Web-based Supplementary Materials for "A Sparse Ising Model with Covariates" by Jie Cheng, Elizaveta Levina, Pei Wang and Ji Zhu

Appendix A

Proof of Theorem 1

For notational convenience, we omit the j indexing each separate regression. Following the literature, we prove the main theorem in two steps: first, we prove the result holds when assumptions **A1** and **A2** hold for $\boldsymbol{I}^n$ and $\boldsymbol{U}^n$, the sample versions of of $\boldsymbol{I}^*$ and $\boldsymbol{U}^*$ defined in (7) and (8) (Proposition 1). Then we show that if **A1** and **A2** hold for the population versions $\boldsymbol{I}^*$ and $\boldsymbol{U}^*$, they also hold for $\boldsymbol{I}^n$ and $\boldsymbol{U}^n$ with high probability (Proposition 2). The sample quantities $\boldsymbol{I}^n$ and $\boldsymbol{U}^n$ are defined as

$$\begin{aligned} \boldsymbol{I}^n &= \nabla^2 \ell(\boldsymbol{\theta}^*, \mathcal{D}_n) = \frac{1}{n}\sum_{i=1}^{n}\left(p_j^i(1-p_j^i)(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})^T\right) \ , \\ \boldsymbol{U}^n &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})^T \ . \end{aligned}$$

PROPOSITION 1: If ***A1*** and ***A2*** are satisfied by $\boldsymbol{I}^n$ and $\boldsymbol{U}^n$, assume moreover that

$$\begin{aligned} M_n &= \sup \|\boldsymbol{x}\|_\infty < \infty \ \text{ a.s.}, \\ \lambda_n &\geqslant \frac{8M_n(2-\alpha)}{\alpha}\sqrt{\frac{\log p + \log q}{n}} \ , \\ n &> Cd^2(\log p + \log q) \ . \end{aligned}$$

Then with probability at least $1 - 2\exp\left(-C\frac{\lambda_n^2 n}{M_n^2}\right)$, the result of Theorem 1 holds.

Proof of Proposition 1. The proof requires several steps. The uniqueness part follows directly from the following lemma:

LEMMA 1: (Shared sparsity and uniqueness of $\hat{\boldsymbol{\theta}}$, Ravikumar et al. (2010)). *Define the* *sign vector* $\boldsymbol{t}$ *for* $\boldsymbol{\theta}$ *to satisfy the following properties,*

$$\begin{cases} \hat{t}_k = sign(\hat{\theta}_k), & \text{if } \hat{\theta}_k \neq 0 \ , \\ |\hat{t}_k| \leqslant 1, & \text{if } \hat{\theta}_k = 0 \ . \end{cases}$$

Suppose there exists an optimal solution $\hat{\boldsymbol{\theta}}$ *with sign* $\hat{\boldsymbol{t}}$ *defined as above, such that,* $\|\hat{\boldsymbol{t}}_{\mathcal{S}^C}\|_\infty < 1$*, then any optimal solution* $\tilde{\boldsymbol{\theta}}$ *must have* $\tilde{\boldsymbol{\theta}}_{\mathcal{S}^C} = 0$*. Furthermore, if the Hessian matrix* $\nabla^2 \ell(\hat{\boldsymbol{\theta}})_{\mathcal{SS}}$ *is strictly positive definite, then* $\hat{\boldsymbol{\theta}}$ *is the unique solution.*

We now proceed to prove the rest of Proposition 1. For $\hat{\boldsymbol{\theta}}$ to be a solution of (9), the sub-gradient at $\hat{\boldsymbol{\theta}}$ must be 0, i.e.,

$$\nabla \ell(\hat{\boldsymbol{\theta}}, \mathcal{D}_n) + \lambda_n \hat{\boldsymbol{t}} = 0 \ . \tag{A.1}$$

Then we can write $\nabla \ell \left(\hat{\boldsymbol{\theta}}, \mathcal{D}_n\right) - \nabla \ell \left(\boldsymbol{\theta}^*, \mathcal{D}_n\right) = -\lambda_n \hat{\boldsymbol{t}} + W^n$, where

$$W^n = -\nabla \ell \left(\boldsymbol{\theta}^*, \mathcal{D}_n\right) = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})(y^i_j - p^i_j(\boldsymbol{\theta}^*)) \ .$$

Let $\tilde{\boldsymbol{\theta}}$ denote a point in the line segment connecting $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$. Applying the mean value theorem gives

$$\boldsymbol{I}^n \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) = W^n - \lambda_n \hat{\boldsymbol{t}} + R^n \ . \tag{A.2}$$

where $R^n = \left(\nabla^2 \ell \left(\boldsymbol{\theta}^*, \mathcal{D}_n\right) - \nabla^2 \ell(\tilde{\boldsymbol{\theta}}, \mathcal{D}_n)\right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$.

Now define $\hat{\boldsymbol{\theta}}$ as follows: let $\mathcal{S}$ be the index set of true non-zeros in $\boldsymbol{\theta}^*$, let $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ be the solution of

$$\min_{(\hat{\boldsymbol{\theta}}_{\mathcal{S}}, 0)} \ell(\hat{\boldsymbol{\theta}}, \mathcal{D}_n) + \lambda_n \|\hat{\boldsymbol{\theta}}_{\mathcal{S}}\|_1 \ , \tag{A.3}$$

and let $\hat{\boldsymbol{\theta}}_{\mathcal{S}^C} = 0$. We will show that this $\hat{\boldsymbol{\theta}}$ is the optimal solution and is sign consistent with high probability.

