Supporting information for "LSMM: A statistical approach to integrating functional annotations with genome-wide association studies"

Jingsi Ming[1], Mingwei Dai[2,5], Mingxuan Cai[1],
Xiang Wan[3], Jin Liu[4*]and Can Yang[5*]

[1]Department of Mathematics, Hong Kong Baptist University, Hong Kong
[2]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
[3]Shenzhen Research Institute of Big Data, Shenzhen, China
[4]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore
[5]Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong

1 The variational EM algorithm

E-step

Let $\boldsymbol{\theta} = \{\alpha, \mathbf{b}, \sigma^2, \omega\}$ be the collection of model parameters. The logarithm of the marginal likelihood is

$$\log \Pr\left(\mathbf{p}|\mathbf{Z}, \mathbf{A}; \boldsymbol{\theta}\right) = \log \sum_{\boldsymbol{\gamma}} \sum_{\boldsymbol{\eta}} \int \Pr\left(\mathbf{p}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}|\mathbf{Z}, \mathbf{A}; \boldsymbol{\theta}\right) d\tilde{\boldsymbol{\beta}}.$$

Using the sigmoid function denoted as $S\left(x\right) = \frac{1}{1+e^{-x}}$, the complete-data likelihood can be written as

$$\Pr\left(\mathbf{p}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}|\mathbf{Z}, \mathbf{A}; \boldsymbol{\theta}\right) = \Pr\left(\mathbf{p}|\boldsymbol{\gamma}; \alpha\right) \Pr\left(\boldsymbol{\gamma}|\mathbf{Z}, \mathbf{A}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}\right) \Pr\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}|\sigma^2, \omega\right),$$

.where

*Correspondence should be addressed to Can Yang (macyang@ust.hk) and Jin Liu (jin.liu@duke-nus.edu.sg)

$$
\begin{aligned}
\Pr\left(\mathbf{p}|\boldsymbol{\gamma};\alpha\right) &= \prod_{j=1}^{M}\Pr\left(p_j|\gamma_j;\alpha\right)=\prod_{j=1}^{M}\left(\alpha p_j^{\alpha-1}\right)^{\gamma_j},\\
\Pr\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b}\right) &= \prod_{j=1}^{M}\Pr\left(\gamma_j|\mathbf{Z}_j,\mathbf{A}_j,\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b}\right)\\
&= \prod_{j=1}^{M}e^{\gamma_j\left(\mathbf{Z}_j\mathbf{b}+\sum_k A_{jk}\eta_k\tilde{\beta}_k\right)}S\left(-\mathbf{Z}_j\mathbf{b}-\sum_k A_{jk}\eta_k\tilde{\beta}_k\right),\\
\Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right) &= \prod_{k=1}^{K}\Pr\left(\tilde{\beta}_k,\eta_k|\sigma^2,\omega\right)=\prod_{k=1}^{K}N\left(\tilde{\beta}_k|0,\sigma^2\right)\omega^{\eta_k}\left(1-\omega\right)^{1-\eta_k}.
\end{aligned}
$$

We can use JJ bound (Jaakkola and Jordan, 2000) to get the tractable lower bound of $\Pr\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b}\right)$ which is denoted by $h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b},\boldsymbol{\xi}\right)$:

$$
\begin{aligned}
&\Pr\left(\gamma_j|\mathbf{Z}_j,\mathbf{A}_j,\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b}\right)\\
=\;& e^{\gamma_j\left(\mathbf{Z}_j\mathbf{b}+\sum_k A_{jk}\eta_k\tilde{\beta}_k\right)}S\left(-\mathbf{Z}_j\mathbf{b}-\sum_k A_{jk}\eta_k\tilde{\beta}_k\right)\\
\geq\;& e^{\gamma_j\left(\mathbf{Z}_j\mathbf{b}+\sum_k A_{jk}\eta_k\tilde{\beta}_k\right)}S\left(\xi_j\right)\exp\left(-\lambda\left(\xi_j\right)\left(\left(\mathbf{Z}_j\mathbf{b}+\sum_k A_{jk}\eta_k\tilde{\beta}_k\right)^2-\xi_j^2\right)-\frac{\mathbf{Z}_j\mathbf{b}+\sum_k A_{jk}\eta_k\tilde{\beta}_k+\xi_j}{2}\right)\\
=\;& h\left(\gamma_j|\mathbf{Z}_j,\mathbf{A}_j,\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b},\xi_j\right),
\end{aligned}
$$

where

$$
\lambda\left(\xi_j\right)=\frac{1}{2\xi_j}\left(S\left(\xi_j\right)-\frac{1}{2}\right).
$$

Let $\boldsymbol{\Theta}=\left\{\alpha,\mathbf{b},\boldsymbol{\xi},\sigma^2,\omega\right\}$. Then

$$
f\left(\mathbf{p},\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{Z},\mathbf{A};\boldsymbol{\Theta}\right)=\Pr\left(\mathbf{p}|\boldsymbol{\gamma};\alpha\right)h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b},\boldsymbol{\xi}\right)\Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right)
$$

is a lower bound of complete-data likelihood.

Next, let $q\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)$ be an approximation of the posterior $\Pr\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{p},\mathbf{Z},\mathbf{A};\boldsymbol{\theta}\right)$. Then we can obtain a lower bound of the logarithm of the marginal likelihood:

$$
\begin{aligned}
& \log \Pr\left(\mathbf{p}|\mathbf{Z},\mathbf{A};\boldsymbol{\theta}\right) \\
= & \log \sum_{\boldsymbol{\gamma}} \sum_{\boldsymbol{\eta}} \int \Pr\left(\mathbf{p},\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{Z},\mathbf{A};\boldsymbol{\theta}\right) d\tilde{\boldsymbol{\beta}} \\
\geq & \log \sum_{\boldsymbol{\gamma}} \sum_{\boldsymbol{\eta}} \int f\left(\mathbf{p},\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{Z},\mathbf{A};\boldsymbol{\Theta}\right) d\tilde{\boldsymbol{\beta}} \\
\geq & \sum_{\boldsymbol{\gamma}} \sum_{\boldsymbol{\eta}} \int q\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right) \log \frac{f\left(\mathbf{p},\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{Z},\mathbf{A};\boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)} d\tilde{\boldsymbol{\beta}} \\
= & \mathbf{E}_q\left[\log f\left(\mathbf{p},\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{Z},\mathbf{A};\boldsymbol{\Theta}\right) - \log q\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)\right] \\
\triangleq & L\left(q\right),
\end{aligned}
$$

where $L(q)$ is the lower bound. The second inequality follows Jensen's inequality. And

$$
\begin{aligned}
& \log \Pr\left(\mathbf{p}|\boldsymbol{\gamma},\alpha\right) \\
= & \sum_{j=1}^{M} \left(\gamma_j \left(\log \alpha + \left(\alpha - 1\right) \log p_j\right)\right),
\end{aligned}
$$

$$
\begin{aligned}
& \log h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta},\mathbf{b},\boldsymbol{\xi}\right) \\
= & \sum_{j=1}^{M} \left(\gamma_j \left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\eta_k\tilde{\beta}_k\right) + \log S\left(\xi_j\right)\right) \\
+ & \sum_{j=1}^{M} \left(-\lambda\left(\xi_j\right)\left(\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\eta_k\tilde{\beta}_k\right)^2 - \xi_j^2\right) - \left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\eta_k\tilde{\beta}_k + \xi_j\right)/2\right),
\end{aligned}
$$

$$
\begin{aligned}
& \log \Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right) \\
= & -\frac{1}{2\sigma^2}\sum_{k=1}^{K}\tilde{\beta}_k^2 - \frac{K}{2}\log\left(2\pi\sigma^2\right) + \sum_{k=1}^{K}\eta_k \log\omega + \sum_{k=1}^{K}\left(1-\eta_k\right)\log\left(1-\omega\right).
\end{aligned}
$$

To make it feasible to evaluate the lower bound, we assume that $q\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)$ can be factorized as

$$
q\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right) = \left(\prod_{k=1}^{K} q\left(\tilde{\beta}_k,\eta_k\right)\right)\left(\prod_{j=1}^{M} q\left(\gamma_j\right)\right),
$$

where $q\left(\tilde{\beta}_k,\eta_k\right) = q\left(\tilde{\beta}_k|\eta_k\right) q\left(\eta_k\right)$,$q\left(\gamma_j = 1\right) = \pi_j$, $q\left(\eta_k = 1\right) = \omega_k$.

We can obtain an approximation according to the mean-field method:

$$
\begin{aligned}
& \log q\left(\tilde{\beta}_i, \eta_i\right) \\
= \quad & \mathbf{E}_{k \neq i} \mathbf{E}_{\boldsymbol{\gamma}}\left[\log f\left(\mathbf{p}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta} | \mathbf{Z}, \mathbf{A}; \boldsymbol{\Theta}\right)\right] \\
= \quad & \left(-\frac{1}{2\sigma^2}-\sum_{j=1}^{M} \lambda\left(\xi_j\right) A_{ji}^2 \eta_i^2\right) \tilde{\beta}_i^2 \\
& +\sum_{j=1}^{M}\left(\left(\pi_j-\frac{1}{2}-2 \lambda\left(\xi_j\right) \mathbf{Z}_j \mathbf{b}\right) A_{ji}-2 \lambda\left(\xi_j\right) A_{ji} \sum_{k \neq i} A_{jk} \mathbf{E}_k\left[\eta_k \tilde{\beta}_k\right]\right) \eta_i \tilde{\beta}_i \\
& +\eta_i \log \omega+\left(1-\eta_i\right) \log (1-\omega)+const,
\end{aligned}
$$

where the expectation is taken under the distribution $q(\boldsymbol{\gamma})$ and $q\left(\tilde{\beta}_{-i}, \eta_{-i}\right)=\prod_{k \neq i} q\left(\tilde{\beta}_k, \eta_k\right)$.

When $\eta_i = 1$, we have

$$
\begin{aligned}
& \log q\left(\tilde{\beta}_i | \eta_i=1\right) \\
= \quad & \left(-\frac{1}{2\sigma^2}-\sum_{j=1}^{M} \lambda\left(\xi_j\right) A_{ji}^2\right) \tilde{\beta}_i^2 \\
& +\sum_{j=1}^{M}\left(\left(\pi_j-\frac{1}{2}-2 \lambda\left(\xi_j\right) \mathbf{Z}_j \mathbf{b}\right) A_{ji}-2 \lambda\left(\xi_j\right) A_{ji} \sum_{k \neq i} A_{jk} \mathbf{E}_k\left[\eta_k \tilde{\beta}_k\right]\right) \tilde{\beta}_i+const,
\end{aligned}
$$

where $\mathbf{E}_k$ denotes the expectation under $q\left(\tilde{\beta}_k, \eta_k\right)$, and the constant doesn't depend on $\tilde{\beta}_i$. Because $\log q\left(\tilde{\beta}_i | \eta_i=1\right)$ is a quadratic form,

$$
q\left(\tilde{\beta}_i | \eta_i=1\right)=N\left(\mu_i, s_i^2\right),
$$

where

$$
\begin{aligned}
\mu_i & = s_i^2 \sum_{j=1}^{M}\left(\pi_j-\frac{1}{2}-2 \lambda\left(\xi_j\right)\left(\mathbf{Z}_j \mathbf{b}+\sum_{k \neq i} A_{jk} \mathbf{E}_k\left[\eta_k \tilde{\beta}_k\right]\right) A_{ji}\right), \\
s_i^2 & = \frac{\sigma^2}{1+2 \sigma^2 \sum_{j=1}^{M} \lambda\left(\xi_j\right) A_{ji}^2}.
\end{aligned}
$$

When $\eta_i = 0$, we have

$$
\log q\left(\tilde{\beta}_i | \eta_i=0\right) \quad = \quad -\frac{1}{2\sigma^2} \tilde{\beta}_i^2+const.
$$

So

$$
q\left(\tilde{\beta}_i | \eta_i=0\right) \quad = \quad N\left(0, \sigma^2\right).
$$

Therefore we have

$$q\left(\tilde{\beta}_i, \eta_i\right) = \left[\omega_i N\left(\mu_i, s_i^2\right)\right]^{\eta_i} \left[\left(1-\omega_i\right) N\left(0, \sigma^2\right)\right]^{1-\eta_i}.$$

Now we evaluate the variational lower bound $L\left(q\right)$.

$$\begin{aligned}
& \mathbf{E}_q\left[\log \Pr\left(\mathbf{p}|\boldsymbol{\gamma}, \alpha\right)\right] \\
= & \sum_{j=1}^{M}\left(\pi_j\left(\log\alpha + \left(\alpha-1\right)\log p_j\right)\right), \\
& \mathbf{E}_q\left[\log h\left(\boldsymbol{\gamma}|\mathbf{Z}, \mathbf{A}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}, \mathbf{b}, \boldsymbol{\xi}\right)\right] \\
= & \sum_{j=1}^{M}\left(\pi_j\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right) + \log S\left(\xi_j\right) - \lambda\left(\xi_j\right)\left(\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right)^2 - \xi_j^2\right)\right) \\
& + \sum_{j=1}^{M}\left(-\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k + \xi_j\right)/2 + \lambda\left(\xi_j\right)\sum_k A_{jk}^2\omega_k^2\mu_k^2 - \lambda\left(\xi_j\right)\sum_k A_{jk}^2\omega_k\left(s_k^2+\mu_k^2\right)\right), \\
& \mathbf{E}_q\left[\log \Pr\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}|\sigma^2, \omega\right)\right] \\
= & -\frac{1}{2\sigma^2}\sum_{k=1}^{K}\left(\omega_k\left(s_k^2+\mu_k^2\right) + \left(1-\omega_k\right)\sigma^2\right) - \frac{K}{2}\log\left(2\pi\sigma^2\right) + \sum_{k=1}^{K}\omega_k\log\omega + \sum_{k=1}^{K}\left(1-\omega_k\right)\log\left(1-\omega\right), \\
& -\mathbf{E}_q\left[\log q\left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}\right)\right] \\
= & \sum_{k=1}^{K}\left(\frac{1}{2}\omega_k\left(\log s_k^2 - \log\sigma^2\right) - \omega_k\log\omega_k - \left(1-\omega_k\right)\log\left(1-\omega_k\right)\right) + \frac{K}{2}\log\sigma^2 + \frac{K}{2} + \frac{K}{2}\log\left(2\pi\right) \\
& - \sum_{j=1}^{M}\left(\pi_j\log\pi_j + \left(1-\pi_j\right)\log\left(1-\pi_j\right)\right).
\end{aligned}$$

We set the partial derivative of the lower bound $L(q)$ w.r.t to ω_k, π_j and ξ_j be 0 to get the variational parameters ω_k, π_j and ξ_j:

$$\begin{aligned}
\omega_k &= \frac{1}{1+\exp\left(-u_k\right)}, \text{ where } u_k = \log\frac{\omega}{1-\omega} + \frac{1}{2}\log\frac{s_k^2}{\sigma^2} + \frac{\mu_k^2}{2s_k^2}, \\
v_j &= \log\alpha + \left(\alpha-1\right)\log p_j + \mathbf{Z}_j\mathbf{b} + \sum_{k=1}^{K} A_{jk}\omega_k\mu_k, \\
\xi_j^2 &= \left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right)^2 + \sum_k A_{jk}^2\left(\omega_k\left(s_k^2+\mu_k^2\right) - \omega_k^2\mu_k^2\right).
\end{aligned}$$

The variational lower bound $L(q)$ is

$$
\begin{aligned}
& L(q) \\
= & \sum_{j=1}^{M} \left(\pi_j \left(\log \alpha + (\alpha - 1) \log p_j\right)\right) \\
& + \sum_{j=1}^{M} \left(\pi_j \left(\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \omega_k \mu_k\right) + \log S\left(\xi_j\right) - \lambda\left(\xi_j\right) \left(\left(\beta_0 + \sum_k A_{jk} \omega_k \mu_k\right)^2 - \xi_j^2\right)\right) \\
& + \sum_{j=1}^{M} \left(-\left(\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \omega_k \mu_k + \xi_j\right)/2 + \lambda\left(\xi_j\right) \sum_k A_{jk}^2 \omega_k^2 \mu_k^2 - \lambda\left(\xi_j\right) \sum_k A_{jk}^2 \omega_k \left(s_k^2 + \mu_k^2\right)\right) \\
& - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(\omega_k \left(s_k^2 + \mu_k^2\right) - \omega_k \sigma^2\right) + \sum_{k=1}^{K} \omega_k \log \omega + \sum_{k=1}^{K} \left(1 - \omega_k\right) \log\left(1 - \omega\right) \\
& + \sum_{k=1}^{K} \left(\frac{1}{2} \omega_k \left(\log s_k^2 - \log \sigma^2\right) - \omega_k \log \omega_k - \left(1 - \omega_k\right) \log\left(1 - \omega_k\right)\right) \\
& - \sum_{j=1}^{M} \left(\pi_j \log \pi_j + \left(1 - \pi_j\right) \log\left(1 - \pi_j\right)\right).
\end{aligned}
$$

M-step

Now we update α, $\mathbf{b}$, σ^2, ω. We set the partial derivative of $L(q)$ w.r.t the parameters to be 0 and get

$$
\begin{aligned}
\alpha &= -\frac{\sum_{j=1}^{M} \pi_j}{\sum_{j=1}^{M} \pi_j \log p_j}, \\
\sigma^2 &= \frac{\sum_{k=1}^{K} \omega_k \left(s_k^2 + \mu_k^2\right)}{\sum_{k=1}^{K} \omega_k}, \\
\omega &= \frac{1}{K} \sum_{k=1}^{K} \omega_k,
\end{aligned}
$$

and use Newton's method to update $\mathbf{b}$:

$$
\mathbf{b} = \mathbf{b}_{old} - \mathbf{H}^{-1} \mathbf{g},
$$

where

$$
\begin{aligned}
\mathbf{g} &= \sum_{j=1}^{M} \mathbf{Z}_j^T \left(\pi_j - 2\lambda\left(\xi_j\right) \left(\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \omega_k \mu_k\right) - \frac{1}{2}\right), \\
\mathbf{H} &= -2 \mathbf{Z}_j^T \lambda\left(\xi_j\right) \mathbf{Z}_j.
\end{aligned}
$$

Implementation

- Initialize α, σ^2, ω, $\mathbf{b}$, $\{\omega_k, \mu_k\}_{k=1,\ldots K}$, $\{\xi_j, \pi_j\}_{j=1,\ldots,M}$. Let $\tilde{y} = \sum_k A_{jk} \omega_k \mu_k$.

- E-step: For $i = 1, ..., K$, first obtain $\tilde{y}_i = \tilde{y} - A_{ji}\omega_i\mu_i$, and then update μ_i, s_i^2, ω_i and $\tilde{y}$ as follows

$$\begin{aligned}
s_i^2 &= \frac{\sigma^2}{1 + 2\sigma^2 \sum_{j=1}^{M} \lambda(\xi_j) A_{ji}^2}, \\
\mu_i &= s_i^2 \sum_{j=1}^{M} \left(\left(\pi_j - \frac{1}{2} - 2\lambda(\xi_j)(\mathbf{Z}_j\mathbf{b} + \tilde{y}_i) \right) A_{ji} \right), \\
\omega_i &= \frac{1}{1 + \exp(-u_i)}, \text{ where } u_i = \log\frac{\omega}{1-\omega} + \frac{1}{2}\log\frac{s_i^2}{\sigma^2} + \frac{\mu_i^2}{2s_i^2}, \\
\tilde{y} &= \tilde{y}_i + A_{ji}\omega_i\mu_i.
\end{aligned}$$

  Then for $j = 1, ..., M$, update π_j, ξ_j as follows

$$\begin{aligned}
\pi_j &= \frac{1}{1 + \exp(-v_j)}, \text{ where } v_j = \log\alpha + (\alpha - 1)\log p_j + \mathbf{Z}_j\mathbf{b} + \tilde{y}, \\
\xi_j^2 &= (\mathbf{Z}_j\mathbf{b} + \tilde{y})^2 + \sum_k A_{jk}^2 \left(\omega_k \left(s_k^2 + \mu_k^2 \right) - \omega_k^2 \mu_k^2 \right).
\end{aligned}$$

  Calculate $L(q)$:

$$\begin{aligned}
& L(q) \\
= \; & \sum_{j=1}^{M} \pi_j (\log\alpha + (\alpha - 1)\log p_j) - \sum_{j=1}^{M} (\pi_j \log\pi_j + (1 - \pi_j)\log(1 - \pi_j)) \\
& + \sum_{j=1}^{M} \left(\pi_j (\mathbf{Z}_j\mathbf{b} + \tilde{y}) + \log S(\xi_j) - \frac{\mathbf{Z}_j\mathbf{b} + \tilde{y} + \xi_j}{2} \right) \\
& - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(\omega_k \left(s_k^2 + \mu_k^2 \right) - \omega_k\sigma^2 \right) + \sum_{k=1}^{K} \omega_k \log\omega + \sum_{k=1}^{K} (1 - \omega_k)\log(1 - \omega) \\
& + \sum_{k=1}^{K} \left(\frac{1}{2}\omega_k \left(\log s_k^2 - \log\sigma^2 \right) - \omega_k \log\omega_k - (1 - \omega_k)\log(1 - \omega_k) \right).
\end{aligned}$$

- M-step

$$
\begin{aligned}
\alpha &= -\frac{\sum_{j=1}^{M} \pi_j}{\sum_{j=1}^{M} \pi_j \log p_j}, \\
\sigma^2 &= \frac{\sum_{k=1}^{K} \omega_k \left(s_k^2 + \mu_k^2\right)}{\sum_{k=1}^{K} \omega_k}, \\
\omega &= \frac{1}{K} \sum_{k=1}^{K} \omega_k, \\
\mathbf{g} &= -\sum_{j=1}^{M} \mathbf{Z}_j^T \left(\pi_j - 2\lambda\left(\xi_j\right)\left(\mathbf{Z}_j \mathbf{b} + \tilde{y}\right) - \frac{1}{2} \right), \\
\mathbf{H} &= 2 \sum_{j=1}^{M} \lambda\left(\xi_j\right) \mathbf{Z}_j^T \mathbf{Z}_j, \\
\mathbf{b} &= \mathbf{b}_{old} - \mathbf{H}^{-1} \mathbf{g}.
\end{aligned}
$$

- Evaluate $L(q)$ to track the convergence of the algorithm.

2 Details of the proposed algorithm

Stage 1: Two-groups model (TGM)

Suppose we have the p-values of M SNPs for a given phenotype. Let γ_j be the latent variables indicating whether the j-th SNP is associated with this phenotype. Here $\gamma_j = 0$ means unassociated and $\gamma_j = 1$ means associated. Then we have the following two-groups model:

$$p_j \sim \begin{cases} U[0,1], & \gamma_j = 0, \\ Beta(\alpha, 1), & \gamma_j = 1, \end{cases}$$

where $\mathbf{p} \in \mathbb{R}^M$ are the p-values, $0 < \alpha < 1$ and $\Pr(\gamma_j = 1) = \pi_1$.

We can use EM algorithm to compute the posterior and parameter estimation.

