Supplementary material for VIMCO: Variational Inference for Multiple Correlated Outcomes in Genome-wide Association Studies

Xingjie Shi[1, 4], Yuling Jiao[2], Yi Yang[3], Ching-Yu Cheng[4], Can Yang[5], Xinyi Lin[*4], and Jin Liu[*4]

[1]Department of Statistics, Nanjing University of Finance and Economics, Nanjing
[2]School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan
[3]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai
[4]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore
[5]Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

December 20, 2018

Contents

*To whom correspondence should be addressed, equal contribution.

A1 Derivation of the Optimal Variational Distribution

The entries in Θ are denoted by $\theta_{st}, s, t = 1, \dots, K$. The logarithm of the joint probability function in the main text is

$$\begin{aligned}
\log \Pr(\mathbf{Y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \Phi) =& \sum_n \log \mathcal{N}(\sum_{j=1}^{p} X_{nj} \boldsymbol{\beta}_j, \Theta^{-1}) + \sum_j \sum_k \log \mathcal{N}(0, \sigma^2_{\beta_k}) \\
&+ \sum_j \sum_k [\gamma_{jk} \log a_k + (1 - \gamma_{jk}) \log(1 - a_k)] \\
=& \sum_n \left[-\frac{K}{2} \log(2\pi) + \frac{1}{2} \log|\Theta| - \frac{1}{2} (\mathbf{y}_n - \sum_{j=1}^{p} X_{nj} \boldsymbol{\beta}_j)^\top \Theta (\mathbf{y}_n - \sum_{j=1}^{p} X_{nj} \boldsymbol{\beta}_j) \right] \\
&+ \sum_j \sum_k \left[-\frac{1}{2} \log(2\pi\sigma^2_{\beta_k}) - \frac{\tilde{\beta}^2_{jk}}{2\sigma^2_{\beta_k}} \right] \\
&+ \sum_j \sum_k [\gamma_{jk} \log a_k + (1 - \gamma_{jk}) \log(1 - a_k)] \\
=& -\frac{1}{2} \sum_n \left[\sum_s \sum_t \theta_{st} \left(Y_{ns} - \sum_j X_{nj} \gamma_{js} \tilde{\beta}_{js} \right) \left(Y_{nt} - \sum_j X_{nj} \gamma_{jt} \tilde{\beta}_{jt} \right) \right] \\
&+ \frac{N}{2} \log|\Theta| - \frac{p}{2} \sum_k \log(\sigma^2_{\beta_k}) - \sum_j \sum_k \frac{\tilde{\beta}^2_{jk}}{2\sigma^2_{\beta_k}} \\
&+ \sum_j \sum_k [\gamma_{jk} \log a_k + (1 - \gamma_{jk}) \log(1 - a_k)] + \text{const.}
\end{aligned} \tag{A1}$$

As stated in the main text, we consider variational distributions of the form

$$q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}) = \prod_{j=1} \prod_{k=1} \left[q(\tilde{\beta}_{jk} | \gamma_{jk}) q(\gamma_{jk}) \right].$$

With this family of variational distributions, the optimal $q^*(\tilde{\beta}_{jk}, \gamma_{jk})$ maximizing the lower bound $\mathcal{L}_q$ has the form

$$\log q^*(\tilde{\beta}_{jk}, \gamma_{jk}) = \mathrm{E}_{(j',k') \neq (j,k)} \left[\log \Pr(\mathbf{Y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \Phi) \right] + \text{const.} \tag{A2}$$

Decomposing (A1) into terms involving and not involving $(\tilde{\beta}_{jk}, \gamma_{jk})$, we get

$$
\begin{aligned}
&\log \Pr(\mathbf{Y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \Phi) \\
=& -\frac{1}{2}\sum_n \sum_s \sum_t \theta_{st} \left(Y_{ns} - \sum_{j' \neq j} X_{nj'}\gamma_{j's}\tilde{\beta}_{j's} \right) \left(Y_{nt} - \sum_{j' \neq j} X_{nj'}\gamma_{j't}\tilde{\beta}_{j't} \right) \\
&+ \frac{1}{2}\sum_n \sum_s \sum_t \theta_{st} \left(Y_{ns} - \sum_{j' \neq j} X_{nj'}\gamma_{j's}\tilde{\beta}_{j's} \right) X_{nj}\gamma_{jt}\tilde{\beta}_{jt} \\
&+ \frac{1}{2}\sum_n \sum_s \sum_t \theta_{st} \left(Y_{nt} - \sum_{j' \neq j} X_{nj'}\gamma_{j't}\tilde{\beta}_{j't} \right) X_{nj}\gamma_{js}\tilde{\beta}_{js} \\
&- \frac{1}{2}\sum_n \sum_s \sum_t \theta_{st} X_{nj}^2 \gamma_{js}\tilde{\beta}_{js}\gamma_{jt}\tilde{\beta}_{jt} \\
&+ \frac{N}{2}\log|\Theta| - \frac{p}{2}\sum_k \log(\sigma^2_{\beta_k}) - \frac{\tilde{\beta}^2_{jk}}{2\sigma^2_{\beta_k}} - \sum_{(j',k') \neq (j,k)} \frac{\tilde{\beta}^2_{j'k'}}{2\sigma^2_{\beta_{k'}}} \\
&+ [\gamma_{jk}\log a_k + (1-\gamma_{jk})\log(1-a_k)] + \sum_{(j',k') \neq (j,k)} [\gamma_{j'k'}\log a_{k'} + (1-\gamma_{j'k'})\log(1-a_{k'})] \\
&+ \text{const.}
\end{aligned} \tag{A3}
$$

When $\gamma_{jk} = 1$, we have

$$
\begin{aligned}
&\log q^*(\tilde{\boldsymbol{\beta}}_j | \gamma_{jk} = 1) \\
=& \left(-\frac{1}{2} X_j^\top X_j \theta_{kk} - \frac{1}{2\sigma^2_{\beta_k}} \right) \tilde{\beta}^2_{jk} \\
&+ \left[\sum_t \theta_{kt} X_j^\top Y_t - \sum_t \theta_{kt} \sum_{j' \neq j} \mathrm{E}(\gamma_{j't}\tilde{\beta}_{j't}) X_{j'}^\top X_j - \sum_{t \neq k} \theta_{kt} \mathrm{E}(\gamma_{jt}\tilde{\beta}_{jt}) X_j^\top X_j \right] \tilde{\beta}_{jk} \\
&+ \text{const},
\end{aligned} \tag{A4}
$$

from which we can see that the posterior of $q(\tilde{\boldsymbol{\beta}}_j | \gamma_{jk} = 1) \sim \mathcal{N}(\mu_{jk}, s^2_{jk})$:

$$
\begin{aligned}
s^2_{jk} &= \frac{1}{X_j^\top X_j \theta_{kk} + \frac{1}{\sigma^2_{\beta_k}}}, \\
\mu_{jk} &= \frac{\sum_t \theta_{kt} X_j^\top Y_t - \sum_t \theta_{kt} \sum_{j' \neq j} \mathrm{E}(\gamma_{j't}\tilde{\beta}_{j't}) X_{j'}^\top X_j - \sum_{t \neq k} \theta_{kt} \mathrm{E}(\gamma_{jt}\tilde{\beta}_{jt}) X_j^\top X_j}{X_j^\top X_j \theta_{kk} + \frac{1}{\sigma^2_{\beta_k}}}.
\end{aligned} \tag{A5}
$$

Similarly, when $\gamma_{jk} = 0$, we have

$$\log q_j^*(\tilde{\boldsymbol{\beta}}_{jk}|\gamma_{jk} = 0) = -\frac{\tilde{\beta}_{jk}^2}{2\sigma_{\beta_k}^2} + \text{const},$$