We set the corresponding sign vector $\hat{\boldsymbol{t}}_{\mathcal{S}}$ for $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ similarly defined as in Lemma 1, and $\hat{\boldsymbol{t}}_{\mathcal{S}^C} = -\frac{1}{\lambda_n}\nabla_{\mathcal{S}^C}\ell(\hat{\boldsymbol{\theta}}_{\mathcal{S}}, \mathcal{D}_n)$ as obtained in (A.1). Now we need to show that with high probability,

$$\|\hat{\boldsymbol{t}}_j\|_\infty < 1, \qquad \text{for } j \in \mathcal{S}^C \tag{A.4}$$

$$\hat{\boldsymbol{t}}_j = sign(\boldsymbol{\theta}^*_j), \quad \text{for } j \in \mathcal{S} \text{ and } \|\theta^*_j\| \geqslant \frac{10\lambda_n\sqrt{d}}{\Delta_{\min}} \tag{A.5}$$

The following three lemmas form the proof.

Lemma 2: (Control the remainder term W^n). *For $\alpha \in (0,1]$, assume $\|\boldsymbol{x}\|_\infty \leqslant M_n$ a.s, then,*

$$P\left(\frac{2-\alpha}{\lambda_n}\|W^n\|_\infty \geqslant \frac{\alpha}{4}\right) \leqslant 4\exp\left(-\frac{\lambda_n^2 n\alpha^2}{32M_n^2(2-\alpha)^2} + \log p + \log q\right) .$$

This probability goes to 0 as long as $\lambda_n \geqslant 8M\frac{2-\alpha}{\alpha}\sqrt{\frac{\log p + \log q}{n}}$.

Proof of Lemma 2. We can write $W^n = \frac{1}{n}\sum_{i=1}^n(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\backslash j})(y^i_j - p^i_j(\boldsymbol{\theta}^*)) = \sum_{i=1}^n Z_i$, where Z_{ik} is bounded by M_n/n. Thus by Azuma-Hoeffding Inequality,

$$\begin{aligned} P\left(\|W^n\|_\infty \geqslant \frac{\lambda_n\alpha}{4(2-\alpha)}\right) &\leqslant 2pqP\left(\|W^n_k\|_\infty \geqslant \frac{\lambda_n\alpha}{4(2-\alpha)}\right) \\ &\leqslant 4\exp\left(-\frac{\lambda_n^2 n\alpha^2}{32M_n^2(2-\alpha)^2} + \log p + \log q\right) . \end{aligned}$$

□

Lemma 3: (ℓ_2-consistency of the sub-vector $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$). *If $\lambda_n d < \frac{\Delta^2_{\min}}{10\Delta_{\max}M_n}$, and, $\|W^n\|_\infty \leqslant \frac{\lambda_n}{4}$, then*

$$\|\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}\|_2 \leqslant \frac{5\lambda_n\sqrt{d}}{\Delta_{\min}} .$$

Proof of Lemma 3. Let $G(u_{\mathcal{S}}) = \ell(\boldsymbol{\theta}^*_{\mathcal{S}} + u_{\mathcal{S}}, \mathcal{D}_n) - \ell(\boldsymbol{\theta}^*_{\mathcal{S}}, \mathcal{D}_n) + \lambda_n(\|\boldsymbol{\theta}^*_{\mathcal{S}} + u_{\mathcal{S}}\|_1 - \|\boldsymbol{\theta}^*_{\mathcal{S}}\|_1)$ be a function $G : \mathbb{R}^d \to \mathbb{R}$. It is easy to see that $G(u_{\mathcal{S}})$ is convex and it achieves its minimum at $\hat{u}_{\mathcal{S}} = \hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}$. Moreover, $G(0) = 0$. Thus if we can show that $G(u_{\mathcal{S}})$ is positive on the set $\|u_{\mathcal{S}}\|_2 = B$, then we will have $\hat{u}_{\mathcal{S}} \leqslant B$ due to convexity of $G(u_{\mathcal{S}})$. Note that

$$G(u_{\mathcal{S}}) = -W^{nT}_{\mathcal{S}}u_{\mathcal{S}} + u^T_{\mathcal{S}}\nabla^2\ell(\boldsymbol{\theta}^*_{\mathcal{S}} + \alpha u_{\mathcal{S}})u_{\mathcal{S}} + \lambda_n(\|\boldsymbol{\theta}^*_{\mathcal{S}} + u_{\mathcal{S}}\|_1 - \|\boldsymbol{\theta}^*_{\mathcal{S}}\|_1) .$$

Further,

$$\begin{aligned} |W_{\mathcal{S}}^{nT} u_{\mathcal{S}}| &\leqslant \|W^n\|_\infty \|u_{\mathcal{S}}\|_1 \leqslant \frac{\lambda_n}{4}\sqrt{d}\|u_{\mathcal{S}}\|_2 \ , \\ \Lambda_{\min}(\nabla^2 \ell(\boldsymbol{\theta}^*_{\mathcal{S}} + \alpha u_{\mathcal{S}})) &\geqslant \Delta_{\min} - \Delta_{\max} M_n \sqrt{d}\|u_{\mathcal{S}}\|_2 \ , \\ |\lambda_n(\|\boldsymbol{\theta}^*_{\mathcal{S}} + u_{\mathcal{S}}\|_1 - \|\boldsymbol{\theta}^*_{\mathcal{S}}\|_1)| &\leqslant \lambda_n \sqrt{d}\|u_{\mathcal{S}}\|_2 \ . \end{aligned}$$

Combining all of the above, we have

$$G(u_{\mathcal{S}}) \geqslant \|u_{\mathcal{S}}\|_2(-\Delta_{\max} M_n \sqrt{d}\|u_{\mathcal{S}}\|_2^2 + \Delta_{\min}\|u_{\mathcal{S}}\|_2 - \frac{5}{4}\lambda_n\sqrt{d}) \ .$$

Easy algebra shows that if $\lambda_n d \leqslant \frac{\Delta_{\min}^2}{10\Delta_{\max} M_n}$ and $B = \frac{5\lambda_n\sqrt{d}}{\Delta_{\min}}$, the result follows. □