Let $\boldsymbol{\theta} = \{\alpha, \pi_1\}$ be the collection of model parameters. The logarithm of the marginal likelihood is

$$\log \Pr(\mathbf{p}|\boldsymbol{\theta}) = \log \sum_{\boldsymbol{\gamma}} \Pr(\mathbf{p}, \boldsymbol{\gamma}|\boldsymbol{\theta}) = \log \sum_{\boldsymbol{\gamma}} \Pr(\mathbf{p}|\boldsymbol{\gamma}; \alpha) \Pr(\boldsymbol{\gamma}|\pi_1),$$

where

$$\begin{aligned} \Pr(\mathbf{p}|\boldsymbol{\gamma}; \alpha) &= \prod_{j=1}^{M} \Pr(p_j|\gamma_j; \alpha) = \prod_{j=1}^{M} \left(\alpha p_j^{\alpha-1}\right)^{\gamma_j}, \\ \Pr(\boldsymbol{\gamma}|\pi_1) &= \prod_{j=1}^{M} \pi_1^{\gamma_j} (1-\pi_1)^{1-\gamma_j}. \end{aligned}$$

In the E step, we compute the posterior:

$$\tilde{\gamma}_j = q(\gamma_j = 1) = \frac{\pi_1 \alpha p_j^{\alpha-1}}{\pi_1 \alpha p_j^{\alpha-1} + 1 - \pi_1},$$

and get the Q function:

$$\begin{aligned} Q &= \mathbf{E}_q \left[\log \Pr(\mathbf{p}|\boldsymbol{\gamma}; \alpha) + \log \Pr(\boldsymbol{\gamma}|\pi_1)\right] \\ &= \sum_{j=1}^{M} \tilde{\gamma}_j (\log \alpha + (\alpha - 1) \log p_j + \log \pi_1) + \sum_{j=1}^{M} (1 - \tilde{\gamma}_j) \log(1 - \pi_1). \end{aligned}$$

The incomplete log likelihood can be evaluated as:

$$L = \sum_{j=1}^{M} \tilde{\gamma}_j (\log \alpha + (\alpha - 1) \log p_j + \log \pi_1 - \log \tilde{\gamma}_j) + \sum_{j=1}^{M} (1 - \tilde{\gamma}_j) (\log(1 - \pi_1) - \log(1 - \tilde{\gamma}_j)).$$

In the M step, we update α and π_1 by maximizing the Q function. We have

$$\begin{aligned}\alpha &= -\frac{\sum_{j=1}^{M}\tilde{\gamma}_j}{\sum_{j=1}^{M}\tilde{\gamma}_j \log p_j},\\ \pi_1 &= \frac{1}{M}\sum_{j=1}^{M}\tilde{\gamma}_j.\end{aligned}$$

Algorithm:

Input: $\mathbf{p}$, Initialize: $\alpha = 0.1$, $\pi_1 = 0.1$, Output: α, π_1, $\{\tilde{\gamma}_j\}_{j=1,...,M}$.

- Initialize $\alpha = 0.1$, $\pi_1 = 0.1$.
- E-step: For $j = 1, ..., M$, calculate $\tilde{\gamma}_j$ as follows

$$\tilde{\gamma}_j = \frac{\pi_1 \alpha p_j^{\alpha-1}}{\pi_1 \alpha p_j^{\alpha-1} + 1 - \pi_1}.$$

  Calculate L:

$$L = \sum_{j=1}^{M}\tilde{\gamma}_j\left(\log\alpha + (\alpha-1)\log p_j + \log\pi_1 - \log\tilde{\gamma}_j\right) + \sum_{j=1}^{M}(1-\tilde{\gamma}_j)\left(\log(1-\pi_1) - \log(1-\tilde{\gamma}_j)\right).$$

- M-step:

$$\begin{aligned}\alpha &= -\frac{\sum_{j=1}^{M}\tilde{\gamma}_j}{\sum_{j=1}^{M}\tilde{\gamma}_j \log p_j},\\ \pi_1 &= \frac{1}{M}\sum_{j=1}^{M}\tilde{\gamma}_j.\end{aligned}$$

- Check convergence.

Stage 2: Latent fixed-effect model (LFM)

Suppose we have the p-values of M SNPs for a given phenotype. Similarly, we assume

$$p_j \sim \begin{cases} U[0,1], & \gamma_j = 0,\\ Beta(\alpha, 1), & \gamma_j = 1,\end{cases}$$

where $\mathbf{p} \in \mathbb{R}^M$ are the p-values, $\gamma_j = 1$ indicates the j-th is associated with this phenotype and $\gamma_j = 0$ otherwise, and $0 < \alpha < 1$.

To integrate more information, we consider the logistic fixed-effect model:

$$\log\frac{\Pr(\gamma_j = 1|\mathbf{Z}_j)}{\Pr(\gamma_j = 0|\mathbf{Z}_j)} = \mathbf{Z}_j\mathbf{b},$$

where $\mathbf{Z} \in \mathbb{R}^{M\times(L+1)}$ and $\mathbf{b} = [b_0, b_1, b_2, ..., b_L]^T$ is an unknown vector of fixed effects, L is the number of covariates.

We can use EM algorithm to compute the posterior and parameter estimation.

Let $\boldsymbol{\theta} = \{\alpha, \mathbf{b}\}$ be the collection of model parameters. The complete data likelihood can be written as

$$\Pr(\mathbf{p}, \boldsymbol{\gamma}|\mathbf{Z}; \boldsymbol{\theta}) = \Pr(\mathbf{p}|\boldsymbol{\gamma}; \alpha) \Pr(\boldsymbol{\gamma}|\mathbf{Z}; \mathbf{b}),$$

where

$$\begin{aligned}
\Pr(\mathbf{p}|\boldsymbol{\gamma}; \alpha) &= \prod_{j=1}^{M} \Pr(p_j|\gamma_j; \alpha) = \prod_{j=1}^{M} \left(\alpha p_j^{\alpha-1}\right)^{\gamma_j}, \\
\Pr(\boldsymbol{\gamma}|\mathbf{Z}; \mathbf{b}) &= \prod_{j=1}^{M} e^{\gamma_j \mathbf{Z}_j \mathbf{b}} S(-\mathbf{Z}_j \mathbf{b}).
\end{aligned}$$

In the E step, we compute the posterior:

$$\tilde{\gamma}_j = q(\gamma_j = 1) = \frac{e^{\mathbf{Z}_j \mathbf{b}} \alpha p_j^{\alpha-1}}{e^{\mathbf{Z}_j \mathbf{b}} \alpha p_j^{\alpha-1} + 1},$$

and get the Q function:

$$Q = \sum_{j=1}^{M} \tilde{\gamma}_j (\log \alpha + (\alpha - 1) \log p_j + \mathbf{Z}_j \mathbf{b}) + \sum_{j=1}^{M} \log S(-\mathbf{Z}_j \mathbf{b}).$$

The incomplete log likelihood can be evaluated as:

$$L = \sum_{j=1}^{M} \tilde{\gamma}_j (\log \alpha + (\alpha - 1) \log p_j + \mathbf{Z}_j \mathbf{b} - \log \tilde{\gamma}_j) - \sum_{j=1}^{M} (1 - \tilde{\gamma}_j) \log(1 - \tilde{\gamma}_j) + \sum_{j=1}^{M} \log S(-\mathbf{Z}_j \mathbf{b}).$$

In the M step, we update α by maximizing the Q function. We have

$$\alpha = -\frac{\sum_{j=1}^{M} \tilde{\gamma}_j}{\sum_{j=1}^{M} \tilde{\gamma}_j \log p_j}.$$

We use Newton's method to update $\mathbf{b}$:

$$\mathbf{b} = \mathbf{b}_{old} - \mathbf{H}^{-1}\mathbf{g},$$

where

$$\begin{aligned}
\mathbf{g} &= \sum_{j=1}^{M} (-\tilde{\gamma}_j + S(\mathbf{Z}_j \mathbf{b})) \mathbf{Z}_j, \\
\mathbf{H} &= \sum_{j=1}^{M} S(\mathbf{Z}_j \mathbf{b}) S(-\mathbf{Z}_j \mathbf{b}) \mathbf{Z}_j^T \mathbf{Z}_j.
\end{aligned}$$

Algorithm:

Input: $\mathbf{p}$, $\mathbf{Z}$, α, $b_0 = \log\frac{\pi_1}{1-\pi_1}$, Output: α, $\mathbf{b}$, $\{\tilde{\gamma}_j\}_{j=1,...,M}$.

- Initialize α, $\mathbf{b} = (b_0, 0, ..., 0)^T$.
- E-step: For $j = 1, ..., M$, calculate $\tilde{\gamma}_j$ as follows

$$\tilde{\gamma}_j = q(\gamma_j = 1) = \frac{e^{\mathbf{Z}_j\mathbf{b}}\alpha p_j^{\alpha-1}}{e^{\mathbf{Z}_j\mathbf{b}}\alpha p_j^{\alpha-1} + 1}.$$

Calculate L:

$$L = \sum_{j=1}^{M} \tilde{\gamma}_j (\log\alpha + (\alpha-1)\log p_j + \mathbf{Z}_j\mathbf{b} - \log\tilde{\gamma}_j) - \sum_{j=1}^{M}(1-\tilde{\gamma}_j)\log(1-\tilde{\gamma}_j) + \sum_{j=1}^{M}\log S(-\mathbf{Z}_j\mathbf{b}).$$

- M-step

$$\begin{aligned}
\alpha &= -\frac{\sum_{j=1}^{M}\pi_j}{\sum_{j=1}^{M}\pi_j\log p_j}, \\
\mathbf{g} &= \sum_{j=1}^{M}(-\tilde{\gamma}_j + S(\mathbf{Z}_j\mathbf{b}))\mathbf{Z}_j, \\
\mathbf{H} &= \sum_{j=1}^{M} S(\mathbf{Z}_j\mathbf{b}) S(-\mathbf{Z}_j\mathbf{b}) \mathbf{Z}_j^T\mathbf{Z}_j, \\
\mathbf{b} &= \mathbf{b}_{old} - \mathbf{H}^{-1}\mathbf{g}.
\end{aligned}$$

- Check convergence.

Stage 3: Logistic sparse mixed model

Suppose we know the latent states $\boldsymbol{\gamma}$ of M SNPs for a given phenotype is given. We consider a logistic mixed model:

$$\log\frac{\Pr(\gamma_j = 1|\mathbf{Z}_j, \mathbf{A}_j)}{\Pr(\gamma_j = 0|\mathbf{Z}_j, \mathbf{A}_j)} = \mathbf{Z}_j\mathbf{b} + \mathbf{A}_j\boldsymbol{\beta} = \sum_{l=0}^{L}\mathbf{Z}_{jl}b_l + \sum_{k=1}^{K}A_{jk}\beta_k,$$

where $\mathbf{Z} \in \mathbb{R}^{M\times(L+1)}$, $A \in \mathbb{R}^{M\times K}$, $\mathbf{b} = [b_0, b_1, b_2, ..., b_L]^T$ is an unknown vector of fixed effects, $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_K]^T$ is an unknown vector of random effects with a spike-slab prior:

$$\beta_k \sim \begin{cases} N(0, \sigma^2), & \eta_k = 1, \\ \delta_0, & \eta_k = 0, \end{cases}$$

where η_k is another latent variable with $\Pr(\eta_k = 1) = \omega$. Here $\eta_k = 1$ means the k-th annotation is relevant to this phenotype and $\eta_k = 0$ otherwise.

To handle the Dirac function, we reparemetrize the spike-slab prior as $\tilde{\beta}_k \sim N\left(0, \sigma^2\right)$, then $\beta_k = \eta_k \tilde{\beta}_k$.

We can use variational EM algorithm to compute the posterior and parameter estimation.

Let $\boldsymbol{\theta} = \left\{\alpha, \mathbf{b}, \sigma^2, \omega\right\}$ be the collection of model parameters. Using the sigmoid function denoted as $S(x) = \frac{1}{1+e^{-x}}$, the complete data likelihood can be written as

$$\Pr\left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta} | \mathbf{Z}, \mathbf{A}; \boldsymbol{\theta}\right) = \Pr\left(\boldsymbol{\gamma} | \mathbf{Z}, \mathbf{A}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}\right) \Pr\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\eta} | \sigma^2, \omega\right),$$

where

$$\begin{aligned}
\Pr\left(\boldsymbol{\gamma} | \mathbf{Z}, \mathbf{A}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}\right) &= \prod_{j=1}^{M} \Pr\left(\gamma_j | \mathbf{Z}_j, \mathbf{A}_j, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}\right) \\
&= \prod_{j=1}^{M} e^{\gamma_j \left(\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \eta_k \tilde{\beta}_k\right)} S\left(-\mathbf{Z}_j \mathbf{b} - \sum_k A_{jk} \eta_k \tilde{\beta}_k\right), \\
\Pr\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\eta} | \sigma^2, \omega\right) &= \prod_{k=1}^{K} \Pr\left(\tilde{\beta}_k, \eta_k | \sigma^2, \omega\right) = \prod_{k=1}^{K} N\left(\tilde{\beta}_k | 0, \sigma^2\right) \omega^{\eta_k} \left(1-\omega\right)^{1-\eta_k}.
\end{aligned}$$

We can use JJ bound (Jaakkola and Jordan, 2000) to bound the sigmoid function by

$$S(x) \geq S(\xi) \exp\left\{(x-\xi)/2 - \lambda(\xi)\left(x^2 - \xi^2\right)\right\},$$

where $\lambda(\xi) = \frac{1}{2\xi}\left[S(\xi) - \frac{1}{2}\right]$. Using this bound, we have a tractable lower bound of $\Pr\left(\boldsymbol{\gamma} | \mathbf{Z}, \mathbf{A}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}\right)$ which is denoted by $h\left(\boldsymbol{\gamma} | \mathbf{Z}, \mathbf{A}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}, \boldsymbol{\xi}\right)$:

$$\begin{aligned}
& h\left(\gamma_j | \mathbf{Z}_j, \mathbf{A}_j, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}; \mathbf{b}, \xi_j\right) \\
= \; & e^{\gamma_j \left(\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \eta_k \tilde{\beta}_k\right)} S(\xi_j) \exp\left(-\lambda(\xi_j)\left(\left(\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \eta_k \tilde{\beta}_k\right)^2 - \xi_j^2\right) - \frac{\mathbf{Z}_j \mathbf{b} + \sum_k A_{jk} \eta_k \tilde{\beta}_k + \xi_j}{2}\right).
\end{aligned}$$

Next, Let $q\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}\right)$ be an approximation of the posterior $\Pr\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\eta} | \mathbf{Z}, \mathbf{A}; \boldsymbol{\theta}\right)$. Then we can obtain a lower bound of the logarithm of the marginal likelihood:

$$
\begin{aligned}
& \log \Pr\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A};\boldsymbol{\theta}\right) \\
= & \log \sum_{\boldsymbol{\eta}} \int \Pr\left(\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\mathbf{Z},\mathbf{A};\boldsymbol{\theta}\right) d\tilde{\boldsymbol{\beta}} \\
= & \log \sum_{\boldsymbol{\eta}} \int \Pr\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b}\right) \Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right) d\tilde{\boldsymbol{\beta}} \\
\geq & \log \sum_{\boldsymbol{\eta}} \int h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b},\boldsymbol{\xi}\right) \Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right) d\tilde{\boldsymbol{\beta}} \\
\geq & \sum_{\boldsymbol{\eta}} \int q\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right) \log \frac{h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b},\boldsymbol{\xi}\right) \Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right)}{q\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)} d\tilde{\boldsymbol{\beta}} \\
= & \mathbf{E}_q\left[\log h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta};\mathbf{b},\boldsymbol{\xi}\right) + \log \Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right) - \log q\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)\right] \\
\triangleq & L\left(q\right),
\end{aligned}
$$

where $L(q)$ is the lower bound. The second inequality follows Jensen's inequality. We can maximize $L(q)$ instead of the marginal likelihood to get parameter estimations. To make it feasible to evaluate the lower bound, we assume that $q\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}\right)$ can be factorized as

$$
q\left(\tilde{\beta},\eta\right) = \prod_{k=1}^{K} q\left(\tilde{\beta}_k,\eta_k\right) = \prod_{k=1}^{K} q\left(\tilde{\beta}_k|\eta_k\right) q\left(\eta_k\right),
$$

where $q\left(\eta_k = 1\right) = \omega_k$.

We can obtain an approximation according to the mean-field method:

$$
\log q\left(\tilde{\beta}_i,\eta_i\right) = \mathbf{E}_{k\neq i}\left[\log h\left(\boldsymbol{\gamma}|\mathbf{Z},\mathbf{A},\tilde{\boldsymbol{\beta}},\boldsymbol{\eta},\mathbf{b},\boldsymbol{\xi}\right) + \log \Pr\left(\tilde{\boldsymbol{\beta}},\boldsymbol{\eta}|\sigma^2,\omega\right)\right],
$$

where the expectation is taken under the distribution $q\left(\tilde{\beta}_{-i},\eta_{-i}\right) = \prod_{k\neq i} q\left(\tilde{\beta}_k,\eta_k\right)$. Then we have

$$
q\left(\tilde{\beta}_i,\eta_i\right) = \left[\omega_i N\left(\mu_i,s_i^2\right)\right]^{\eta_i}\left[\left(1-\omega_i\right) N\left(0,\sigma^2\right)\right]^{1-\eta_i},
$$

where

$$
\begin{aligned}
\mu_i &= s_i^2 \sum_{j=1}^{M} \left(\pi_j - \frac{1}{2} - 2\lambda\left(\xi_j\right)\left(\mathbf{Z}_j\mathbf{b} + \sum_{k\neq i} A_{jk}\mathbf{E}_k\left[\eta_k\tilde{\beta}_k\right]\right)\right) A_{ji}, \\
s_i^2 &= \frac{\sigma^2}{1 + 2\sigma^2\sum_{j=1}^{M}\lambda\left(\xi_j\right)A_{ji}^2}.
\end{aligned}
$$

Then we maximize $L\left(q\right)$ with respect to ω_k and ξ_j and get

$$
\begin{aligned}
\omega_k &= \frac{1}{1+\exp\left(-u_k\right)}, \text{ where } u_k = \log\frac{\omega}{1-\omega} + \frac{1}{2}\log\frac{s_k^2}{\sigma^2} + \frac{\mu_k^2}{2s_k^2}, \\
\xi_j^2 &= \left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right)^2 + \sum_k A_{jk}^2\left(\omega_k\left(s_k^2+\mu_k^2\right) - \omega_k^2\mu_k^2\right).
\end{aligned}
$$

Now we have evaluate $L(q)$:

$$
\begin{aligned}
& L(q) \\
= & \sum_{j=1}^{M}\left(\gamma_j\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right) + \log S\left(\xi_j\right) - \lambda\left(\xi_j\right)\left(\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right)^2 - \xi_j^2\right)\right) \\
& + \sum_{j=1}^{M}\left(-\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k + \xi_j\right)/2 + \lambda\left(\xi_j\right)\sum_k A_{jk}^2\omega_k^2\mu_k^2 - \lambda\left(\xi_j\right)\sum_k A_{jk}^2\omega_k\left(s_k^2+\mu_k^2\right)\right) \\
& - \frac{1}{2\sigma^2}\sum_{k=1}^{K}\left(\omega_k\left(s_k^2+\mu_k^2\right) - \omega_k\sigma^2\right) + \sum_{k=1}^{K}\omega_k\log\omega + \sum_{k=1}^{K}\left(1-\omega_k\right)\log\left(1-\omega\right) \\
& + \sum_{k=1}^{K}\left(\frac{1}{2}\omega_k\left(\log s_k^2 - \log\sigma^2\right) - \omega_k\log\omega_k - \left(1-\omega_k\right)\log\left(1-\omega_k\right)\right).
\end{aligned}
$$

With $q\left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\eta}\right)$ obtained, we can evaluate the lower bound and then update the model parameters by maximizing $L(q)$.

In the M step, we update σ^2 and ω by maximizing $L(q)$. We have

$$
\begin{aligned}
\sigma^2 &= \frac{\sum_{k=1}^{K}\omega_k\left(s_k^2+\mu_k^2\right)}{\sum_{k=1}^{K}\omega_k}, \\
\omega &= \frac{1}{K}\sum_{k=1}^{K}\omega_k.
\end{aligned}
$$

We use Newton's method to update $\mathbf{b}$:

$$
\mathbf{b} = \mathbf{b}_{old} - \mathbf{H}^{-1}\mathbf{g},
$$

where

$$
\begin{aligned}
\mathbf{g} &= -\sum_{j=1}^{M}\mathbf{Z}_j^T\left(\gamma_j - 2\lambda\left(\xi_j\right)\left(\mathbf{Z}_j\mathbf{b} + \sum_k A_{jk}\omega_k\mu_k\right) - \frac{1}{2}\right), \\
\mathbf{H} &= 2\sum_{j=1}^{M}\lambda\left(\xi_j\right)\mathbf{Z}_j^T\mathbf{Z}_j.
\end{aligned}
$$

Algorithm:

Input: $\mathbf{Z}$, $\mathbf{A}$, $\{\gamma_j = \tilde{\gamma}_j\}_{j=1,...,M}$, $\mathbf{b}$, Initialize: $\sigma^2 = 1$, $\omega = 0.5$, $\{\omega_k = 0, \mu_k = 0\}_{k=1,...K}$, $\boldsymbol{\xi} = \mathbf{Zb}$, Output: $\mathbf{b}$, $\boldsymbol{\xi}$, σ^2, ω, $\{\omega_k, \mu_k\}_{k=1,...K}$.

- Initialize $\mathbf{b}$, $\boldsymbol{\xi} = \mathbf{Zb}$, $\sigma^2 = 1$, $\omega = 0.5$, $\{\omega_k = 0, \mu_k = 0\}_{k=1,...K}$. Let $\tilde{y} = \sum_k A_{jk}\omega_k\mu_k$.
- E-step: For $i = 1, ..., K$, first obtain $\tilde{y}_i = \tilde{y} - A_{ji}\omega_i\mu_i$, and then update μ_i, s_i^2, ω_i and $\tilde{y}$ as follows

$$\begin{aligned}
s_i^2 &= \frac{\sigma^2}{1 + 2\sigma^2 \sum_{j=1}^{M} \lambda(\xi_j) A_{ji}^2}, \\
\mu_i &= s_i^2 \sum_{j=1}^{M} \left(\left(\gamma_j - \frac{1}{2} - 2\lambda(\xi_j)(\mathbf{Z}_j\mathbf{b} + \tilde{y}_i) \right) A_{ji} \right), \\
\omega_i &= \frac{1}{1 + \exp(-u_i)}, \text{ where } u_i = \log\frac{\omega}{1-\omega} + \frac{1}{2}\log\frac{s_i^2}{\sigma^2} + \frac{\mu_i^2}{2s_i^2}, \\
\tilde{y} &= \tilde{y}_i + A_{ji}\omega_i\mu_i.
\end{aligned}$$

  Then for $j = 1, ..., M$, update ξ_j as follows

$$\xi_j^2 = (\mathbf{Z}_j\mathbf{b} + \tilde{y})^2 + \sum_k A_{jk}^2 \left(\omega_k \left(s_k^2 + \mu_k^2 \right) - \omega_k^2\mu_k^2 \right).$$

  Calculate $L(q)$:

$$\begin{aligned}
& L(q) \\
= & \sum_{j=1}^{M} \left(\gamma_j (\mathbf{Z}_j\mathbf{b} + \tilde{y}) + \log S(\xi_j) - \frac{\mathbf{Z}_j\mathbf{b} + \tilde{y} + \xi_j}{2} \right) \\
& - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(\omega_k \left(s_k^2 + \mu_k^2 \right) - \omega_k\sigma^2 \right) + \sum_{k=1}^{K} \omega_k \log\omega + \sum_{k=1}^{K} (1 - \omega_k) \log(1 - \omega) \\
& + \sum_{k=1}^{K} \left(\frac{1}{2}\omega_k \left(\log s_k^2 - \log\sigma^2 \right) - \omega_k \log\omega_k - (1 - \omega_k) \log(1 - \omega_k) \right).
\end{aligned}$$

- M-step

$$
\begin{aligned}
\mathbf{g} &= -\sum_{j=1}^{M} \mathbf{Z}_j^T \left(\pi_j - 2\lambda\left(\xi_j\right)\left(\mathbf{Z}_j\mathbf{b} + \tilde{y}\right) - \frac{1}{2} \right), \\
\mathbf{H} &= 2\sum_{j=1}^{M} \lambda\left(\xi_j\right) \mathbf{Z}_j^T \mathbf{Z}_j, \\
\mathbf{b} &= \mathbf{b}_{old} - \mathbf{H}^{-1}\mathbf{g}, \\
\sigma^2 &= \frac{\sum_{k=1}^{K} \omega_k \left(s_k^2 + \mu_k^2\right)}{\sum_{k=1}^{K} \omega_k}, \\
\omega &= \frac{1}{K}\sum_{k=1}^{K} \omega_k.
\end{aligned}
$$

- Check convergence.

