Biometrics 000, 000–000 DOI: 000
000 0000

Web Appendix for

"Robust Gaussian Graphical Modeling via l_1 Penalization"

Hokeun Sun and Hongzhe Li

Department of Biostatistics and Epidemiology,

University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.

**email:* hsunk@mail.med.upenn.edu

***email:* hongzhe@upenn.edu

Web Appendix A - Derivation of $l_\beta(\theta)$, $\nabla l_\beta(\theta)$, and $\nabla^2 l_\beta(\theta)$

We derive the robustified log-likelihood $l_\beta(\theta)$, $\nabla l_\beta(\theta)$, and $\nabla^2 l_\beta(\theta)$ for the Gaussian graphical model. We first introduce the vec-operators and the duplication matrices. Let $vec(A)$ of a $p \times p$ matrix $A = \{a_{ij}\}$ denote a $p^2 \times 1$ vector formed from stacking column vectors of A and $vech(A)$ denote a $p^* \times 1$ vector formed from all elements of lower triangular part of A including diagonals, where $p^* = p(p+1)/2$. The duplication matrix D_p is then defined the relation:

$$D_p \cdot vech(A) = vec(A)$$

for any symmetric matrix A of order p, and $D_p^+ = (D_p^\top D_p)^{-1} D_p^\top$ denote the Moore-Penrose generalized inverse of D_p. The Kronecker product $A \otimes B$ of matrices A and B is defined as a partitioned matrix with the (i,j)-th block equal to $a_{ij}B$.

The probability density function of multivariate normal distribution is

$$f_\Omega(y_k) = \frac{|\Omega|^{1/2}}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2} y_k^\top \Omega y_k\right),$$

and the robustified log-likelihood is then

$$\begin{aligned}
l_\beta(\Omega) &= \frac{1}{n\beta} \sum_{k=1}^n f_\Omega(y_k)^\beta - b_\beta(\Omega) \\
&= \frac{1}{n\beta} \sum_{k=1}^n \frac{|\Omega|^{\beta/2}}{(2\pi)^{p\beta/2}} \exp\left(-\frac{\beta}{2} y_k^\top \Omega y_k\right) - \frac{|\Omega|^{\beta/2}}{(1+\beta)^{p/2+1}(2\pi)^{p\beta/2}} \\
&= \frac{|\Omega|^{\beta/2}}{(2\pi)^{p\beta/2}} \left[\frac{1}{n\beta} \sum_{k=1}^n \exp\left(-\frac{\beta}{2} y_k^\top \Omega y_k\right) - \frac{1}{(1+\beta)^{p/2+1}}\right] \\
&= c_\beta(\Omega) \left[\frac{1}{n\beta} \sum_{k=1}^n e^{\beta z_k(\Omega)} - \frac{1}{(1+\beta)^{p/2+1}}\right],
\end{aligned}$$

where

$$c_\beta(\Omega) = \frac{|\Omega|^{\beta/2}}{(2\pi)^{p\beta/2}}, \qquad z_k(\Omega) = -\frac{1}{2} y_k^\top \Omega y_k,$$

and

$$\begin{aligned} b_\beta(\Omega) &= \frac{1}{1+\beta}\int \left(\frac{|\Omega|^{1/2}}{(2\pi)^{p/2}}\right)^{1+\beta} \exp\left(-\frac{1+\beta}{2} y_k^\top \Omega y_k\right) dy_k \\ &= \frac{1}{1+\beta}\frac{|\Omega|^{(1+\beta)/2}}{(2\pi)^{p(1+\beta)/2}}\frac{(2\pi)^{p/2}}{|\Omega|^{1/2}(1+\beta)^{p/2}} \\ &= \frac{|\Omega|^{\beta/2}}{(1+\beta)^{p/2+1}(2\pi)^{p\beta/2}}. \end{aligned}$$