LEMMA 4: (Control the remainder term R^n). *If* $\lambda_n d \leqslant \frac{\Delta_{\min}^2}{100 M_n \Delta_{\max}}\frac{\alpha}{2-\alpha}$, $\|W^n\|_\infty \leqslant \frac{\lambda_n}{4}$, *then*

$$\frac{\|R^n\|_\infty}{\lambda_n} \leqslant \frac{25\Delta_{\max}}{\Delta_{\min}^2} M_n \lambda_n d \leqslant \frac{\alpha}{4(2-\alpha)} \ .$$

Proof of Lemma 4. Recall that

$$\begin{aligned} R^n &= \left(\nabla^2 \ell\left(\boldsymbol{\theta}^*, \mathcal{D}_n\right) - \nabla^2 \ell\left(\tilde{\boldsymbol{\theta}}, \mathcal{D}_n\right)\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \\ &= \frac{1}{n}\sum_{i=1}^n \left(p_j^i(\boldsymbol{\theta}^*)(1-p_j^i(\boldsymbol{\theta}^*)) - p_j^i(\tilde{\boldsymbol{\theta}})(1-p_j^i(\tilde{\boldsymbol{\theta}}))\right)(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \ . \end{aligned}$$

Let $\omega_j^i(\boldsymbol{\theta}) = p_j^i(\boldsymbol{\theta})(1-p_j^i(\boldsymbol{\theta}))$. The k-th element of R^n has the form

$$\begin{aligned} R_k^n &= \frac{1}{n}\sum_{i=1}^n (\omega_j^i(\boldsymbol{\theta}^*) - \omega_j^i(\tilde{\boldsymbol{\theta}})) Z_k^i (\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \\ &= \frac{1}{n}\sum_{i=1}^n \dot{\omega}_j^i(\bar{\boldsymbol{\theta}}) Z_k^i \left(\boldsymbol{\theta}^* - \tilde{\boldsymbol{\theta}}\right)^T (\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \ , \end{aligned}$$

where $Z_k^i = x_l^i y_m^i$, for some (l, m). By **A1** and Lemma 3, we have

$$|R_k^n| \leqslant M_n \Delta_{\max}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2^2 \leqslant M_n \Delta_{\max}\left(\frac{5\lambda_n\sqrt{d}}{\Delta_{\min}}\right)^2 \ .$$

□

Putting all the lemmas together, we are ready to prove Proposition 1.

Proof of Proposition 1. Set $\lambda_n = \frac{8M_n(2-\alpha)}{\alpha}\sqrt{\frac{\log p + \log q}{n}}$. By Lemma 2, we have $\|W^n\|_\infty \leqslant$

$\frac{\lambda_n \alpha}{4(2-\alpha)} \leqslant \frac{\lambda_n}{4}$ with probability at least $1-4\exp(C\lambda_n^2 n/M_n^2)$. Choosing $n \geqslant \frac{100^2 \Delta_{\max}^2 (2-\alpha)^2}{\Delta_{\min}^4 \alpha^2} d^2(\log p + \log q))$, we have $\lambda_n d \leqslant \frac{\Delta_{\min}^2}{100 M_n \Delta_{\max}} \frac{\alpha}{2-\alpha}$, thus the conditions of Lemmas 3 and 4 hold.

By rewriting (A.2) and utilizing the fact that $\hat{\boldsymbol{\theta}}_{\mathcal{S}^C} = \boldsymbol{\theta}^*_{\mathcal{S}^C} = 0$, we have

$$\boldsymbol{I}^n_{\mathcal{S}^C\mathcal{S}}(\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}) = W^n_{\mathcal{S}^C} - \lambda_n \hat{\boldsymbol{t}}_{\mathcal{S}^C} + R^n_{\mathcal{S}^C} \ , \qquad (A.6)$$

$$\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}}(\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}) = W^n_{\mathcal{S}} - \lambda_n \hat{\boldsymbol{t}}_{\mathcal{S}} + R^n_{\mathcal{S}} \ . \qquad (A.7)$$

Since $\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}}$ is invertible by assumption, combining (A.6) and (A.7) gives

$$\boldsymbol{I}^n_{\mathcal{S}^C\mathcal{S}}(\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}})^{-1}(W^n_{\mathcal{S}} - \lambda_n \hat{\boldsymbol{t}}_{\mathcal{S}} + R^n_{\mathcal{S}}) = W^n_{\mathcal{S}^C} - \lambda_n \hat{\boldsymbol{t}}_{\mathcal{S}^C} + R^n_{\mathcal{S}^C} \ . \qquad (A.8)$$

To show (A.4), we reorganize (A.8) and use results from Lemmas 2 and 4:

$$\begin{aligned} \lambda_n \|\hat{\boldsymbol{t}}_{\mathcal{S}^C}\|_\infty &= \|\boldsymbol{I}^n_{\mathcal{S}^C\mathcal{S}}(\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}})^{-1}(W^n_{\mathcal{S}} - \lambda_n \hat{\boldsymbol{t}}_{\mathcal{S}} + R^n_{\mathcal{S}}) - W^n_{\mathcal{S}^C} - R^n_{\mathcal{S}^C}\|_\infty \\ &\leqslant \|\boldsymbol{I}^n_{\mathcal{S}^C\mathcal{S}}(\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}})^{-1}\|_\infty(\|W^n\|_\infty + \lambda_n + \|R^n\|_\infty) + \|W^n\|_\infty + \|R^n\|_\infty \\ &\leqslant \lambda_n(1 - \frac{\alpha}{2}) \ . \end{aligned}$$

To show (A.5), it suffices to show that $\|\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}\|_\infty \leqslant \frac{\boldsymbol{\theta}^*_{\min}}{2}$. By Lemma 3,

$$\|\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}\|_\infty \leqslant \frac{5\lambda_n \sqrt{d}}{\Delta_{\min}} \leqslant \frac{\boldsymbol{\theta}^*_{\min}}{2} \ .$$

The last inequality follows as long as $\boldsymbol{\theta}^*_{\min} \geqslant \frac{10\lambda_n\sqrt{d}}{\Delta_{\min}}$. This completes the proof of Proposition 1. □

