Stable Prediction in High-Dimensional Linear Models-Supplementary Materials

Bingqing Lin[1], Qihua Wang[1,2], Jun Zhang[1] and Zhen Pang[3]

[1]Institute of Statistical Sciences, Shenzhen University, Shenzhen 518060 China

[2]Academy of Mathematics and Systems Science, Chinese Academy of Sciences Beijing 100190, China

[3]Department of Applied Mathematics, the Hong Kong Polytechnic University Hong Kong

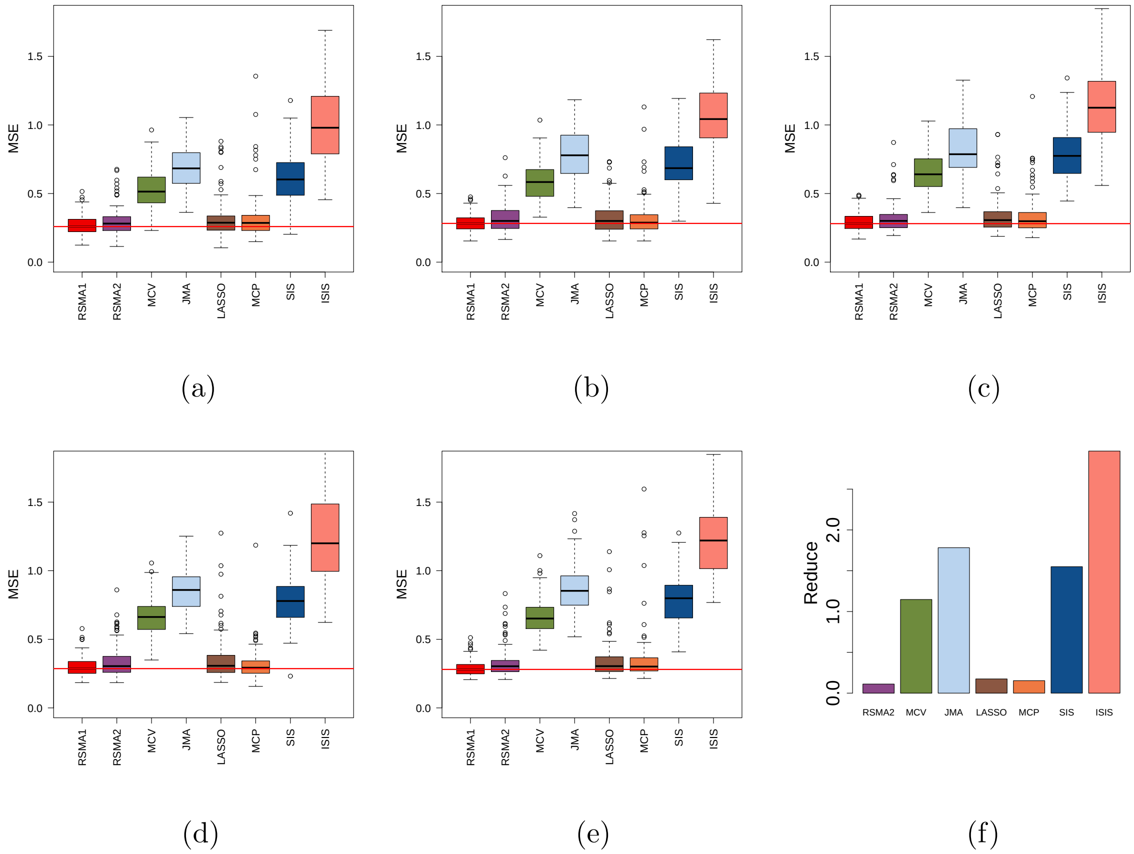
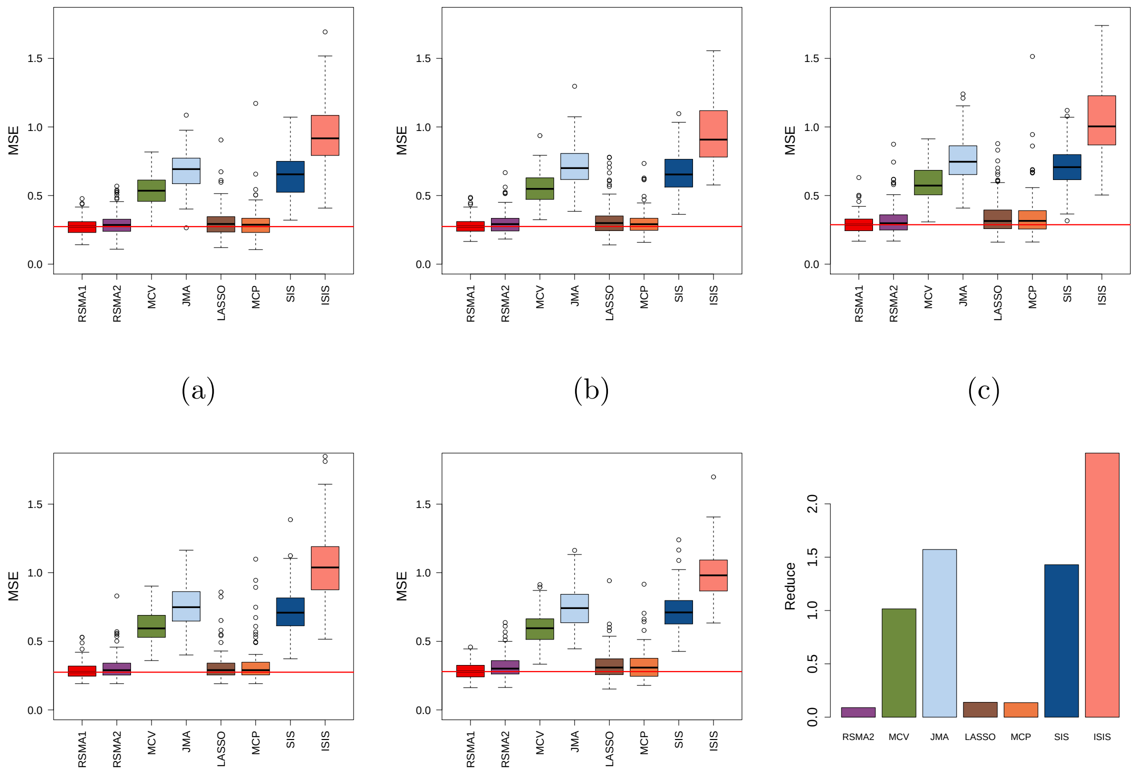


