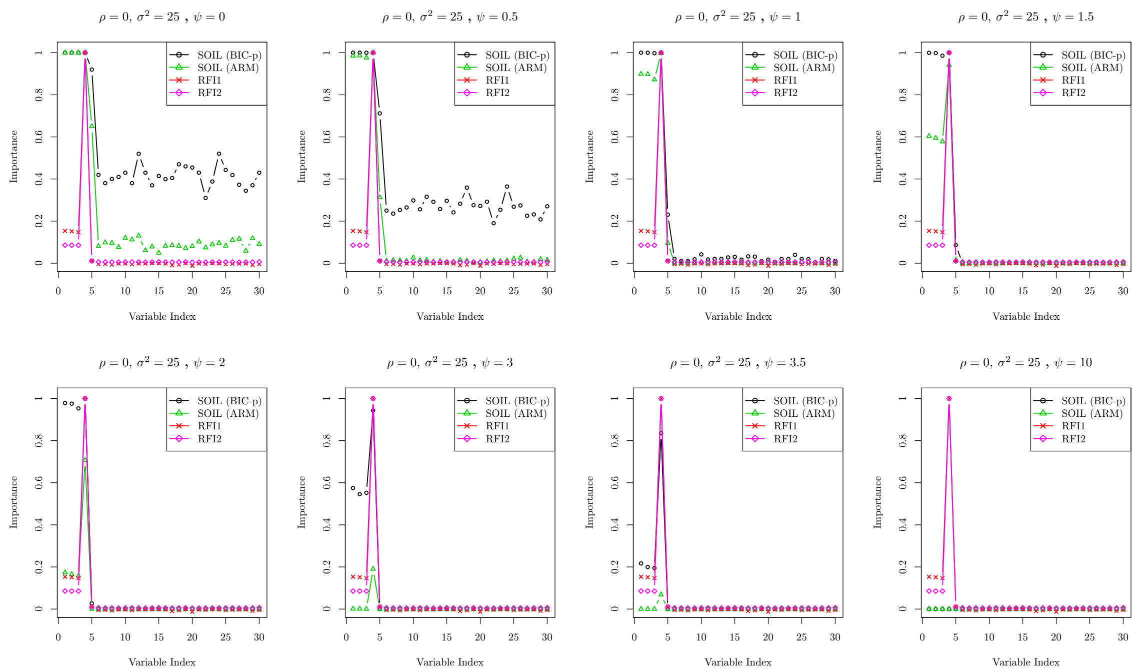

Figure S6: Sensitivity analysis of ψ, where $n = 100$, $p = 200$. The true coefficients $\boldsymbol{\beta}^* = (4, 4, 4, -6\sqrt{2}, \frac{3}{4}, 0, ..., 0)$.

Part C: Comparison with stability selection.

In this subsection, we present a comparison of SS (Meinshausen & Bühlmann 2010) importance and our SOIL importance.

The simulation data $\{y_i, \mathbf{x}_i\}_{i=1}^n$ is generated from the linear model $y_i = \mathbf{x}_i^\intercal \boldsymbol{\beta}^* + \epsilon_i$, $\epsilon \sim N(0, \sigma^2)$. We generate $\mathbf{x}_i$ from multivariate normal distribution $N_p(0, \Sigma)$. For each element

Σ_{ij} of Σ, $\Sigma_{ij} = \rho^{|i-j|}$, i.e. the correlation of X_i and X_j is $\rho^{|i-j|}$. We consider two cases, the settings of which are listed in Table S1.

Example	n	p	ρ	σ^2	Coefficients
1	100	20	0	0.01	$\boldsymbol{\beta}^* = (4, 4, 4, -6\sqrt{2}, \frac{3}{4}, 0, ..., 0)^\intercal$
2	100	20	0.7	0.1	$\boldsymbol{\beta}^* = (4, 0, 4, -6\sqrt{2}, \frac{3}{4}, 0, ..., 0)^\intercal$

Table S1: Simulation settings for SS

It can be seen from Tables S2 and S3 that SS does not give enough importance to the true variable X_5 in Example 1 while it more strongly supports the noise variable X_2 than the true variable X_5 in Example 2, which leads to unavoidable incorrect variable selection regardless of the cutoff to be used to decide if a variable is in or out based on its importance. In contrast, SOIL-ARM and SOIL-BIC-p pick all the important variables and leave noise variables out. From these results, together with the fact that the main goal of SS is not on variable importance, we have not considered stability selection in the main simulations in this work.

Method/Variable	X_1	X_2	X_3	X_4	X_5	max of rest
SOIL-ARM	1.00	1.00	1.00	1.00	1.00	0.12
SOIL-BIC-p	1.00	1.00	1.00	1.00	1.00	0.07
Stability Selection	0.99	0.99	0.99	1.00	0.02	0.002

Table S2: Variable importance for Example 1.

Method/Variable	X_1	X_2	X_3	X_4	X_5	max of rest
SOIL-ARM	1.00	0.15	1.00	1.00	1.00	0.14
SOIL-BIC-p	1.00	0.06	1.00	1.00	1.00	0.05
Stability Selection	1.00	0.44	0.94	1.00	0.26	0.05

Table S3: Variable importance for Example 2.

Part D: Stability comparison of SOIL and Lasso.