Proposition 2: If $\boldsymbol{I}^*$ and $\boldsymbol{U}^*$ satisfy $\boldsymbol{A1}$ and $\boldsymbol{A2}$, and $M_n = \sup\|\boldsymbol{x}\|_\infty < \infty$ a.s., the following hold for any $\delta > 0$. A and B are some positive constants.

$$P\left\{\Lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})(\boldsymbol{x}^i \otimes \boldsymbol{y}^i_{\setminus j})^T\right) \geqslant D_{\max} + \delta\right\} \leqslant 2\exp\left(-A\frac{\delta^2 n}{M_n^2 d^2} + B(\log p + \log q)\right)$$

$$\begin{aligned} P\left(\Lambda_{\min}(\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}}) \leqslant C_{\min} - \delta\right) &\leqslant 2\exp\left(-A\frac{\delta^2 n}{M_n^2 d^2} + B\log d\right) \\ P\left(|||\boldsymbol{I}^n_{\mathcal{S}^c\mathcal{S}}\left(\boldsymbol{I}^n_{\mathcal{S}\mathcal{S}}\right)^{-1}|||_\infty \geqslant 1 - \frac{\alpha}{2}\right) &\leqslant \exp\left(-A\frac{n}{M_n^2 d^3} + B(\log p + \log q)\right) \end{aligned}$$

We omit the proof of Proposition 2, which is very similar to Lemmas 5 and 6 in Ravikumar et al. (2010).

Proof of Theorem 1. With Propositions 1 and 2, the proof of Theorem 1 is straightforward. Given that A1 and A2 are satisfied by $\boldsymbol{I}^*$ and $\boldsymbol{U}^*$ and that conditions (13) and (14) hold, on the set $\mathcal{A} = \{\boldsymbol{x} : M_n = \sup \|\boldsymbol{x}\| < \infty\}$ the assumptions in Proposition 2 are satisfied. Thus with probability at least $1 - \exp(-\frac{C\lambda_n^2 n}{M_n^2})$, the conditions of Proposition 1 hold, and therefore the results in Theorem 1 hold. Finally, let $\mathcal{T}$ stand for the set where the results of Theorem 1 hold. Then by (11) and (12), we have

$$P(\mathcal{T}^c) \leqslant P(\mathcal{T}^c \mid \mathcal{A}) + P(\mathcal{A}^c) \leqslant \exp(-\frac{C\lambda_n^2 n}{M_n^2}) + \exp(-M_n^\delta) \leqslant \exp -(C'\lambda_n^2 n)^{\delta^*}, \text{where } 0 < \delta^* < 1.$$

□

Appendix B

Proof of Result (14). This result follows from that

$$\sqrt{\sum_{k:k\in\mathcal{S}_j} (\hat{\boldsymbol{\theta}}_{jk} - \boldsymbol{\theta}^*_{jk})^2} = \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}^*_j\|_2 \leqslant \frac{5\lambda_n\sqrt{d}}{\Delta_{\min}}, \quad \text{for any } j = 1, \dots q,$$

$$\Longrightarrow \quad \|\hat{\boldsymbol{\theta}}_{jk} - \boldsymbol{\theta}^*_{jk}\|_2 \leqslant \frac{5\lambda_n\sqrt{d}}{\Delta_{\min}} \quad \text{for any } j \neq k,$$

$$\Longrightarrow \quad \max(\|\hat{\boldsymbol{\theta}}_j^{\max} - \boldsymbol{\theta}^*_j\|_2, \|\hat{\boldsymbol{\theta}}_j^{\min} - \boldsymbol{\theta}^*_j\|_2) \leqslant \sqrt{d \max_{j\neq k} \|\hat{\boldsymbol{\theta}}_{jk} - \boldsymbol{\theta}^*_{jk}\|_2^2} \leqslant \frac{5\lambda_n d}{\Delta_{\min}}.$$

Appendix C

Simulation results for section 4.1 and section 4.2 with stability selection

We also investigated our model selection performance when coupled with stability selection for section 4.1 and 4.2. The settings are the same as in those sections and the stability selection results are shown in red curves. As shown from Figures 1 and 2, stability selection does not show much stronger model selection performance under these settings.

[Figure 1 about here.]

[Figure 2 about here.]

Appendix D

Simulation results for high-dimensional responses.

We also investigated the model selection performance of our methods when the dimension of the response is large. We set $p = 5$, $n = 100$ and $q = \{50, 100, 200\}$. The results are shown in Figure 3. As the dimension q increases, coupling stability selection with the joint approach and the separate-max approach significantly improve model selection performance over the respective original methods. Also the separate-min approach performs the best among the three approaches because it is conservative enough to eliminate many of the false positives.

[Figure 3 about here.]

Appendix E

Simulation results for KNN graphs.

We performed the sparsity effect analysis similar to section 4.1 on k-nearest neighbor graphs. The results are shown in Figure 4. The pattern is similar to section 4.1, where the model selection performance deteriorates as the model sparsity decreases both in terms of edges and effective number of covariates. The red curves represent the results of stability selection. Since the dimension of the variables are relatively low in this case, the improvement of model selection by applying stability selection is not significant.

[Figure 4 about here.]

Appendix F

Numerical results for the tumor suppressor gene study.

We applied stability selection to infer the stable set of important covariates for each pairwise conditional association. Table 1 records the rank list of the edges depending on different covariates. The first two columns of each covariate-related columns are the node names and the third column records the selection frequency. Further, for each node, each covariate, and each stability selection subsample, we computed the "covariate-specific" degree of a node. A ranking of nodes can then be produced for each covariate and each replication. Then we computed the median rank across all stability selection subsamples and ordered nodes by rank for each covariate. The results are listed in Table 2.

[Table 1 about here.]

[Table 2 about here.]

