*Manuscript: March 4, 2016*

# SUPPLEMENT TO "BAYESIAN LARGE-SCALE MULTIPLE REGRESSION WITH SUMMARY STATISTICS FROM GENOME-WIDE ASSOCIATION STUDIES"

By Xiang Zhu and Matthew Stephens

*University of Chicago*

## APPENDIX A: PROOFS

We first provide a summary of assumptions made in the main text.

- The centered genotypes of $n$ individuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{\text{i.i.d.}}{\sim} \boldsymbol{x}$, where $\boldsymbol{x} := (x_1, \ldots, x_p)^\intercal$, $\mathrm{E}(\boldsymbol{x}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{x}) = \Sigma_x = \mathrm{diag}(\boldsymbol{\sigma}_x) R \mathrm{diag}(\boldsymbol{\sigma}_x)$ is finite, $|\mathrm{E}(x_j x_k x_l x_m)| < \infty$ for any $j, k, l, m \in [p]$. Note that these moment assumptions are satisfied by default for genotype data.
- The additive errors $\epsilon_1, \ldots, \epsilon_n \overset{\text{i.i.d.}}{\sim} \epsilon$, where $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \tau^{-1} < \infty$.
- The centered phenotypes of $n$ individuals $y_1, \ldots, y_n \overset{\text{i.i.d.}}{\sim} y$, where $y = \boldsymbol{x}^\intercal \boldsymbol{\beta} + \epsilon$. For each individual $i \in [n]$, $y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \epsilon_i$, where $\boldsymbol{x}_i$, $\boldsymbol{\beta}$ and $\epsilon_i$ are mutually independent.

**A.1. Proof of Proposition 2.1.** Notice that $\widehat{\boldsymbol{\beta}} = D^{-2} X^\intercal \mathbf{y}$, $\widehat{S} = \sqrt{n^{-1} \mathbf{y}^\intercal \mathbf{y}} \cdot D^{-1}$ and

$$\begin{aligned} -2 \log L_{\mathsf{rss}}(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}, \widehat{S}, \widehat{R}) &= p \log(2\pi) + \log |\widehat{S}\widehat{R}\widehat{S}| + \widehat{\boldsymbol{\beta}}^\intercal (\widehat{S}\widehat{R}\widehat{S})^{-1} \widehat{\boldsymbol{\beta}} - 2\widehat{\boldsymbol{\beta}}^\intercal \widehat{S}^{-2} \boldsymbol{\beta} + \boldsymbol{\beta}^\intercal \widehat{S}^{-1} \widehat{R} \widehat{S}^{-1} \boldsymbol{\beta}, \\ -2 \log L_{\mathsf{mvn}}(\boldsymbol{\beta}; \mathbf{y}, X, \tau) &= p \log(2\pi\tau^{-1}) + \tau \mathbf{y}^\intercal \mathbf{y} - 2\tau \mathbf{y}^\intercal X \boldsymbol{\beta} + \tau \boldsymbol{\beta}^\intercal X^\intercal X \boldsymbol{\beta}. \end{aligned}$$

If $\tau^{-1} = n^{-1} \mathbf{y}^\intercal \mathbf{y}$ and $\widehat{R} = \widehat{R}^{\mathsf{sam}}$, then $\widehat{S}^{-2} \widehat{\boldsymbol{\beta}} = \tau X^\intercal \mathbf{y}$ and $\widehat{S}^{-1} \widehat{R} \widehat{S}^{-1} = \tau X^\intercal X$, further implying that

$$-2[\log L(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}, S, R) - \log L(\boldsymbol{\beta}; \mathbf{y}, X)] = \log |D^{-1} \widehat{R} D^{-1}| - \tau \mathbf{y}^\intercal [I - X(X^\intercal X) X^\intercal] \mathbf{y}. \tag{A.1}$$

**A.2. Proof of Proposition 2.2.** First define the statistic $T_n \in \mathbb{R}^{2p \times 1}$,

$$T_n := n^{-1} \left( \textstyle\sum_{i=1}^n x_{i1} y_i, \ldots, \sum_{i=1}^n x_{ip} y_i, \sum_{i=1}^n x_{i1}^2, \ldots, \sum_{i=1}^n x_{ip}^2 \right)^\intercal. \tag{A.2}$$

The asymptotic distribution of $T_n$ is given by the multivariate Central Limit Theorem

$$\sqrt{n}(T_n - \boldsymbol{\mu}_T) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_T), \tag{A.3}$$

where $\boldsymbol{\mu}_T := \mathrm{E}(\mathbf{t})$, $\Sigma_T := \mathrm{Var}(\mathbf{t})$ and $\mathbf{t} := (x_1 y, \ldots, x_p y, x_1^2, \ldots, x_p^2)^\intercal$. Note that $\Sigma_T$ has finite entries because $\tau^{-1}$, $\Sigma_x$ and $\mathrm{E}(x_j x_k x_l x_m)$ are finite.