Figure 1: MSE for Example 2 with $\rho = -0.7$ and $p = 500$. Subfigures (a)-(e) are results at h from 1 to 5. The red horizontal lines represent the median values of MSE for RSMA1. Subfigure (f) shows the average of relative MSE reduction of RSMA1 compared with other methods considered.



(a) (b) (c)

(d) (e) (f)

Figure 2: MSE for Example 2 with $\rho = 0$ and $p = 500$. Subfigures (a)-(e) are results at h from 1 to 5. The red horizontal lines represent the median values of MSE for RSMA1. Subfigure (f) shows the average of relative MSE reduction of RSMA1 compared with other methods considered.

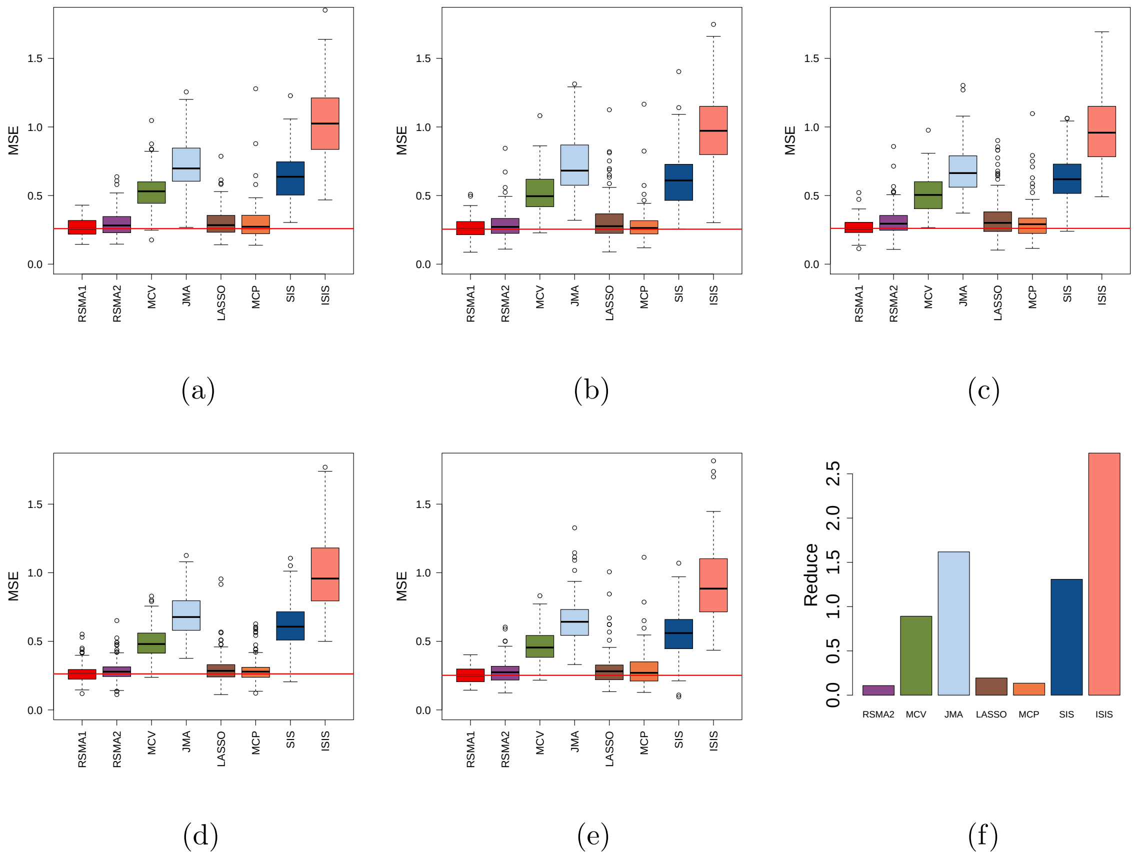


Figure 3: MSE for Example 2 with $\rho = 0.7$ and $p = 500$. Subfigures (a)-(e) are results at h from 1 to 5. The red horizontal lines represent the median values of MSE for RSMA1. Subfigure (f) shows the average of relative MSE reduction of RSMA1 compared with other methods considered.