Thus $q^*(\tilde{\boldsymbol{\beta}}_{jk}|\gamma_{jk} = 0) = \mathcal{N}(0, \sigma_{\beta_k}^2)$. This implies that the posterior distribution of $\tilde{\beta}_{jk}$ will be the same as its prior if this variable is irrelevant ($\gamma_{jk} = 0$). Note that γ_{jk} is a binary variable and then denote $\alpha_{jk} = q(\gamma_{jk} = 1)$. Therefore we have

$$q^*(\tilde{\boldsymbol{\beta}}, \gamma) = \prod_j \prod_k \left[\alpha_{jk}^{\gamma_{jk}}(1-\alpha_{jk})^{1-\gamma_{jk}} \mathcal{N}(\mu_{jk}, s_{jk}^2)^{\gamma_{jk}} \mathcal{N}(0, \sigma_{\beta_k}^2)^{1-\gamma_{jk}}\right]. \tag{A6}$$

A2 Evaluation of the Lower Bound

Now we evaluate the first part of the lower bound $\mathcal{L}(q)$.

$$\begin{aligned} & -\mathrm{E}_q\left[\log q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})\right] \\ =& -\sum_j \sum_k \left[\mathrm{E}_q(\gamma_{jk}) \log \alpha_{jk} + (1 - \mathrm{E}_q(\gamma_{jk})) \log(1-\alpha_{jk})\right] \\ & -\sum_j \sum_k \left[\mathrm{E}_q\left(\gamma_{jk} \log \mathcal{N}(\mu_{jk}, s_{jk}^2)\right) + \mathrm{E}_q\left((1-\gamma_{jk}) \log \mathcal{N}(0, \sigma_{\beta_k}^2)\right)\right]. \end{aligned} \tag{A7}$$

The first expectation on the previous line is

$$\begin{aligned} & \mathrm{E}_q\left[\gamma_{jk} \log \mathcal{N}(\mu_{jk}, s_{jk}^2)\right] \\ =& \mathrm{E}_q\left[\gamma_{jk} \log q(\tilde{\beta}_{jk}|\gamma_{jk} = 1)\right] \\ =& \sum_{\gamma_{jk}} \int_{\tilde{\beta}_{jk}} \gamma_{jk} \log q(\tilde{\beta}_{jk}|\gamma_{jk} = 1) q(\tilde{\beta}_{jk}|\gamma_{jk}) q(\gamma_{jk}) d\tilde{\beta}_{jk} \\ =& \alpha_{jk} \int_{\tilde{\beta}_{jk}} \log q(\tilde{\beta}_{jk}|\gamma_{jk} = 1) q(\tilde{\beta}_{jk}|\gamma_{jk} = 1) d\tilde{\beta}_{jk} \\ =& \alpha_{jk}\left[-\frac{1}{2}(1 + \log 2\pi) - \frac{1}{2} \log s_{jk}^2\right], \end{aligned}$$

where the last equality is based on the entropy of a Gaussian distribution. Similarly, $\mathrm{E}_q\left((1-\gamma_{jk})\log\mathcal{N}(0,\sigma^2_{\beta_k})\right) = (1-\alpha_{jk})\left[-\frac{1}{2}(1+\log 2\pi) - \frac{1}{2}\log\sigma^2_{\beta_k}\right]$. Overall, we have

$$
\begin{aligned}
& -\mathrm{E}_q\left[\log q(\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma})\right] \\
=& -\sum_j\sum_k\left[\alpha_{jk}\log\alpha_{jk} + (1-\alpha_{jk})\log(1-\alpha_{jk})\right] \\
& + \frac{1}{2}\sum_j\sum_k\left[\alpha_{jk}\left(1+\log 2\pi+\log s^2_{jk}\right) + (1-\alpha_{jk})\left(1+\log 2\pi+\log\sigma^2_{\beta_k}\right)\right] \\
=& -\sum_j\sum_k\left[\alpha_{jk}\log\alpha_{jk} + (1-\alpha_{jk})\log(1-\alpha_{jk})\right] \\
& + \frac{1}{2}\sum_j\sum_k\alpha_{jk}\log\frac{s^2_{jk}}{\sigma^2_{\beta_k}} + \frac{p}{2}\sum_k\log\sigma^2_{\beta_k} + \frac{Kp}{2}(1+\log 2\pi)
\end{aligned}
\tag{A8}
$$

The second part of $\mathcal{L}(q)$:

$$
\begin{aligned}
& \mathrm{E}_q\left[\log\Pr(\mathbf{Y},\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma}|\mathbf{X};\Phi)\right] \\
=& -\frac{1}{2}\sum_n\left[\sum_s\sum_t\theta_{st}\mathrm{E}_q\left(Y_{ns}-\sum_j X_{nj}\gamma_{js}\tilde{\beta}_{js}\right)\left(Y_{nt}-\sum_j X_{nj}\gamma_{jt}\tilde{\beta}_{jt}\right)\right] \\
& + \frac{N}{2}\log|\Theta| - \frac{p}{2}\sum_k\log(\sigma^2_{\beta_k}) - \sum_j\sum_k\frac{\mathrm{E}_q\tilde{\beta}^2_{jk}}{2\sigma^2_{\beta_k}} \\
& + \sum_j\sum_k\left[\mathrm{E}_q\gamma_{jk}\log a_k + (1-\mathrm{E}_q\gamma_{jk})\log(1-a_k)\right] + \text{const.}
\end{aligned}
\tag{A9}
$$

The expectation in the first term

$$
\begin{aligned}
&\mathrm{E}_q\left(Y_{ns}-\sum_j X_{nj}\gamma_{js}\tilde{\beta}_{js}\right)\left(Y_{nt}-\sum_j X_{nj}\gamma_{jt}\tilde{\beta}_{jt}\right)\\
=&Y_{ns}Y_{nt}-Y_{ns}\sum_j X_{nj}\mathrm{E}_q(\gamma_{jt}\tilde{\beta}_{jt})-Y_{nt}\sum_j X_{nj}\mathrm{E}_q(\gamma_{js}\tilde{\beta}_{js})+\mathrm{E}_q\left(\sum_j X_{nj}\gamma_{js}\tilde{\beta}_{js}\sum_j X_{nj}\gamma_{jt}\tilde{\beta}_{jt}\right)\\
=&Y_{ns}Y_{nt}-Y_{ns}\sum_j X_{nj}\mathrm{E}_q(\gamma_{jt}\tilde{\beta}_{jt})-Y_{nt}\sum_j X_{nj}\mathrm{E}_q(\gamma_{js}\tilde{\beta}_{js})\\
&+\mathrm{E}_q\left(\sum_j X_{nj}\gamma_{js}\tilde{\beta}_{js}\sum_{j'\neq j} X_{nj'}\gamma_{j't}\tilde{\beta}_{j't}\right)+\mathrm{E}_q\left(\sum_j X_{nj}^2\gamma_{js}\gamma_{jt}\tilde{\beta}_{js}\tilde{\beta}_{jt}\right)\\
=&Y_{ns}Y_{nt}-Y_{ns}\sum_j X_{nj}\mathrm{E}_q(\gamma_{jt}\tilde{\beta}_{jt})-Y_{nt}\sum_j X_{nj}\mathrm{E}_q(\gamma_{js}\tilde{\beta}_{js})\\
&+\sum_j X_{nj}\mathrm{E}_q\left(\gamma_{js}\tilde{\beta}_{js}\right)\sum_{j'\neq j} X_{nj'}\mathrm{E}_q\left(\gamma_{j't}\tilde{\beta}_{j't}\right)+\sum_j X_{nj}^2\mathrm{E}_q\left(\gamma_{js}\gamma_{jt}\tilde{\beta}_{js}\tilde{\beta}_{jt}\right).
\end{aligned}
\tag{A10}
$$