Next, for any $\boldsymbol{\xi} \in \mathbb{R}^{2p \times 1}$, define the following function $g(\boldsymbol{\xi}) \in \mathbb{R}^{p \times 1}$:

$$g(\boldsymbol{\xi}) := (\xi_1 / \xi_{p+1}, \ldots, \xi_p / \xi_{2p})^\intercal. \tag{A.4}$$

Note that $g(T_n) = \widehat{\boldsymbol{\beta}}$ and $g(\boldsymbol{\mu}_T) = \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x) \boldsymbol{\mu}_{xy} = S R S^{-1} \boldsymbol{\beta}$.

Use the multivariate Delta method to show that

$$\sqrt{n}(g(T_n) - g(\boldsymbol{\mu}_T)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \nabla^\intercal g(\boldsymbol{\mu}_T) \Sigma_T \nabla g(\boldsymbol{\mu}_T)) \tag{A.5}$$

where $\nabla g(\boldsymbol{\mu}_T) \in \mathbb{R}^{2p \times p}$ is the gradient matrix of $g$ at $\boldsymbol{\mu}_T$. A straightforward calculation yields that

$$\nabla^\intercal g(\boldsymbol{\mu}_T) \Sigma_T \nabla g(\boldsymbol{\mu}_T) = \sigma_y^2 \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)(R + \Delta(\mathbf{c})) \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x) = nS(R + \Delta(\mathbf{c}))S. \tag{A.6}$$

The explicit form of $\Delta(\mathbf{c})$ is given by

$$\Delta(\mathbf{c}) := \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x) \cdot [G_1(\mathbf{c}) + G_2(\mathbf{c}) + G_2^\intercal(\mathbf{c}) + G_3(\mathbf{c})] \cdot \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x), \tag{A.7}$$

where functions $G_i(\mathbf{c}) : \mathbb{R}^{p\times 1} \mapsto \mathbb{R}^{p\times p}$ are defined as follows:

$$\begin{aligned}
G_1(\mathbf{c}) &:= -(\mathbf{c}^\intercal R^{-1}\mathbf{c})\Sigma_x - \mathrm{diag}(\boldsymbol{\sigma}_x)\mathbf{c}\mathbf{c}^\intercal\mathrm{diag}(\boldsymbol{\sigma}_x) + \mathrm{E}[(\boldsymbol{x}^\intercal\mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)R^{-1}\mathbf{c})^2\boldsymbol{x}\boldsymbol{x}^\intercal], \\
G_2(\mathbf{c}) &:= \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\mathrm{diag}(\mathbf{c})W(\mathbf{c}), \;\; [W(\mathbf{c})]_{ij} := \sigma_{x,i}\sigma_{x,j}^2 c_i - \mathbf{c}^\intercal R^{-1}\mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\mathrm{E}(x_i x_j^2\boldsymbol{x}), \\
G_3(\mathbf{c}) &:= \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\ \mathrm{diag}(\mathbf{c})\ \Sigma_{xx}\ \mathrm{diag}\ (\mathbf{c})\ \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x), \;\; [\Sigma_{xx}]_{ij} := \mathrm{Cov}(x_i^2, x_j^2).
\end{aligned}$$

Notice that $G_i(\mathbf{c})$ are continuous functions of $\mathbf{c}$, $G_i(\mathbf{0}) = \mathbf{0}$, and $G_i(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$ for $i = 1, 2, 3$.

**A.3. Proof of Proposition 2.3.** First note that

$$\begin{aligned}
&\log\mathcal{N}(\widehat{\boldsymbol{\beta}}; SRS^{-1}\boldsymbol{\beta}, SRS) - \log\mathcal{N}(\widehat{\boldsymbol{\beta}}; SRS^{-1}\boldsymbol{\beta}, n^{-1}\Sigma) \\
= \;& \frac{1}{2}\left\{\log|R + \Delta(\mathbf{c})| - \log|R| + \sigma_y^{-2}\boldsymbol{\lambda}^\intercal\mathrm{diag}(\boldsymbol{\sigma}_x)[(R + \Delta(\mathbf{c}))^{-1} - R^{-1}]\mathrm{diag}(\boldsymbol{\sigma}_x)\boldsymbol{\lambda}\right\},
\end{aligned} \tag{A.8}$$

where $\boldsymbol{\lambda} := \sqrt{n}(\widehat{\boldsymbol{\beta}} - SRS^{-1}\boldsymbol{\beta})$. Since the determinant and inverse of a matrix are both continuous, we invoke Proposition 2.2, that is, $\boldsymbol{\lambda} = \mathcal{O}_p(1)$ and $\Delta(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$, to complete the proof.