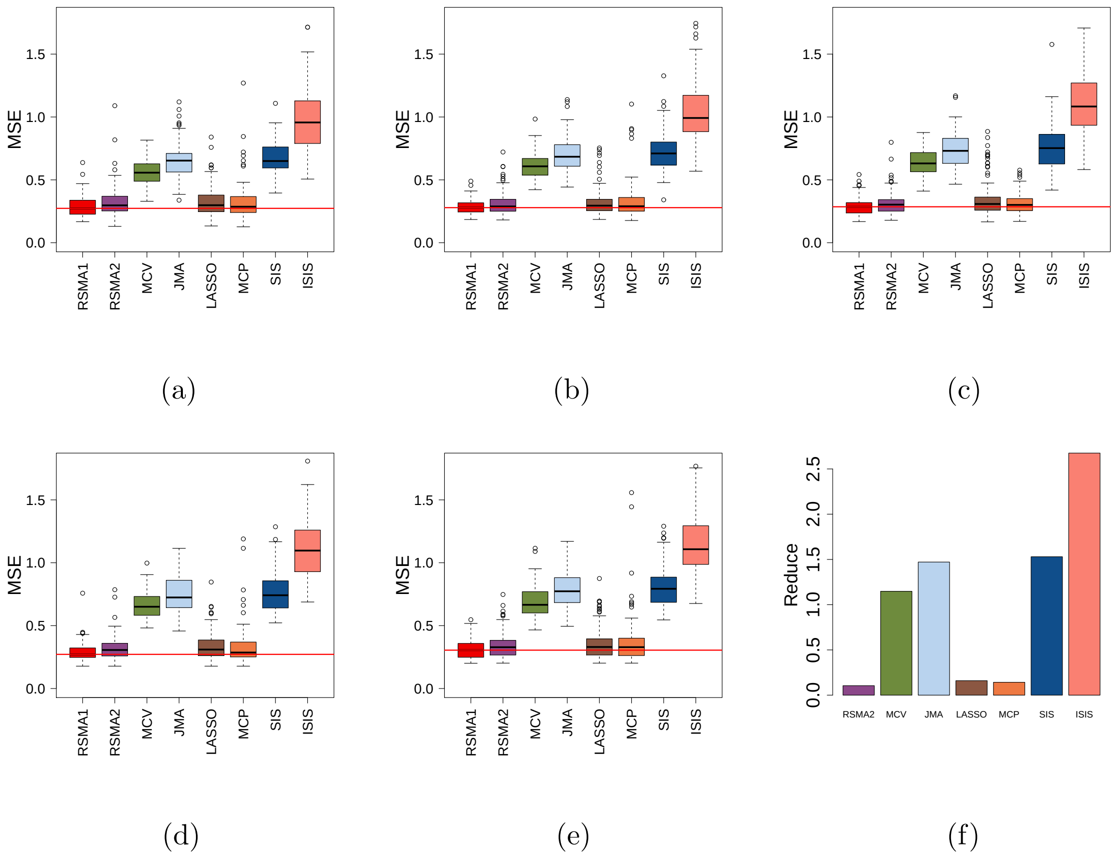


Figure 4: MSE for Example 2 with $\rho = -0.7$ and $p = 2000$. Subfigures (a)-(e) are results at h from 1 to 5. The red horizontal lines represent the median values of MSE for RSMA1. Subfigure (f) shows the average of relative MSE reduction of RSMA1 compared with other methods considered.