Let us denote $\theta = vech(\Omega)$ and then the first derivative of the log-likelihood is

$$\nabla l_\beta(\theta) = D_p^+ \cdot vec\left[\frac{\partial l_\beta(\Omega)}{\partial \Omega}\right],$$

and

$$\begin{aligned} \frac{\partial l_\beta(\Omega)}{\partial \Omega} &= \frac{1}{n\beta}\sum_{k=1}^n \frac{\partial}{\partial \Omega}\left\{c_\beta(\Omega) e^{\beta z_k(\Omega)}\right\} - \frac{1}{(1+\beta)^{p/2+1}}\frac{\partial}{\partial \Omega}\left\{c_\beta(\Omega)\right\} \\ &= \frac{1}{n\beta}\sum_{k=1}^n \left[e^{\beta z_k(\Omega)}\frac{\partial}{\partial \Omega}\left\{c_\beta(\Omega)\right\} + c_\beta(\Omega)\frac{\partial}{\partial \Omega}\left\{e^{\beta z_k(\Omega)}\right\}\right] \\ &\quad - \frac{1}{(1+\beta)^{p/2+1}}\frac{\partial}{\partial \Omega}\left\{c_\beta(\Omega)\right\}, \end{aligned} \tag{1}$$

where

$$\begin{aligned} \frac{\partial}{\partial \Omega}\left\{c_\beta(\Omega)\right\} &= \frac{\partial}{\partial \Omega}\left\{\frac{|\Omega|^{\beta/2}}{(2\pi)^{p\beta/2}}\right\} = \frac{1}{(2\pi)^{p\beta/2}}\frac{\partial}{\partial \Omega}\left\{|\Omega|^{\beta/2}\right\} \\ &= \frac{\beta}{2(2\pi)^{p\beta/2}}|\Omega|^{\beta/2}\Omega^{-1} \\ &= \frac{\beta}{2}c_\beta(\Omega)\Omega^{-1}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \Omega}\left\{e^{\beta z_k(\Omega)}\right\} &= \frac{\partial}{\partial \Omega}\left\{\exp\left(-\frac{\beta}{2} y_k^\top \Omega y_k\right)\right\} \\ &= -\frac{\beta}{2} y_k y_k^\top \exp\left(-\frac{\beta}{2} y_k^\top \Omega y_k\right) \\ &= -\frac{\beta}{2} e^{\beta z_k(\Omega)} y_k y_k^\top. \end{aligned}$$

Then, we can easily complete (1) after plug-ins, i.e.,

$$\frac{\partial l_\beta(\Omega)}{\partial \Omega} = \frac{1}{n}\sum_{k=1}^n \frac{c_\beta(\Omega)}{2}\left[e^{\beta z_k(\Omega)}\left(\Omega^{-1} - y_k y_k^\top\right) - \frac{\beta}{(1+\beta)^{p/2+1}}\Omega^{-1}\right]. \tag{2}$$

Notice that setting (2) to 0 gives the robustified estimating equation of the Gaussian graphical model.

Next, the second derivative of the log-likelihood is

$$\nabla^2 l_\beta(\theta) = D_p^+ \cdot \left[\frac{\partial}{\partial \Omega} \left\{ \frac{\partial l_\beta(\Omega)}{\partial \Omega} \right\} \right] \cdot D_p,$$

and

$$\begin{aligned} \frac{\partial}{\partial \Omega} \left\{ \frac{\partial l_\beta(\Omega)}{\partial \Omega} \right\} &= \frac{1}{n} \sum_{k=1}^{n} \frac{\partial}{\partial \Omega} \left\{ \frac{c_\beta(\Omega)}{2} \left[e^{\beta z_k(\Omega)} \left(\Omega^{-1} - y_k y_k^\top \right) - \frac{\beta}{(1+\beta)^{p/2+1}} \Omega^{-1} \right] \right\} \\ &= \frac{1}{n} \sum_{k=1}^{n} \frac{h_k(\Omega)}{2}, \end{aligned}$$