It's easy to verify that

$$
\begin{aligned}
\mathrm{E}_q(\gamma_{js}\tilde{\beta}_{js}) &= \alpha_{js}\mu_{js},\\
\mathrm{E}_q\tilde{\beta}_{jk}^2 &= \alpha_{jk}(\mu_{jk}^2+s_{jk}^2)+(1-\alpha_{jk})\sigma_{\beta_k}^2,\\
\mathrm{E}_q(\gamma_{js}\gamma_{jt}\tilde{\beta}_{js}\tilde{\beta}_{jt}) &= \alpha_{js}\alpha_{jt}\mu_{js}\mu_{jt},\\
\mathrm{E}_q(\gamma_{js}^2\tilde{\beta}_{js}^2) &= \alpha_{js}(\mu_{js}^2+s_{js}^2).
\end{aligned}
\tag{A11}
$$

Plugging these expectations in (A9) and combing with (A8), the lower bound is given by

$$\begin{aligned}
\mathcal{L}_q =& -\frac{1}{2}\sum_n\sum_s\sum_t \theta_{st}\Bigg[Y_{ns}Y_{nt} - Y_{ns}\sum_j X_{nj}\mathrm{E}_q(\gamma_{jt}\tilde{\beta}_{jt}) - Y_{nt}\sum_j X_{nj}\mathrm{E}_q(\gamma_{js}\tilde{\beta}_{js}) \\
&+ \sum_j X_{nj}\mathrm{E}_q\left(\gamma_{js}\tilde{\beta}_{js}\right)\sum_{j'\neq j} X_{nj'}\mathrm{E}_q\left(\gamma_{j't}\tilde{\beta}_{j't}\right) + \sum_j X_{nj}^2\mathrm{E}_q\left(\gamma_{js}\gamma_{jt}\tilde{\beta}_{js}\tilde{\beta}_{jt}\right)\Bigg] \\
&+ \frac{N}{2}\log|\Theta| - \frac{p}{2}\sum_k \log(\sigma^2_{\beta_k}) - \sum_j\sum_k \frac{\mathrm{E}_q\tilde{\beta}^2_{jk}}{2\sigma^2_{\beta_k}} \\
&+ \sum_j\sum_k \left[\mathrm{E}_q\gamma_{jk}\log a_k + (1-\mathrm{E}_q\gamma_{jk})\log(1-a_k)\right] \\
&- \sum_j\sum_k \left[\alpha_{jk}\log\alpha_{jk} + (1-\alpha_{jk})\log(1-\alpha_{jk})\right] \\
&+ \frac{1}{2}\sum_j\sum_k \alpha_{jk}\log\frac{s^2_{jk}}{\sigma^2_{\beta_k}} + \frac{p}{2}\sum_k \log\sigma^2_{\beta_k} + \frac{Kp}{2}(1+\log 2\pi) \\
=& -\frac{1}{2}\sum_s\sum_t \theta_{st}(Y_s - \sum_j X_j\alpha_{js}\mu_{js})^\top(Y_t - \sum_j X_j\alpha_{jt}\mu_{jt}) \\
&- \frac{1}{2}\sum_s \theta_{ss}\sum_j X_j^\top X_j[\alpha_{js}(\mu^2_{js}+s^2_{js}) - \alpha^2_{js}\mu^2_{js}] \\
&- \sum_j\sum_k \left[\alpha_{jk}\log\frac{\alpha_{jk}}{a_k} + (1-\alpha_{jk})\log\frac{1-\alpha_{jk}}{1-a_k}\right] \\
&+ \frac{N}{2}\log|\Theta| + \frac{1}{2}\sum_j\sum_k \alpha_{jk}\left(1 + \log\frac{s^2_{jk}}{\sigma^2_{\beta_k}} - \frac{\mu^2_{jk}+s^2_{jk}}{\sigma^2_{\beta_k}}\right) + \text{const.}
\end{aligned} \tag{A12}$$

A3 E-step

Now we consider the iterative update of parameters. Initialize $(a_1, \ldots, a_K, \sigma^2_{\beta_1}, \ldots, \sigma^2_{\beta_K}, \Theta, \alpha_{jk}, \mu_{jk}, s^2_{jk})$, $j = 1, \ldots, p$.

To update μ_{jk} and s^2_{jk}, $j = 1, \ldots, p$, we use

$$\begin{aligned}
s^2_{jk} &= \frac{1}{X_j^\top X_j\theta_{kk} + \frac{1}{\sigma^2_{\beta_k}}}, \\
\mu_{jk} &= \frac{\sum_t \theta_{kt}X_j^\top Y_t - \sum_t \theta_{kt}\sum_{j'\neq j}\alpha_{j't}\mu_{j't}X_{j'}^\top X_j - \sum_{t\neq k}\theta_{kt}\alpha_{jt}\mu_{jt}X_j^\top X_j}{X_j^\top X_j\theta_{kk} + \frac{1}{\sigma^2_{\beta_k}}}.
\end{aligned} \tag{A13}$$

To update α_{jk}, we let

$$\frac{\partial \mathcal{L}_q}{\partial \alpha_{jk}} = \log \frac{a_k}{1-a_k} - \log \frac{\alpha_{jk}}{1-\alpha_{jk}} + \frac{1}{2}\left(\frac{\mu_{jk}^2}{s_{jk}^2} + \log \frac{s_{jk}^2}{\sigma_{\beta_k}^2}\right) = 0, \tag{A14}$$

and obtain

$$\alpha_{jk} = \frac{1}{1+\exp(-v_{jk})}, \tag{A15}$$

where

$$v_{jk} = \log \frac{a_k}{1-a_k} + \frac{1}{2}\left(\frac{\mu_{jk}^2}{s_{jk}^2} + \log \frac{s_{jk}^2}{\sigma_{\beta_k}^2}\right)$$

A4 M-Step

By taking derivatives with respect to model parameters and setting them to zero,

$$\begin{aligned}
\frac{\partial \mathcal{L}_q}{\partial \theta_{kk}} =& \frac{N}{2}(\Theta^{-1})_{kk} - \frac{1}{2}(Y_k - \sum_j X_j \alpha_{jk}\mu_{jk})^\top (Y_k - \sum_j X_j \alpha_{jk}\mu_{jk}) \\
&- \frac{1}{2}\sum_j X_j^\top X_j \left[\alpha_{jk}(\mu_{jk}^2 + s_{jk}^2) - \alpha_{jk}^2 \mu_{jk}^2\right] = 0, \\
\frac{\partial \mathcal{L}_q}{\partial \theta_{kt}} =& \frac{N}{2}(\Theta^{-1})_{kt} - \frac{1}{2}(Y_k - \sum_j X_j \alpha_{jk}\mu_{jk})^\top (Y_t - \sum_j X_j \alpha_{jt}\mu_{jt}) = 0 \text{ for } k \neq t, \\
\frac{\partial \mathcal{L}_q}{\partial \sigma_{\beta_k}^2} =& \sum_j \frac{\alpha_{jk}(\mu_{jk}^2 + s_{jk}^2)}{\sigma_{\beta_k^4}} - \sum_j \frac{\alpha_{jk}}{\sigma_{\beta_k}^2} = 0, \\
\frac{\partial \mathcal{L}_q}{\partial a_k} =& \sum_j \left[\alpha_{jk}(1-a_k) - (1-\alpha_{jk})a_k\right] = 0,
\end{aligned} \tag{A16}$$

we can obtain

$$\begin{aligned}
(\Theta^{-1})_{kk} &= \frac{(Y_k - \sum_j X_j \alpha_{jk}\mu_{jk})^\top (Y_k - \sum_j X_j \alpha_{jk}\mu_{jk}) + \sum_j X_j^\top X_j \left[\alpha_{jk}(\mu_{jk}^2 + s_{jk}^2) - \alpha_{jk}^2 \mu_{jk}^2\right]}{N}, \\
(\Theta^{-1})_{kt} &= \frac{(Y_k - \sum_j X_j \alpha_{jk}\mu_{jk})^\top (Y_t - \sum_j X_j \alpha_{jt}\mu_{jt})}{N}, \\
\sigma_{\beta_k}^2 &= \frac{\sum_j \alpha_{jk}(\mu_{jk}^2 + s_{jk}^2)}{\sum_j \alpha_{jk}}, \text{ for } k = 1, \ldots, K, \\
a_k &= \frac{\sum_j \alpha_{jk}}{p}.
\end{aligned} \tag{A17}$$