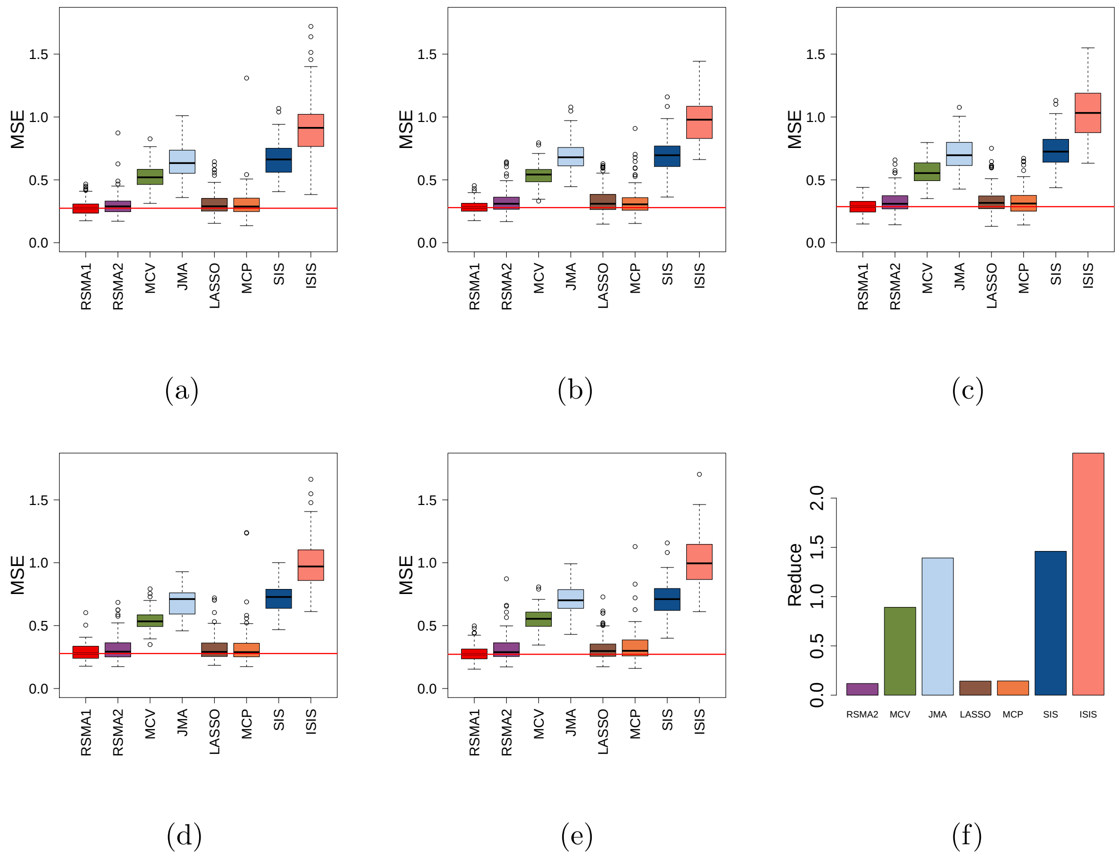


Figure 5: MSE for Example 2 with $\rho = 0$ and $p = 2000$. Subfigures (a)-(e) are results at h from 1 to 5. The red horizontal lines represent the median values of