where

$$\begin{aligned} h_k(\Omega) &= \frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) e^{\beta z_k(\Omega)} \Omega^{-1} \right\} - y_k y_k^\top \frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) e^{\beta z_k(\Omega)} \right\} - \frac{\beta}{(1+\beta)^{p/2+1}} \frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) \Omega^{-1} \right\} \\ &= \left(\Omega^{-1} - y_k y_k^\top \right) \frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) e^{\beta z_k(\Omega)} \right\} + c_\beta(\Omega) e^{\beta z_k(\Omega)} \frac{\partial}{\partial \Omega} \left\{ \Omega^{-1} \right\} \\ &\quad - \frac{\beta}{(1+\beta)^{p/2+1}} \left[\Omega^{-1} \frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) \right\} + c_\beta(\Omega) \frac{\partial}{\partial \Omega} \left\{ \Omega^{-1} \right\} \right]. \end{aligned}$$

Since we have $\frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) e^{\beta z_k(\Omega)} \right\}$ and $\frac{\partial}{\partial \Omega} \left\{ c_\beta(\Omega) \right\}$ in (1), we only need

$$\frac{\partial}{\partial \Omega} \left\{ \Omega^{-1} \right\} = - \left(\Omega^{-1} \otimes \Omega^{-1} \right)$$

to complete $h_k(\Omega)$.

Finally, we obtain

$$\begin{aligned} h_k(\Omega) &= \frac{\beta}{2} c_\beta(\Omega) vec \left[e^{\beta z_k(\Omega)} \left(\Omega^{-1} - y_k y_k^\top \right) - \frac{\beta}{(1+\beta)^{p/2+1}} \Omega^{-1} \right] \cdot vec \left[\Omega^{-1} \right]^\top \\ &\quad - \frac{\beta}{2} c_\beta(\Omega) e^{\beta z_k(\Omega)} vec \left[\Omega^{-1} - y_k y_k^\top \right] \cdot vec \left[y_k y_k^\top \right]^\top \\ &\quad - c_\beta(\Omega) \left(e^{\beta z_k(\Omega)} - \frac{\beta}{(1+\beta)^{p/2+1}} \right) \left[\Omega^{-1} \otimes \Omega^{-1} \right]. \end{aligned} \tag{3}$$

Remark: $\theta = vech(\Omega)$ includes the diagonal elements of Ω for convenience purpose here. So, the elements of $\nabla l_\beta(\theta)$ and $\nabla^2 l_\beta(\theta)$ corresponding to the diagonals should be excluded when they are applied to the coordinate gradient descent method in Section 3.

Web Appendix B: Simulation with Different Concentration Matrices

In the final set of simulations, we consider four different models where the outliers are generated from models with different concentration matrices while differing the magnitude of μ's. These models mimic the scenarios where the outliers come from models with different

graphical structures. In particularly, each sample is generated from the following mixture distribution,

$$y_k \sim (1-p_0)N_p(0,\Omega^{-1}) + \frac{p_0}{2}N_p(-\mu,\Omega_0^{-1}) + \frac{p_0}{2}N_p(\mu,\Omega_0^{-1}), \quad k=1,\ldots,n.$$

where $p_0 = 0.1$, $n = 100$, and $p = 150$. The four different outlier distributions are then considered in the following models:

$$\begin{aligned}
&\text{Model V}: && \mu = (1,\ldots,1)^\top, \Omega_0 = \Omega' \\
&\text{Model VI}: && \mu = (1.5,\ldots,1.5)^\top, \Omega_0 = \Omega' \\
&\text{Model VII}: && \mu = (1,\ldots,1)^\top, \Omega_0 = I_p \\
&\text{Model VIII}: && \mu = (1.5,\ldots,1.5)^\top, \Omega_0 = I_p,
\end{aligned}$$

where $\Omega' \neq \Omega$ is a randomly generated precision matrix in the same way that Ω was generated in the previous simulations, and I_p is a p-dimensional identity matrix. Since the dimension and sample size are same as the second set of simulations, we still use the same tuning parameter $\beta = 0.005$, 0.01, 0.02, and 0.03 for the robust estimation method.