Stage 4: LSMM

Input: $\mathbf{p}$, $\mathbf{Z}$, $\mathbf{A}$, α,$\mathbf{b}$, $\boldsymbol{\xi}$, σ^2, ω, $\{\omega_k, \mu_k\}_{k=1,...K}$, Initialize: $\{\pi_j = \tilde{\gamma}_j\}_{j=1,...,M}$, Output: α,$\mathbf{b}$, σ^2, ω, $\{\omega_k, \beta_k = \mu_k\omega_k\}_{k=1,...K}$, $\{\pi_j\}_{j=1,...,M}$

Algorithm:

- Initialize α, σ^2, ω, $\mathbf{b}$, $\{\omega_k, \mu_k\}_{k=1,...K}$, $\{\xi_j, \pi_j\}_{j=1,...,M}$. Let $\tilde{y} = \sum_k A_{jk}\omega_k\mu_k$.
- E-step: For $i = 1, ..., K$, first obtain $\tilde{y}_i = \tilde{y} - A_{ji}\omega_i\mu_i$, and then update μ_i, s_i^2, ω_i and $\tilde{y}$ as follows

$$
\begin{aligned}
s_i^2 &= \frac{\sigma^2}{1 + 2\sigma^2 \sum_{j=1}^{M} \lambda\left(\xi_j\right) A_{ji}^2}, \\
\mu_i &= s_i^2 \sum_{j=1}^{M} \left(\left(\pi_j - \frac{1}{2} - 2\lambda\left(\xi_j\right)\left(\mathbf{Z}_j\mathbf{b} + \tilde{y}_i\right) \right) A_{ji} \right), \\
\omega_i &= \frac{1}{1 + \exp\left(-u_i\right)}, \text{ where } u_i = \log\frac{\omega}{1-\omega} + \frac{1}{2}\log\frac{s_i^2}{\sigma^2} + \frac{\mu_i^2}{2s_i^2}, \\
\tilde{y} &= \tilde{y}_i + A_{ji}\omega_i\mu_i.
\end{aligned}
$$

Then for $j = 1, ..., M$, update π_j, ξ_j as follows

$$
\begin{aligned}
\pi_j &= \frac{1}{1 + \exp\left(-v_j\right)}, \text{ where } v_j = \log\alpha + (\alpha - 1)\log p_j + \mathbf{Z}_j\mathbf{b} + \tilde{y}, \\
\xi_j^2 &= \left(\mathbf{Z}_j\mathbf{b} + \tilde{y}\right)^2 + \sum_k A_{jk}^2 \left(\omega_k \left(s_k^2 + \mu_k^2\right) - \omega_k^2\mu_k^2 \right).
\end{aligned}
$$

Calculate $L\left(q\right)$:

$$
\begin{aligned}
& L(q) \\
= \quad & \sum_{j=1}^{M} \pi_j \left(\log \alpha + (\alpha - 1) \log p_j \right) - \sum_{j=1}^{M} \left(\pi_j \log \pi_j + (1 - \pi_j) \log (1 - \pi_j) \right) \\
& + \sum_{j=1}^{M} \left(\pi_j \left(\mathbf{Z}_j \mathbf{b} + \tilde{y} \right) + \log S \left(\xi_j \right) - \frac{\mathbf{Z}_j \mathbf{b} + \tilde{y} + \xi_j}{2} \right) \\
& - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(\omega_k \left(s_k^2 + \mu_k^2 \right) - \omega_k \sigma^2 \right) + \sum_{k=1}^{K} \omega_k \log \omega + \sum_{k=1}^{K} (1 - \omega_k) \log (1 - \omega) \\
& + \sum_{k=1}^{K} \left(\frac{1}{2} \omega_k \left(\log s_k^2 - \log \sigma^2 \right) - \omega_k \log \omega_k - (1 - \omega_k) \log (1 - \omega_k) \right).
\end{aligned}
$$

- M-step

$$
\begin{aligned}
\alpha &= -\frac{\sum_{j=1}^{M} \pi_j}{\sum_{j=1}^{M} \pi_j \log p_j}, \\
\sigma^2 &= \frac{\sum_{k=1}^{K} \omega_k \left(s_k^2 + \mu_k^2 \right)}{\sum_{k=1}^{K} \omega_k}, \\
\omega &= \frac{1}{K} \sum_{k=1}^{K} \omega_k, \\
\mathbf{g} &= -\sum_{j=1}^{M} \mathbf{Z}_j^T \left(\pi_j - 2\lambda \left(\xi_j \right) \left(\mathbf{Z}_j \mathbf{b} + \tilde{y} \right) - \frac{1}{2} \right), \\
\mathbf{H} &= 2 \sum_{j=1}^{M} \lambda \left(\xi_j \right) \mathbf{Z}_j^T \mathbf{Z}_j, \\
\mathbf{b} &= \mathbf{b}_{old} - \mathbf{H}^{-1} \mathbf{g}.
\end{aligned}
$$

- Evaluate $L(q)$ to track the convergence of the algorithm.

3 More simulation results

3.1 Performance in identification of risk SNPs

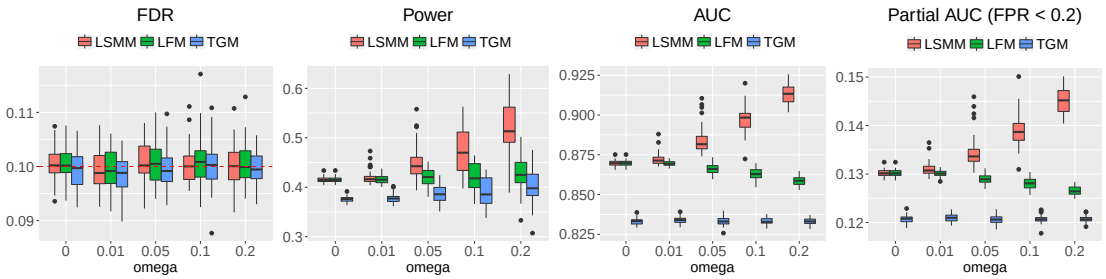

Figure S1: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.2$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

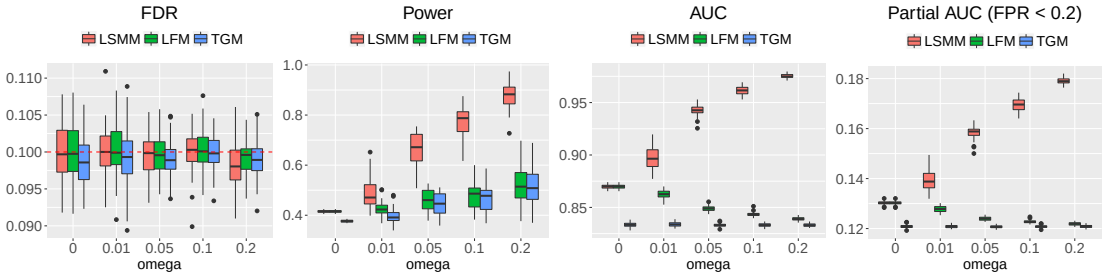

Figure S2: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.2$ and $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

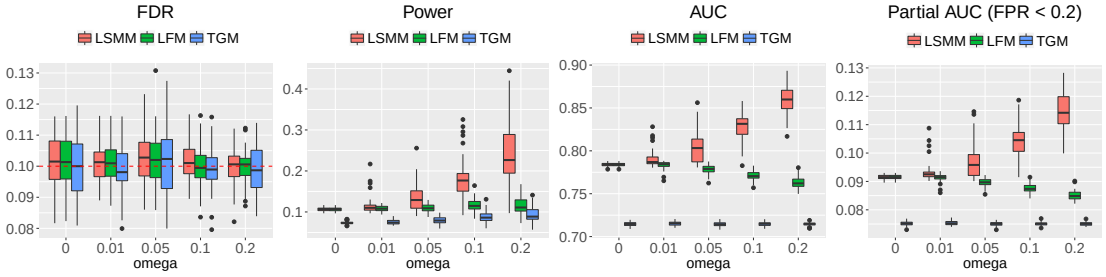

Figure S3: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.4$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

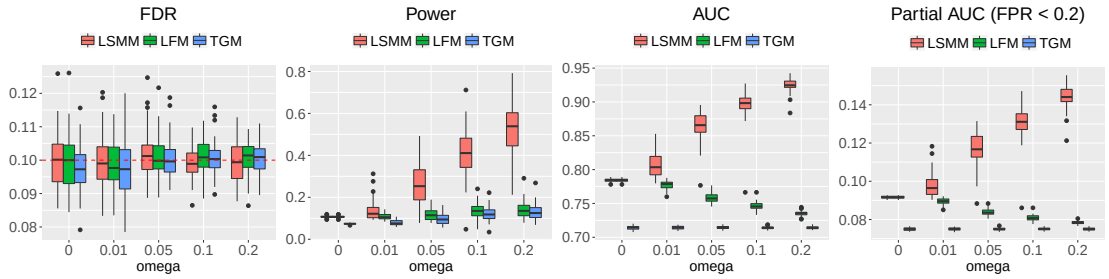

Figure S4: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.4$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

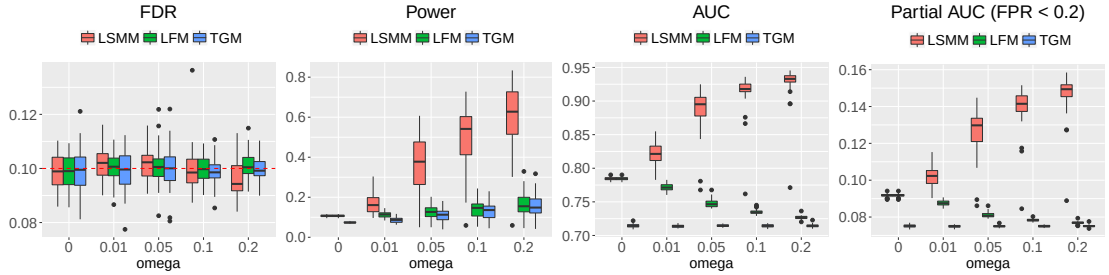

Figure S5: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.4$ and $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

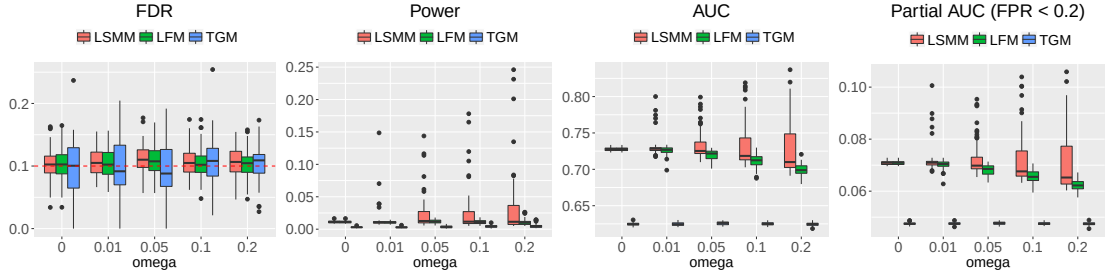

Figure S6: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.6$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

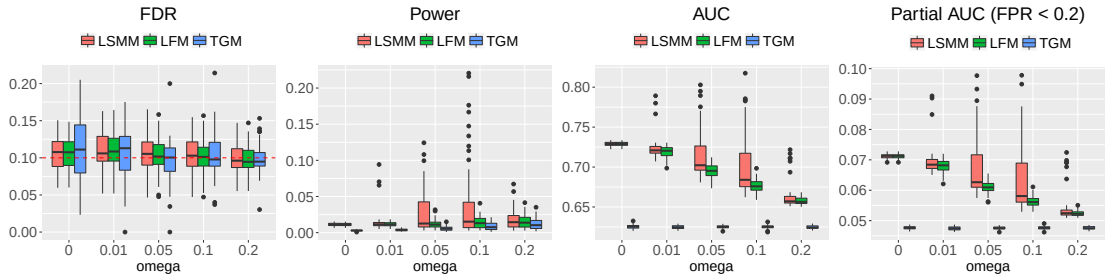

Figure S7: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.6$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

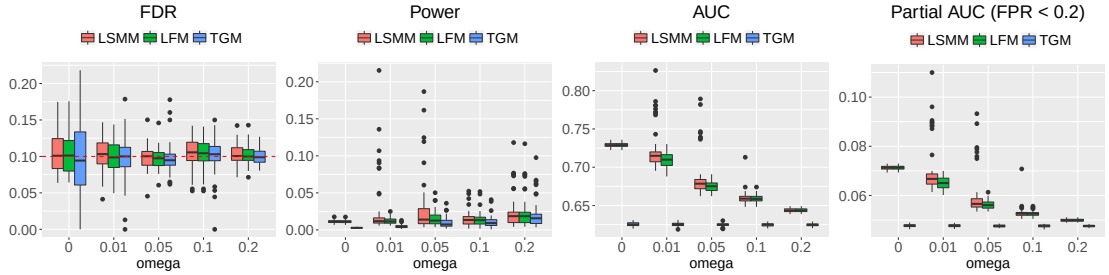

Figure S8: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs with $\alpha = 0.6$ and $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.2 Performance in identification of risk SNPs if treat the effects of all covariates as fixed effects

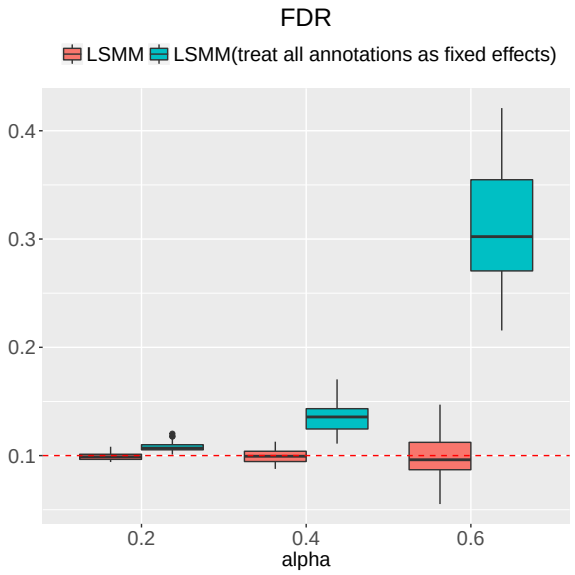

Figure S9: FDR of LSMM and LSMM (treat the effects of all covariates as fixed effects) for identification of risk SNPs with $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

3.3 Performance in identification of relevant annotations

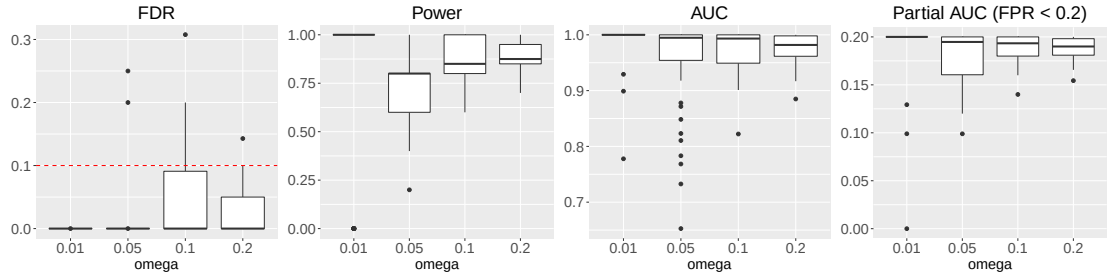

Figure S10: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.2$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

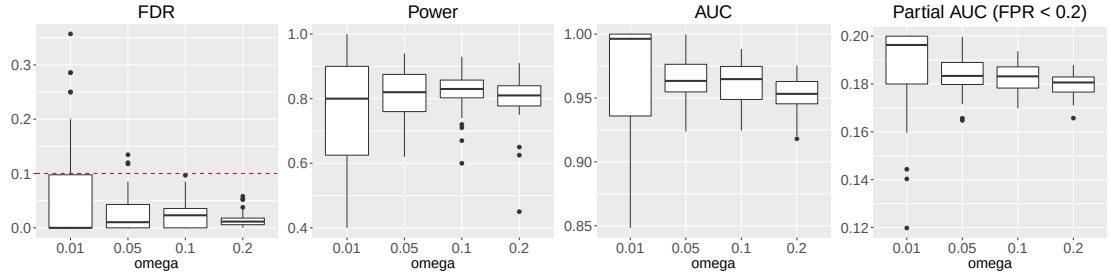

Figure S11: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.2$ and $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

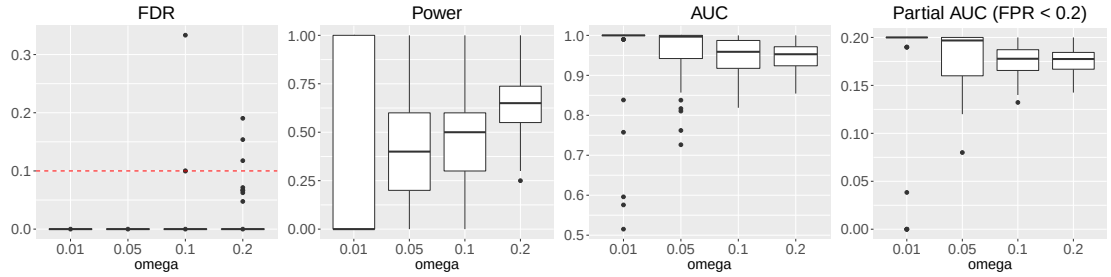

Figure S12: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.4$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

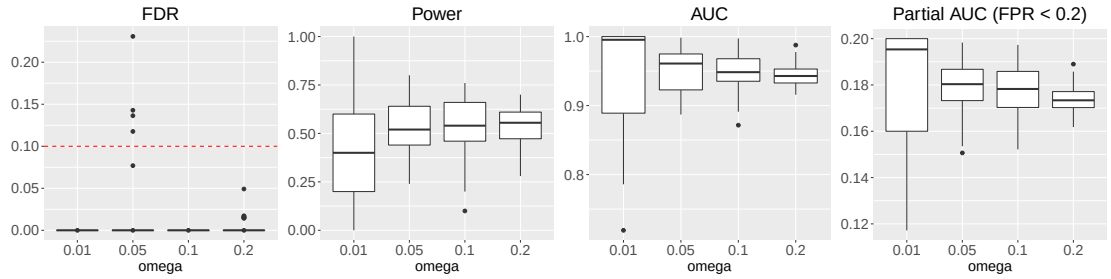

Figure S13: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.4$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

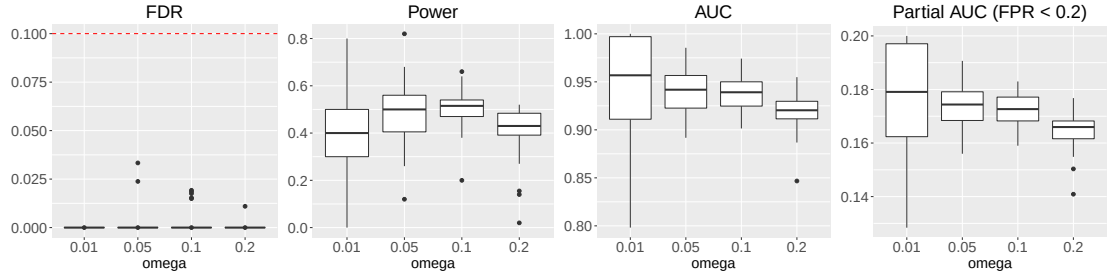

Figure S14: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.4$ and $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

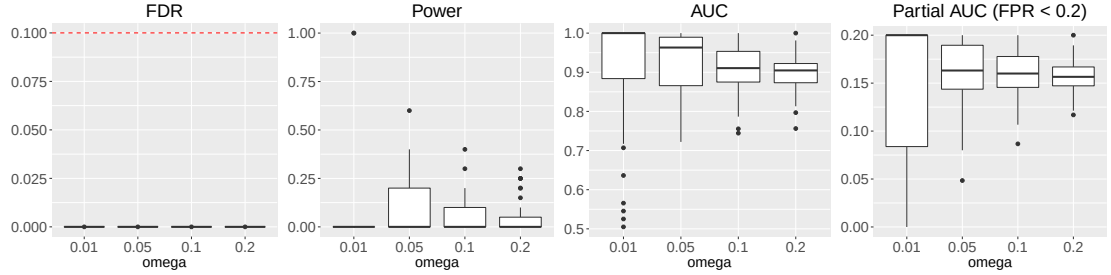

Figure S15: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.6$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

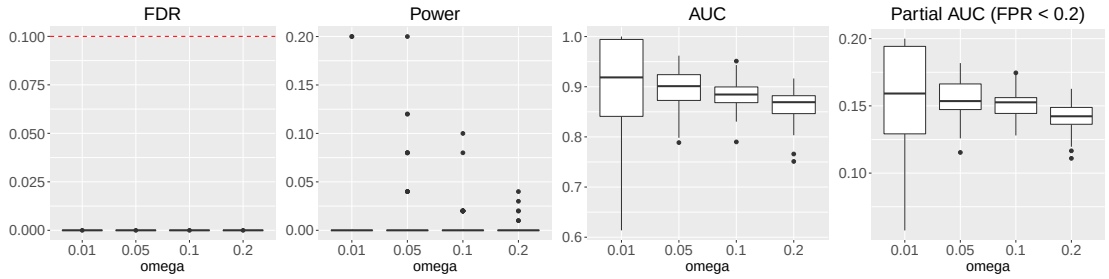

Figure S16: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.6$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

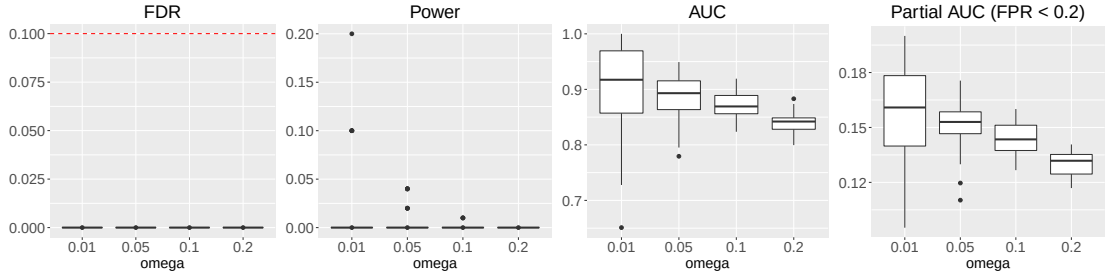

Figure S17: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.6$ and $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.4 Performance in identification of relevant annotations when the number of SNPs is large