**A.4. Proof of Proposition 3.1.** Since the matrix $X$ is column-centered,

$$V(X\boldsymbol{\beta}) = n^{-1}\textstyle\sum_{i=1}^n(\boldsymbol{x}_i^\intercal\boldsymbol{\beta})^2 = n^{-1}\mathrm{trace}[(X\boldsymbol{\beta})(X\boldsymbol{\beta})^\intercal] = n^{-1}\boldsymbol{\beta}^\intercal X^\intercal X\boldsymbol{\beta}, \tag{A.9}$$

and therefore,

$$\mathrm{E}[V(X\boldsymbol{\beta})|S, X] = \boldsymbol{\mu}_\beta^\intercal \cdot (n^{-1}X^\intercal X) \cdot \boldsymbol{\mu}_\beta + \mathrm{trace}[(n^{-1}X^\intercal X) \cdot \Sigma_\beta], \tag{A.10}$$

where $\boldsymbol{\mu}_\beta := \mathrm{E}(\boldsymbol{\beta}|S) = \mathbf{0}$ and $\Sigma_\beta := \mathrm{Var}(\boldsymbol{\beta}|S) = (\pi\sigma_B^2 + \sigma_P^2) \cdot I_p$. Hence,

$$\mathrm{E}[V(X\boldsymbol{\beta})] = \mathrm{E}[\mathrm{E}[V(X\boldsymbol{\beta})|S, X]] = (\pi\sigma_B^2 + \sigma_P^2) \cdot \textstyle\sum_{j=1}^p \mathrm{E}[V(X_j)] = \dfrac{h}{\sum_{j=1}^p n^{-1}s_j^{-2}} \cdot \sum_{j=1}^p \mathrm{E}[V(X_j)].$$

From the definition of $\{s_j\}$ we can see that $\mathrm{E}[V(X_j)] = n^{-1}s_j^{-2}\mathrm{E}[V(\mathbf{y})]$, implying that

$$\mathrm{E}[V(X\boldsymbol{\beta})] = \frac{h}{\sum_{j=1}^p n^{-1}s_j^{-2}} \cdot \textstyle\sum_{j=1}^p n^{-1}s_j^{-2}\mathrm{E}[V(\mathbf{y})] = h \cdot \mathrm{E}[V(\mathbf{y})]. \tag{A.11}$$

## APPENDIX B: DETAILS OF POSTERIOR SAMPLING SCHEME

We describe the Markov chain Monte Carlo (MCMC) algorithms in terms of $\{S, R\}$, and then replace the unknown $\{S, R\}$ with their estimates $\{\widehat{S}, \widehat{R}\}$ in practice. This is similar to the likelihood derivation and prior specification in the main text.

**B.1. Rank-based strategy.** When locally updating the SNP-specific parameters (e.g. genetic effect $\beta_j$ and sparsity indicator $\gamma_j$ for each SNP $j$) in the MCMC algorithms, we allocate more computational resources to SNPs with larger marginal association signals, using the rank-based strategy (Guan and Stephens, 2011). In particular, we first rank all the variants based on the single-SNP $p$-values and draw one SNP to update according to some probability distributions with decreasing probability. In our current implementation, we use a mixture distribution $q_p = 0.3u_p + 0.7g_p$, where $u_p$ is a discrete uniform distribution and $g_p$ is a geometric distribution truncated to $1, \ldots, p$ with its parameter chosen to give a mean of 2000.

Based on $q_p$, we introduce $Q(\cdot|\boldsymbol{\gamma})$, a proposal for the indicator $\boldsymbol{\gamma}$. To propose a new value $\boldsymbol{\gamma}^*$ given the current value $\boldsymbol{\gamma}$, we start by setting $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}$ and then randomly choose one of the following:

1. With probability $P_a$, draw SNP $r$ according to $q_p$ until $\gamma_r = 0$ and set $\gamma_r^* = 1$.
2. With probability $P_r$, draw SNP $r$ uniformly from $\{j : \gamma_j = 1\}$ and set $\gamma_r^* = 0$.
3. With probability $P_e$, sample two SNPs by the above two steps and switch their indicators.

The default setting in our software is $P_a = P_r = 0.4, P_e = 0.2$.