[Figure 1 about here.]

Figure 1 presents the average ROC curves of *glasso* and the robust methods over 100 simulation data sets for each model as the tuning parameter λ varies. It is clear that our proposed robust method still recovers many more true edges than *glasso* for the same false positive rates even when the Markov structures are blurred by the outliers. Particularly, the sensitivities from the robust procedures with $\beta = 0.02$ and 0.03 are almost twice higher than those from *glasso* for models VI and $VIII$. Similar to the second set of simulations, the robust tuning parameter $\beta = 0.02$ and 0.03 show the best selection performances.

Finally, we consider the scenario where the outliers are not symmetric about the mean and simulate data from the following model,

$$\text{Model IX: } y_k \sim (1-p_0)N_p(0,\Omega^{-1}) + p_0 N_p(2, I_p), \quad k=1,\ldots,n.$$

where $p_0 = 0.1, n = 100$ and $p = 150$. The results are shown in Figure 2 for the proposed robust estimation procedure with different β values. We observe that our proposed robust method still recovers more true edges than *glasso* for the same false positive rates and results in a much higher area under the receiver-operator curve than *glasso*.

[Figure 2 about here.]

Web Appendix C - Additional Analysis and Figures For Real Data Analysis

We include two additional Figures that are related to the analysis of real data set. Figure 3 shows the illustrative KEGG MAPK pathway from the KEGG database http://www.genome.jp/kegg/pathway/sce/sce04011.html.

Figure 4 shows the histograms of the gene expression data of genes from the KEGG MAPK pathway, where the left two plots are the histograms of all the genes in simulated and real MAPK data sets and the middle and right plots are the histograms for one gene in simulated and MAPK data sets. Overall, these plots show that the simulated data sets have similar characteristics as the real MAPK gene expression data.

We calculate the sample skewness of the expression level of each of the 54 genes in the MAPK pathway and present a summary of the skewness in Table 1, indicating that some genes have skewed distributions and some have outliers. Figure 5 presents the histograms of the expression levels of 12 MAPK genes with the smallest skewness. These plots indicate that that outliers are not always symmetric to the means.