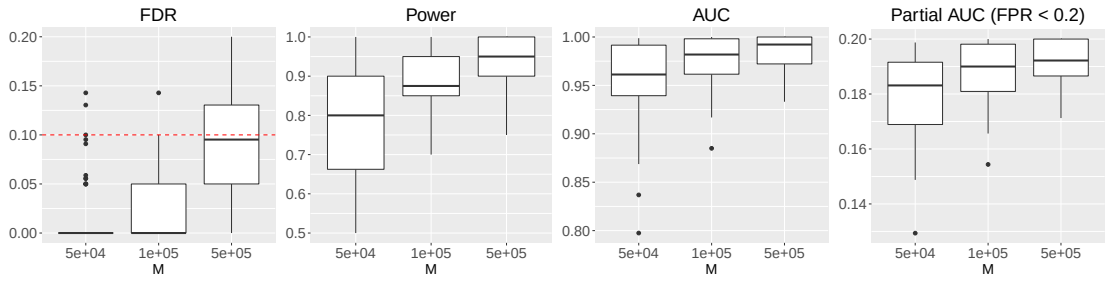

Figure S18: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.2$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

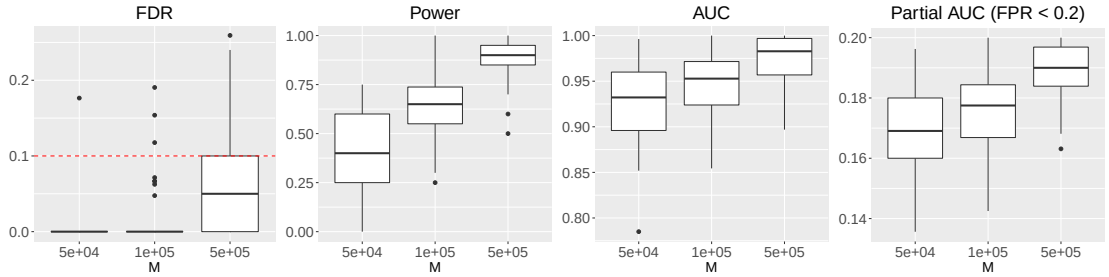

Figure S19: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.4$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

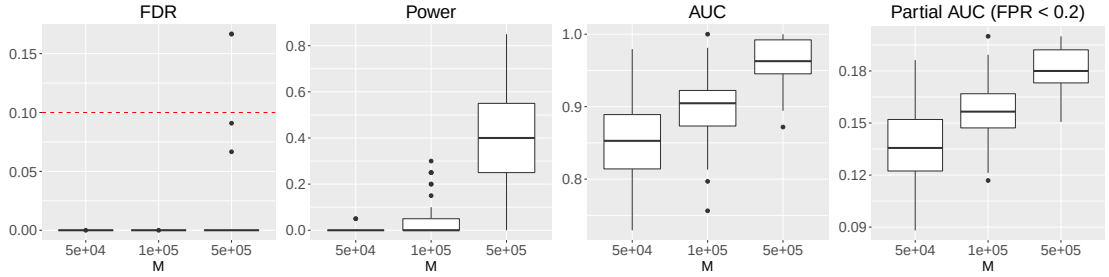

Figure S20: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations with $\alpha = 0.6$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.5 Performance of LSMM when the number of SNPs is small

We conducted simulations with the number of SNPs M varied from 1,000 to 100,000 to evaluate the performance of LSMM. In the simulation, we set $L = 10$, $K = 100$, $\alpha = 0.2$ and $\omega = 0.2$. To easily control signal-noise ratio, we used the probit model:

$$y_j = \mathbf{Z}_j\mathbf{b} + \mathbf{A}_j\boldsymbol{\beta} + e_j, \tag{S1}$$

where $e_j \sim N\left(0, \sigma_e^2\right)$. And we set $\gamma_j = 1$ if $y_j > 0$, $\gamma_j = 0$ if $y_j \leq 0$. The first entry of the coefficients of fixed effects $\mathbf{b}$, i.e. the intercept term, was fixed at -1 and other entries were generated from $N\left(0, 1\right)$ and fixed during multiple replications. We varied the signal-noise ratio $r = \frac{var(\mathbf{Zb}+\mathbf{A}\boldsymbol{\beta})}{var(\mathbf{e})} = \{4:1, 1:1, 1:4\}$. The results are given in Figures S21 and S22. As the number of SNPs becomes smaller, the performance of LSMM for both identification of risk SNPs and detection of relevant annotations become worse, as indicated by power, AUC and partial AUC. With a large signal-noise ratio, the performance of LSMM becomes better, especially when the number of SNPs is small. In order to obtain reliable results using LSMM, the number of SNPs should not be very small. To summarize, LSMM could be applied to a subset of SNPs when the number of SNPs is not too small and signals from annotations are not too weak.

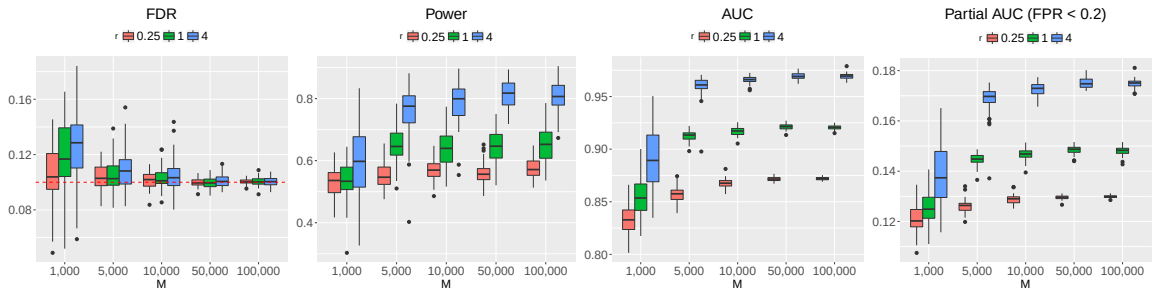

Figure S21: FDR, power, AUC and partial AUC of LSMM for identification of risk SNPs based on probit model. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

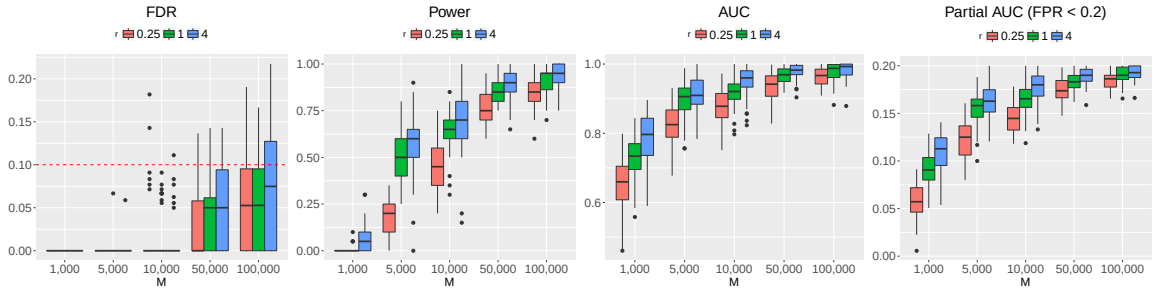

Figure S22: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations based on probit model. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.6 Performance in identification of relevant annotations when the annotations in design matrix of fixed effects and random effects are not independent

We simulated a case that 10 genic category annotations and first 50 cell-type specific annotations are correlated with correlation coefficient varied at at $\{0, 0.2, 0.4, 0.6, 0.8\}$ and the remaining annotations are generated independently. To simulate the design matrices for genic category and cell-type specific annotations, we first simulated M samples from a multivariate normal distribution with the correlation matrix among annotations and then made a cutoff so that 10% of the entries would be 1 and the others be 0.

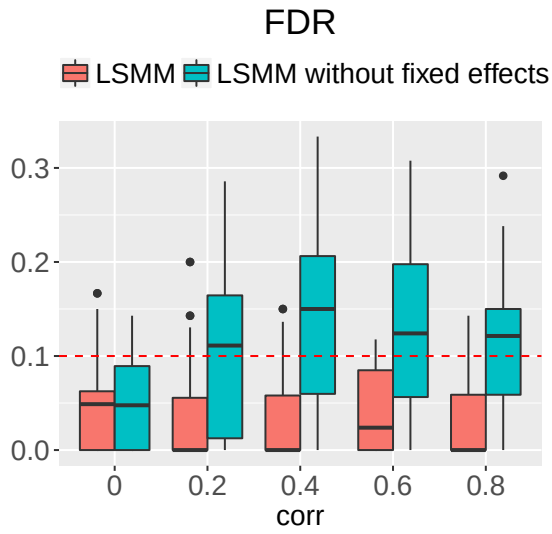

Figure S23: FDR of LSMM and LSMM without fixed effects for detection of relevant annotations with $\alpha = 0.2$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

3.7 Simulations to investigate the sensitivity of LSMM to initial parameter specification

The parameters which need initialization in LSMM are only α and π_1 in the first stage (Two-group model, in short, TGM), and their estimates naturally give the starting point of the second stage. Here we used the TGM to generate data such that we can evaluate whether the estimates converge to their true values. In the simulation, we set the numbers of SNPs $M = 100,000$ and varied the true value of $\pi_1 \in \{0.01, 0.05, 0.1, 0.15, 0.2\}$. To check whether LSMM could give accurate estimates when using different initial values, we considered two cases, default setting and random setting. In the default setting, both α and π_1 are initialized at 0.1. In the random setting, we randomly generated the initial values of α from $U[0.1, 0.6]$ and the initial values of π_1 from $U[0, 0.3]$.

The results of the estimation $\hat{\pi}_1$ using LSMM (default setting and random setting) are shown in the upper panel of Figure S24. The true values are indicated by dotted lines with different colors. Comparing the performance of difference initial value settings, default setting and random setting, we note that LSMM is not sensitive to initial parameter specification in most situations except when the true proportion of risk SNPs is small and the signal of GWAS data is weak (e.g., $\pi_1 = 0.01$ for $\alpha = 0.6$). However, LSMM can still provide a valid FDR control which is shown in the lower panel of Figure S24. In the context of GWAS, the proportion of risk variants is not very small due to the polygenic effect. Therefore, we believe LSMM with default setting will work well in practice, without suffering much from initialization.

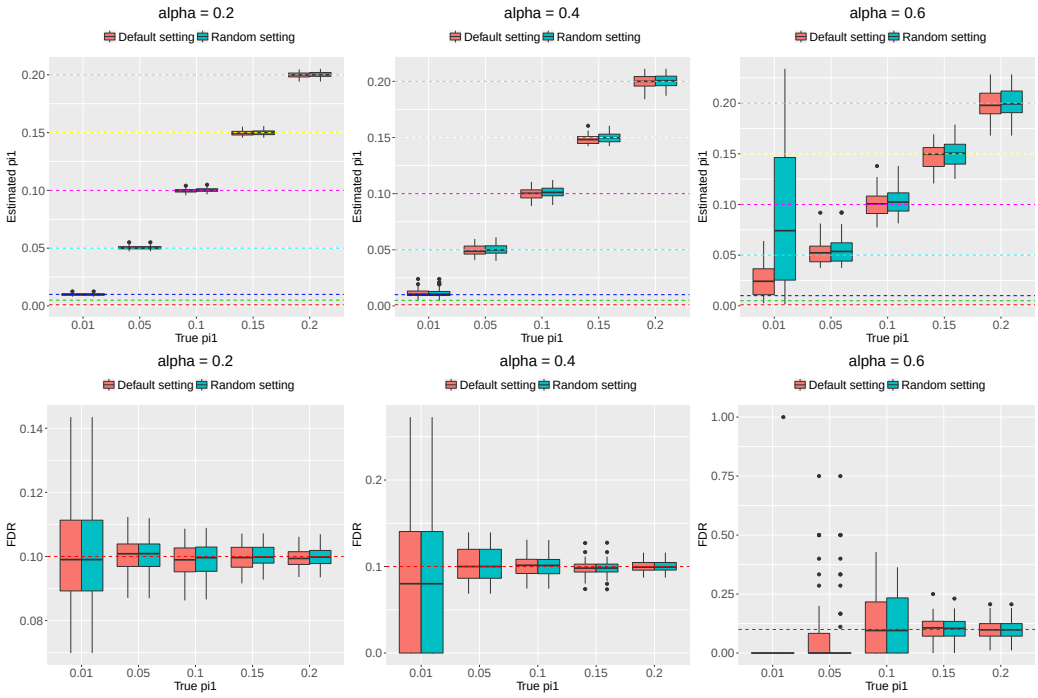

Figure S24: Upper panel: parameter estimation ($\hat{\pi}_1$ v.s. true π_1) using LSMM (default setting and random setting). Lower panel: FDR for identification of risk SNPs using LSMM (default setting and random setting). We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

3.8 Estimation of parameters

3.8.1 Estimation of α

We evaluate the performance of LSMM in estimation of parameter α in the beta distribution. We compare LSMM with the other three methods, TGM (without fixed effects and random effects), LFM (with only fixed effects) and LSMM without fixed effects. We varied ω at $\{0, 0.25, 0.5, 0.75, 1\}$. Figures S25-S27 show the comparison among these methods with $\alpha = 0.2$, 0.4 and 0.6 respectively.

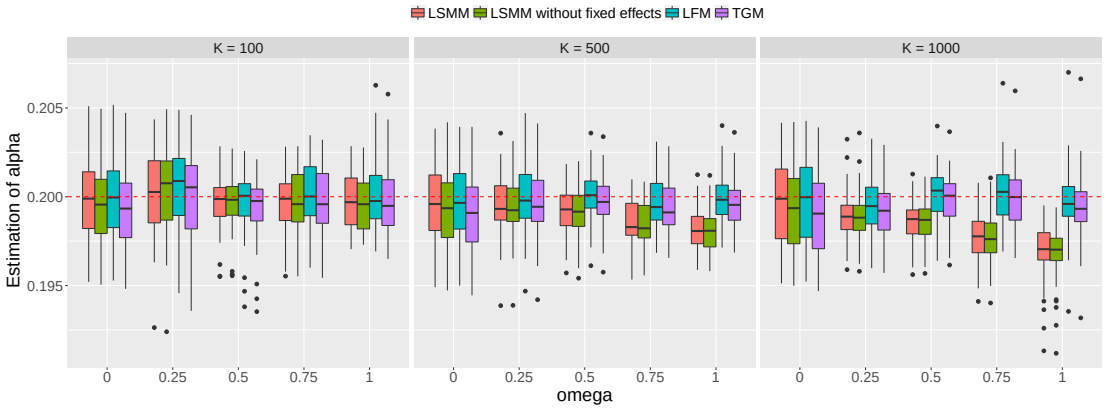

Figure S25: Performance in estimation of parameter α when the true $\alpha = 0.2$.

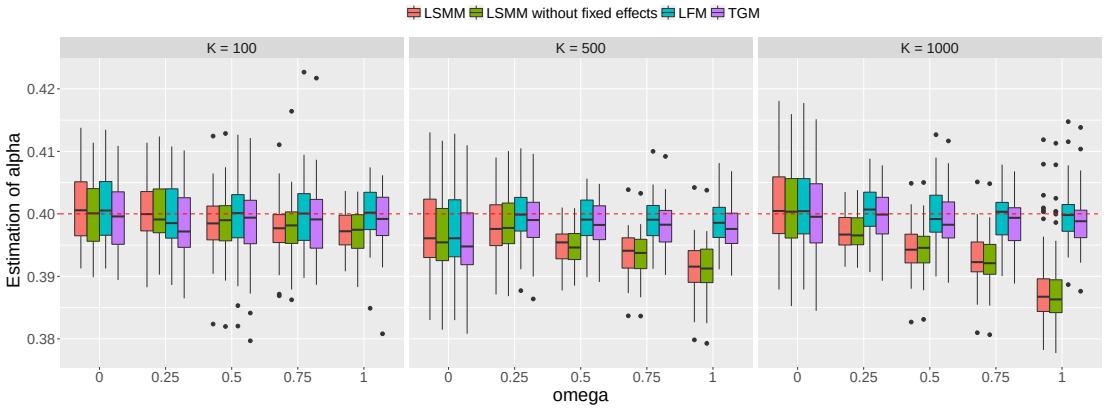

Figure S26: Performance in estimation of parameter α when the true $\alpha = 0.4$.

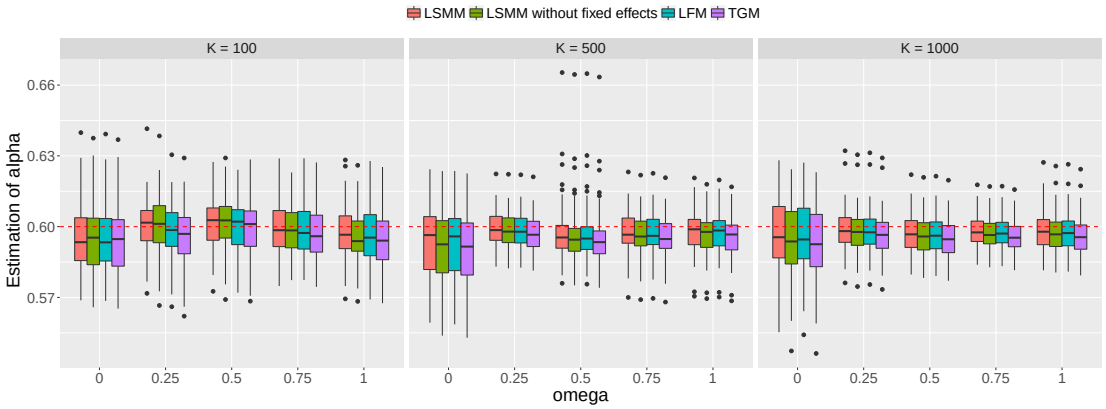

Figure S27: Performance in estimation of parameter α when the true $\alpha = 0.6$.

3.8.2 Estimation of b

We evaluate the performance of LSMM in estimation of parameter b_0 and $\boldsymbol{b}$. We varied ω at $\{0, 0.25, 0.5, 0.75, 1\}$. Figures S28-S38 show the comparison between LSMM and LFM (with only fixed effects) with $\alpha = 0.2$, 0.4 and 0.6.

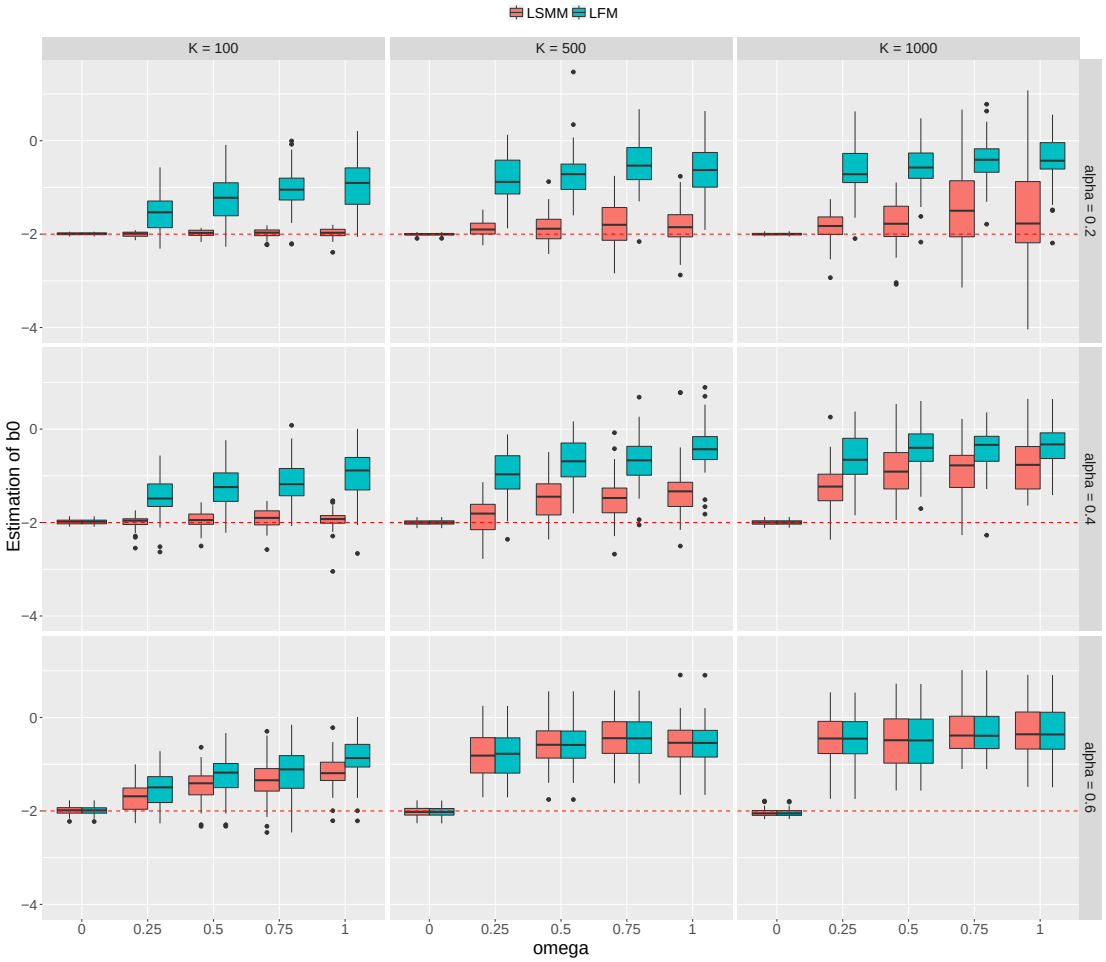

Figure S28: Performance in estimation of parameter b_0.

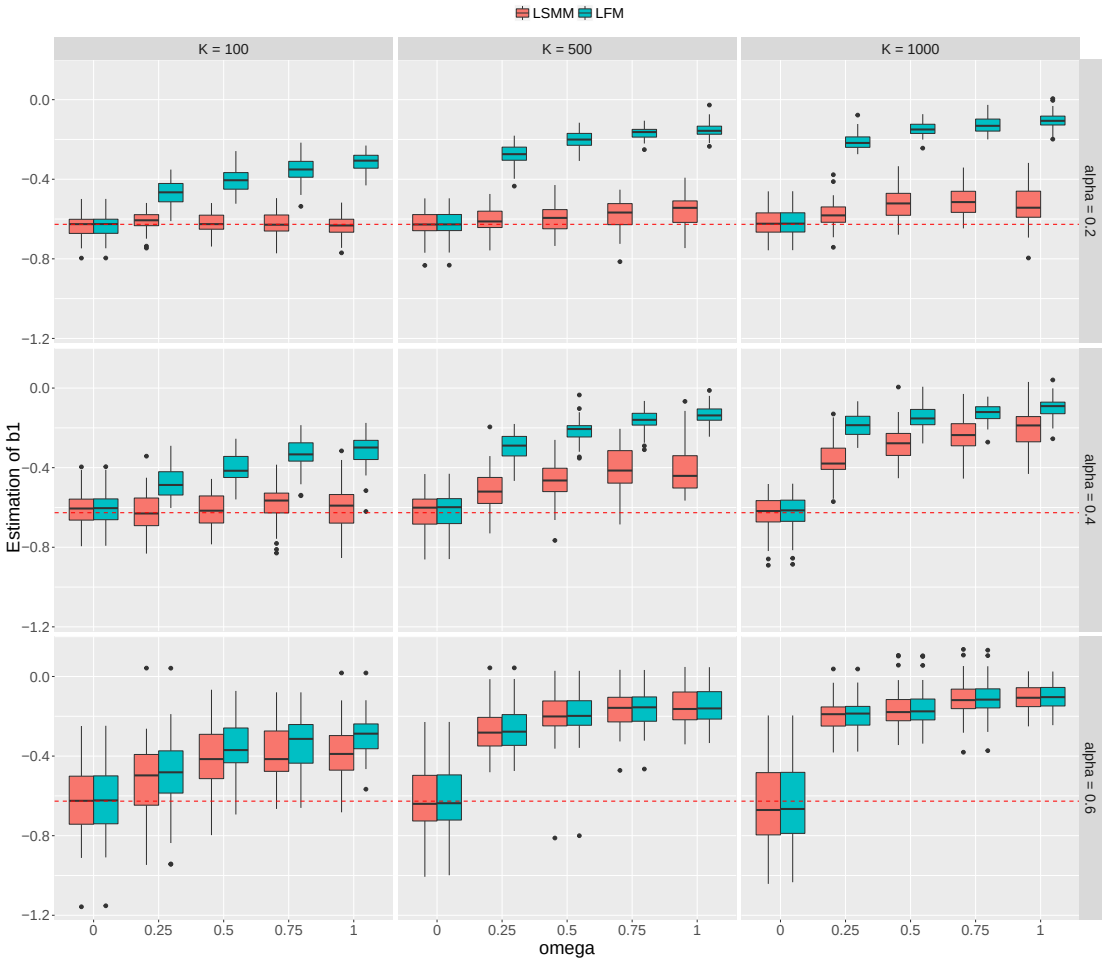

Figure S29: Performance in estimation of parameter b_1.

