Additional numerical examples to “The Sparse MLE for Ultra-High-Dimensional Feature Screening”

Blinded 1 and Blinded 2

A. Testing Under Marginally Independent Data Structure

In the main article, we have empirically tested the proposed SMLE on the datasets where the covariates have independent, auto-correlated or compound structures. Fan et al. (2009) considered an interesting setup in their numerical studies, where one relevant feature is marginally independent with the response but not jointly independent in the presence of other features. The marginal-effect-based screening methods (e.g. SIS) may hence fail to identify such important features. Since SMLE is a joint-effect-based method, it is interesting to examine its performance under this particular situation.

To this end, we first generate features according to the following setting, which is the same as setup S3 in the main context of the paper:

S3: Candidate features $x_1, \ldots, x_p$ are joint normal, marginally $N(0,1)$, with $\mathrm{cov}(x_j, x_h) = 0.15$ for $j, h \in s^*$ and $\mathrm{cov}(x_j, x_h) = 0.3$ for j or $h \in s_c^*$.

The desired situation is created by the deliberate choices on the relevant features and their effects, which is given as follows.

Linear: $s^* = \{1, 2, 3, 4\}$ with $\boldsymbol{\beta}_{s^*} = (1, 1, 1, -0.45)^T$, $\boldsymbol{\beta}_{s_c^*} = 0$, and $(n, p, \sigma) = (100, 1000, 1)$.

Logistic: $s^* = \{1, 2, 3, 4\}$ with $\boldsymbol{\beta}_{s^*} = (1.5, 1.5, 1.5, -0.675)^T$, $\boldsymbol{\beta}_{s_c^*} = 0$, and $(n, p) = (400, 1000)$.

Poisson: $s^* = \{1, 2, 3, 4\}$ with $\boldsymbol{\beta}_{s^*} = (0.6, 0.6, 0.6, -0.27)^T$, $\boldsymbol{\beta}_{s_c^*} = 0$, and $(n, p) = (200, 1000)$.

Under the normal linear regression, it is seen that

$$\mathrm{cov}(X_4, Y) = \mathrm{cov}(X_4, X_1 + X_2 + X_3 - 0.45X_4) = 0.15 + 0.15 + 0.15 - 0.45 = 0.$$

The joint normality implies X_4 and Y are independent. Clearly, X_4 is still an important feature, but a marginal-effect-based method is likely to miss it. See Fan et al. (2009) for interpretations of the other two settings.

Table A.1: Simulation results under the setup that X_4 is marginally uncorrelated with Y.

Models	1st Stage	2nd Stage	RC	PSR	FDR	CSR	AMS	P.err	TIME
Linear	SIS	–	.00	.63	.88	.00	21.0	.45	.70
		LASSO	.00	.58	.47	.00	4.8	.62	.87
		SCAD	.00	.61	.27	.00	3.7	.37	1.02
	ISIS	–	.67	.91	.83	.00	21.0	.46	4.67
		LASSO	.27	.76	.37	.02	5.38	.52	4.81
		SCAD	.60	.88	.25	.29	5.01	.15	5.31
	FR	–	.66	.91	.83	.00	21.0	.55	6.23
		LASSO	.45	.85	.27	.06	5.1	.41	6.35
		SCAD	.63	.90	.18	.37	4.6	.11	7.19
	LASSO	–	.79	.95	.81	.00	19.7	.34	.02
		LASSO	.15	.71	.41	.01	5.3	.60	.16
		SCAD	.69	.90	.30	.19	5.5	.14	.46
	SMLE	–	.92	.98	.81	.00	21.0	.44	.23
		LASSO	.28	.80	.32	.02	5.1	.53	.37
		SCAD	.74	.92	.23	.30	5.1	.14	1.18
Logistic	SIS	–	.00	.74	.80	.00	15.0	.20	2.91
		LASSO	.00	.73	.06	.00	3.19	.20	3.57
		SCAD	.00	.74	.03	.00	3.09	.19	3.63
	ISIS	–	.58	.90	.76	.00	15.	.22	16.2
		LASSO	.31	.83	.13	.03	4.1	.19	16.6
		SCAD	.54	.88	.12	.39	4.1	.18	17.2
	LASSO	–	.76	.94	.73	.00	14.1	.20	.06
		LASSO	.13	.78	.09	.02	3.55	.20	.51
		SCAD	.70	.93	.05	.57	3.96	.18	.86
	SMLE	–	.98	.98	.73	.00	15.0	.22	3.75
		LASSO	.31	.83	.09	.06	3.9	.19	4.21
		SCAD	.77	.94	.08	.57	4.2	.18	4.68
Poisson	SIS	–	.00	.64	.88	.00	21.0	.17	3.01
		LASSO	.00	.61	.48	.00	5.25	.33	3.49
		SCAD	.00	.62	.46	.00	5.23	.26	4.97
	ISIS	–	.79	.95	.82	.00	21.0	.16	17.62
		LASSO	.48	.86	.32	.07	5.48	.17	18.02
		SCAD	.70	.92	.22	.33	5.02	.03	20.42
	LASSO	–	.81	.95	.79	.00	19.0	.09	.03
		LASSO	.19	.75	.42	.02	5.74	.28	.38
		SCAD	.72	.91	.30	.22	5.69	.07	1.61
	SMLE	–	.90	.97	.81	.00	21.0	.19	.60
		LASSO	.54	.87	.21	.12	4.72	.14	.97
		SCAD	.79	.94	.11	.59	4.40	.03	3.94