[Table 1 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

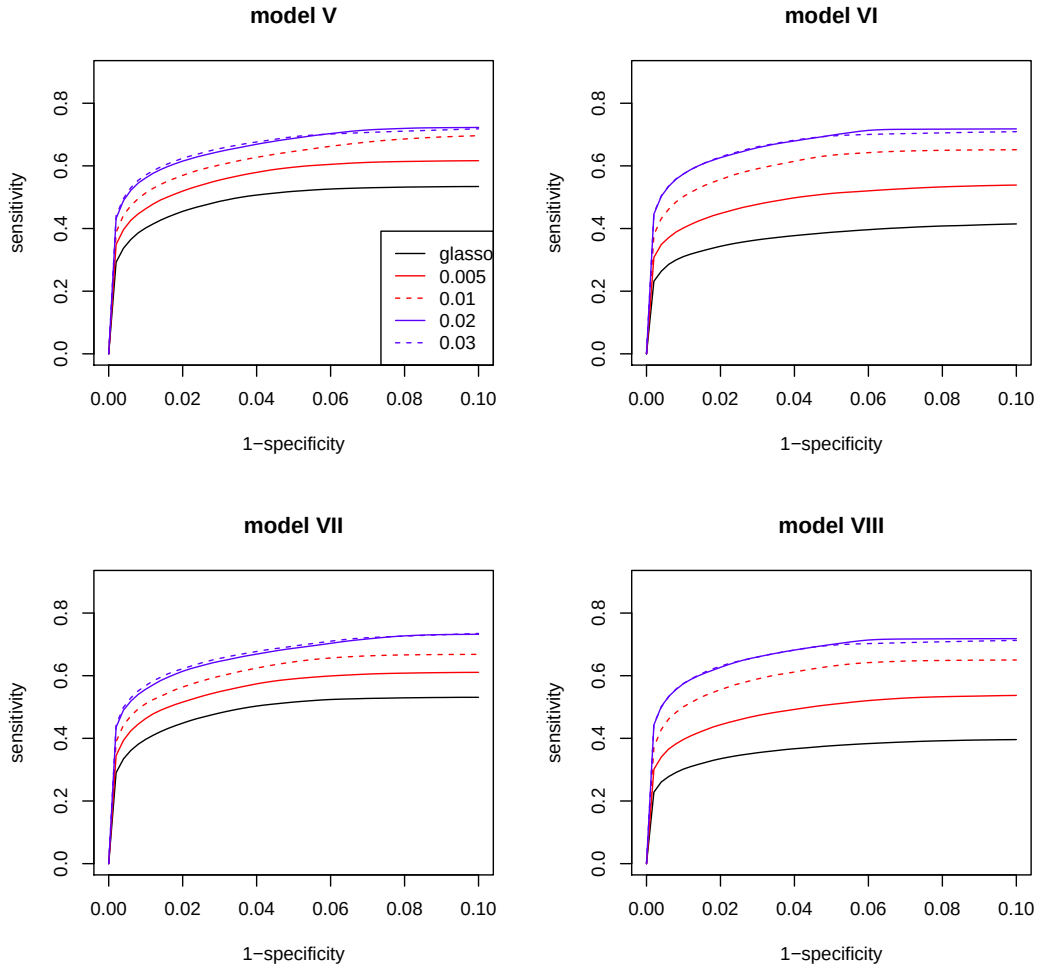

Figure 1. The ROC curves of *glasso* and the robust method with different robustness tuning parameters, $\beta = 0.005$, 0.01, 0.02, and 0.03 for the third set of simulations with $p = 150, n = 100$. The outliers in Models V and VI are generated randomly from multivariate normal distributions with different concentration matrices from those of the main distributions and the outliers in Models VII and VIII are generated from the multivariate normal distribution with an identity concentration matrix. Models V and VII have 10% small magnitudes of outliers and and Models VI and VIII have 10% of medium magnitudes of outliers. Each curve is an average over 100 simulated data sets.

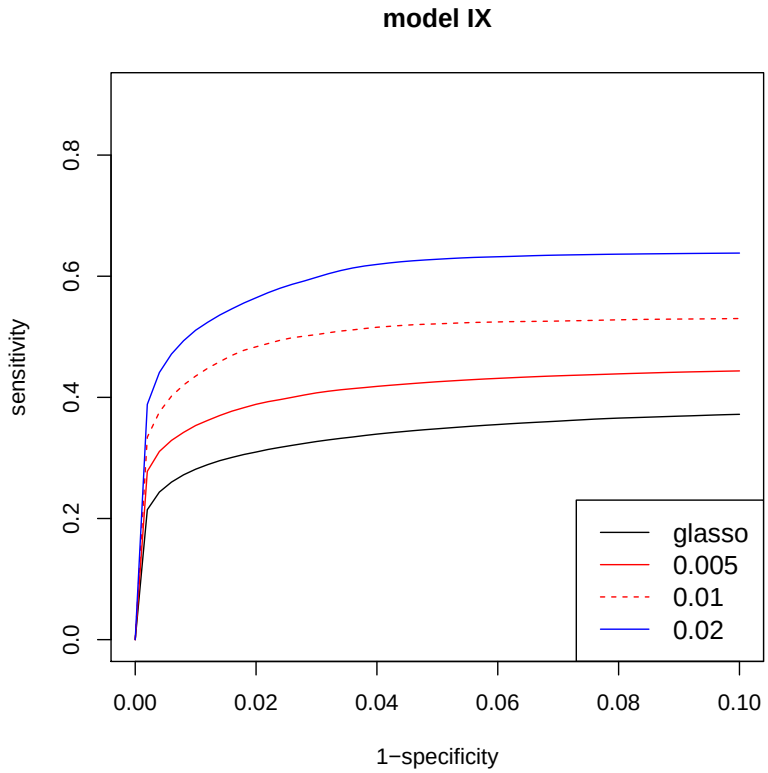

Figure 2. The ROC curves of *glasso* and the robust method with different robustness tuning parameters, $\beta = 0.005$, 0.01, 0.02 for Model IX with $p = 150, n = 100$. The outliers are simulated to have non-symmetric mean. Each curve is an average over 100 simulated data sets.