A5 VBEM Algorithm

Let $\circ$ denote the element-wise product, $\alpha_{j\bullet} = [\alpha_{j1}, \dots, \alpha_{jk}]$ and $\theta_{\bullet k} = [\theta_{1k}, \dots, \theta_{kk}]^\top$. Implementation details are summarized in Algorithm 1 for clarity.

Algorithm 1: VBEM

Initialize $\{\alpha_{jk}, \mu_{jk}, s^2_{jk}, \sigma^2_{\beta_k}\}_{j=1,\dots,p;k=1,\dots,K}, \Theta$. Compute $a_k = \frac{\sum_j \alpha_{jk}}{p}$. Let $\tilde{Y} = [\sum_j X_j \alpha_{j1}\mu_{j1}, \dots, \sum_j X_j \alpha_{jK}\mu_{jK}]$.

repeat

E-step:

for $j = 1, \dots, p$ **do**

$\tilde{Y}_{-j} = \tilde{Y} - [(\alpha_{j\bullet} \circ \mu_{j\bullet}) \otimes X_j]$;

for $k = 1, \dots, K$ **do**

$s^2_{jk} = \frac{1}{X_j^\top X_j \theta_{kk} + \frac{1}{\sigma^2_{\beta_k}}}$;

$\mu_{jk} = \left[X_j^\top (Y - \tilde{Y}_{-j})\theta_{\bullet k} - \sum_{t \neq k} \theta_{kt}\alpha_{jt}\mu_{jt} X_j^\top X_j \right] s^2_{jk}$;

$\alpha_{jk} = \frac{1}{1+\exp(-v_{jk})}$, where $v_{jk} = \log \frac{a_k}{1-a_k} + \frac{1}{2}\left(\frac{\mu^2_{jk}}{s^2_{jk}} + \log \frac{s^2_{jk}}{\sigma^2_{\beta_k}} \right)$;

end

end

M-step:

for $k = 1, \dots, K$ **do**

$\sigma^2_{\beta_k} = \frac{\sum_j \alpha_{jk}(\mu^2_{jk} + s^2_{jk})}{\sum_j \alpha_{jk}}$;

$(\Theta^{-1})_{kk} = \frac{(Y_k - \sum_j X_j \alpha_{jk}\mu_{jk})^\top (Y_k - \sum_j X_j \alpha_{jk}\mu_{jk}) + \sum_j X_j^\top X_j \left[\alpha_{jk}(\mu^2_{jk} + s^2_{jk}) - \alpha^2_{jk}\mu^2_{jk} \right]}{N}$;

for $t = k+1, \dots, K$ **do**

$(\Theta^{-1})_{kt} = \frac{(Y_k - \sum_j X_j \alpha_{jk}\mu_{jk})^\top (Y_t - \sum_j X_j \alpha_{jt}\mu_{jt})}{N}$;

end

end

$a_k = \frac{\sum_j \alpha_{jk}}{p}, k = 1, \dots, K$;

until *the change of lower bound* $\mathcal{L}_q$ *is smaller than a threshold*;

return $\{\alpha_{jk}, \mu_{jk}, s^2_{jk}, \sigma^2_{\beta_k}\}_{j=1,\dots,p;k=1,\dots,K}, \Theta, a_1, \dots, a_K$.

A6 Positive-definiteness of Θ

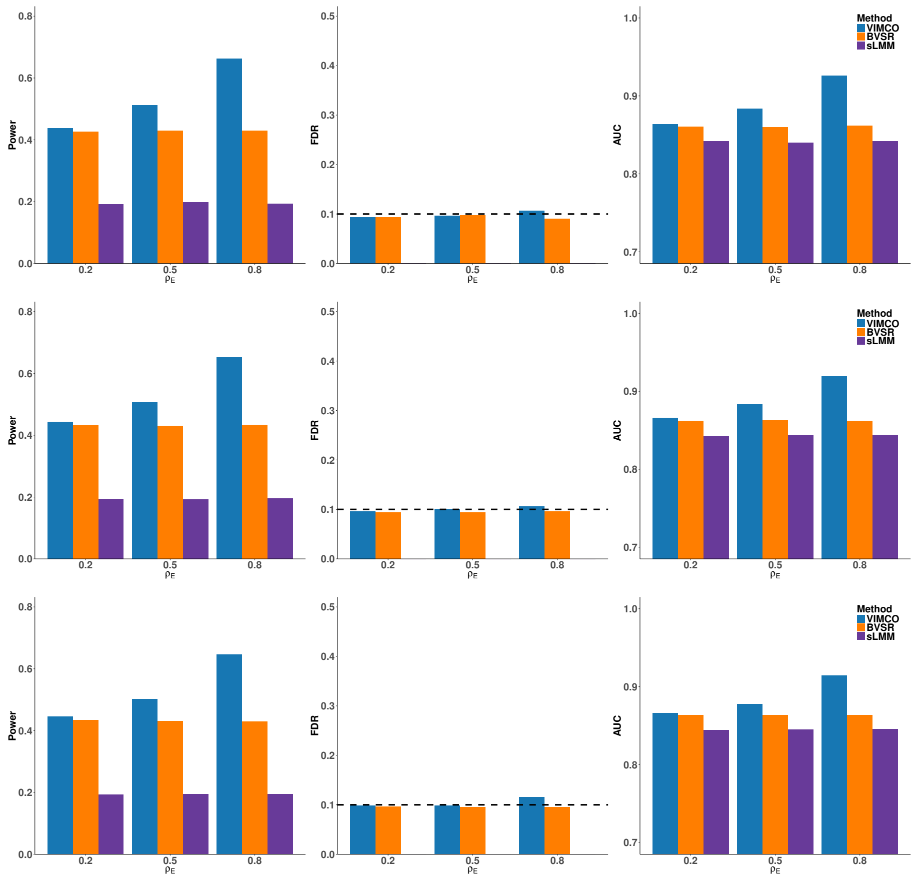
Lemma 0.1. *If* $K < N$, Θ *is a positive-definite matrix.*

Proof of Lemma 0.1 It is equivalent to prove Θ^{-1} is positive definite. Let $e = (e_1, \ldots, e_K)$ with $e_k = Y_k - \sum_j X_j \alpha_{jk} \mu_{jk}$, and $Z_j = (\gamma_{j1} \tilde{\beta}_{j1}, \ldots, \gamma_{j1} \tilde{\beta}_{jK})^\top$. From the first two equations in (A17), we have