**B.2. BVSR prior.** For RSS with BVSR prior, we use Metropolis-Hastings (MH) algorithm to obtain posterior samples of $(\gamma, \pi, h)$ on the product space of $\{0,1\}^p \times (0,1) \times (0,1)$,

$$p(\gamma, \pi, h|\widehat{\beta}, S, R) \propto p(\widehat{\beta}|S, R, \gamma, \pi, h)p(\gamma|\pi)p(\pi)p(h). \tag{B.1}$$

Here we are exploiting the fact that $\beta$ can be integrated out analytically to compute $p(\widehat{\beta}|S, R, \gamma, \pi, h)$:

$$\widehat{\beta}|S, R, \gamma, \pi, h \sim \mathcal{N}(\mathbf{0}, SRS + \sigma_B^2 M_\gamma M_\gamma^\intercal), \tag{B.2}$$

where $M := SRS^{-1}$ and $M_\gamma$ denotes the sub-matrix of $M$ restricted to those columns $j$ for which $\gamma_j = 1$. We update $\gamma$ using the rank-based proposal $Q(\cdot|\gamma)$. We update $\log \pi$ by adding a random number from $\mathcal{U}(-0.05, 0.05)$ to the current value, and update $h$ by adding a random number from $\mathcal{U}(-0.1, 0.1)$ to the current value. New values of $\log \pi$ and $h$ outside boundaries are reflected back.

For each simulated posterior draw of $(\gamma, \pi, h)$, we sample $\beta$ according to its conditional distributions given $(\gamma, \pi, h)$ and $(\widehat{\beta}, S, R)$:

$$\beta_\gamma|\widehat{\beta}, S, R, \gamma, \pi, h \sim \mathcal{N}(\mu, \Omega^{-1}), \tag{B.3}$$

$$\beta_{-\gamma}|\widehat{\beta}, S, R, \gamma, \pi, h \sim \delta_0, \tag{B.4}$$

where $\beta_\gamma$ and $\beta_{-\gamma}$ denote the subsets of $\beta$ corresponding to the entries that $\gamma_j = 1$ and 0 respectively, $\delta_0$ denotes the point mass at zero and,

$$\Omega := M_\gamma^\intercal (SRS)^{-1} M_\gamma + \sigma_B^{-2}(\gamma, \pi, h) I_{|\gamma|}, \tag{B.5}$$

$$\mu := \Omega^{-1} M_\gamma^\intercal (SRS)^{-1} \widehat{\beta}. \tag{B.6}$$

The marginal likelihood (B.2), up to some constant, can be written in terms of $(\Omega, \mu)$,

$$p(\widehat{\beta}|S, R, \gamma, \pi, h) \propto \sigma_B^{-|\gamma|} |\Omega|^{-1/2} \exp\{\mu^\intercal \mathbf{q}_\gamma / 2\}, \tag{B.7}$$

where $\mathbf{q}_\gamma$ denotes the subset of $\mathbf{q} := S^{-1}\beta$ corresponding to the entries that $\gamma_j = 1$. The matrix computation in a single step of the MCMC algorithm above involves one Cholesky decomposition of $\Omega$ and three triangular linear systems. Hence, the computational cost for each iteration of MCMC is $\mathcal{O}(|\gamma|^3 + 3|\gamma|^2)$, where $|\gamma|$ denotes the number of non-zero entries in $\gamma$.

To improve precision, we can use Rao-Blackwellized estimates. For SPIP, we have

$$\Pr(\gamma_j = 1|\widehat{\beta}, S, R) = \mathrm{E}(\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j})) \approx M^{-1} \textstyle\sum_{i=1}^M \Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j}^{(i)})$$

where $\xi_{-j}$ stands for $\{\beta_{-j}, \gamma_{-j}, \pi, h\}$, $\gamma_{-j}$ and $\beta_{-j}$ denote the vectors $\gamma$ and $\beta$ excluding the $j$th coordinate and $\xi_{-j}^{(i)}$ denotes the $i$th MCMC sample from the posterior distribution of $\xi_{-j}$. For the posterior mean of the multiple-SNP effect at SNP $j$, we have

$$\mathrm{E}(\beta_j|\widehat{\beta}, S, R) = \mathrm{E}(\mathrm{E}(\beta_j|\widehat{\beta}, S, R, \xi_{-j})) \approx M^{-1} \textstyle\sum_{i=1}^M \mathrm{E}(\beta_j|\widehat{\beta}, S, R, \gamma_j = 1, \xi_{-j}^{(i)}) \Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j}^{(i)}).$$