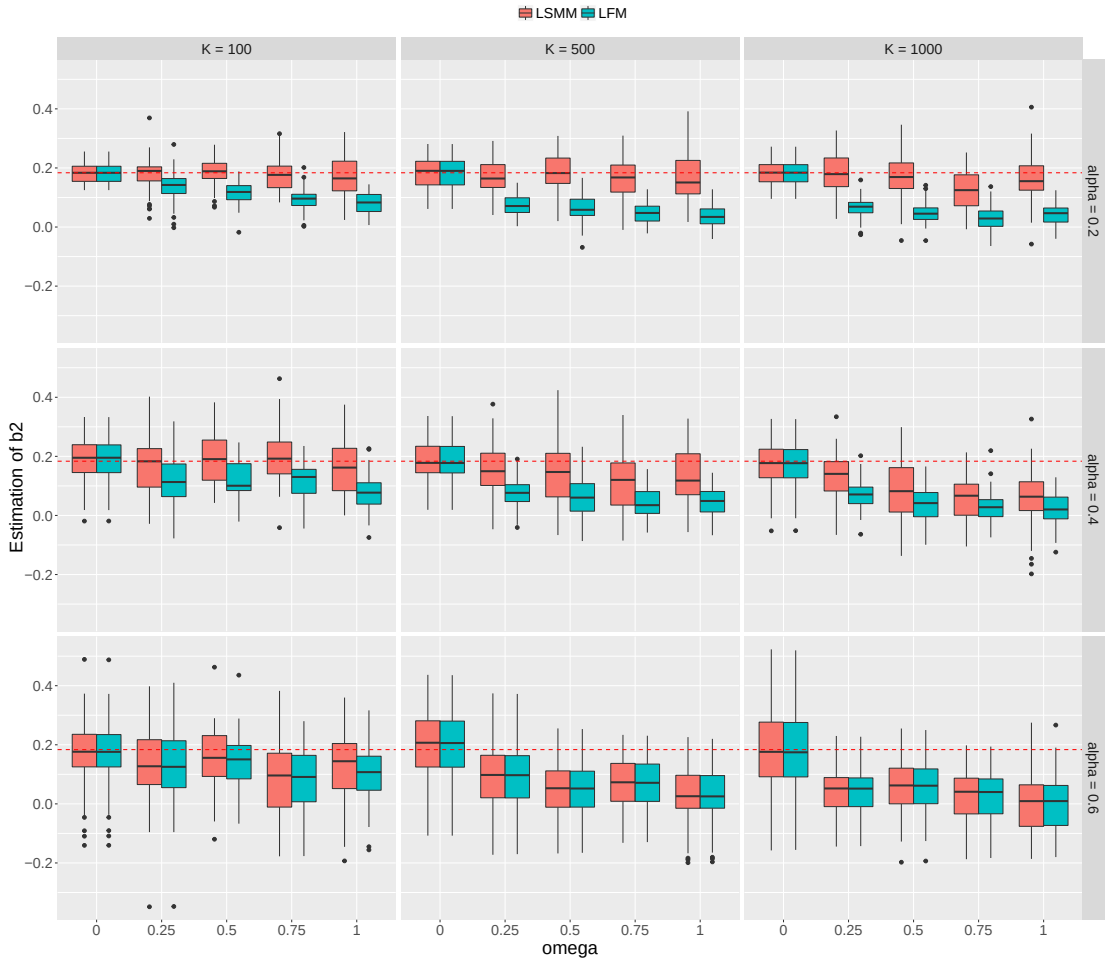

Figure S30: Performance in estimation of parameter b_2.

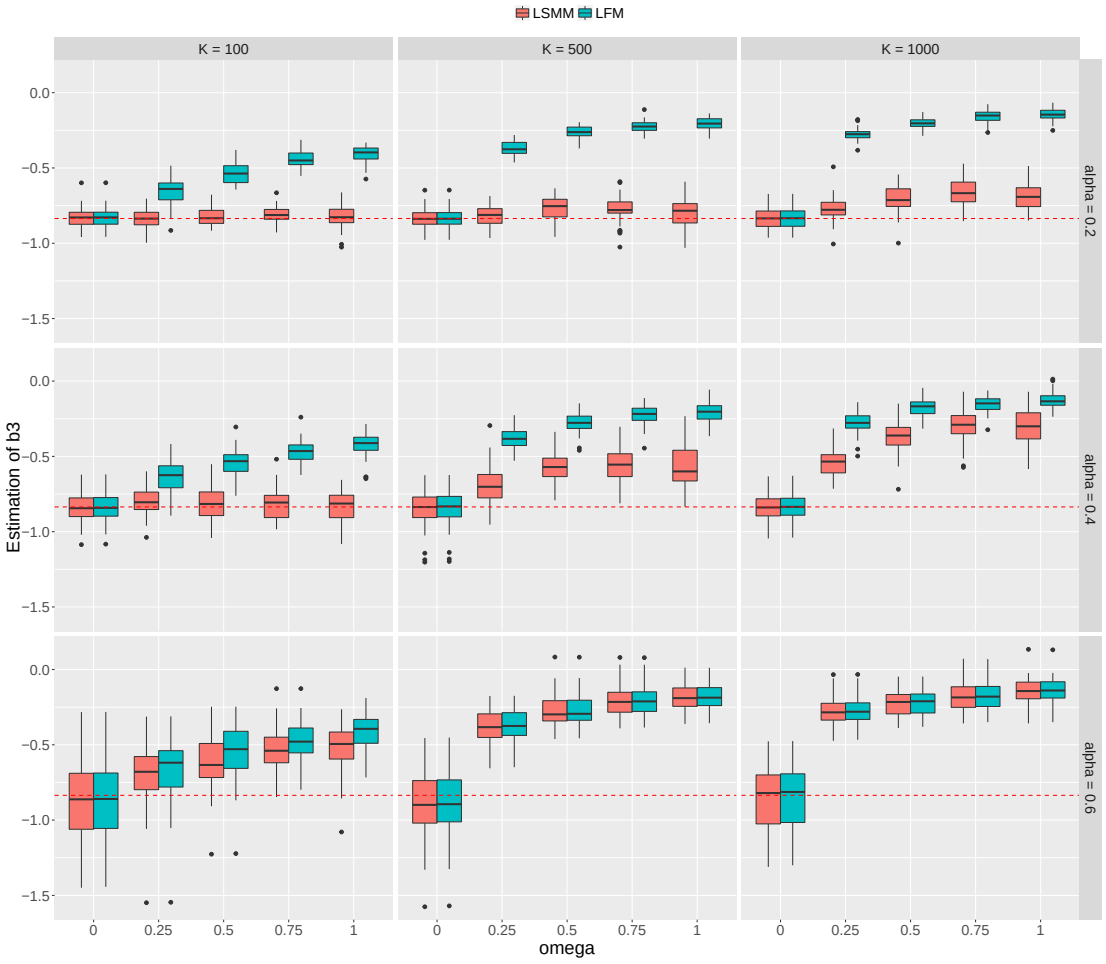

Figure S31: Performance in estimation of parameter b_3.

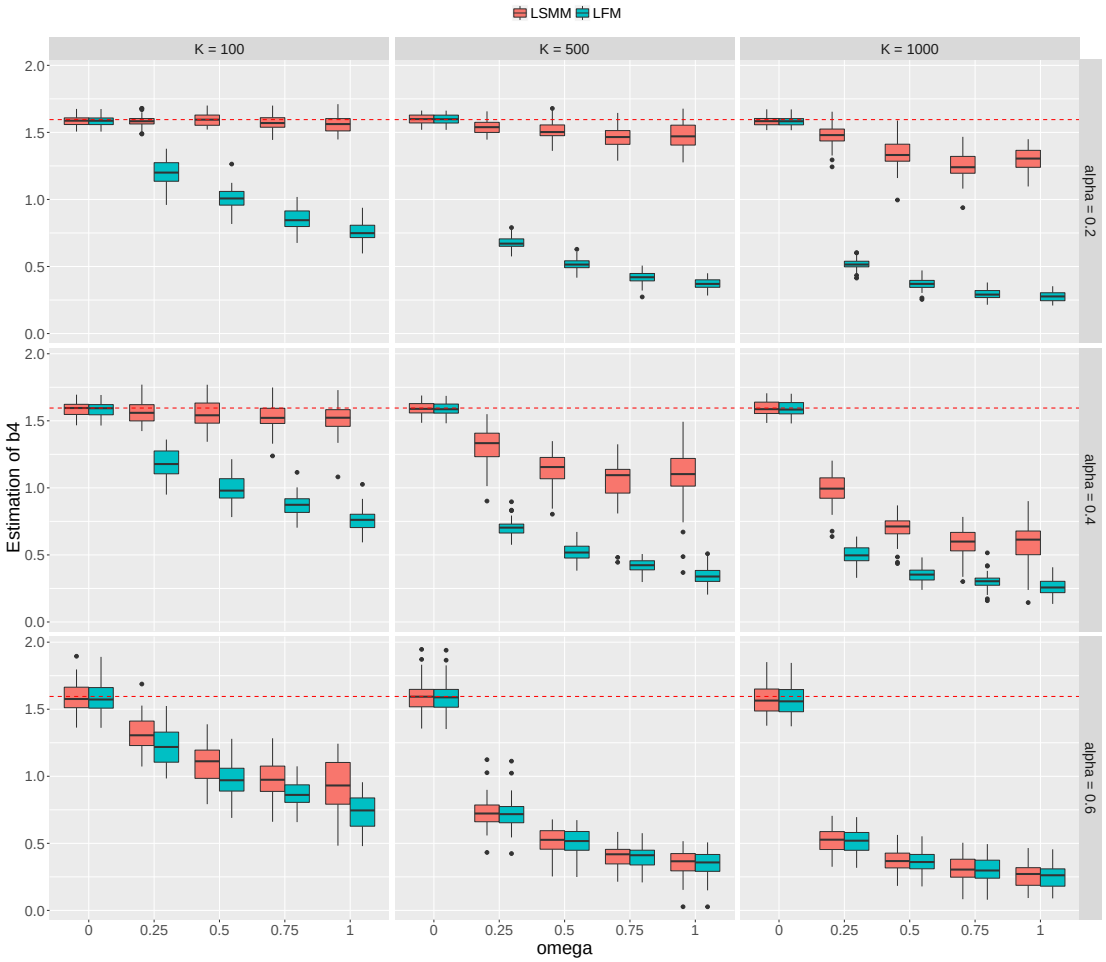

Figure S32: Performance in estimation of parameter b_4.

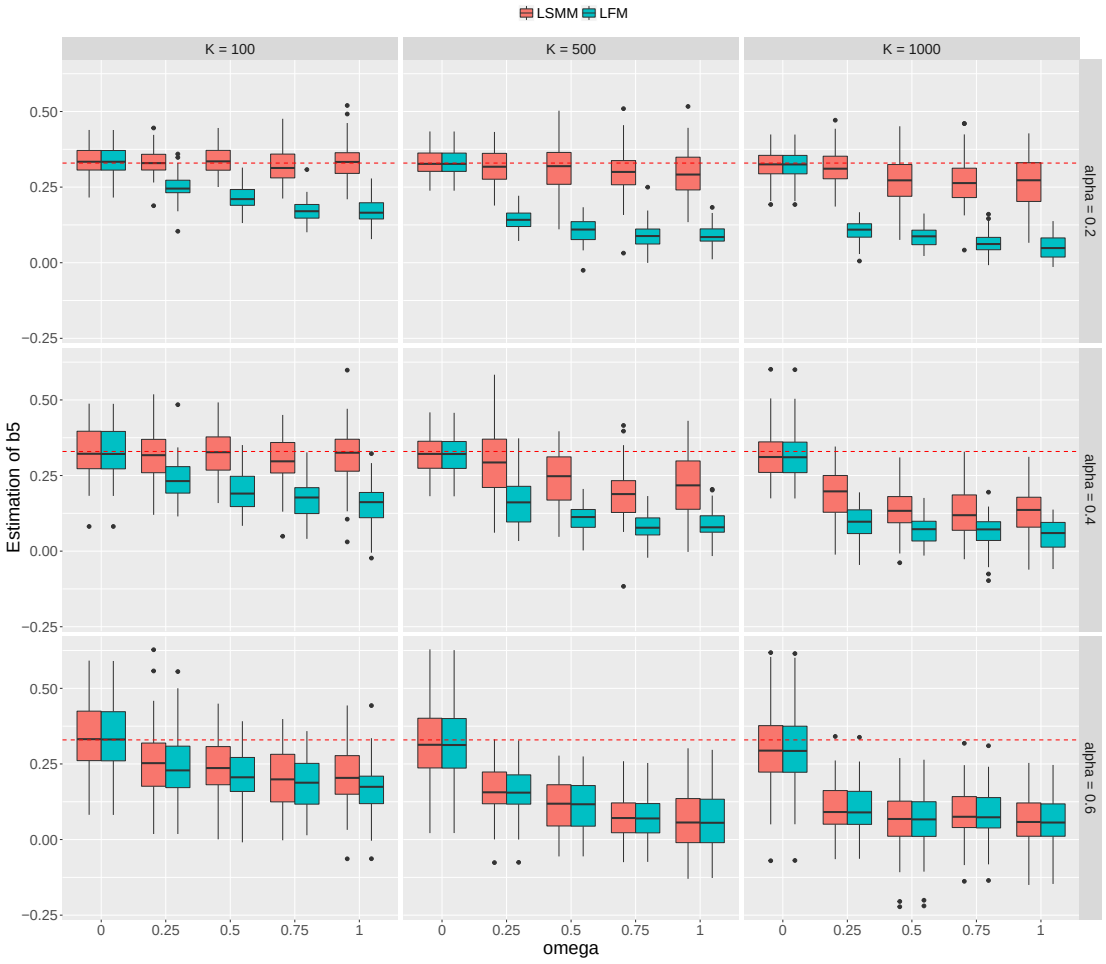

Figure S33: Performance in estimation of parameter b_5.

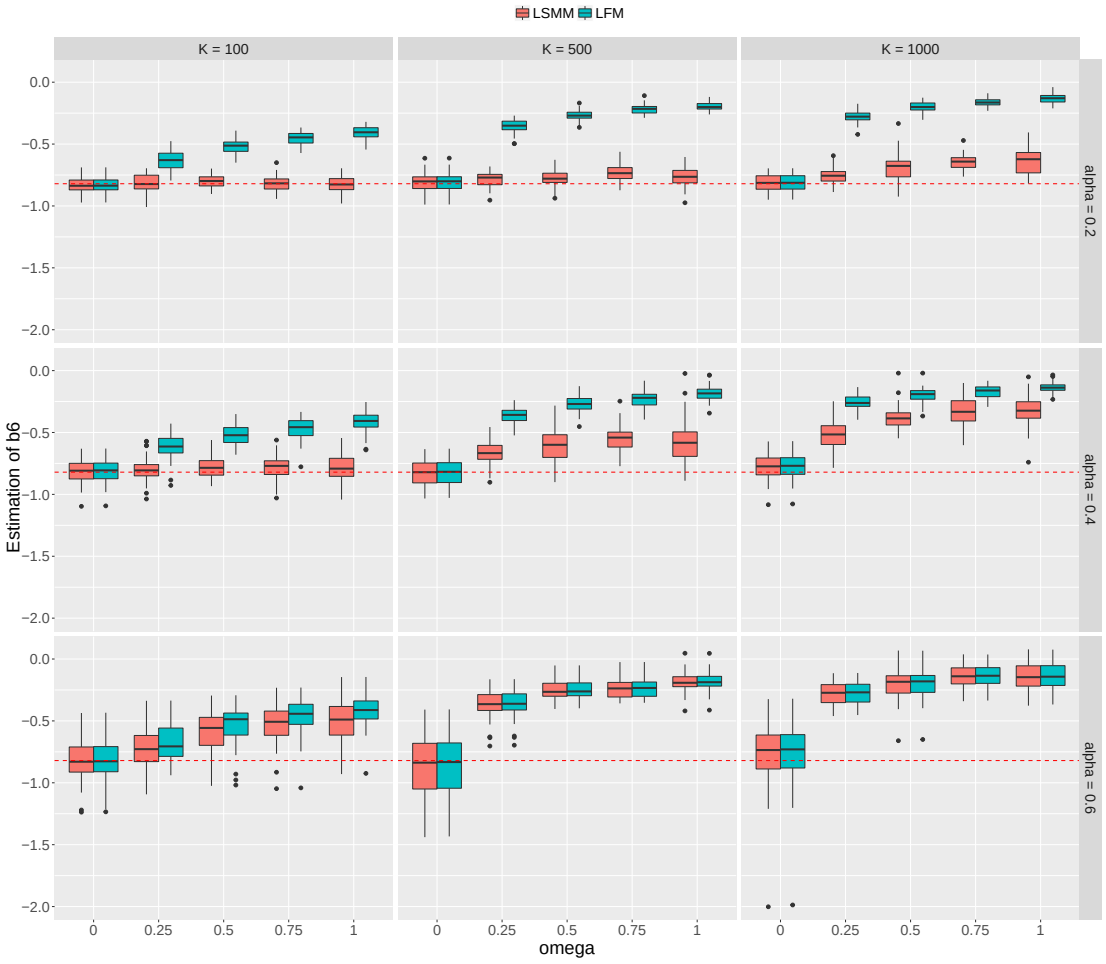

Figure S34: Performance in estimation of parameter b_6.

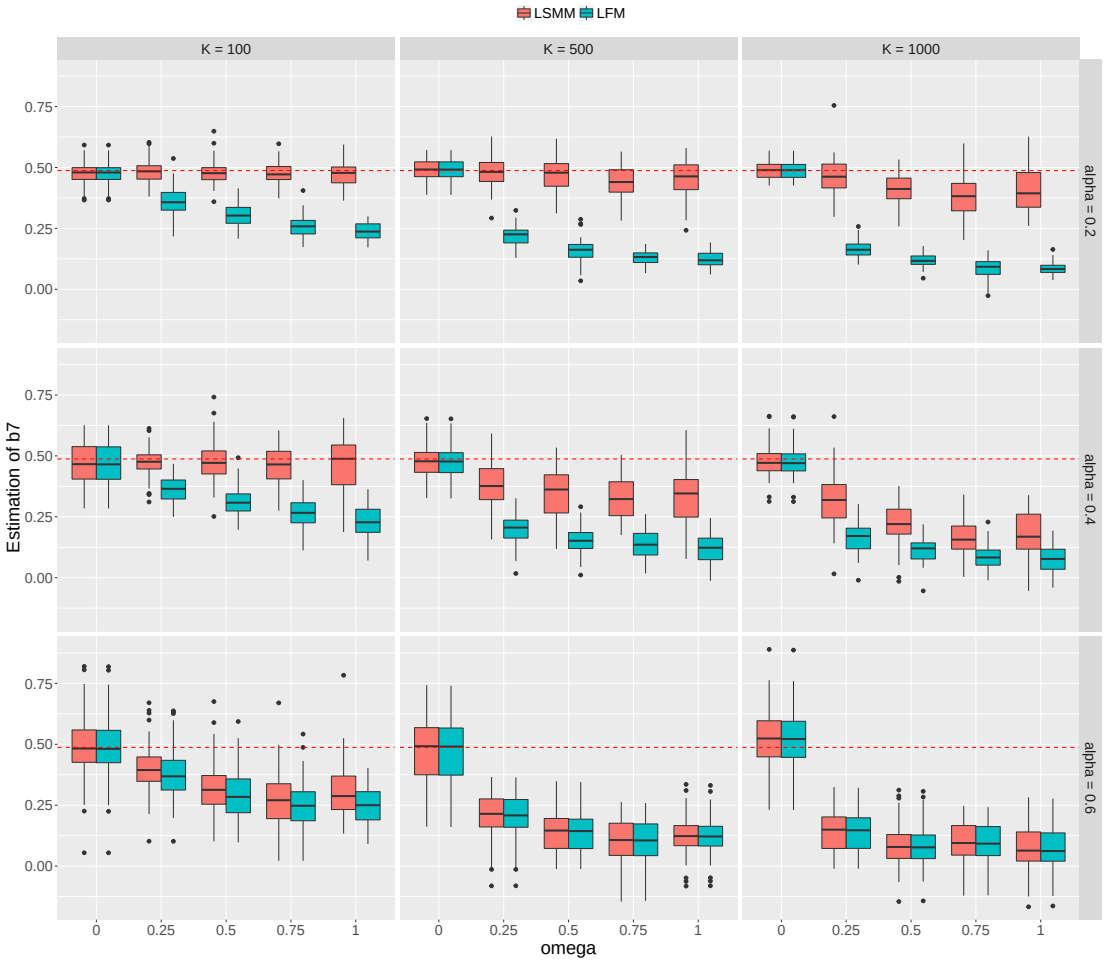

Figure S35: Performance in estimation of parameter b_7.