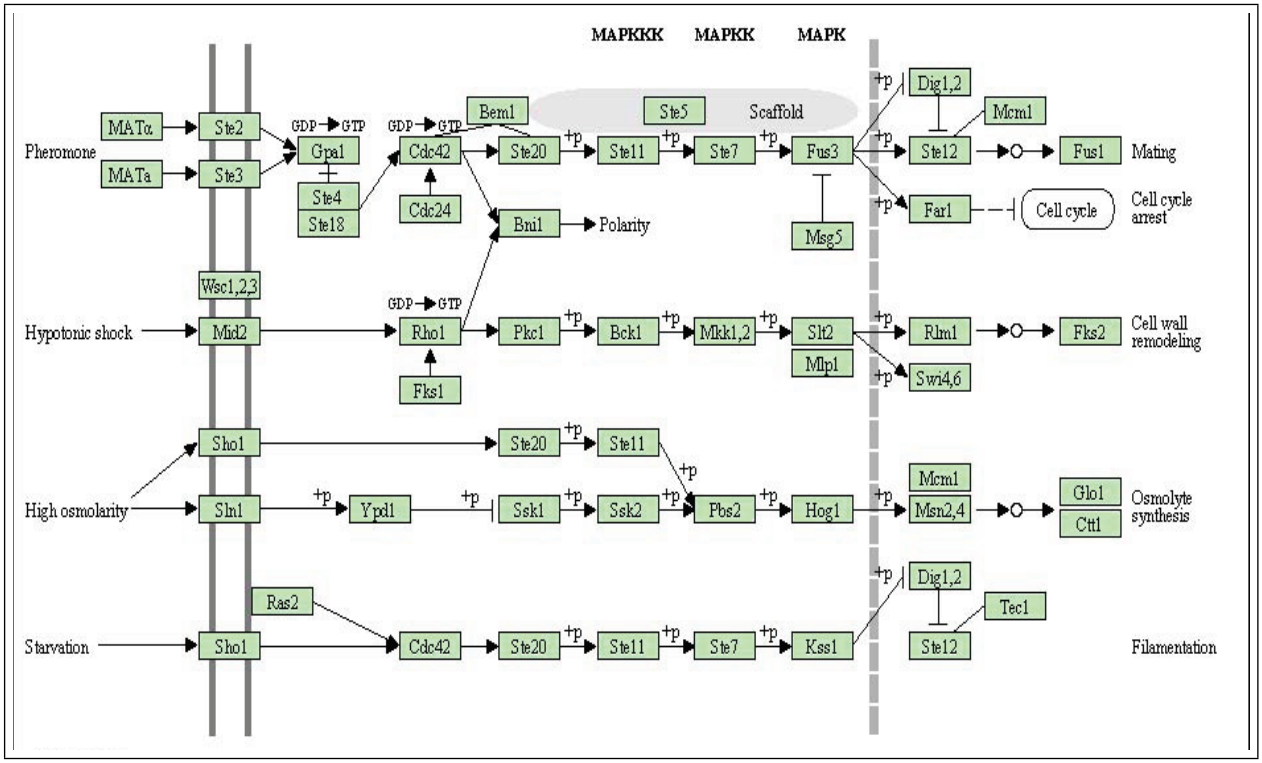

Figure 3. The yeast MAPK pathway from the KEGG database http://www.genome.jp/kegg/pathway/sce/sce04011.html.