MSE for RSMA1. Subfigure (f) shows the average of relative MSE reduction of RSMA1 compared with other methods considered.

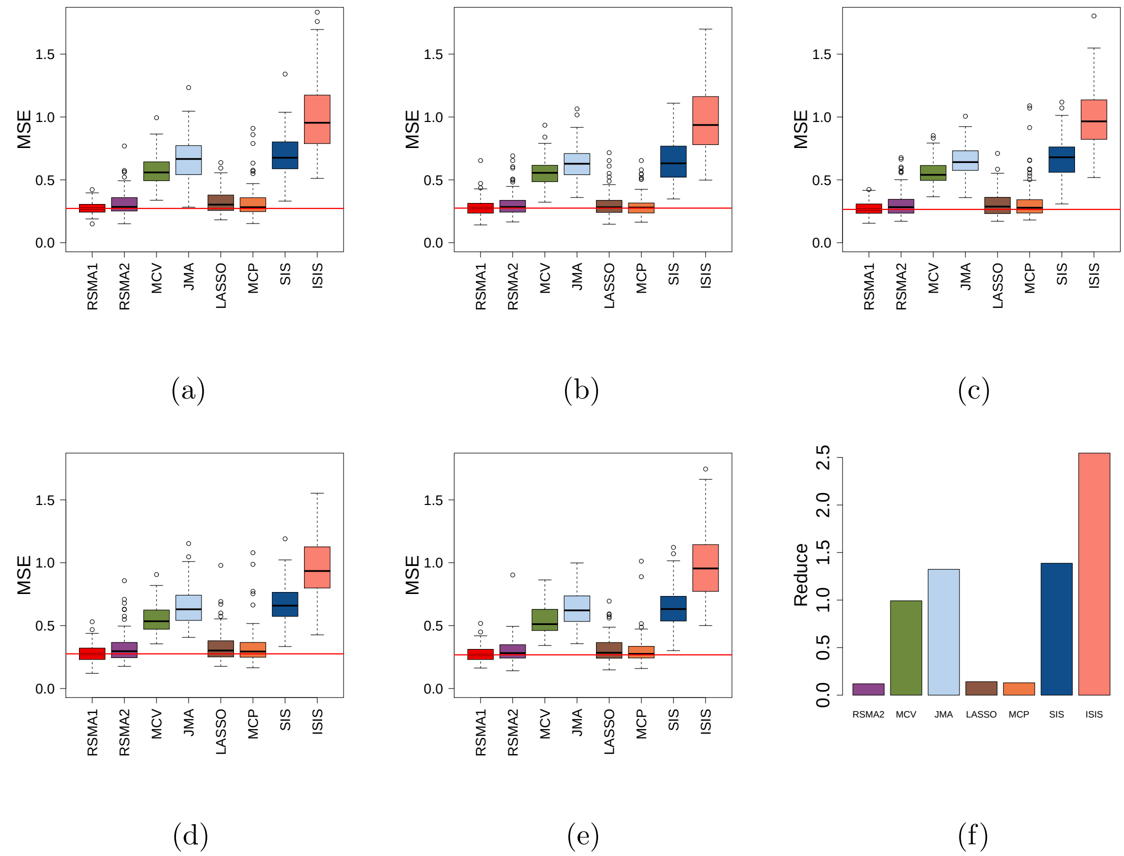
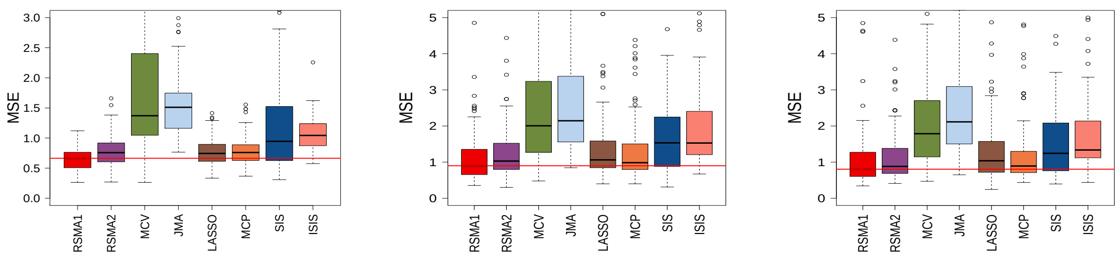