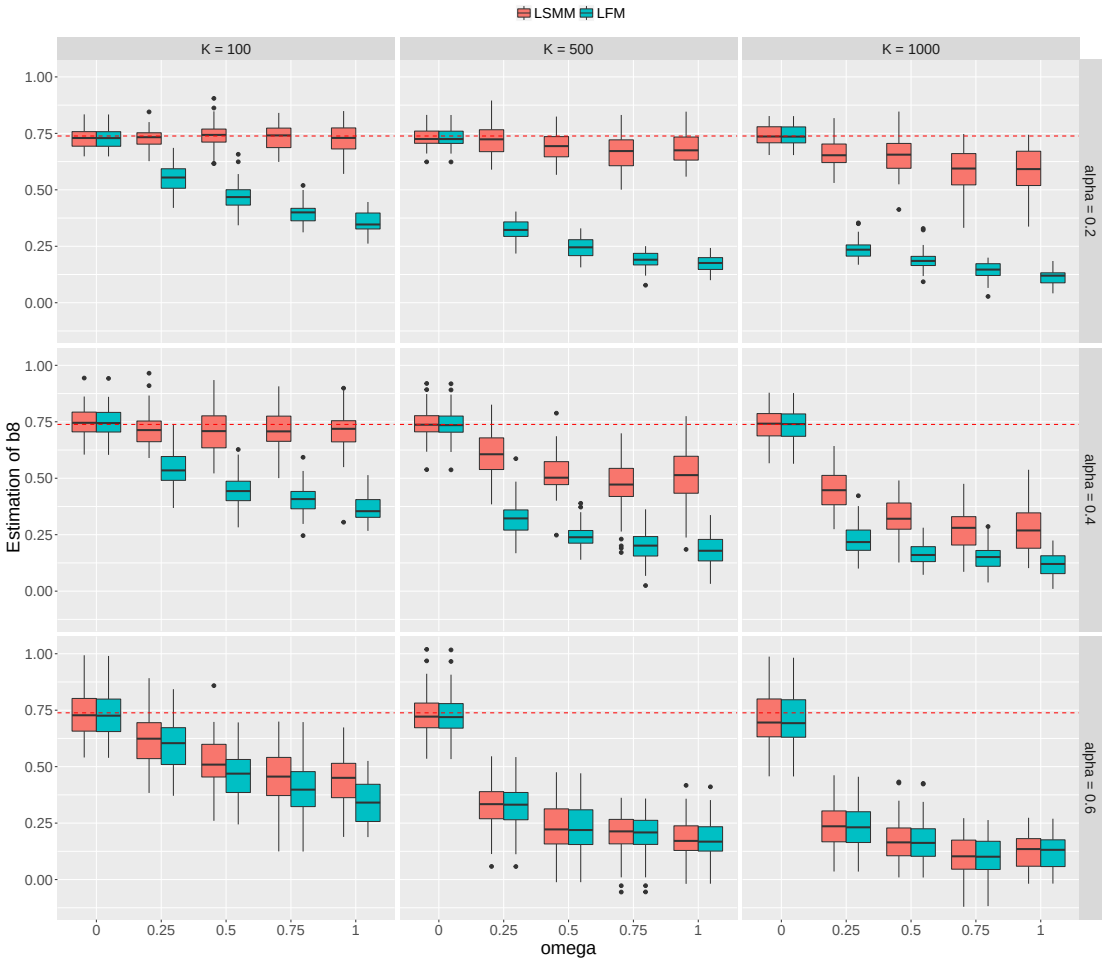

Figure S36: Performance in estimation of parameter b_8.

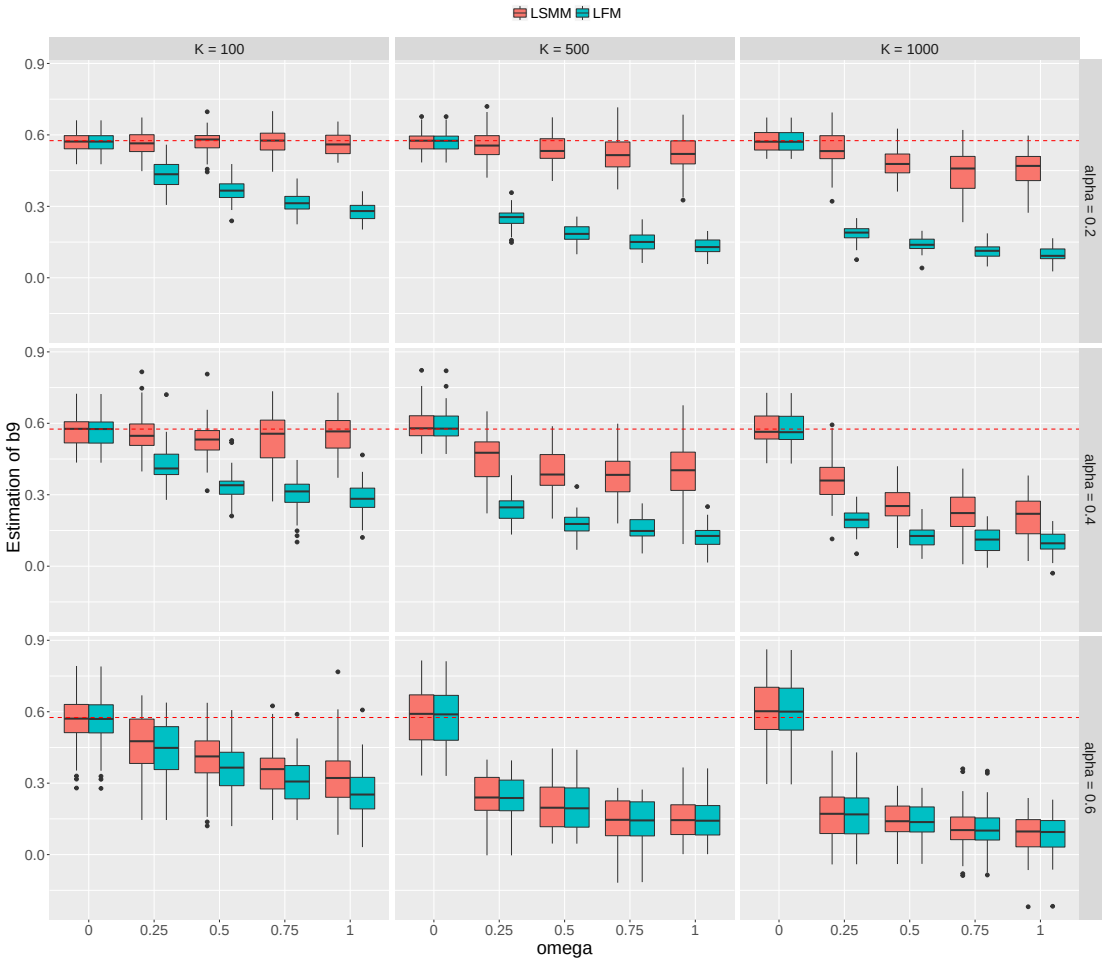

Figure S37: Performance in estimation of parameter b_9.

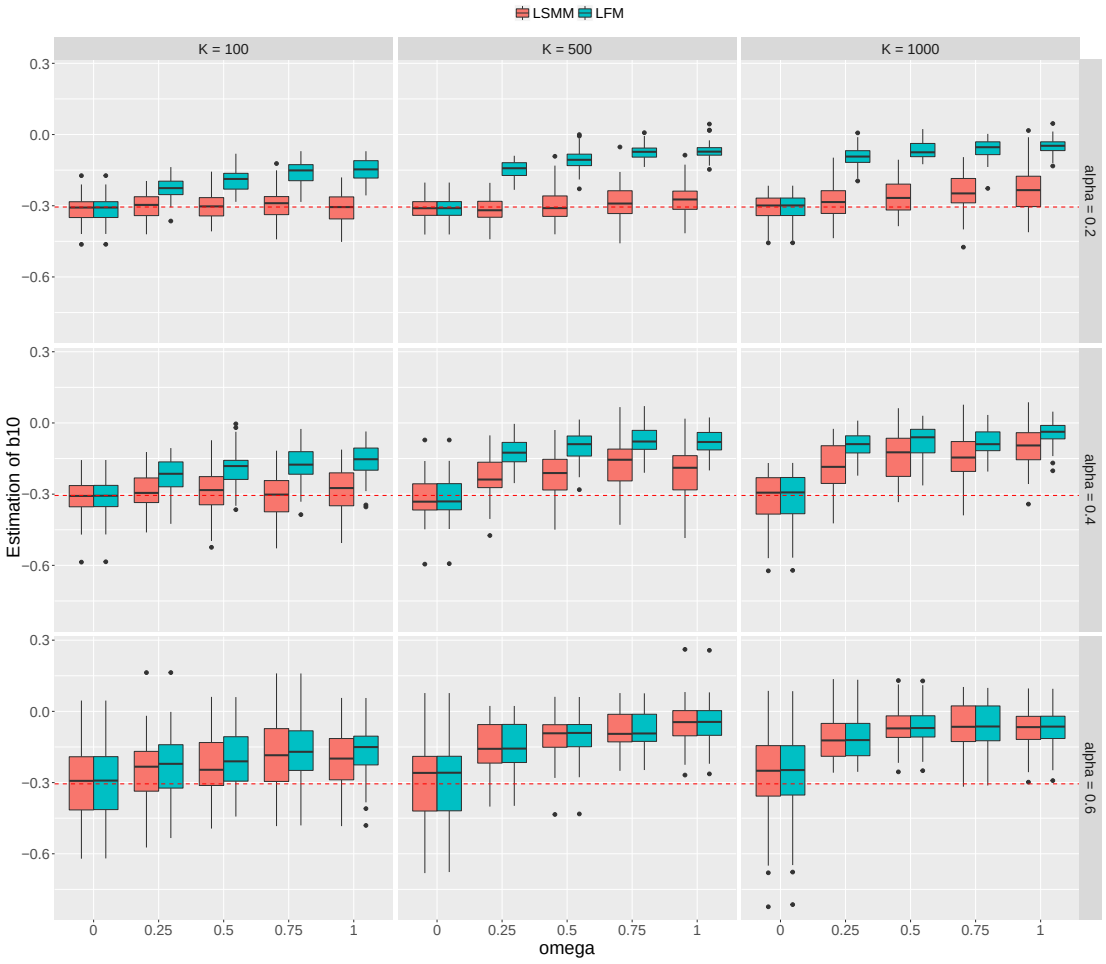

Figure S38: Performance in estimation of parameter b_{10}.

3.8.3 Estimation of ω

We evaluate the performance of LSMM in estimation of parameter ω which measures the proportion of relevant annotations. We varied ω at $\{0, 0.25, 0.5, 0.75, 1\}$. Figure S39 shows the results with $\alpha = 0.2$, 0.4 and 0.6.

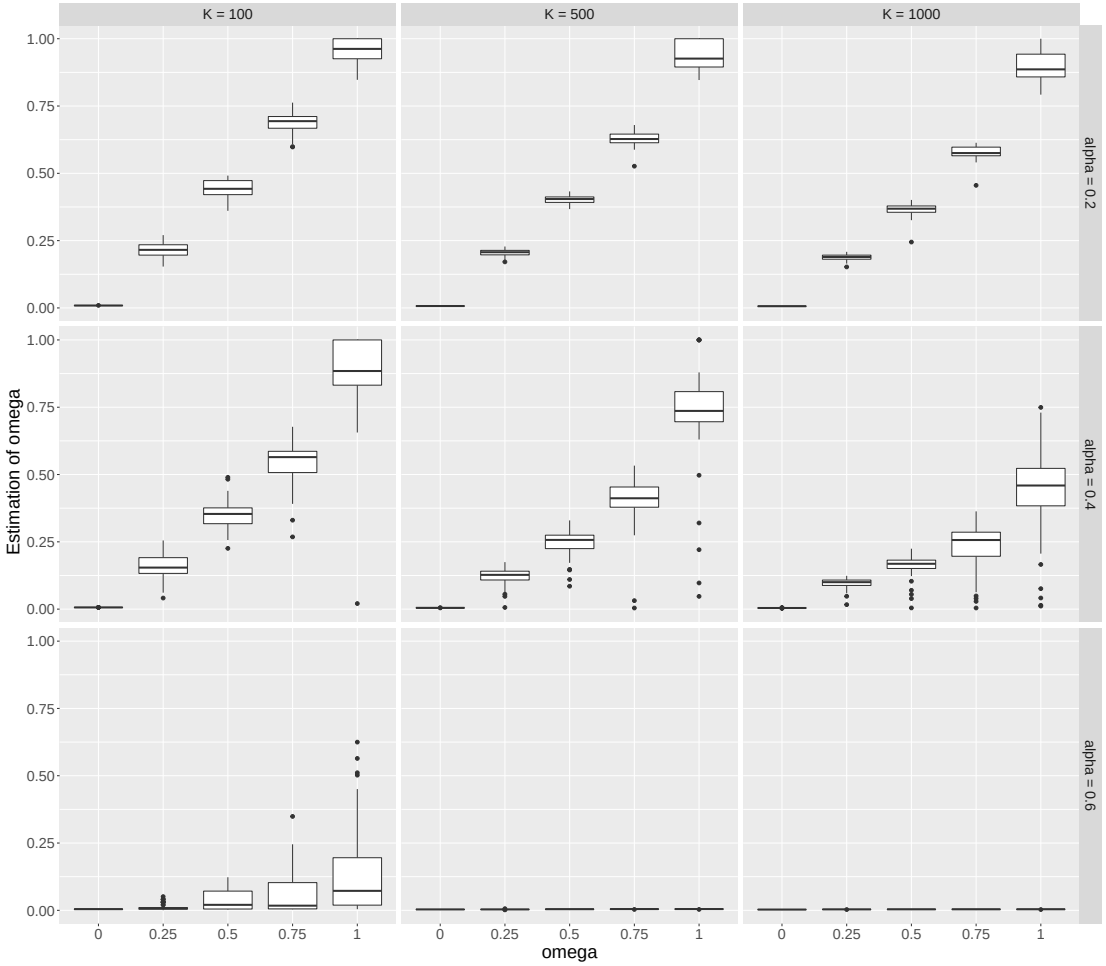

Figure S39: Performance in estimation of parameter ω.

3.9 Estimation of α using individual-level data

To provide a reference for the relationship between heritability and α, here we conducted simulations when the p-values for SNPs are obtained from individual-level data instead of directly simulating from the generative model (1). The simulation data was generated as follows. To simulate the genotype matrix $\mathbf{X}$ for N individuals with M independent SNPs, we first draw the minor allele frequencies (MAFs) of these SNPs from $U[0,1]$. Based on the MAFs, the entries in the genotype matrix $\mathbf{X}$, which were encoded by $\{0,1,2\}$, were generated according to the Hardy-Weinberg principle. Given $\boldsymbol{\gamma}$, which was simulated as what we described in the paper, the corresponding nonzero entries of effect sizes $\boldsymbol{\beta}_{SNP}$ were simulated from $N(0,1)$. The noise level σ_e^2 was specified to control heritability $h^2 = \frac{var(\mathbf{X}\boldsymbol{\beta}_{SNP})}{var(\mathbf{X}\boldsymbol{\beta}_{SNP})+\sigma_e^2}$ at given levels. The phenotype data $\mathbf{y}$ was generated based on $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{SNP} + \mathbf{e}$, where $e_i \sim N(0, \sigma_e^2)$ for $i = 1, ..., N$. Then we conducted univariate linear regression to obtain the summary statistics (p-value) for each SNP.

In the simulation, we set $M = 20,000$, $L = 10$, $K = 100$ and $\omega = 0.1$. We varied heritability $h^2 \in \{0.2, 0.4, 0.6, 0.8\}$ and the sample size $N \in$\{10,000, 5,000\}. Figure S40 shows the estimation of α using LSMM, indicating that the value of α is determined by both heritability and sample size. When $N = 10,000$, heritability $h^2 = 0.6$ and $h^2 = 0.2$ are approximately corresponding to $\alpha = 0.4$ and $\alpha = 0.6$, respectively. When the sample size reduces to $N = 5,000$, the corresponding estimation of α becomes larger. To conclude, given fixed sample sizes and nonzero proportion, smaller alpha corresponds to larger heritability. Hence, we used alpha to indicate the strength of GWAS in our paper.

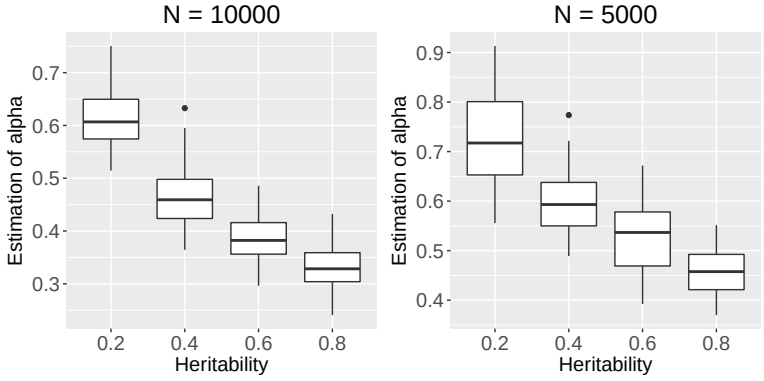

Figure S40: The estimation of parameter α using individual-level data.

3.10 Simulations if p-values are not from beta distribution

In the model setting of the LSMM, we assume that p-values are from the mixture of uniform and Beta distributions. To check the robustness of our method, we conducted simulations as follows. We first generated z-scores and then converted them to p-values. Here z-values from the null group follow the standard normal distribution and z-values from the non-null group follow the alternative distributions in Table S1. In these simulations, the p-values in non-null group converted from z-scores will not from Beta distribution. Instead of using generative model (2), we conducted simulations based on probit model:

$$y_j = \mathbf{Z}_j\mathbf{b} + \mathbf{A}_j\boldsymbol{\beta} + e_j, \tag{S2}$$

where $e_j \sim N\left(0, \sigma_e^2\right)$. And we set $\gamma_j = 1$ if $y_j > 0$, $\gamma_j = 0$ if $y_j \leq 0$. The first entry of the coefficients of fixed effects $\mathbf{b}$, i.e. the intercept term, was fixed at -1 and other entries were generated from $N\left(0, 1\right)$ and fixed during multiple replications. We set $\alpha = 0.2$, $\omega = 0.2$ and varied the signal-noise ratio $r = \{4:1, 1:1, 1:4\}$. The empirical FDRs are shown in Figures S41-S43.

Scenario	Distribution
spiky	$0.4N\left(0, 0.25^2\right) + 0.2N\left(0, 0.5^2\right) + 0.2N\left(0, 1^2\right) + 0.2N\left(0, 2^2\right)$
near normal	$\frac{2}{3}N\left(0, 1^2\right) + \frac{1}{3}N\left(0, 2^2\right)$
skew	$\frac{1}{4}N\left(-2, 2^2\right) + \frac{1}{4}N\left(-1, 1.5^2\right) + \frac{1}{3}N\left(0, 1^2\right) + \frac{1}{6}N\left(1, 1^2\right)$
big-normal	$N\left(0, 4^2\right)$

Table S1: Alternative distributions for z-scores.

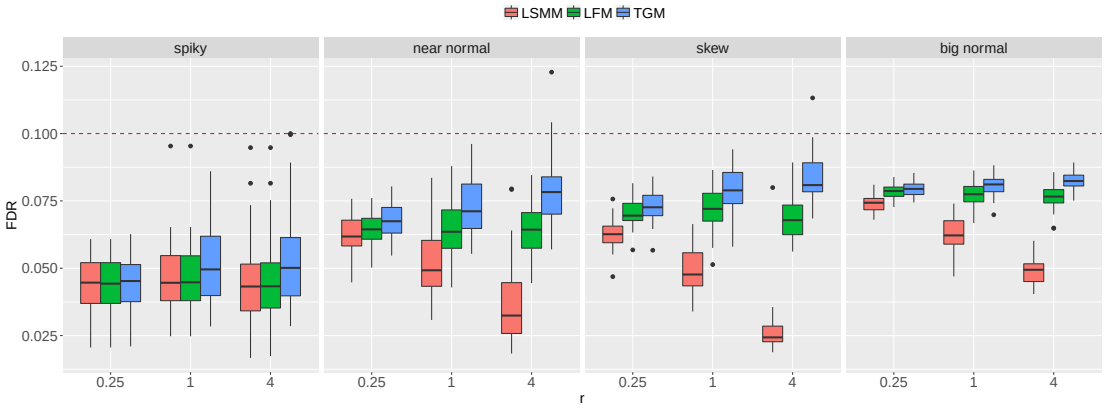

Figure S41: FDR of LSMM, LFM and TGM with $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

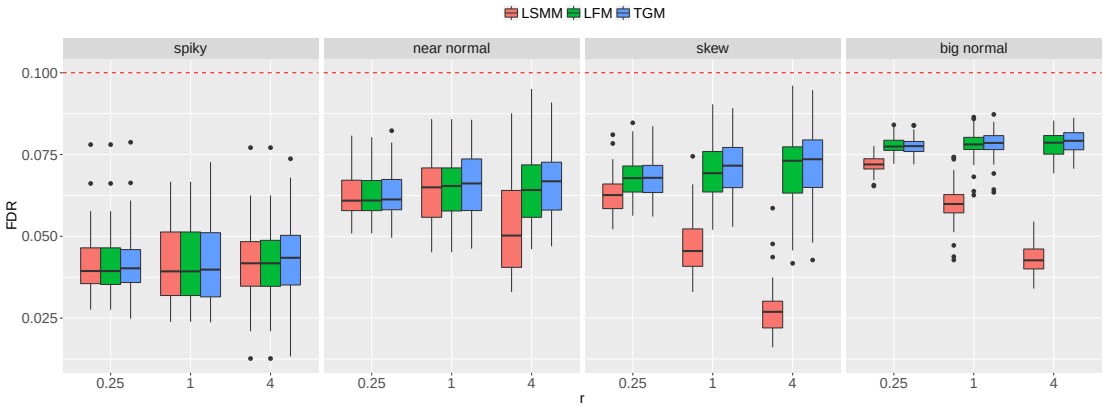

Figure S42: FDR of LSMM, LFM and TGM with $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

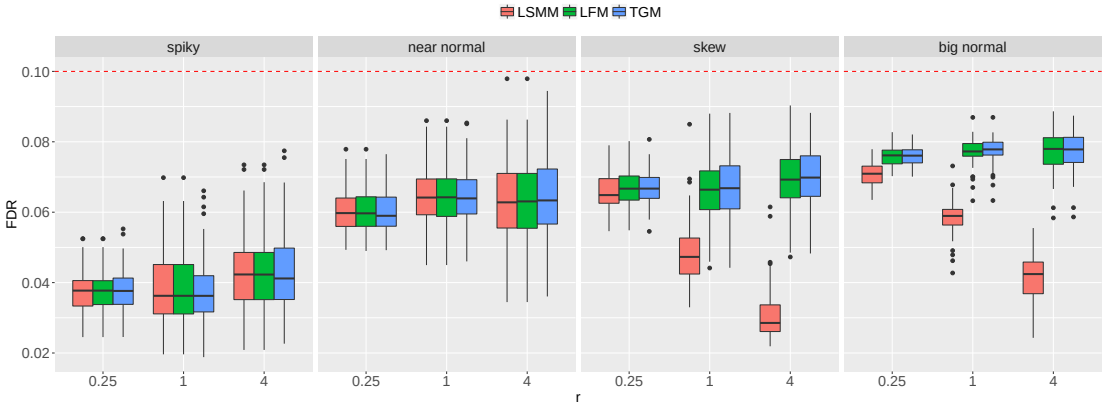

Figure S43: FDR of LSMM, LFM and TGM with $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

3.11 Simulation study for evaluating the LD effects on LSMM

To study the influence of LD effects on our LSMM, we used the observed genotype data (1,500 individuals from the 1958 British Birth Cohort (58C)) from WTCCC (The Wellcome Trust Case Control Consortium, 2007). For simplicity, we only consider 23874 SNPs in chromosome 1 after quality control. We simulated a risk SNP every 1000 SNPs. So we had 24 risk SNPs. We assumed the 24 risk SNPs can explain 5% phenotypic variance. We used GCTA to simulation phenotypes and used PLINK to get p-values for SNPs. Then we applied LSMM and detect risk SNPs.

As the presence of LD effects, SNPs in a local genomic region would be correlated and detection of risk SNPs would be difficult. We are just expected to identify the region which contains the risk SNPs. Here we used different distance threshold to define the region around true risk SNPs. The identified risk SNPs which in the region of true risk SNPs were considered as true positive.