Does SMLE has the promised utility of taking joint effect into consideration? We carried out numerical studies on 500 datasets from above models, with all other simulation settings kept unchanged. The results are summarized in Table A.1. As anticipated, we observe that the marginal-effect-based SIS fails completely in terms of retaining s^*. In comparison, other screening methods perform reasonably well, as they all take the joint effects of features in the screening process. The proposed SMLE has the top performances by achieving the highest retaining capability with a relative low computation cost.

B. Testing Under Model Mis-specification

Up to this moment, the feature selections are carried out when the data-generating models are provided without errors. This is unrealistic in applications, especially in the high-dimensional situations. The sensitivity of model mis-specification is always an important issue for an analytic method. For this purpose, we designed additional simulation studies in this section.

For easy comparison, we adopt the same correlation and parameter settings S1-S3 as described in Sections 6.1-6.4. We generated the response variables under two scenarios. The first one is the same as Example 5 of Wang (2009) - the presumed model is a linear regression with normal error distribution but the data are generated with a heavy-tail error distribution. The second one is when the data set is contaminated by outliers, where the effect of one relevant feature is reversed in a percentage of observations. More specifically,

Scenario 1: $\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{\beta} + e$, where the error e follows a centered exponential distribution with rate parameter $\sigma > 0$. That is, $P(e + \sigma^{-1} > t) = \exp(-\sigma t)$ for $t \geq 0$

Scenario 2: 90% of observations are generated from the linear model $\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{\beta} + e$ with $e \sim N(0,1)$, and the other 10% from $\boldsymbol{Y} = \boldsymbol{X}^T\tilde{\boldsymbol{\beta}} + e$ such that $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$ except for $\tilde{\beta}_{s_3^*} = -\beta_{s_3^*}$.

Under Scenario 2, the effect of the 3rd relevant feature is substantially weakened. It is therefore of interest to see how the performances of different methods are affected.

We analyze the data as if they were generated from a presumed normal linear model, and apply the same set of methods for feature screening. The results are summarized in Tables B.1-B.2. Based on Table B.1, we notice that the heavy-tail error distribution has little impact on all five methods. There is a close match between Table B.1 and Table 1 (section 6.4) in all aspects. Based on Table B.2, the presence of outliers under Scenario 2 deteriorates the retaining capability for all five methods. There is no winner under this situation. Nevertheless, the proposed SMLE is reasonably robust, by achieving relatively high RC/PSR in all three correlation settings.

REFERENCES

Fan, J., Samworth, R., and Wu, Y. (2009). Ultrahigh dimentional variable selection: beyond the linear model. *Journal of Machine Learning Research*, 10:1829–1853.

Wang, H. (2009). Forward regression for ultra-high dimentional varibale screening. *Journal of the American Statistical Association*, 104:1512–1524.

Table B.1: Simulation results for linear regression with heavy-tail random errors

Setup	1st Stage	2nd Stage	RC	PSR	FDR	CSR	AMS	P.err	TIME
S1	SIS	–	.23	.84	.78	.00	31.0	.49	10.64
		LASSO	.23	.84	.11	.17	7.7	.62	10.88
		SCAD	.23	.84	.12	.17	7.8	.42	11.04
	ISIS	–	.92	.99	.74	.00	31.0	.41	62.74
		LASSO	.82	.97	.11	.41	8.9	.57	63.01
		SCAD	.92	.99	.07	.56	8.6	.14	64.08
	FR	–	.99	.99	.74	.00	31.0	.53	162.86
		LASSO	.98	.99	.07	.61	8.75	.44	163.11
		SCAD	.98	.99	.09	.47	8.94	.12	163.23
	LASSO	–	.97	.99	.73	.00	29.1	.35	.39
		LASSO	.81	.97	.11	.39	8.9	.58	.64
		SCAD	.95	.99	.07	.53	8.7	.20	1.82
	SMLE	–	.98	.99	.74	.00	31.0	.45	2.62
		LASSO	.91	.99	.09	.46	8.9	.55	2.89
		SCAD	.97	.99	.08	.52	8.8	.14	3.06
S2	SIS	–	.58	.91	.81	.00	24.0	.34	4.51
		LASSO	.34	.83	.09	.23	4.7	.41	4.73
		SCAD	.28	.78	.14	.22	4.7	.26	4.95
	ISIS	–	.66	.93	.81	.00	24.0	.41	23.70
		LASSO	.39	.85	.11	.26	4.85	.39	23.93
		SCAD	.31	.79	.18	.20	5.00	.30	24.25
	FR	–	.38	.78	.84	.00	24.0	.57	45.24
		LASSO	.32	.77	.23	.23	5.24	.33	45.41
		SCAD	.20	.73	.28	.12	5.31	.25	45.78
	LASSO	–	.89	.98	.78	.00	22.4	.34	.12
		LASSO	.43	.86	.09	.29	4.82	.41	.33
		SCAD	.34	.82	.16	.24	5.01	.30	.83
	SMLE	–	.77	.95	.80	.00	24.0	.44	1.17
		LASSO	.46	.87	.10	.33	4.9	.39	1.38
		SCAD	.30	.80	.19	.19	5.1	.27	1.93
S3	SIS	–	.00	.32	.94	.00	21.0	.89	.74
		LASSO	.00	.22	.91	.00	9.8	.95	.95
		SCAD	.00	.24	.90	.01	9.8	.94	.97
	ISIS	–	.73	.84	.84	.00	21.0	.57	5.69
		LASSO	.52	.70	.66	.02	8.96	.87	5.90
		SCAD	.73	.78	.30	.51	5.92	.26	6.24
	FR	–	.86	.87	.83	.00	21.0	.60	6.55
		LASSO	.75	.83	.54	.03	8.07	.83	6.76
		SCAD	.85	.87	.22	.57	5.29	.14	7.35
	LASSO	–	.26	.61	.88	.00	19.8	.71	.01
		LASSO	.03	.33	.87	.00	9.88	.95	.22
		SCAD	.26	.47	.66	.24	8.38	.70	.24
	SMLE	–	.99	1.00	.81	.00	21.0	.49	.34
		LASSO	.66	.88	.55	.02	8.5	.91	.57
		SCAD	.99	.99	.08	.69	4.5	.01	.92