Figure 6: MSE for Example 2 with $\rho = 0.7$ and $p = 2000$. Subfigures (a)-(e) are results at h from 1 to 5. The red horizontal lines represent the median values of MSE for RSMA1. Subfigure (f) shows the average of relative MSE reduction of RSMA1 compared with other methods considered.

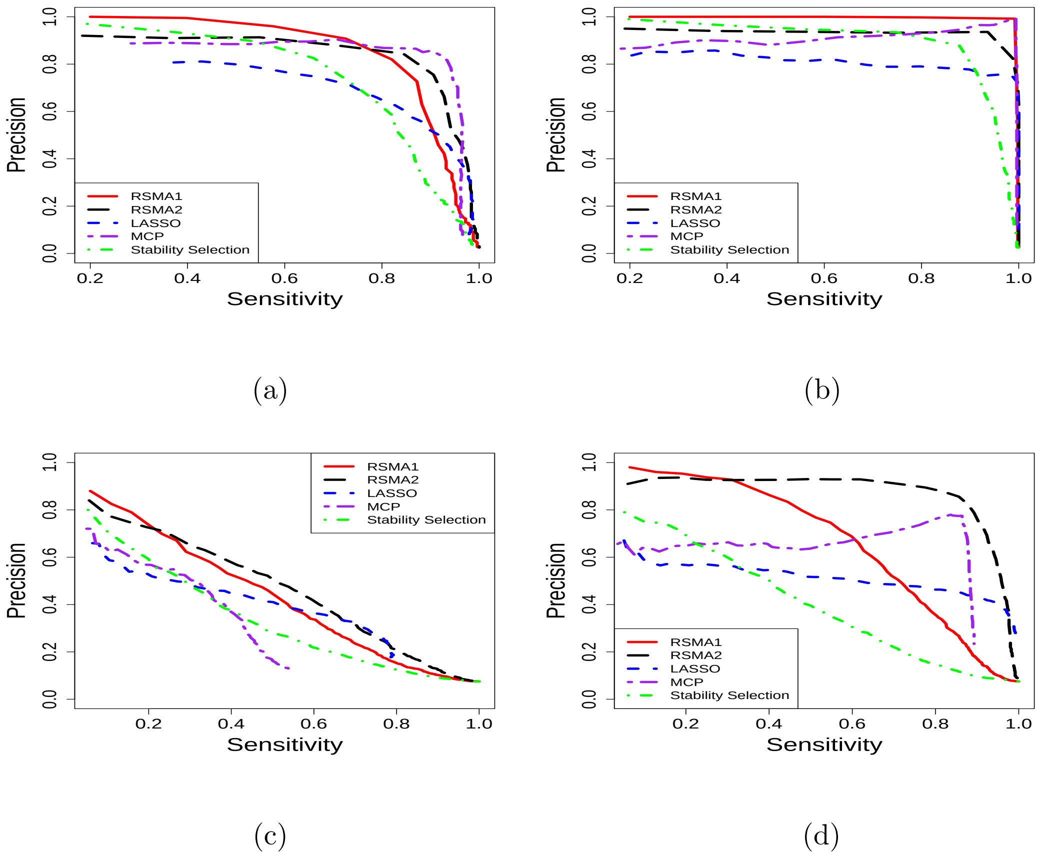
Table 1: Simulation results for Example 4, $p = 500, 1000$. Scenario A: the error term has a t distribution with 3 degrees of freedom; Scenario B: the error term has a normal distribution $N(0, \sigma_i^2)$ with $\sigma_i = x_{1i}^2$. $10\times$ standard errors are reported in parentheses. The smallest MSEs and standard errors are marked in bold.