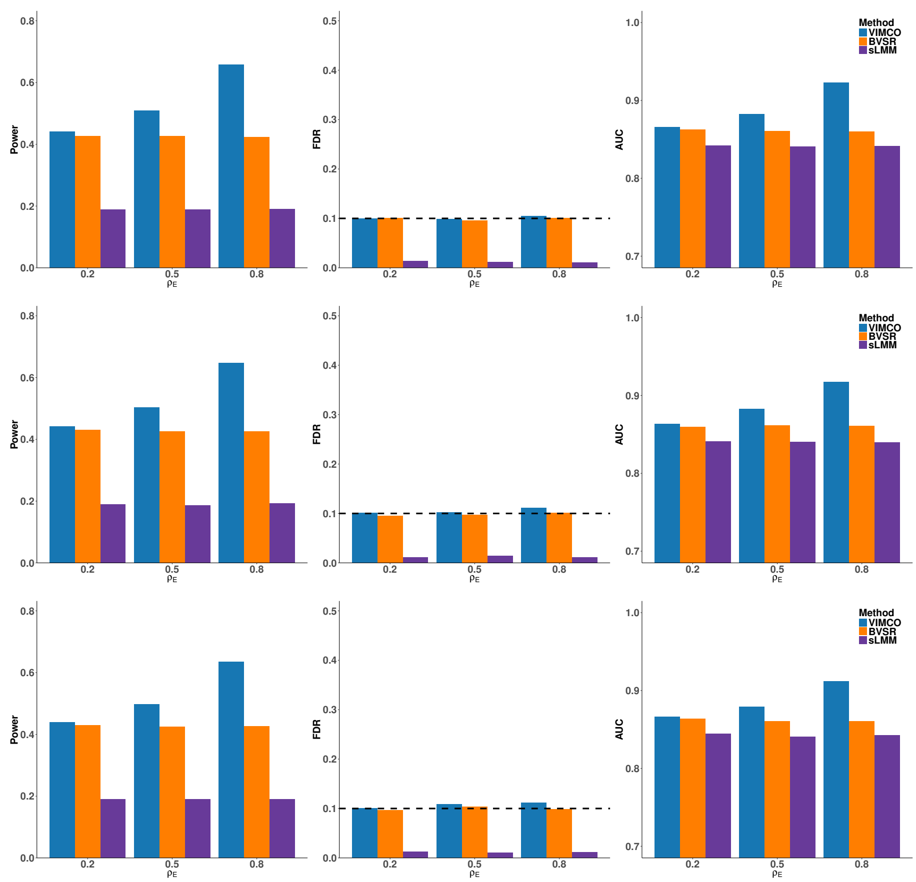
$$\Theta^{-1} = \frac{1}{N} \left[e^\top e + \sum_j X_j^\top X_j \mathrm{Cov}(Z_j) \right]. \tag{A18}$$

Since $X_j^\top X_j$ is positive, and both $e^\top e$ and covariance matrix $\mathrm{Cov}(Z_j)$ are positive definite, Θ^{-1} is also positive definite.

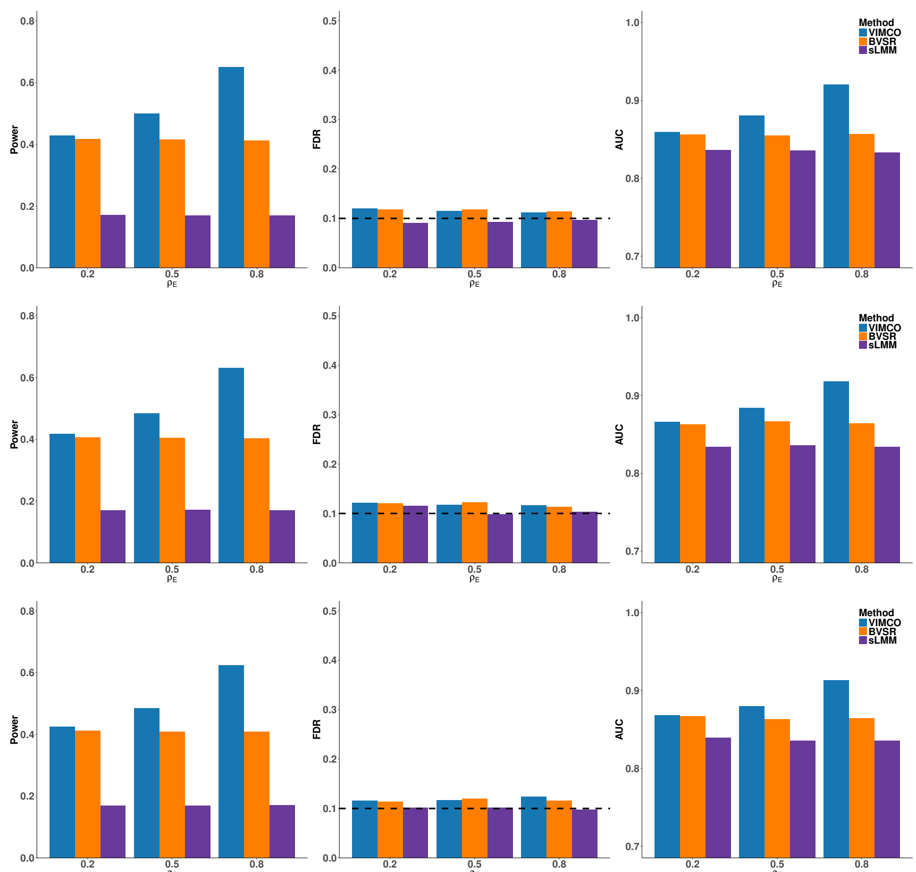
B1 Additional results on simulation studies



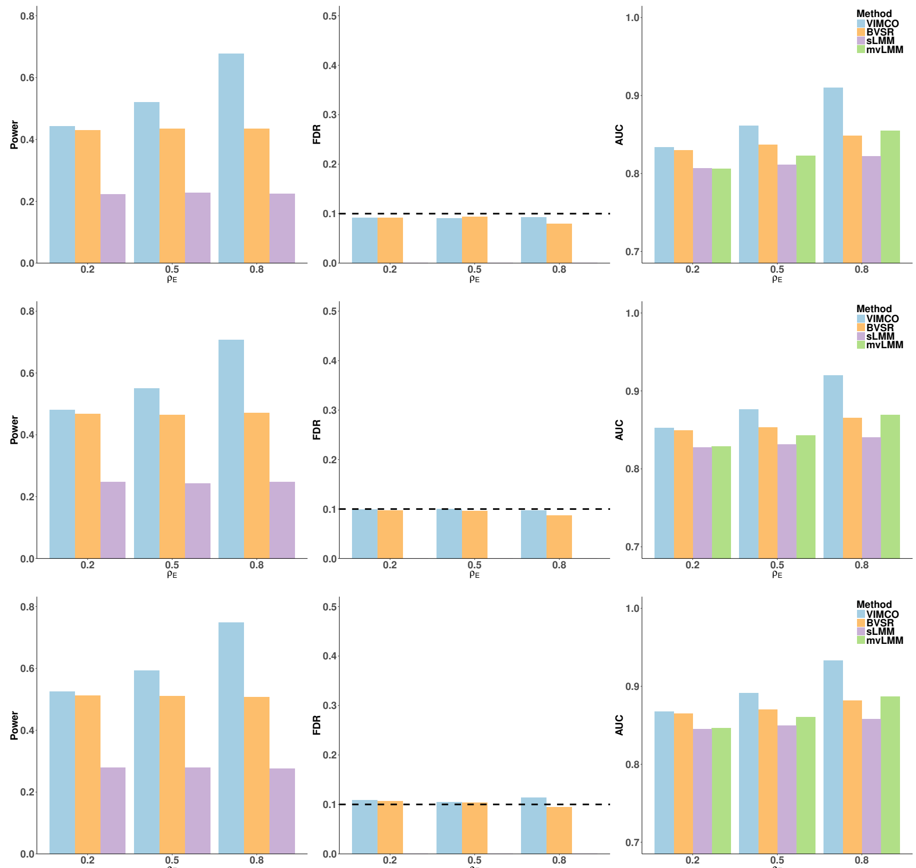
Supplementary Figure B1: Power, FDR and AUC for evaluating $H_{0a} : \beta_{jk} = 0$, for simulations with $\rho_x = 0.2$. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