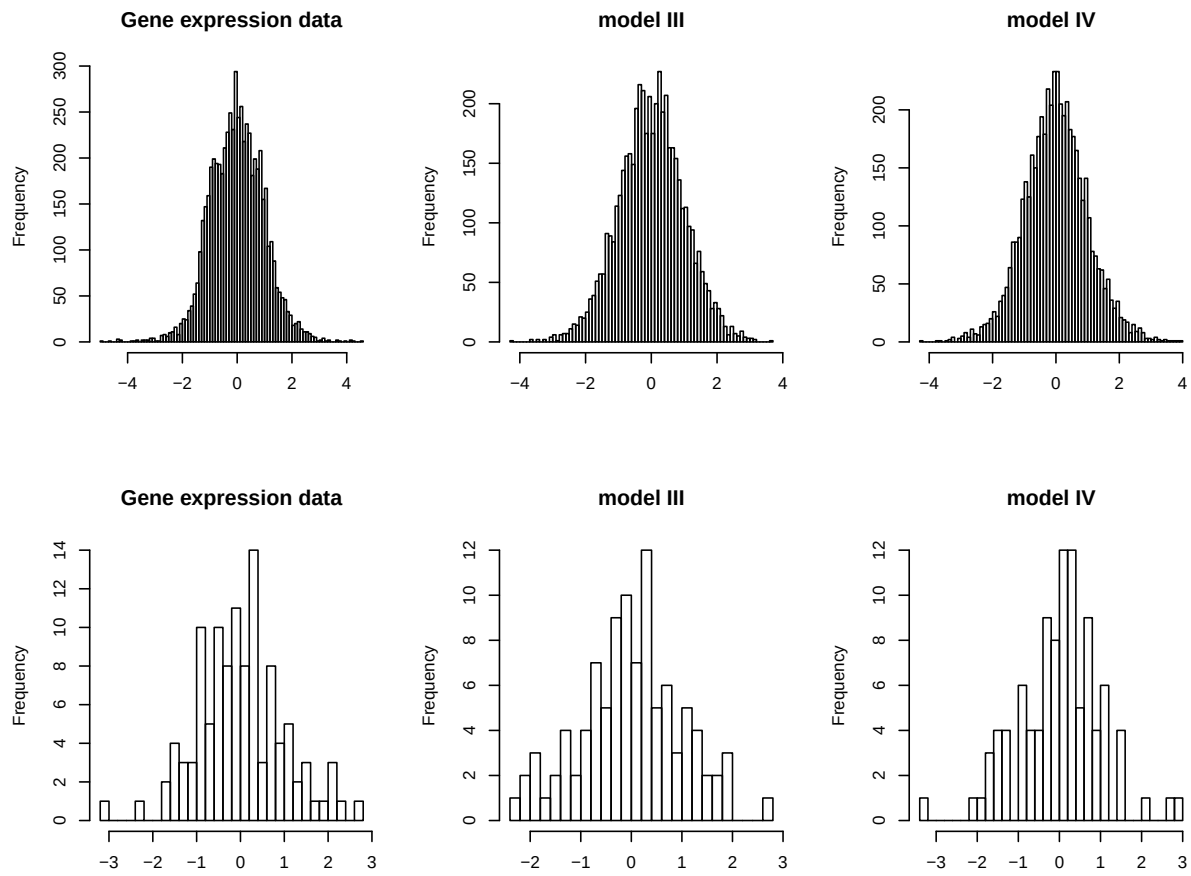

Figure 4. The histograms of yeast gene expression data (left), simulated data based on model III (middle) and model IV (right). The histograms in the upper panel are for the entire data set, and those in the lower panel are for the expression of gene SWI4 and simulated data for two genes. All data are re-scaled so that each gene has a mean of 0 and standard deviation of 1. Models III and IV have 10% of medium and large magnitudes of outliers, respectively.