We considered four cases. The first case, no effects, means we only used the p-values and didn't use fixed effects and random effects. In the second case, fixed effects, we only add 10 fixed effects. In the fixed effects, SNPs within 1Mb of true risk SNPs are annotated with a probability of 0.6. In the third case, fixed + random effects, we further add 100 random effects in which SNPs are annotated randomly. In the fourth case, fixed + relevant random effects, we assume 20% of random effects are relevant to the phenotype and SNPs within 1Mb of true risk SNPs are annotated with a probability of 0.6 in the relevant random effects. The results of observed FDR were shown in Figure S44 based on 50 simulations. In the first case, when we used no effects, the observed FDR was quite stable at 0.1. When we added fixed effects and random effects, the observed FDR was just inflated a little with the smallest distance threshold and became conservative as the distance threshold increased. As a result, we believe that LSMM can provide a satisfactory FDR control in detecting a local genomic region of risk SNPs.

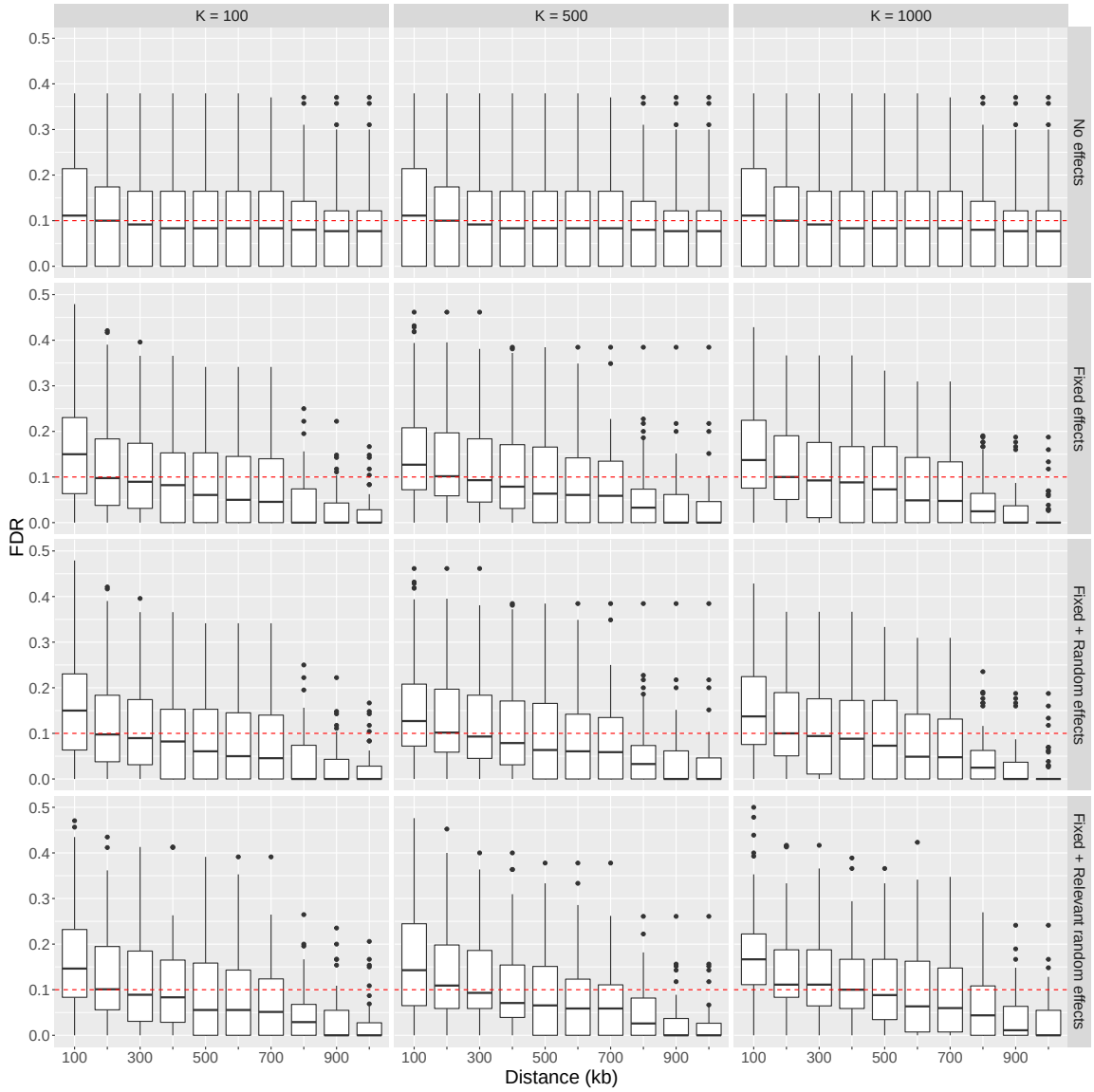

Figure S44: FDR of LSMM for identification of risk SNPs with different distance thresholds. The red line indicates the threshold of global FDR $\tau = 0.1$.

3.12 Performance of LSMM when the proportion of risk SNPs π_1 was extremely small

Here we used the TGM to generate data such that we can evaluate whether the estimates converge to their true values. In the simulation, we set the numbers of SNPs $M = 100,000$ and varied the true value of $\pi_1 \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2\}$. We also used Higher Criticism to estimate the proportion of non-null effects as a comparison. The software for Higher Criticism was downloaded from http://www.stat.cmu.edu/ jiashun/Research/software/NullandProp/.

The results of the estimation $\hat{\pi}_1$ using LSMM and Higher Criticism are shown in the upper panel of Figure S45. The true values are indicated by dotted lines with different colors. When the true proportion of risk SNPs is extremely small (e.g., $\pi_1 \leq 0.001$ for $\alpha = 0.4$) and the signal of GWAS data is weak (e.g., $\pi_1 \leq 0.01$ for $\alpha = 0.6$), the estimation using LSMM is not very accurate. However, LSMM can still provide a valid FDR control (See lower panel of Figure S45). The performance of Higher Criticism is quite opposite. Although it can provide stable estimation when the true proportion of risk SNPs is small ($\pi_1 \leq 0.01$), its performance for larger π_1 is not as well as LSMM when π_1 is relatively large, e.g., $\pi_1 \geq 0.05$. In the context of GWAS, the proportion of risk variants is not very small due to the polygenic effect. Therefore, we believe LSMM will work well in practice.

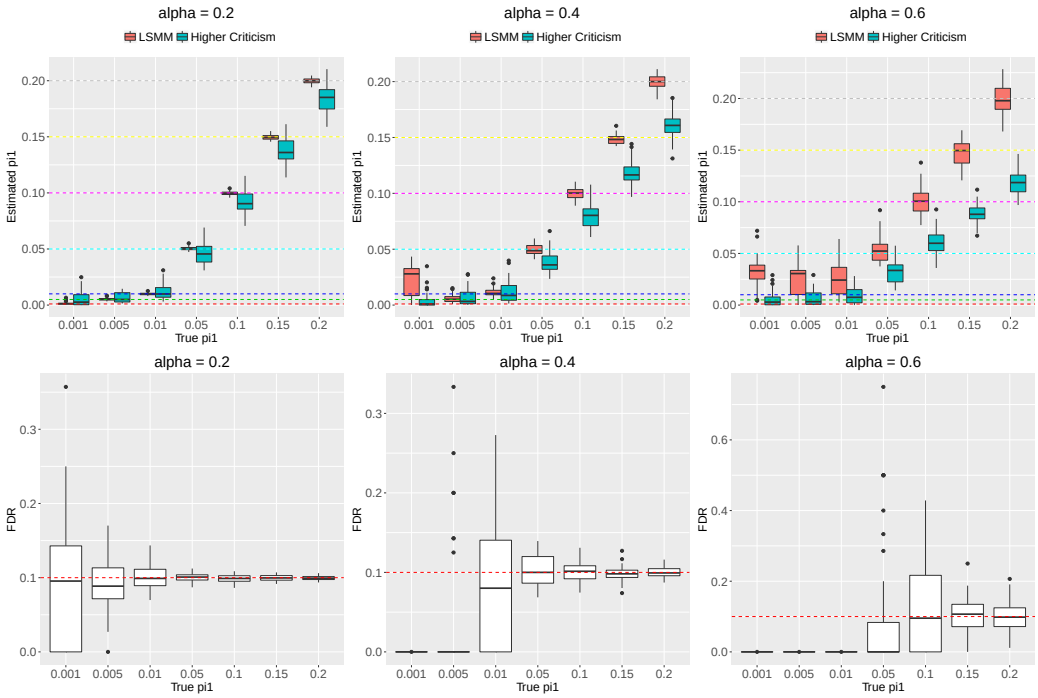

Figure S45: Upper panel: parameter estimation ($\hat{\pi}_1$ v.s. true π_1) using LSMM and Higher Criticism. Lower panel: FDR for identification of risk SNPs using LSMM. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

3.13 Simulations based on probit model

To test the robustness of LSMM, instead of using generative model (2), we conducted simulations based on probit model:

$$y_j = \mathbf{Z}_j\mathbf{b} + \mathbf{A}_j\boldsymbol{\beta} + e_j, \tag{S3}$$

where $e_j \sim N\left(0, \sigma_e^2\right)$. And we set $\gamma_j = 1$ if $y_j > 0$, $\gamma_j = 0$ if $y_j \leq 0$. The first entry of the coefficients of fixed effects $\mathbf{b}$, i.e. the intercept term, was fixed at -1 and other entries were generated from $N\left(0, 1\right)$ and fixed during multiple replications. We set $\alpha = 0.2$ and varied the signal-noise ratio $r = \{4:1, 1:1, 1:4\}$. The performance in identification of risk SNPs is provided in Figures S36-S38. The performance of LSMM in the detection of relevant functional annotations is provided in Figures S46-S51.

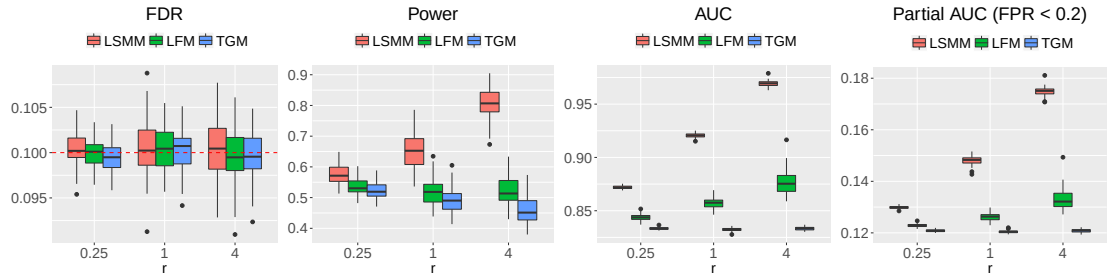

Figure S46: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs based on probit model with $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

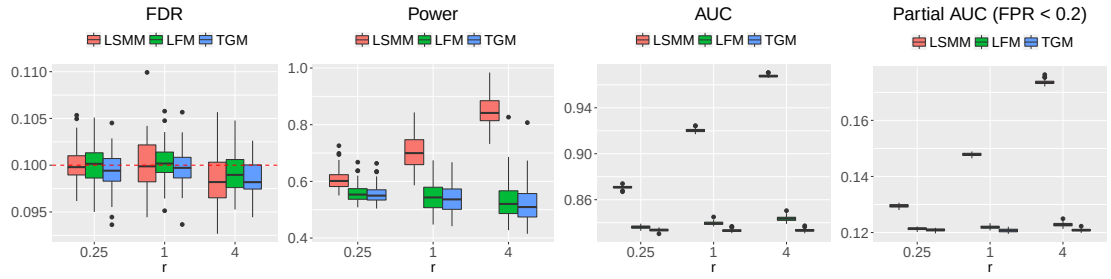

Figure S47: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs based on probit model with $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

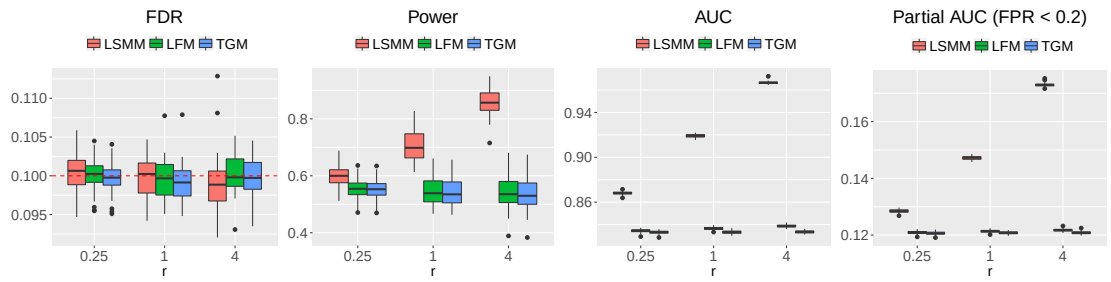

Figure S48: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs based on probit model with $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

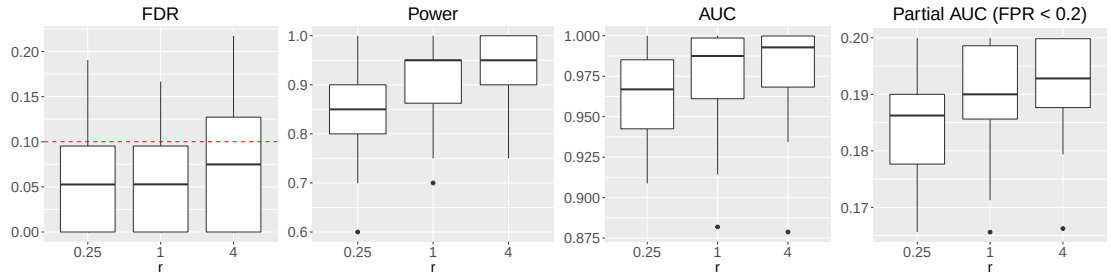

Figure S49: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for detection of relevant annotations based on probit model with $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

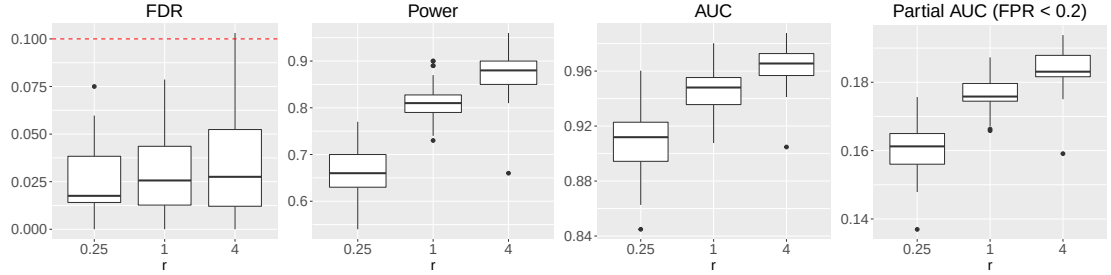

Figure S50: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for detection of relevant annotations based on probit model with $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

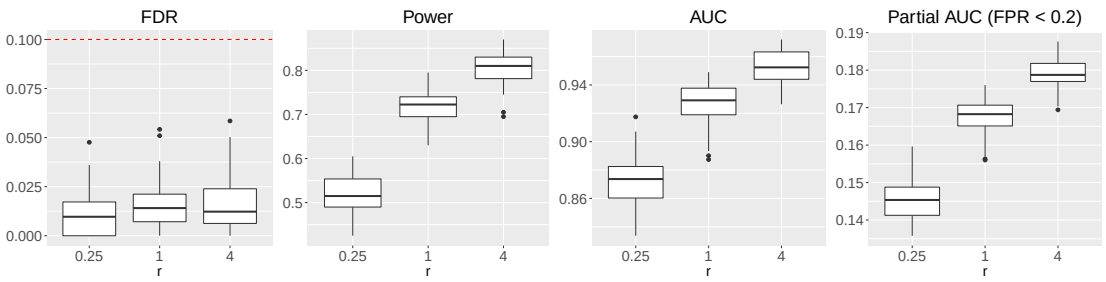

Figure S51: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for detection of relevant annotations based on probit model with $K = 1000$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.14 Performance of LSMM when the random effects are correlated

We generated $\boldsymbol{\beta}$ from a multivariate normal distribution $MVN(\mathbf{0}, \Sigma)$, where Σ is an autocorrelation matrix with ρ varied at $\{0, 0.2, 0.4, 0.6, 0.8\}$. Here we set $\omega = 0.2$. The results are shown in Figure S52 and Figure S53.

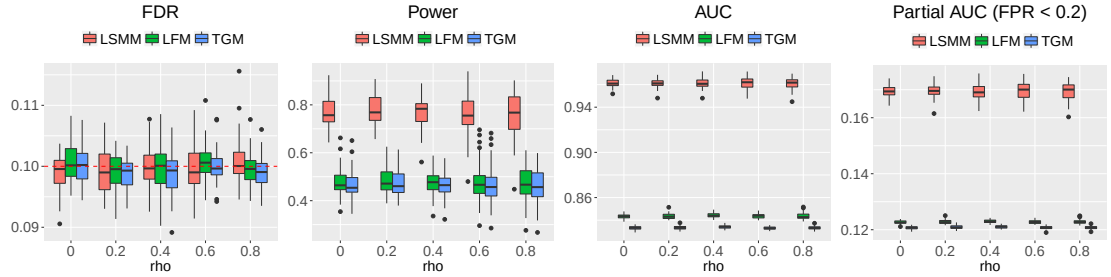

Figure S52: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs when random effects are correlated with $\alpha = 0.2$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

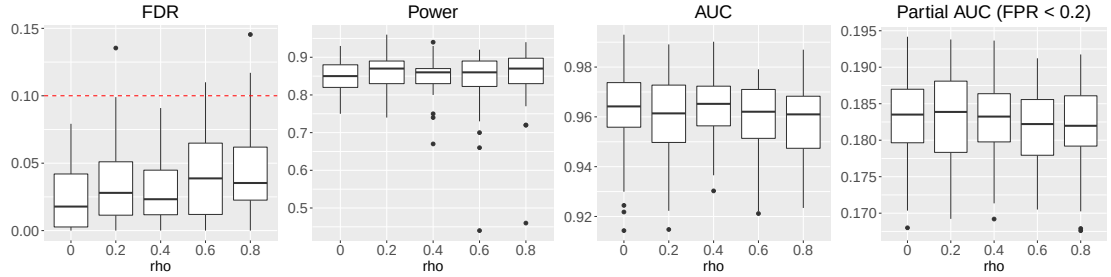

Figure S53: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations when random effects are correlated with $\alpha = 0.2$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.15 Performance of LSMM when the random effects don't share the same variance

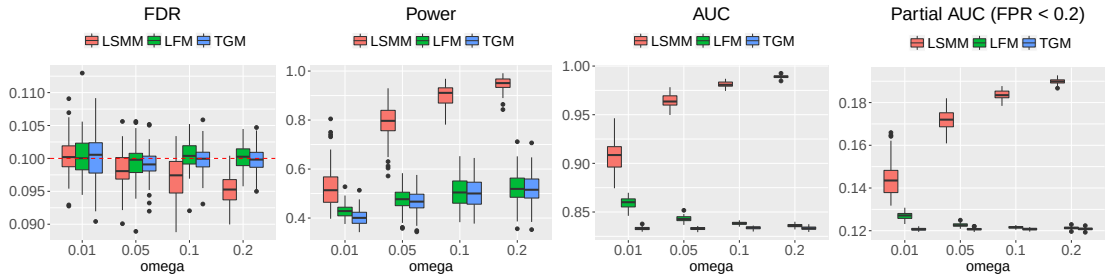

Figure S54: FDR, power, AUC and partial AUC of LSMM, LFM and TGM for identification of risk SNPs when the variance of random effects are from $U[1, 10]$ with $\alpha = 0.2$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

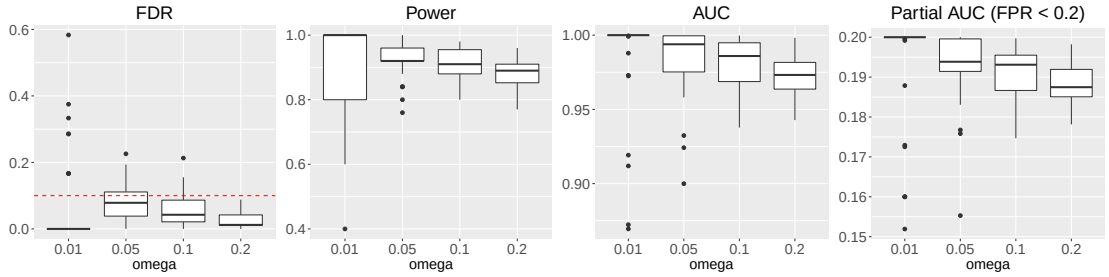

Figure S55: FDR, power, AUC and partial AUC of LSMM for detection of relevant annotations when the variance of random effects are from $U[1, 10]$ with $\alpha = 0.2$ and $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.16 Comparison between LSMM and GPA (Chung et al., 2014)

To check the influence of correlated functional annotations, we simulated a case that the first 10 functional annotations were correlated and all the others were independent. We set $\alpha = 0.2$ and varied the correlation among annotations *corr* at $\{0, 0.2, 0.4, 0.6, 0.8\}$. To simulate the design matrices for correlated functional annotations, we first simulated M samples from a multivariate normal distribution with the correlation matrix among annotations and then made a cutoff so that 10% of the entries would be 1 and the others be 0.