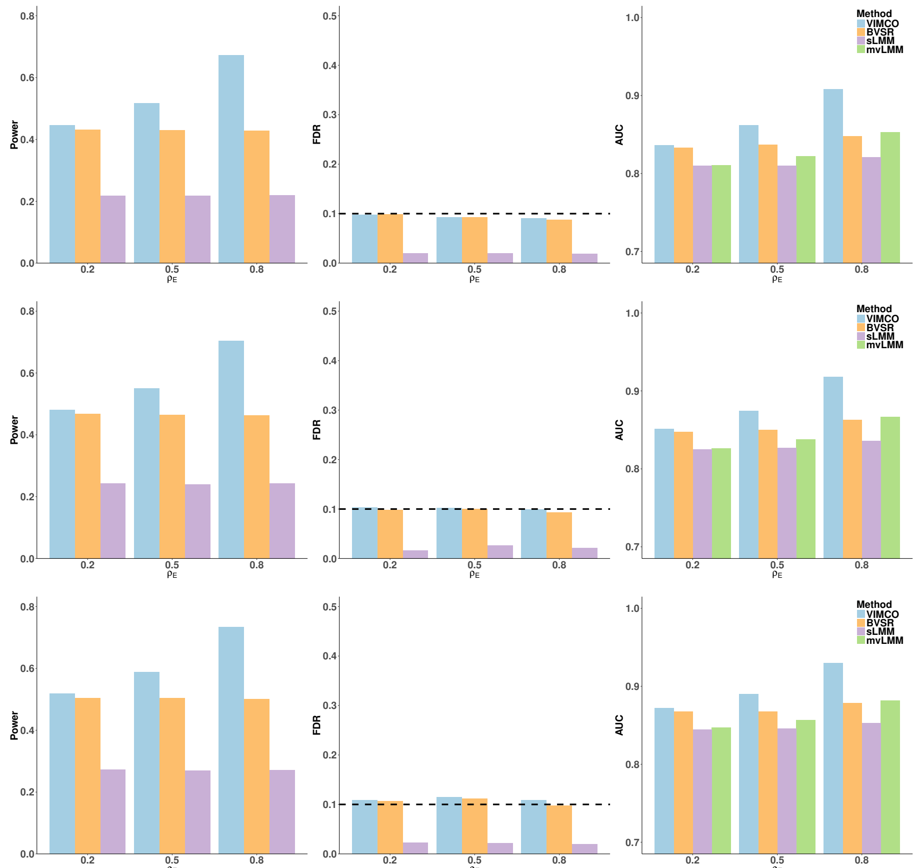
Supplementary Figure B2: Power, FDR and AUC for evaluating $H_{0a} : \beta_{jk} = 0$, for simulations with $\rho_x = 0.5$. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



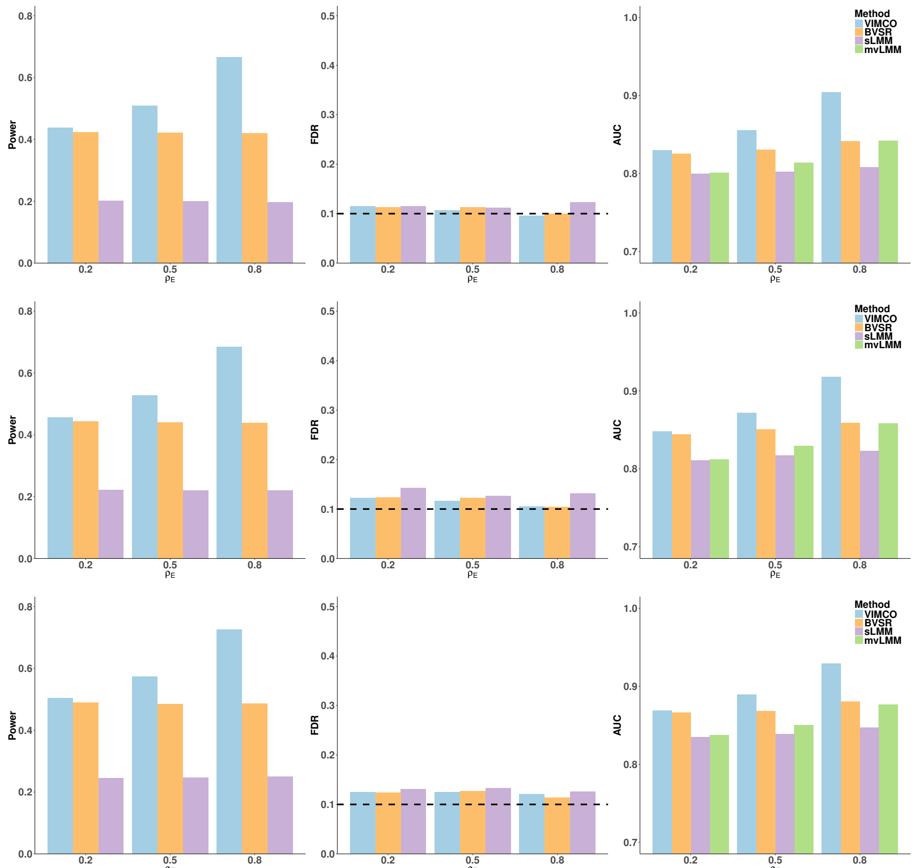
Supplementary Figure B3: Power, FDR and AUC for evaluating $H_{0a} : \beta_{jk} = 0$, for simulations with $\rho_x = 0.8$. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



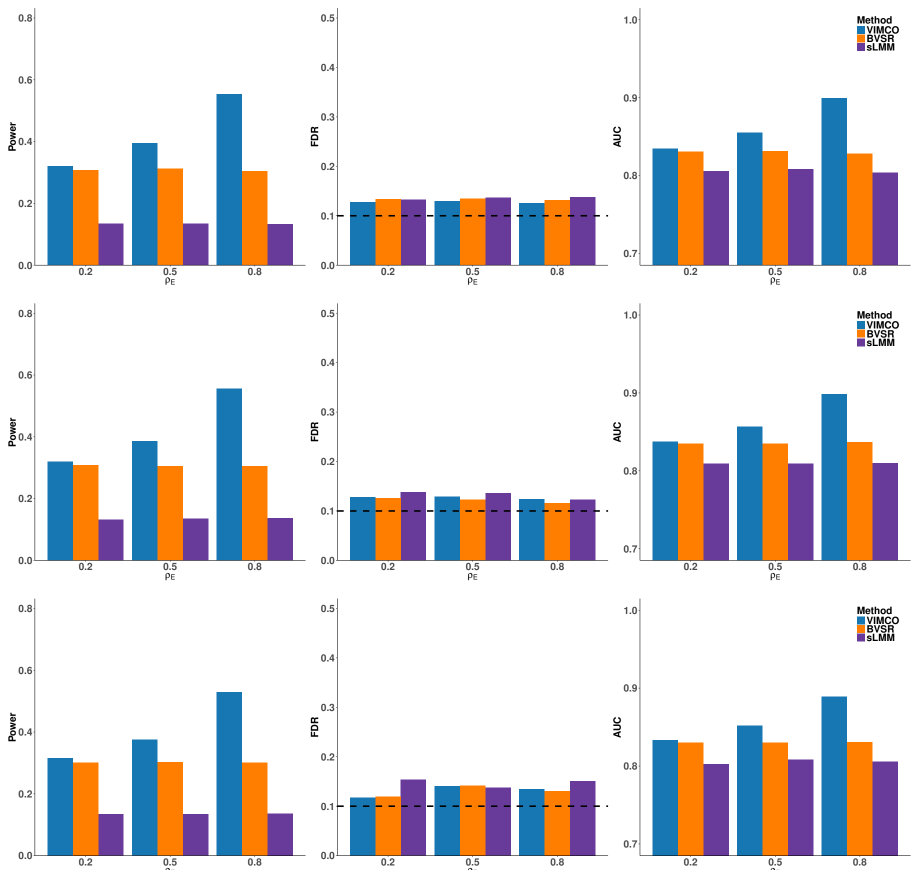
Supplementary Figure B4: Power, FDR and AUC for evaluating $H_{0b} : \boldsymbol{\beta}_j = 0$, for simulations with $\rho_x = 0.2$. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