	p	RSMA1	RSMA2	MCV	JMA	LASSO	MCP	SIS	ISIS
Scenario A	500	0.84	**0.78**	1.03	2.23	1.10	1.37	1.00	1.94
		(0.57)	**(0.53)**	(0.69)	(1.37)	(0.67)	(0.78)	(1.00)	(1.20)
	1000	0.92	**0.81**	1.59	2.29	1.24	1.40	1.13	1.88
		(0.59)	**(0.57)**	(0.97)	(1.54)	(0.74)	(0.91)	(1.02)	(1.20)
Scenario B	500	0.77	**0.71**	0.98	2.09	1.04	1.21	0.88	1.67
		(0.42)	**(0.42)**	(0.51)	(0.69)	(0.61)	(0.68)	(0.69)	(0.93)
	1000	0.91	**0.85**	1.52	2.25	1.27	1.56	1.12	1.90
		(0.42)	(0.44)	(0.51)	(1.06)	(0.61)	(0.68)	(0.69)	(0.93)



(a) (b) (c)

Figure 7: Simulation results for Example 5 with three distributions: (a) normal distribution $N(0, 1.5)$; (b) t distribution with 3 degrees of freedom; (c) mixture distribution of normal and t distribution with degrees of freedom 2.



(a) (b) (c) (d)

Figure 8: The precision-sensitivity curves comparing the ranking ability of all methods on the semi-real dataset for Leukaemia data with $p = 200$ and error term following normal distribution. (a) SNR=2 and s=5; (b) SNR=8 and s=5; (c) SNR=2 and s=15; (d) SNR=8 and s=15.