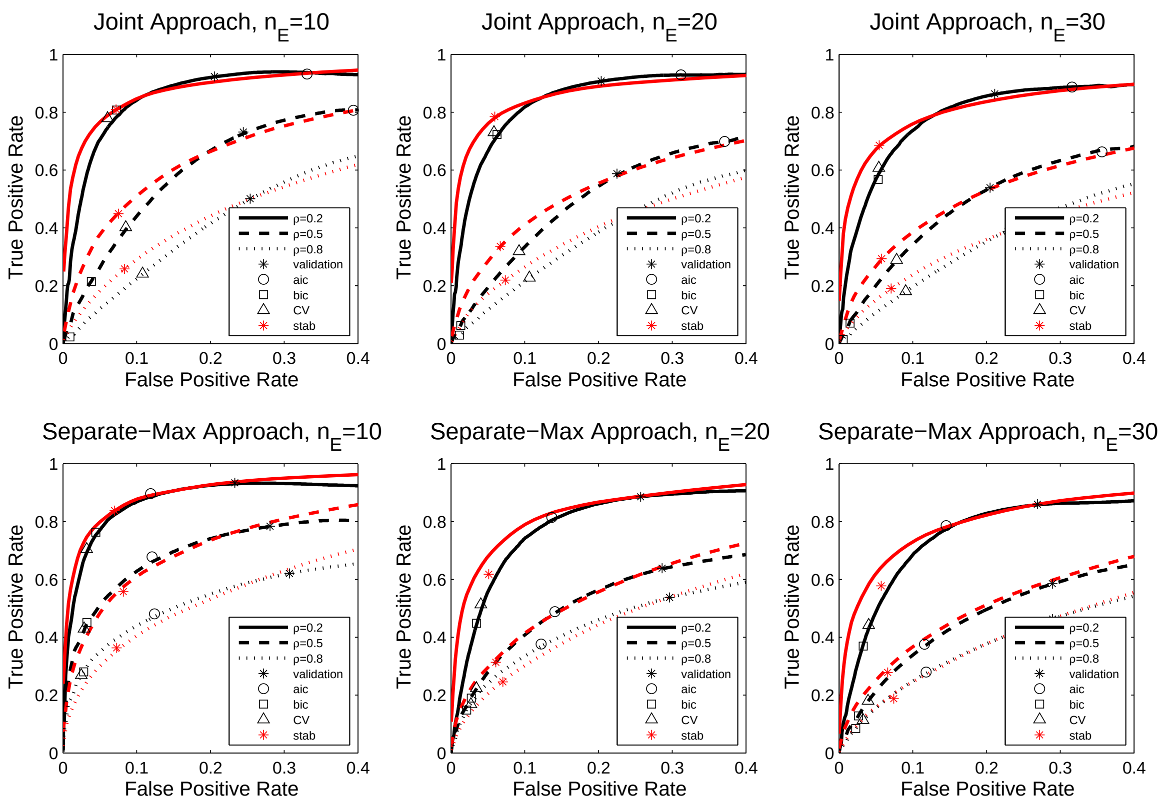

Figure 1. ROC curves for different sparsity settings.

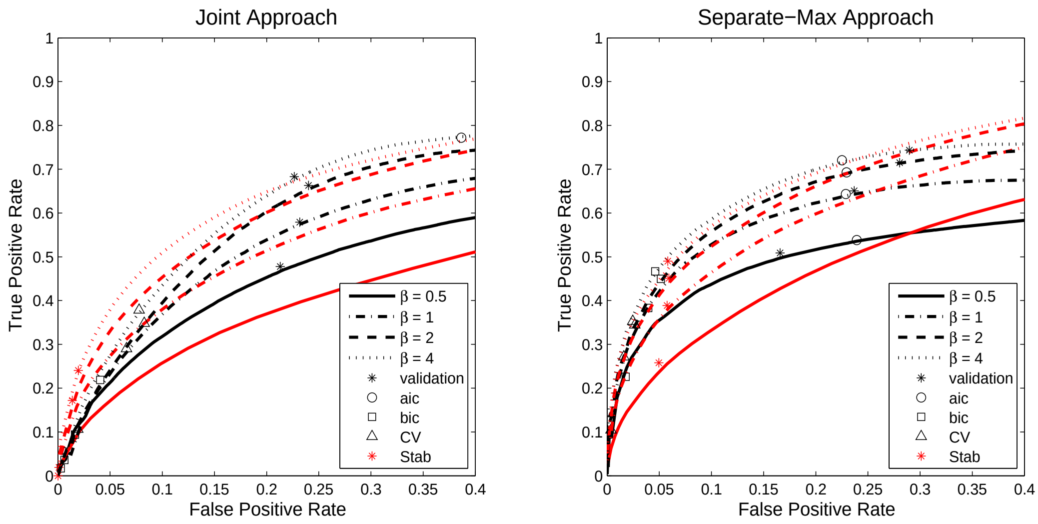

Figure 2. ROC curves for different scale settings.

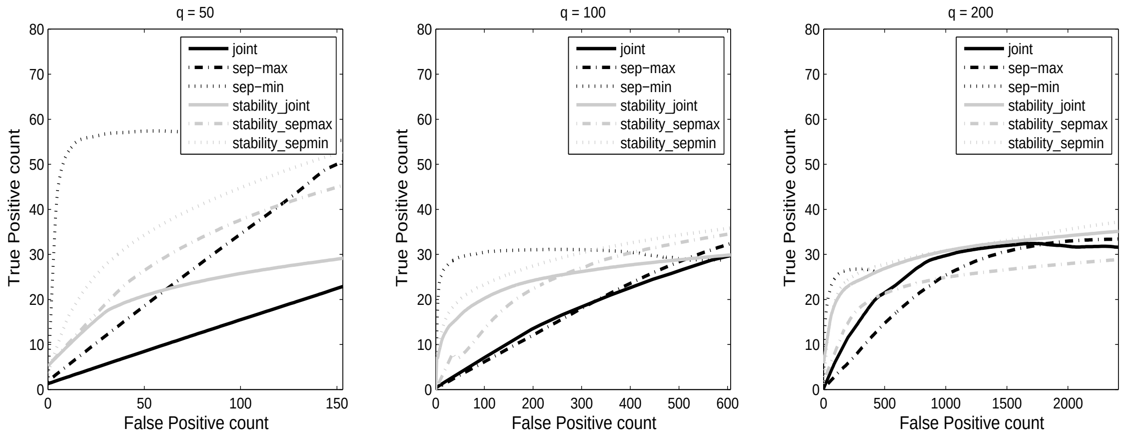

Figure 3. ROC curves for high-dimensional responses.