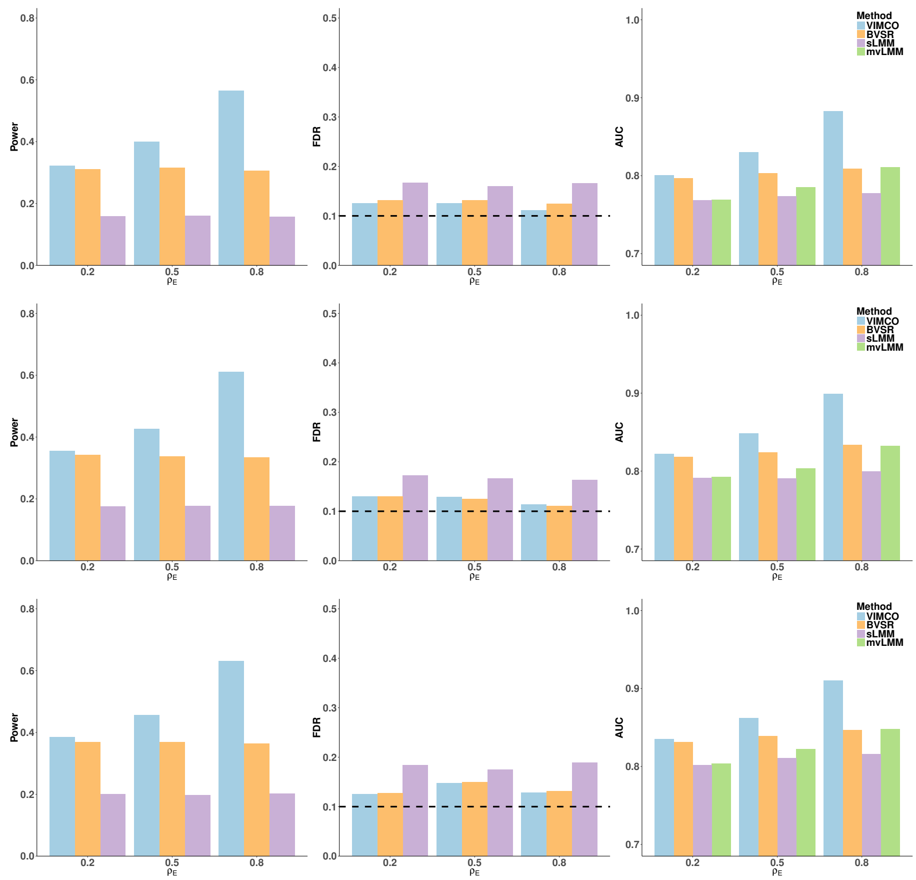
Supplementary Figure B5: Power, FDR and AUC for evaluating $H_{0b} : \boldsymbol{\beta}_j = 0$, for simulations with $\rho_x = 0.5$. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



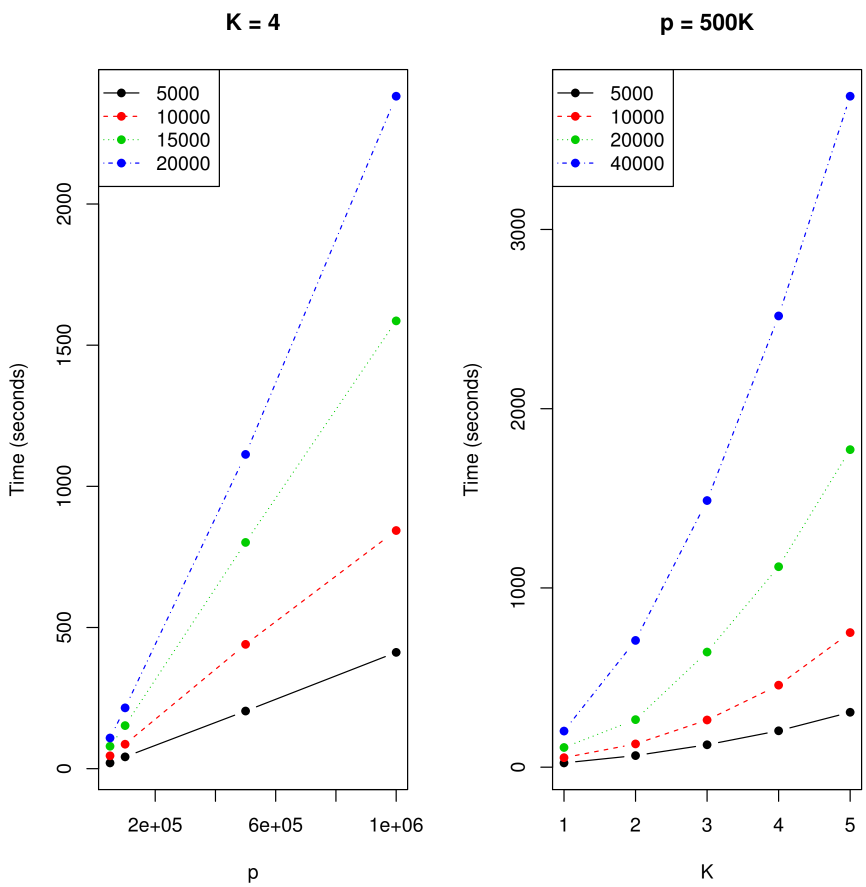
Supplementary Figure B6: Power, FDR and AUC for evaluating $H_{0b} : \boldsymbol{\beta}_j = 0$, for simulations with $\rho_x = 0.8$. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



Supplementary Figure B7: Power, FDR and AUC for evaluating $H_{0a} : \beta_{jk} = 0$, for simulations where the genotypes were sampled from the NFBC1966 study. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.