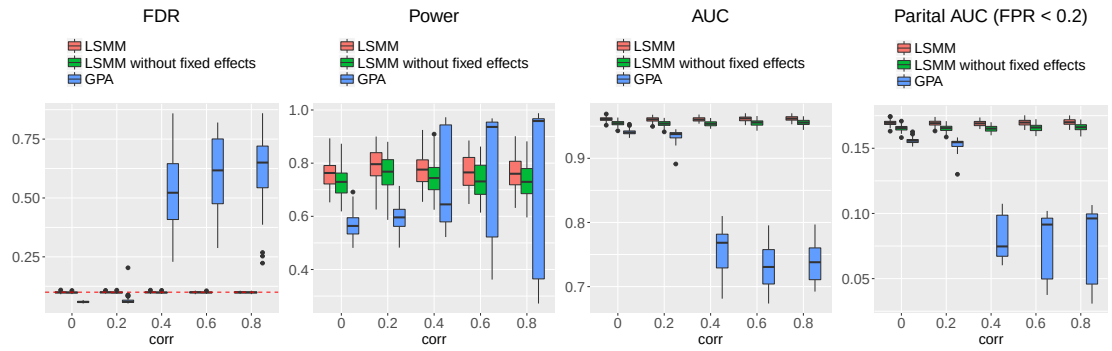

Figure S56: FDR, power, AUC and partial AUC of LSMM, LSMM without fixed effects and GPA for identification of risk SNPs with $K = 500$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

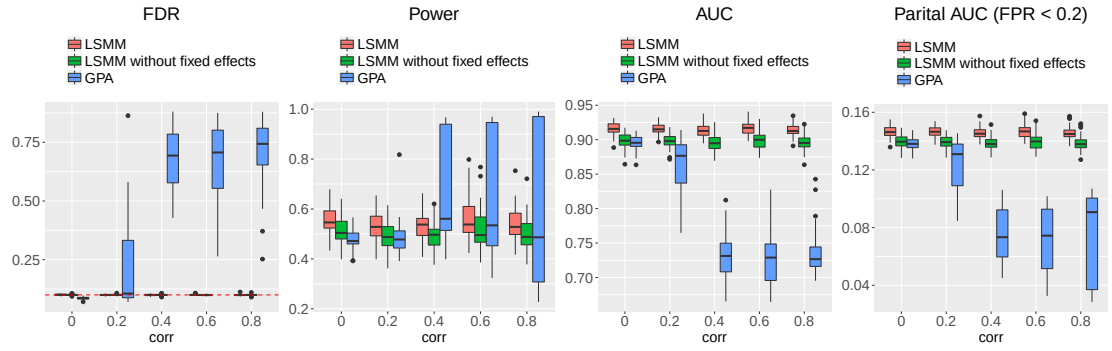

Figure S57: FDR, power, AUC and partial AUC of LSMM, LSMM without fixed effects and GPA for identification of risk SNPs with $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

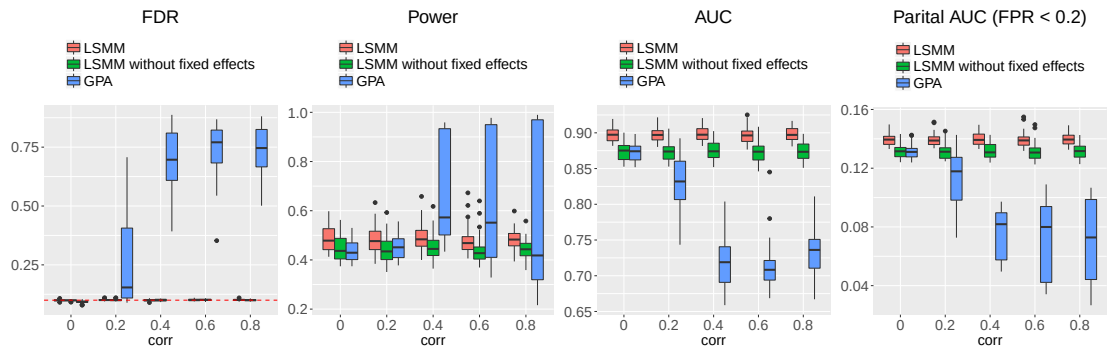

Figure S58: FDR, power, AUC and partial AUC of LSMM, LSMM without fixed effects and GPA for identification of risk SNPsn with $K = 50$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

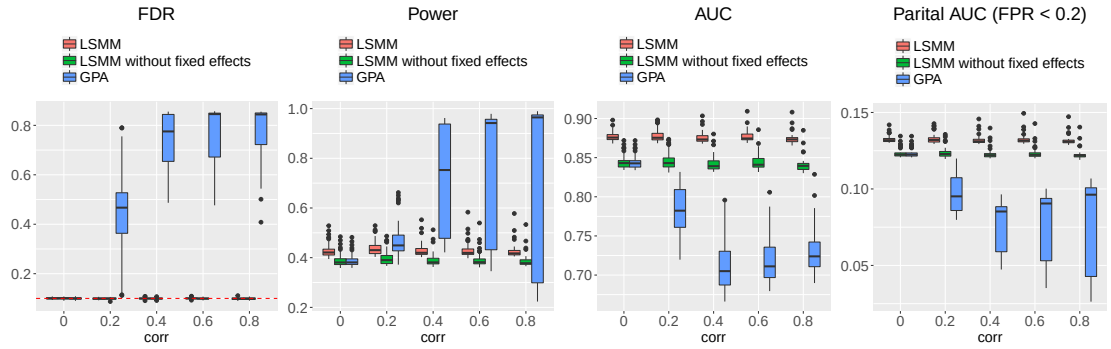

Figure S59: FDR, power, AUC and partial AUC of LSMM, LSMM without fixed effects and GPA for identification of risk SNPs with $K = 10$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.17 Comparison between LSMM and cmfdr (Zablocki et al., 2014)

We compared LSMM with cmfdr. As cmfdr is not able to handle a large number of covariates and the MCMC sampling algorithm it derived is time-consuming, we set $M = 5000$, $L = 5$, $K = 5$ and run 2500 iterations with 2000 retained draws for cmfdr. The comparison between LSMM and cmfdr are shown in Figure S60.

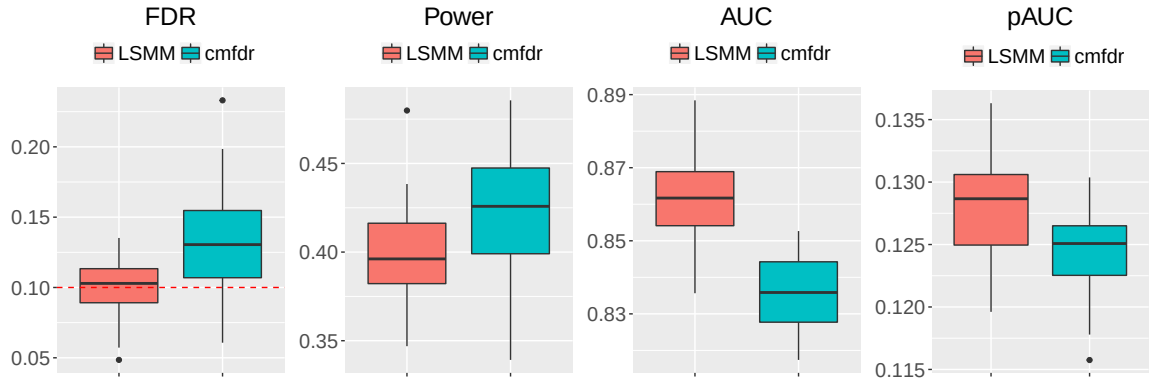

Figure S60: FDR, power, AUC and partial AUC of LSMM and cmfdr for identification of risk SNPs. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

3.18 Comparison between LSMM and GenoWAP (Lu et al., 2016)

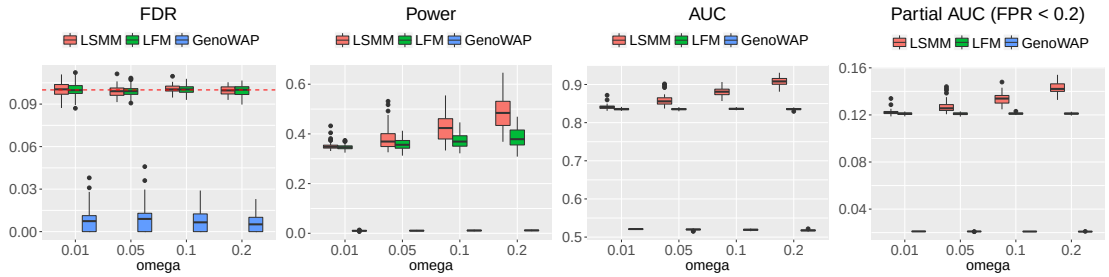

Figure S61: FDR, power, AUC and partial AUC of LSMM, LFM and GenoWAP for identification of risk SNPs with $\alpha = 0.2$ and $K = 100$. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

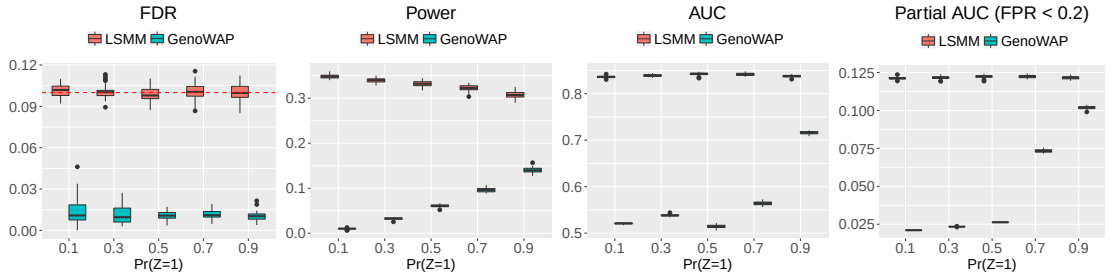

Figure S62: FDR, power, AUC and partial AUC of LSMM and GenoWAP for identification of risk SNPs. We controlled global FDR at 0.1 to evaluate empirical FDR and power. The results are summarized from 50 replications.

4 More about real data analysis

4.1 The source of the 30 GWAS

Alzheimer	Lambert et al., 2013, Nature Genetics. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
BMI	Speliotes et al., 2010, Nature Genetics. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Bipolar Disorder	Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011, Nature Genetics https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Coronary Artery Disease	Schunkert et al., 2011, Nature Genetics. http://www.cardiogramplusc4d.org/data-downloads
Crohns Disease	Jostins et al., 2012, Nature. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Height	Wood et al., 2014, Nature Genetics http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
High-density Lipoprotein	Global Lipids Genetics Consortium, 2013, Nature Genetics http://csg.sph.umich.edu//abecasis/public/lipids2013/
HIV	McLaren et al., 2013, PLoS Pathogens http://journals.plos.org/plospathogens/article?id=10.1371%2Fjournal.ppat.1003515
Inflammatory Bowel Disease	Jostins et al., 2012, Nature. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Low-density Lipoprotein	Global Lipids Genetics Consortium, 2013, Nature Genetics http://csg.sph.umich.edu//abecasis/public/lipids2013/
Lupus	Bentham et al., 2015, Nature Genetics https://www.immunobase.org/downloads/protected_data/GWAS_Data/
Mean Cell Haemoglobin	Pickrell, 2014, The American Journal of Human Genetics https://ega-archive.org/studies/EGAS00000000132
Mean Cell Volume	Pickrell, 2014, The American Journal of Human Genetics https://ega-archive.org/studies/EGAS00000000132
Menopause	Day et al., 2015, Nature Genetics. http://www.reprogen.org/data_download.html
Multiple Sclerosis	Sawcer et al., 2011, Nature. https://www.immunobase.org/downloads/protected_data/GWAS_Data/
Neuroticism	Okbay et al., 2016a, Nature Genetics. http://ssgac.org/documents/Neuroticism_Full.txt.gz
Primary Biliary Cirrhosis	Cordell et al., 2015, Nature Communications https://www.immunobase.org/downloads/protected_data/GWAS_Data/
Red Cell Count	Pickrell, 2014, The American Journal of Human Genetics https://ega-archive.org/studies/EGAS00000000132
Rheumatoid Arthritis	Okada et al., 2014, Nature. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Schizophrenia1	Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013, The Lancet. https://www.med.unc.edu/pgc/results-and-downloads (SCZ subset)
Schizophrenia2	Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011, Nature Genetics. https://www.med.unc.edu/pgc/results-and-downloads (SCZ1)
Schizophrenia3	Ripke et al., 2013, Nature Genetics. https://www.med.unc.edu/pgc/results-and-downloads (Sweden+SCZ1)
Schizophrenia4	Ripke et al., 2014, Nature. https://www.med.unc.edu/pgc/results-and-downloads (SCZ2)
Total Cholesterol	Global Lipids Genetics Consortium, 2013, Nature Genetics http://csg.sph.umich.edu//abecasis/public/lipids2013/
Triglycerides	Global Lipids Genetics Consortium, 2013, Nature Genetics http://csg.sph.umich.edu//abecasis/public/lipids2013/
Type 1 Diabetes	Bradfield et al., 2011, PLoS Genetics https://www.immunobase.org/downloads/protected_data/GWAS_Data/
Type 2 Diabetes	Morris et al., 2012, Nature Genetics. http://diagram-consortium.org/downloads.html
Ulcerative Colitis	Jostins et al., 2012, Nature. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Years of Education1	Rietveld et al., 2013, Science. https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
Years of Education2	Okbay et al., 2016b, Nature. http://ssgac.org/documents/EduYears_Main.txt.gz

Table S2: The source of the 30 GWAS.

4.2 Four Schizophrenia GWAS with different sample sizes

Table S3: Summary of results for Schizophrenia.

	$\hat{\alpha}$	No. of risk SNPs			
		Bonferroni correction	TGM	LFM	LSMM
Schizophrenia1	0.677	2	470	527	527
Schizophrenia2	0.633	7	2,107	2,404	2,405
Schizophrenia3	0.562	126	6,811	7,541	7,545
Schizophrenia4	0.413	1110	48,802	50,481	50,990

a. The estimate $\hat{\alpha}$ is obtained using LSMM.
b. The number of risk SNPs is reported based on global $FDR \leq 0.1$.

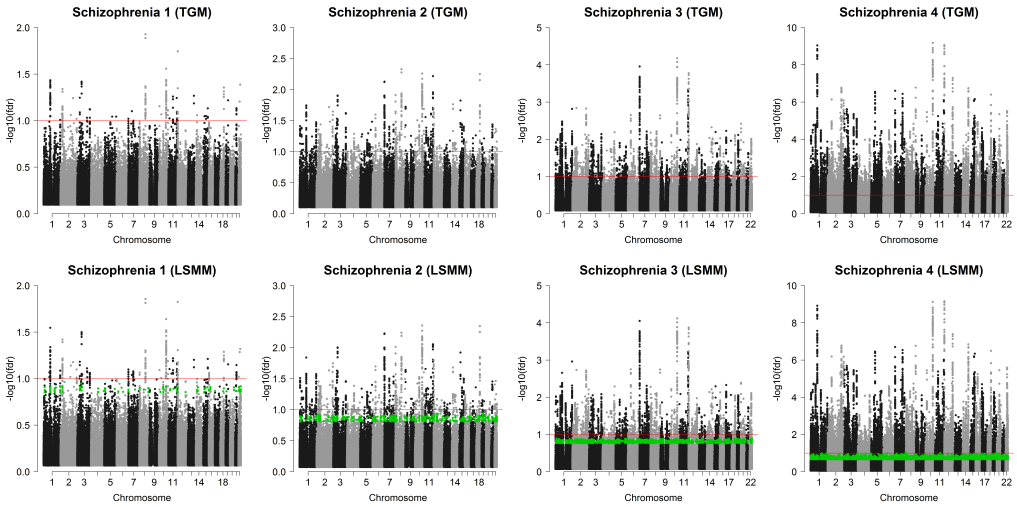

Figure S63: Manhattan plots of Schizophrenia1-4 using TGM and LSMM. The red lines indicate local $fdr = 0.1$. The green points denote the additional SNPs LSMM identied with $FDR \leq 0.1$.

		No. of risk SNPs	
		Schizophrenia3	Schizophrenia4
GenoWAP+	upstream	11	65
	downstream	12	42
	exonic	27	143
	intergenic	142	1,531
	intronic	624	3,541
	ncRNA_exonic	11	30
	ncRNA_intronic	63	272
	UTR3	29	156
	UTR5	4	21
	Total	922	5,798
TGM		3,092	24,575
LFM		3,395	25,383
LSMM		3,396	25,612

Table S4: The number of risk SNPs identified by GenoWAP, TGM, LFM and LSMM for Schizophrenia with the nominal local FDR controlled at 0.1.

4.3 Computational time for 30 GWAS

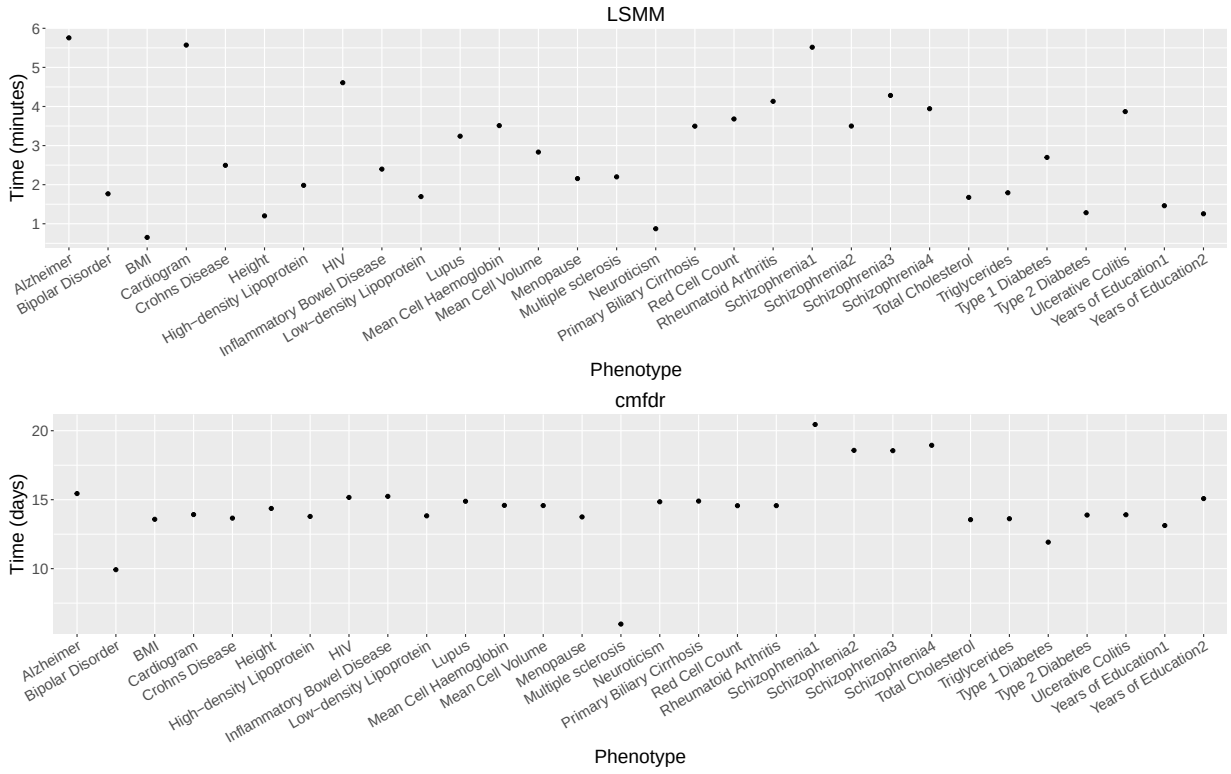

Figure S64: Computational time using LSMM and cmfdr for 30 GWAS.

4.4 Relevant functional annotations for 30 GWAS without fixed effects

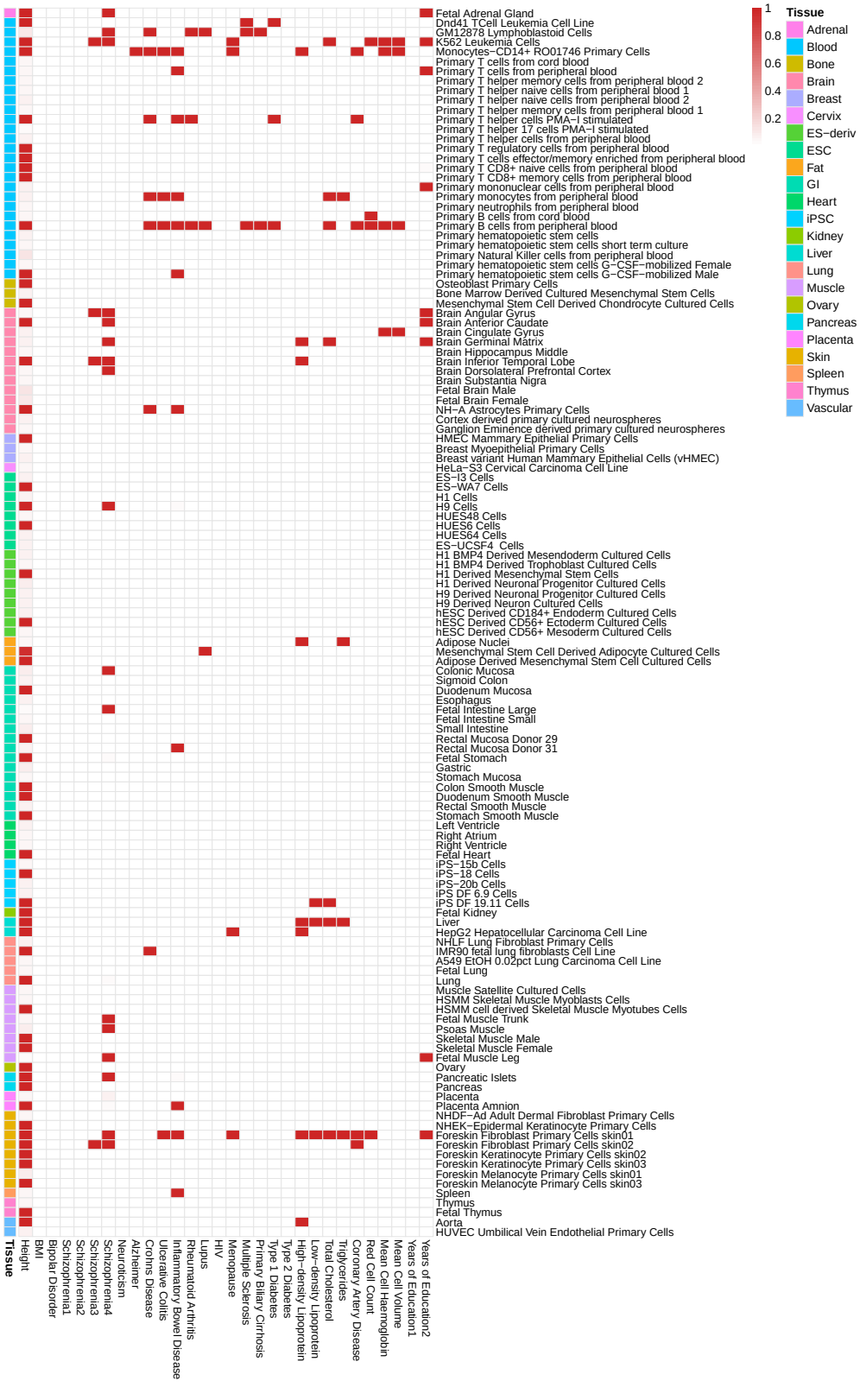

Figure S65: Relevant functional annotations for 30 GWAS without integrating genic category annotations.

References

Bentham, J. et al. (2015, 12). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature Genetics 47*(12), 1457–1464.

Bradfield, J. P. et al. (2011, 09). A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLOS Genetics 7*(9), 1–8.

Chung, D. et al. (2014, 11). GPA: A statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLOS Genetics 10*(11), 1–14.

Consortium, T. W. T. C. C. (2007, June). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature 447*(7145), 661–678.

Cordell, H. J. et al. (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nature Communications 6*, 8019.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet 381*(9875), 1371–1379.

Day, F. R. et al. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature Genetics 47*(11), 1294–1303.

Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics 45*(11), 1274–1283.

Jaakkola, T. S. and M. I. Jordan (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing 10*(1), 25–37.

Jostins, L. et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature 491*(7422), 119–124.

Lambert, J. C. et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics 45*(12), 1452–1458.

Lu, Q. et al. (2016). GenoWAP: GWAS signal prioritization through integrated analysis of genomic functional annotation. *Bioinformatics 32*(4), 542–548.

McLaren, P. J. et al. (2013, 07). Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLOS Pathogens 9*(7), 1–9.

Morris, A. P. et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics 44*(9), 981–990.

Okada, Y. et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature 506*(7488), 376–381.

Okbay, A. et al. (2016a). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics 48*(6), 624–633.

Okbay, A. et al. (2016b). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature 533*(7604), 539–542.

Pickrell, J. K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics 94*(4), 559–573.

Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics 43*(10), 977–983.

Rietveld, C. A. et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science 340*(6139), 1467–1471.

Ripke, S. et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics 45*(10), 1150–1159.

Ripke, S. et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature 511*(7510), 421–427.

Sawcer, S. et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature 476*(7359), 214–219.

Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics 43*(10), 969–976.

Schunkert, H. et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics 43*(4), 333–338.

Speliotes, E. K. et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics 42*(11), 937–948.

Wood, A. R. et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics 46*(11), 1173–1186.

Zablocki, R. W. et al. (2014). Covariate-modulated local false discovery rate for genome-wide association studies. *Bioinformatics 30*(15), 2098–2104.