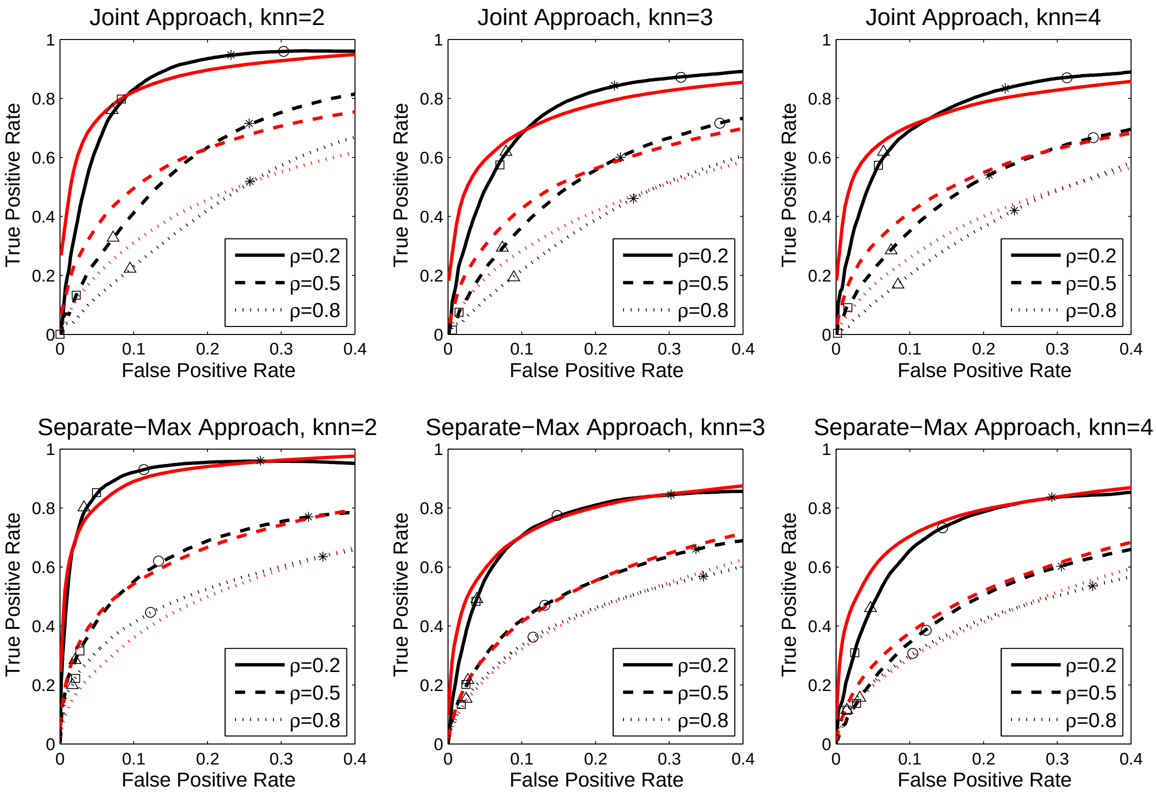

Figure 4. ROC curves for varying levels of sparsity, as measured by the parameter β. The red curve corresponds to stability selection over a grid of threshold for selection frequencies.

Table 1

Frequency-based ranked list of covariate-dependent inter-chromosomal interactions