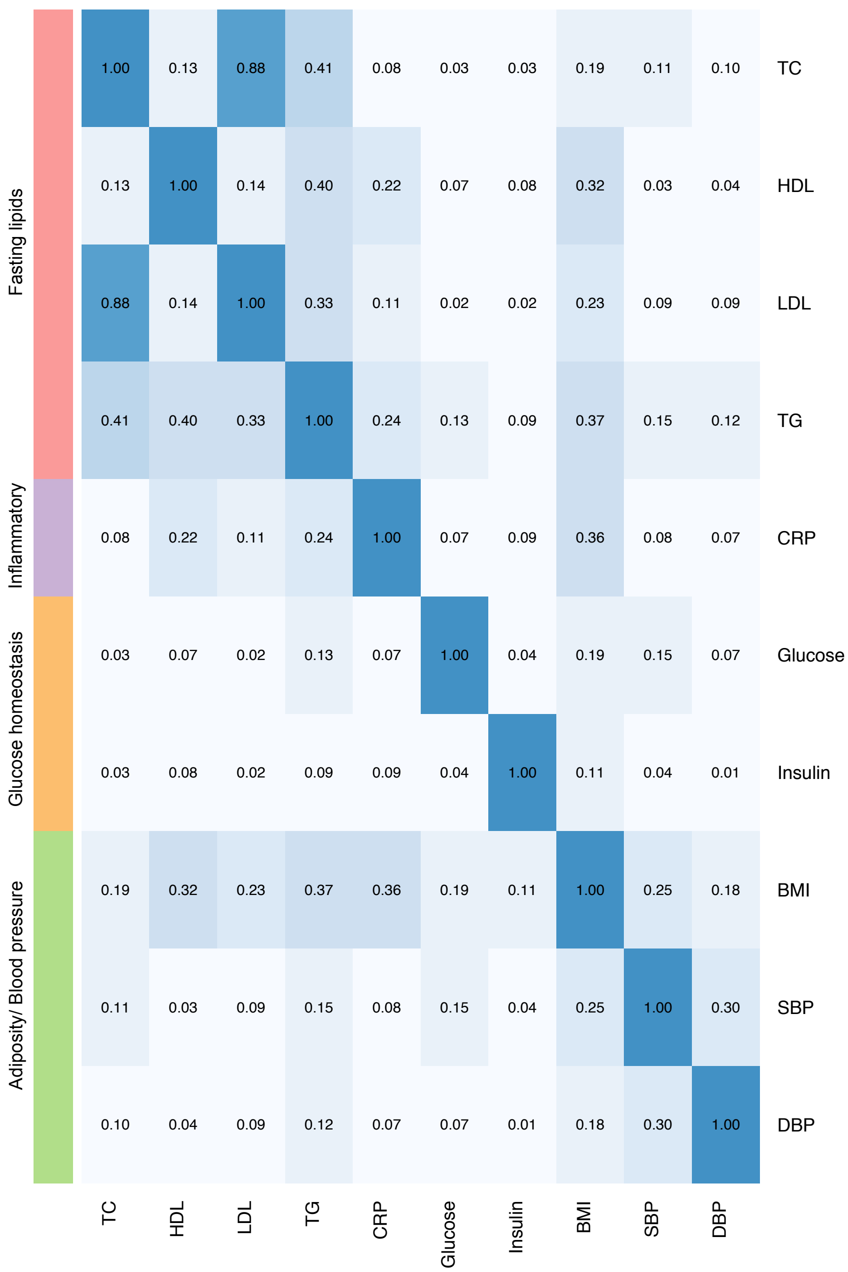


Supplementary Figure B8: Power, FDR and AUC for evaluating $H_{0b} : \boldsymbol{\beta}_j = 0$, for simulations where the genotypes were sampled from the NFBC1966 study. The top, middle and bottom panels are for $g = 0, 0.15, 0.3$ respectively.



Supplementary Figure B9: Average running times (CPU seconds) of 10 iterations with respect to different number of SNPs p (the left subfigure) and traits K (the right subfigure). In the two subfigures, the four lines denotes for different number of sample sizes.

C1 Additional Results in the NFBC1966 Study



Supplementary Figure C1: Heatmap shows the absolute pairwise correlation among the 10 traits in the NFBC1966 study.

Supplementary Table C1: SNPs identified by VIMCO, BVSR and sLMM to be associated with the four lipid traits from the NFBC1966 study. For VIMCO and BVSR, we report SNPs while controlling the global FDR at 0.1 (local FDRs are shown). For sLMM, we report SNPs with p-values the $< 1.25E-8$ (p-values are shown). For SNPs identified by either VIMCO, BVSR or sLMM, we also report the corresponding p-values from mvLMM. The null hypothesis tested by VIMCO, BVSR and sLMM (H_{0a}: $\beta_{jk} = 0$) is distinct from that tested by mvLMM (H_{0b}:$\boldsymbol{\beta}_j = \mathbf{0}$).

	H_{0a}												H_{0b}
	VIMCO (ldfr)				BVSR (lfdr)				sLMM (p-value)				mvLMM
SNP	TC	HDL	LDL	TG	TC	HDL	LDL	TG	TC	HDL	LDL	TG	p-value
rs2149038	1.0E-08		1.0E-08				3.7E-02						9.1E-05
rs2479394	1.0E-08		1.0E-08				2.0E-05						9.5E-02
rs2479418	1.0E-08		1.0E-08				1.0E-08						3.5E-03
rs611917	1.0E-08		1.0E-08		8.7E-02		1.0E-08						1.6E-06
rs14000	1.0E-08		1.0E-08				7.8E-04						8.7E-02
rs4844614	1.0E-08		1.0E-08				2.2E-03						8.8E-06
rs2802955	1.0E-08		1.0E-08				3.3E-01						3.6E-04
rs2031373					1.2E-01								1.2E-05
rs6750540	7.2E-01												5.1E-05
rs693	1.0E-08		1.0E-08				5.3E-05		3.1E-09		7.5E-10		1.4E-08
rs780094				4.7E-01				5.6E-04				2.8E-09	1.1E-07
rs3923037					8.0E-04								3.2E-07
rs1713222					6.8E-03								1.8E-04
rs2255831			6.5E-01										5.6E-05
rs12179536								2.4E-01					2.5E-05
rs12670798	1.0E-08		1.0E-08				2.5E-01						5.3E-05
rs10953541	1.0E-08		1.0E-08				2.8E-01						1.2E-03
rs7826963	1.1E-01												6.8E-06
rs10096633		2.0E-04		1.0E-08		2.5E-01		1.7E-02					5.6E-09
rs1863593			6.0E-01										1.6E-04
rs945559					3.6E-01								2.5E-04
rs174450					3.4E-01								1.4E-05
rs7120118						6.4E-03							1.5E-05
rs1532085	5.5E-07	1.0E-08				1.1E-08				4.0E-13			3.6E-17
rs1532624	1.0E-08	1.0E-08				1.0E-08				4.7E-26			1.4E-25
rs7499892						3.5E-03				3.2E-19			1.5E-21
rs255049						7.5E-03				2.8E-09			2.9E-08
rs9989419										4.3E-10			9.5E-10
rs1877031					1.7E-01								4.1E-05
rs11668477	1.0E-08		1.0E-08		4.4E-03		2.5E-05				3.9E-09		4.0E-07
rs4803750	1.0E-08		1.0E-08				1.4E-01						6.9E-06
rs157580	1.0E-08		1.0E-08		9.9E-03		1.2E-03						1.1E-07
rs7254919			7.0E-01										6.1E-04
rs1054623						3.0E-01							1.4E-04
rs846856							2.5E-01						1.5E-04
rs1800961						1.1E-01				3.3E-09			3.4E-07
rs2837324	2.7E-02												1.9E-07

C2 Additional Results in the SINDI Study

Supplementary Figure C2: Heatmap shows the absolute pairwise correlation among the 3 eye measurements in the SINDI study.

Supplementary Table C2: SNPs identified by VIMCO, BVSR and sLMM to be associated with the eye outcomes in the SINDI study. For VIMCO and BVSR, we report SNPs while controlling the global FDR at 0.1 (local FDRs are shown). For sLMM, we report SNPs with p-value $< 1.67E-8$ (p-values are shown). For SNPs identified by either VIMCO, BVSR or sLMM, we also report the corresponding p-values from mvLMM. The null hypothesis tested by VIMCO, BVSR and sLMM (H_{0a}: $\beta_{jk} = 0$) is distinct from that tested by mvLMM (H_{0b}:$\boldsymbol{\beta}_j = \mathbf{0}$).

	H_{0a}									H_{0b}
	VIMCO (ldfr)			BVSR (lfdr)			sLMM (p-value)			mvLMM
SNP	BP/IOP	CCT	VCDR	BP/IOP	CCT	VCDR	BP/IOP	CCT	VCDR	p-value
rs12074848		2.8E-02			2.1E-02					5.8E-07
rs12447690		4.8E-03			6.1E-03			5.5E-09		1.5E-07