We conduct a stability comparison of our methods and Lasso at a reduced sample size to show that our method is more stable than Lasso against small changes in the data. The simulation data $\{y_i, \mathbf{x}_i\}_{i=1}^n$ is generated from the linear model $y_i = \mathbf{x}_i^\intercal \boldsymbol{\beta}^* + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$ and $\sigma^2 = 0.01$. $\mathbf{x}_i$ is generated from $N_p(0, \Sigma)$, where $\Sigma_{ij} = \rho^{|i-j|}$ and $\rho = 0.5$. We set $n = 50$, $p = 200$ and

$\boldsymbol{\beta}^* = (4, 4, -6\sqrt{2}, 4/3, 0, 0, 4, 0, 1, 0, \ldots, 0)^\intercal$. We randomly remove 10 observations from the dataset and use the remaining data to compute the corresponding SOIL-BIC-p importances and the Lasso coefficients. The results are recorded over 100 replications and shown in Figure S7. We can see that, for each run with the reduced sample size, the result for the SOIL importance is pretty consistent, while the result for the Lasso coefficients varies considerably, indicating that the SOIL importance has the continuity property with respect to a reduced sample size and is more stable than Lasso.

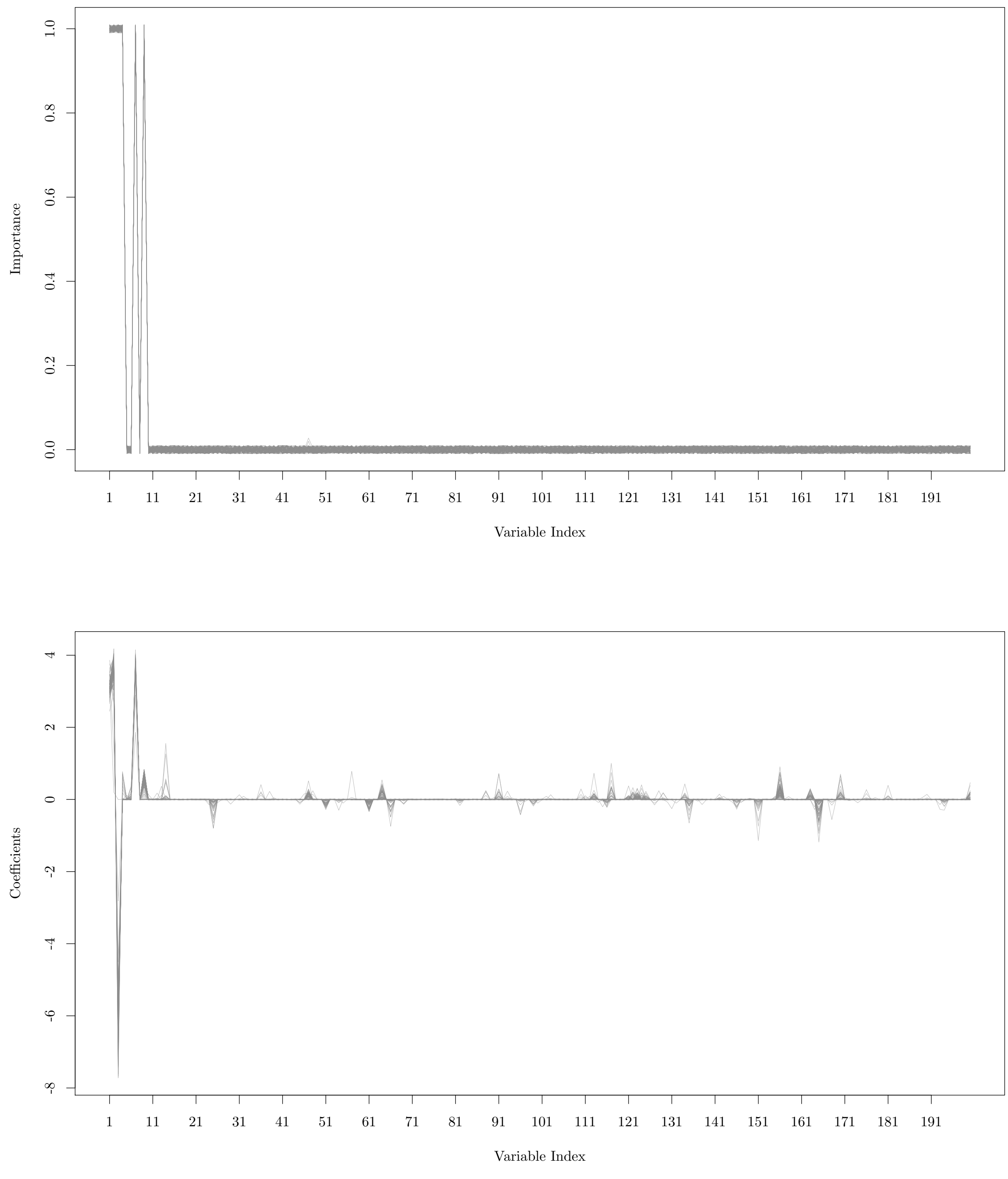

Figure S7: Stability comparison of SOIL-BIC-p and Lasso at a reduced sample size for 100 replications. Top panel: SOIL-BIC-p importances. Bottom panel: Lasso coefficients. Each grey line represents the result from one replication.

References

Lai, R. C., Hannig, J. & Lee, T. C. (2015), 'Generalized fiducial inference for ultrahigh-dimensional regression', *Journal of the American Statistical Association* **110**(510), 760–772.

Meinshausen, N. & Bühlmann, P. (2010), 'Stability selection', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**(4), 417–473.