Main effect			TP53 mutation status			ER status			Tumor Stage		
Gene1	Gene2	Freq	Gene1	Gene2	Freq	Gene1	Gene2	Freq	Gene1	Gene2	Freq
4q31.3	18q23	0.83	1p13.1	21q21.1	0.5	4q34.3	5q32	0.79	11q23.3	16q24.1	0.58
6p21.32	13q31.2	0.81	2p12	7p21.1	0.5	2p16.1	4q28.3	0.51	3p21.31	5q21.3	0.57
11p15.1	14q22.2	0.71	8p21.3	11q23.3	0.5	4q12	22q11.23	0.49	9q33.1	16q24.1	0.54
6q16.2	Xp11.23	0.68	3p22.1	6p21.31	0.48	10q11.21	16q23.2	0.49	2q24.1	16q24.1	0.46
2q34	3q13.31	0.64	4q25	8p11.22	0.48	11q23.3	15q13.1	0.48	3q23	8p11.22	0.46
1p36.11	2p21	0.61	4q32.1	17p13.1	0.46	5q22.2	9q21.13	0.46	10q26.3	17p11.2	0.45
1p31.1	2q32.2	0.6	6p21.32	8p11.22	0.46	9p24.1	10q22.1	0.46	3p12.1	12q23.1	0.44
2q32.1	12q12	0.59	16q23.1	18q21.32	0.46	8p21.3	12q22	0.45	7q21.11	22q11.21	0.44
6p21.32	Xp11.4	0.58	12q23.1	Xq23	0.45	2q22.1	9q33.3	0.44	4q35.1	15q22.1	0.43
9q31.3	14q24.3	0.58	2p21	Xp11.22	0.44	2q23.1	4q28.3	0.44	16q23.3	17p13.1	0.43
6p21.32	9q31.3	0.57	18q21.2	Xq23	0.44	3q24	22q11.23	0.44	3p12.3	7p21.3	0.42
2p25.2	15q26.2	0.56	2q24.3	4p14	0.43	1p31.1	8q21.13	0.43	4q22.3	7p21.3	0.41
6p21.32	13q21.1	0.54	6q26	9p21.3	0.43	1p13.1	8p11.22	0.43	9q34.13	10q11.21	0.41
10q25.3	12p13.31	0.54	12q23.1	Xq26.1	0.43	4q21.1	9q33.3	0.42	2q24.2	9q33.1	0.4
3p21.1	17p13.2	0.53	15q13.1	Xp22.32	0.43	4q25	10q22.1	0.42	2q24.2	16q24.1	0.4
12p13.31	17q11.2	0.53	4q13.1	5q23.3	0.42	4q32.2	7p21.1	0.41	13q33.1	Xq13.3	0.4
2q32.1	6q14.1	0.52	4q35.1	15q23	0.42	2p25.2	3p22.2	0.38	2q32.1	15q15.1	0.39
5q33.1	11p15.4	0.52	6q14.1	13q22.1	0.42	10q25.1	12q23.3	0.38	3q13.13	7p21.3	0.39
9q34.13	22q11.21	0.52	8p11.22	15q21.2	0.42	4p14	10q22.1	0.37	4q31.22	9q22.33	0.39
1p34.2	2p24.1	0.51	2p16.2	13q34	0.41	4q12	12p12.3	0.37	11q23.2	13q21.1	0.39
2q24.1	3q13.31	0.51	3p21.31	17q21.2	0.41	4q12	15q13.1	0.37	12p12.1	15q14	0.39
3p22.1	6p21.31	0.51	5q12.1	9p21.1	0.41	7q21.11	16q22.1	0.37	2q32.1	7p21.3	0.38
3p22.1	15q25.3	0.51	9p24.2	16q24.1	0.41	9q31.2	12q15	0.37	9q22.33	21q21.1	0.38
9q21.11	16q21	0.51	8p23.3	9q22.31	0.39	3p12.1	9p21.3	0.36	15q15.1	22q13.1	0.38
6q23.3	14q24.3	0.5	10p12.31	11q14.1	0.39	5q11.2	10q25.1	0.36	4q31.22	9q34.13	0.37
7q21.13	8q21.13	0.5	2q32.3	4p16.1	0.38	14q24.1	18p11.21	0.36	4q31.3	15q22.1	0.37
10q11.21	12p13.32	0.5	4p14	5q12.3	0.38	3p12.3	7p21.1	0.35	5q33.3	16q12.1	0.37
2p16.1	6p12.3	0.49	3p12.3	12p13.1	0.37	4q31.3	7q21.3	0.35	6q12	13q14.12	0.37
6p21.32	16q12.2	0.49	5q22.1	18q21.32	0.37	15q25.1	22q11.23	0.34	6q15	14q24.3	0.37
6q12	12p11.22	0.49	12q22	15q14	0.37	4p15.33	6q27	0.33	3p14.3	8p11.22	0.36
9q33.1	14q12	0.49	2p21	Xp22.13	0.36	8p22	11q24.2	0.33	10q26.3	16q12.2	0.36
11q22.2	17q21.31	0.49	3q23	9p21.1	0.36	12q22	22q11.23	0.33	1q31.1	4q28.2	0.35
11q24.3	22q13.2	0.49	4q35.1	17p13.2	0.36	3p12.1	9p22.2	0.32	6q14.1	11q23.2	0.35
3p21.1	12q23.3	0.48	5q15	15q23	0.36	9p21.2	10q25.1	0.32	9q21.12	17q12	0.34
4q35.1	8p21.2	0.47	8p11.22	15q14	0.36	9q31.1	12q15	0.32	12q23.1	22q11.23	0.34
7p21.1	14q24.3	0.47	18q21.33	21q21.1	0.36	12p12.3	22q11.23	0.32	6q27	7q31.31	0.33
11p15.2	18q12.1	0.47	2q34	8q21.13	0.35	3q26.1	14q32.13	0.31	10q23.33	12p13.2	0.33
2q32.2	5q23.2	0.46	3p13	9p24.2	0.35	4p15.2	12q12	0.31	18q21.1	20p12.2	0.33
4q22.3	15q26.3	0.46	4q28.1	15q15.3	0.35	5q22.1	7p21.1	0.31	5q11.2	9q34.13	0.32
6p21.31	15q25.3	0.46	5q12.3	12q21.2	0.35	5q22.2	10q22.1	0.31	5q33.3	16q12.2	0.32
11q23.1	12q21.31	0.45	7q31.33	16q12.1	0.35	5q33.3	13q14.13	0.31	6q27	11q14.1	0.32
12p13.2	15q21.3	0.44	8p21.3	12p13.1	0.35	3p22.1	17p13.1	0.3	1p13.2	3q25.2	0.31
2p16.2	9q31.3	0.43	3q26.1	11p13	0.34	4q12	11q23.3	0.3	2p25.2	5q33.3	0.31
3p21.31	17p11.2	0.43	5q11.1	12p13.1	0.34	5q14.1	12p11.22	0.3	3p12.1	22q11.23	0.31
4q35.2	9p21.2	0.43	12q15	13q12.12	0.34	5q21.3	6q26	0.3	4p12	6q24.2	0.31
12p11.22	16q21	0.43	1q31.1	5q23.3	0.33	8p23.1	10q22.1	0.3	4q12	9p21.3	0.31
14q23.3	22q11.21	0.43	3p11.1	9q34.13	0.33	9p21.2	10q23.2	0.3	6p21.32	10q11.21	0.31
2q23.3	6p12.1	0.42	9p24.1	18q22.3	0.33	13q33.1	22q11.23	0.3	6q16.2	7q31.32	0.31
6p21.31	17q11.2	0.42	9q22.32	12p13.31	0.33	15q13.1	Xq24	0.3	7q31.31	8p11.22	0.31
6q24.3	10q11.21	0.42	11q24.3	13q32.1	0.33	18q11.2	22q11.23	0.3	11q24.2	17p11.2	0.31
14q24.3	17p13.1	0.42	2q34	12q21.2	0.32				2p25.2	16q23.2	0.3
1p21.1	7q21.12	0.41	4q22.3	Xp11.22	0.32				2p24.1	11q14.1	0.3
2p16.1	4q13.3	0.41	6p21.32	12q21.1	0.32				2q24.2	6q26	0.3
3p21.2	8p12	0.41	6q14.1	12q23.1	0.32				4q22.3	11q23.1	0.3
4q21.3	15q26.3	0.41	6q14.3	11q14.1	0.32				6q13	9q21.2	0.3
8p21.3	21q21.1	0.41	6q27	13q12.13	0.32				7p21.3	12p11.23	0.3
8p11.22	15q21.2	0.41	7q21.3	22q12.3	0.32				14q23.1	18q22.1	0.3
1p21.3	4q31.22	0.4	8p22	14q22.1	0.32						
3p26.3	4q28.1	0.4	2q21.3	6q22.1	0.31						
4p13	5q22.1	0.4	3p21.1	10q21.3	0.31						
5q32	11p15.4	0.4	3q23	15q25.3	0.31						
6q16.3	13q14.13	0.4	4q34.3	10q21.2	0.31						
11p15.4	17q21.31	0.4	5q21.3	10q26.3	0.31						
11q22.2	12q21.33	0.4	5q22.2	12q12	0.31						