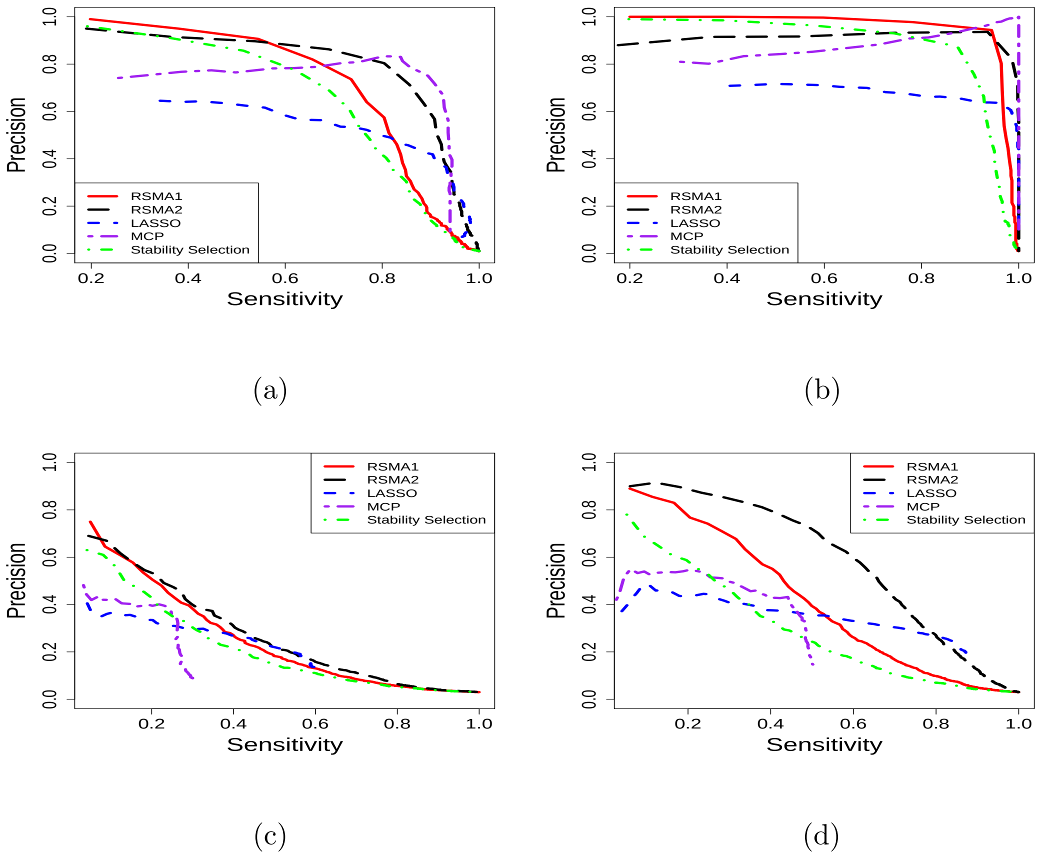


Figure 9: The precision-sensitivity curves comparing the ranking ability of all methods on the semi-real dataset for Leukaemia data with $p = 500$ and error term following normal distribution. (a) SNR=2 and s=5; (b) SNR=8 and s=5; (c) SNR=2 and s=15; (d) SNR=8 and s=15.

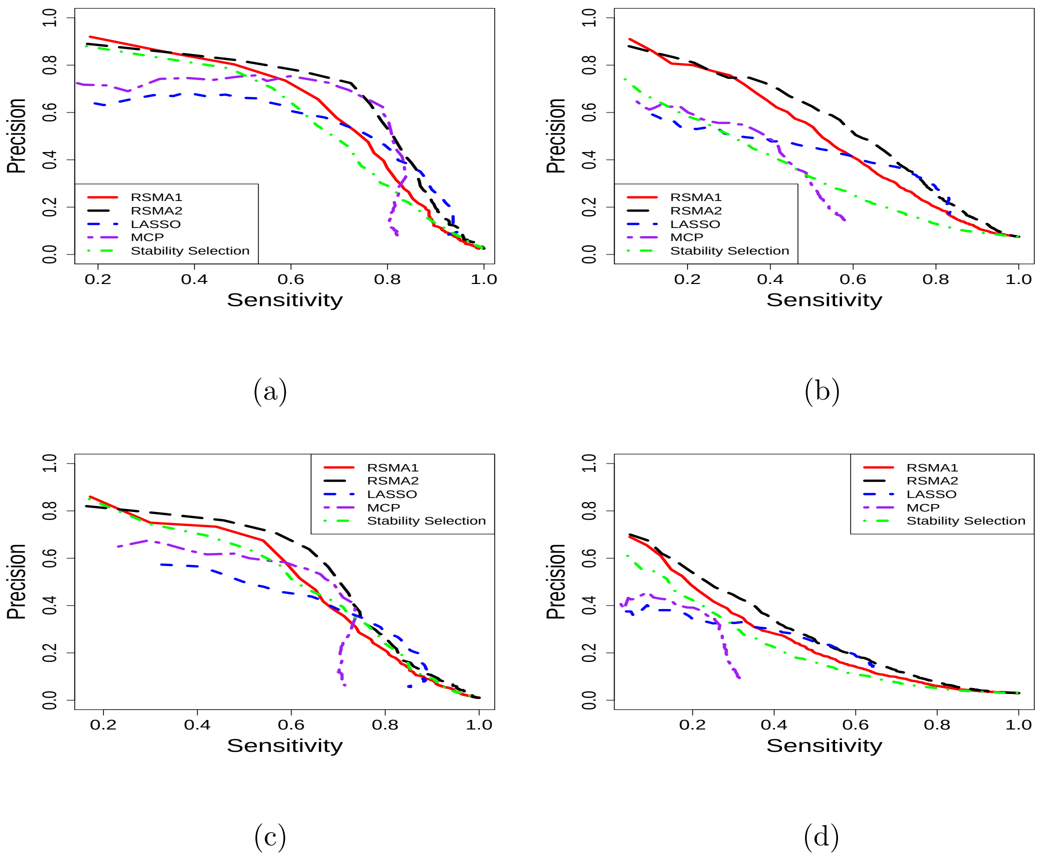


Figure 10: The precision-sensitivity curves comparing the ranking ability of all methods on the semi-real dataset for Leukaemia data with $p = 200, 500$ and error term following t distribution with degrees of freedom 3. (a) p=200 and s=5; (b) p=200 and s=15; (c) p=500 and s=5; (d) p=500 and s=15.