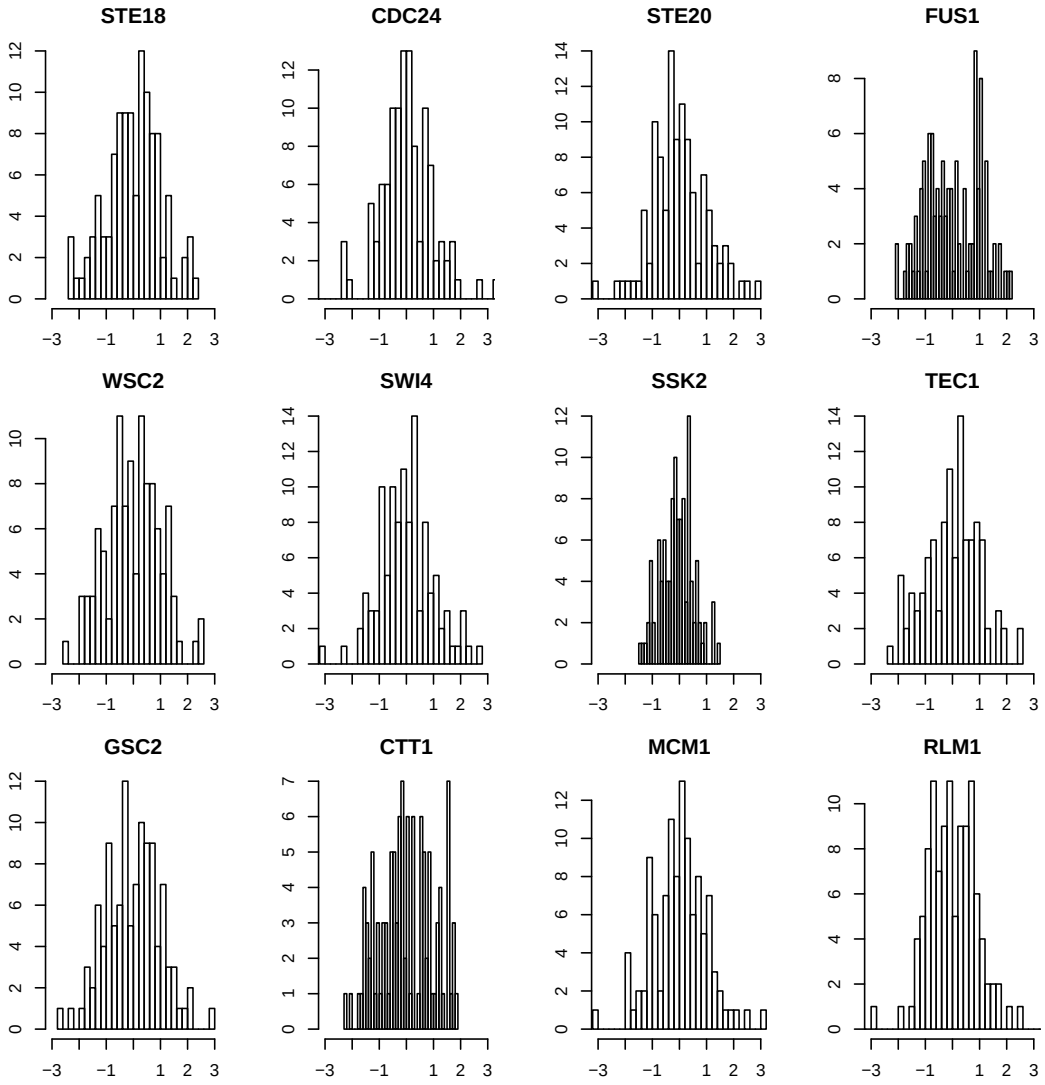

Figure 5. The histograms of yeast gene expression data of 12 genes in the KEGG MAPK pathway.

Table 1
Summary of the skewness of gene expressions of 54 genes in the MAPK pathway.

Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
-2.19300	-0.36860	0.05865	-0.01706	0.37830	1.71000