Table 2
Degree-based ranking of nodes

Main effect		TP53 mutation status		ER status		Tumor stage	
Gene	Median rank	Gene	Median rank	Gene	Median rank	Gene	Median rank
6p21.32	21.5	4q13.3	19.75	10q22.1	8.75	22q11.23	15.5
4q21.1	33.25	8q21.13	24.75	8q21.13	15.5	9q34.13	17
17q21.31	44.25	12q23.1	32	9p22.3	24.75	8p11.22	18.5
2q24.2	48.25	15q23	40.5	5q32	25	10q25.2	24.75
2q32.1	55	8p11.22	41.25	12q23.3	26	16q24.1	33.25
17p13.2	57.25	2p12	47.25	22q11.23	28.25	11q14.1	35.75
6q22.31	62	2q24.3	55.5	12p11.22	39	6p21.32	39
11q14.1	66	2p21	55.75	12q22	46.25	6q12	46.5
11q22.3	66.5	3q26.1	57.25	12p12.3	47.25	2q24.1	47.5
10q22.1	66.75	21q21.1	57.5	8p21.3	47.5	10p12.31	50
17q11.2	67.25	4q24	57.75	3q24	51.5	14q32.12	50.5
9p21.3	69.75	1p21.2	62.5	5q33.3	54.75	6q14.1	51.25
12q21.32	70.75	17p13.2	62.75	10q25.1	54.75	16q23.1	51.25
2q33.1	73	12p13.1	64.5	8p11.22	55.25	10p12.2	51.75
1p36.11	76.25	9p24.2	66.5	10q23.2	55.5	6q13	57.25
6q14.3	79.5	14q32.12	69.25	4q12	55.75	9q33.1	59
6p21.31	81.75	4q21.1	70	5q22.2	56	4q21.1	59.25
11q23.1	83	Xp11.23	71.25	3p12.1	56.5	11q22.2	59.25
12p11.22	85.75	4q13.2	71.5	3p14.2	58.25	16q23.3	59.5
14q24.3	87.5	6p21.32	71.75	14q32.12	58.75	9p21.3	63
12p11.23	87.75	12q21.32	75.25	8p22	59.5	11q24.2	63.25
7q21.11	88.25	4q31.1	76	7p21.1	59.75	2q36.3	64.5
4q13.3	89.25	8p22	76	Xp11.23	65	12p11.23	65.25
11q23.3	89.5	12q21.2	77.75	17q21.31	65.25	10q26.3	65.75
15q25.3	90	9q33.3	79.75	5q14.1	69	15q15.1	66.75
6q13	93	17q21.31	81	9p21.2	75	7p21.3	69
2q32.2	95.75	9p21.3	83	3p22.1	75.5	4q21.23	72.5
10q26.3	98.25	2q34	83.25	4p13	75.75	16q12.2	73
21q21.1	99.5	6q23.3	84.5	3p12.3	76.75	2q32.1	73.25
Xp11.23	99.5	8p21.1	84.75	21q21.1	78	4q23	75
4p13	100	4q12	85	3p22.2	79	17p13.1	75
9q31.1	102.25	1q31.1	85.75	5q13.3	80.25	3p21.31	77.5
13q12.3	103.75	12p11.22	87.5	5q14.2	82.75	14q24.3	80.25
8q21.13	106	21q21.3	87.75	12q12	83.25	6q27	82.5
9q21.11	106.5	17p13.1	89.25	4p16.1	84.5	11q23.1	84.75
7q21.12	107.5	8p12	90	5q13.2	87.25	7p21.1	85.5
1p12	107.75	6q14.1	92	3p24.3	88	8p12	85.75
2q23.2	108	2q24.2	92.25	16q21	88	12p11.22	86
6q22.32	109.25	15q22.31	92.75	9p21.3	88.25	11q14.3	87
3q13.31	109.75	4q25	94.5	14q32.13	89.25	8p21.1	88.5
6q26	111.25	9p21.2	94.75	3p14.1	90.5	9q22.31	90.25
8p12	112	4p16.1	96.25	15q13.1	90.75	9p21.2	90.75
12q12	114.25	22q13.2	97	5q23.3	91.5	4q25	91.5
11q22.2	115	15q25.3	101.75	6q14.3	91.5	18q21.32	91.75
12p13.1	115	6q13	104	9p22.2	91.5	6q24.3	94
11p12	115.5	4q35.2	104.25	11q23.3	92	17p11.2	94
12q23.1	115.75	3p25.3	106.25	2p25.2	93	3p25.3	94.25
2p21	116.25	17p12	106.25	1p36.11	93.75	16q22.1	94.25
16q21	117.5	1p31.3	107.25	3q26.1	94.5	10q21.3	95.75