Table B.2: Simulation results for linear regression with 10% outliers

Setup	1st Stage	2nd Stage	RC	PSR	FDR	CSR	AMS	P.err	TIME
S1	SIS	–	.16	.82	.79	.00	31.0	.48	9.63
		LASSO	.16	.81	.16	.08	7.9	.58	9.84
		SCAD	.16	.81	.17	.09	8.0	.42	10.08
	ISIS	–	.83	.97	.75	.00	31.0	.41	51.86
		LASSO	.65	.95	.16	.19	9.26	.54	52.11
		SCAD	.80	.97	.14	.25	9.21	.16	53.23
	FR	–	.96	.99	.74	.00	31.0	.53	153.71
		LASSO	.95	.99	.17	.21	9.8	.39	153.94
		SCAD	.95	.99	.20	.13	10.1	.14	153.10
	LASSO	–	.92	.99	.73	.00	29.0	.35	.35
		LASSO	.65	.95	.17	.17	9.4	.55	.57
		SCAD	.88	.98	.15	.23	9.5	.21	1.79
	SMLE	–	.89	.99	.75	.00	31.0	.46	2.45
		LASSO	.80	.97	.16	.22	9.5	.51	2.69
		SCAD	.89	.99	.16	.22	9.6	.17	2.89
S2	SIS	–	.55	.90	.81	.00	24.0	.34	3.93
		LASSO	.21	.77	.11	.14	4.45	.41	4.10
		SCAD	.15	.74	.16	.10	4.54	.27	4.26
	ISIS	–	.56	.90	.81	.00	24.0	.40	18.36
		LASSO	.25	.79	.14	.13	4.7	.40	18.53
		SCAD	.19	.75	.24	.08	5.1	.27	18.87
	FR	–	.20	.75	.84	.00	24.0	.58	41.35
		LASSO	.17	.73	.29	.08	5.38	.30	41.50
		SCAD	.13	.70	.33	.05	5.44	.22	41.77
	LASSO	–	.78	.95	.79	.00	22.5	.34	.10
		LASSO	.28	.80	.13	.15	4.7	.40	.27
		SCAD	.17	.75	.23	.08	5.0	.29	.82
	SMLE	–	.60	.90	.81	.00	24.0	.43	.95
		LASSO	.27	.80	.14	.14	4.7	.38	1.12
		SCAD	.14	.73	.24	.06	5.0	.24	1.58
S3	SIS	–	.00	.39	.93	.00	21.0	.69	.69
		LASSO	.00	.30	.88	.00	9.8	.85	.87
		SCAD	.00	.31	.87	.00	9.8	.81	.89
	ISIS	–	.65	.82	.84	.00	21.0	.52	4.49
		LASSO	.15	.54	.77	.00	9.8	.81	4.66
		SCAD	.58	.72	.66	.01	9.1	.40	5.02
	FR	–	.85	.87	.83	.00	21.0	.54	6.07
		LASSO	.72	.83	.62	.00	9.2	.63	6.22
		SCAD	.84	.86	.62	.00	9.4	.23	6.86
	LASSO	–	.25	.66	.87	.00	20.0	.52	.01
		LASSO	.02	.38	.84	.00	9.8	.85	.18
		SCAD	.24	.51	.73	.02	9.1	.62	.21
	SMLE	–	.92	.97	.81	.00	21.0	.43	.25
		LASSO	.48	.81	.64	.01	9.5	.76	.41
		SCAD	.76	.90	.59	.01	9.1	.23	.72