To obtain the Rao-Blackwellized estimates, we only need $p(\gamma_j|\widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\xi}_{-j})$ and $p(\beta_j|\widehat{\boldsymbol{\beta}}, S, R, \gamma_j, \boldsymbol{\xi}_{-j})$:

$$\begin{aligned}
\frac{\Pr(\gamma_j = 1|\widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\xi}_{-j})}{\Pr(\gamma_j = 0|\widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\xi}_{-j})} &= \frac{\pi}{1-\pi}\sqrt{\frac{s_j^2}{s_j^2+\sigma_B^2}}\exp\left\{\frac{1}{2(\sigma_B^{-2}+s_j^{-2})}\left(\frac{\hat{\beta}_j}{s_j^2}-\sum_{i\neq j}\frac{r_{ij}\beta_i}{s_i s_j}\right)^2\right\} \\
\beta_j|\widehat{\boldsymbol{\beta}}, S, R, \gamma_j = 1, \boldsymbol{\xi}_{-j} &\sim \mathcal{N}\left(\frac{1}{\sigma_B^{-2}+s_j^{-2}}\left(\frac{\hat{\beta}_j}{s_j^2}-\sum_{i\neq j}\frac{r_{ij}\beta_i}{s_i s_j}\right), \frac{1}{\sigma_B^{-2}+s_j^{-2}}\right) \\
\beta_j|\widehat{\boldsymbol{\beta}}, S, R, \gamma_j = 0, \boldsymbol{\xi}_{-j} &\sim \delta_0
\end{aligned}$$

where $r_{ij}$ is the $(i,j)$-th entry of $R$.

**B.3. BSLMM prior.** We propose a component-wise MCMC algorithm for RSS with BSLMM prior. First, we re-parameterize the multiple-SNP effect sizes $\beta_j$ as follows

$$\beta_j|\gamma_j = 1, \pi, h, \rho, S = \sqrt{\sigma_B^2+\sigma_P^2}\cdot\tilde{\beta}_j \tag{B.8}$$

$$\beta_j|\gamma_j = 0, \pi, h, \rho, S = \sigma_P\cdot\tilde{\beta}_j \tag{B.9}$$

where the standardized effect sizes $\tilde{\beta}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, for $j \in \{1,\ldots,p\}$. Equivalently,

$$\boldsymbol{\beta} = B\widetilde{\boldsymbol{\beta}},\ \ \widetilde{\boldsymbol{\beta}} \sim \mathcal{N}(\mathbf{0}, I_p) \tag{B.10}$$

where the scaling matrix $B$ is diagonal with the $j$th diagonal $b_j$ defined as

$$b_j = \sigma_P\mathbf{1}\{\gamma_j = 0\} + \sqrt{\sigma_B^2+\sigma_P^2}\mathbf{1}\{\gamma_j = 1\}. \tag{B.11}$$

The new parameterization could help speed up the convergence of MCMC, since $\tilde{\boldsymbol{\beta}}$ are independent with $(\boldsymbol{\gamma}, \pi, h, \rho)$ *a priori*. We then draw posterior samples of $(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \pi, h, \rho)$ iteratively.

- Given $(\widetilde{\boldsymbol{\beta}}, \pi, h, \rho)$, we update $\boldsymbol{\gamma}$ by a standard MH algorithm, where the proposal is $Q(\cdot|\boldsymbol{\gamma})$.
- Given $(\boldsymbol{\gamma}, \pi, h, \rho)$, we update $\widetilde{\boldsymbol{\beta}}$ by a mixture of global and local moves. With probability $P_{\text{g}}$, we draw a new value of $\widetilde{\boldsymbol{\beta}}$ from its full conditional,

  $$\widetilde{\boldsymbol{\beta}}|\widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\gamma}, \pi, h, \rho \sim \mathcal{N}((BS^{-1}RS^{-1}B+I)^{-1}BS^{-2}\widehat{\boldsymbol{\beta}}, (BS^{-1}RS^{-1}B+I)^{-1}). \tag{B.12}$$

  With probability $1-P_{\text{g}}$, we randomly pick a SNP $j$ according to the distribution $q_p$ and draw $\tilde{\beta}_j$ from its full conditional

  $$\tilde{\beta}_j|\widehat{\boldsymbol{\beta}}, S, R, \tilde{\boldsymbol{\beta}}_{-j}, \boldsymbol{\gamma}, \pi, h, \rho \sim \mathcal{N}\left(\frac{b_j s_j \ell_j}{s_j^2+b_j^2}, \frac{s_j^2}{s_j^2+b_j^2}\right),\ \ell_j := \frac{\hat{\beta}_j}{s_j}-\sum_{i\neq j}\frac{r_{ij}b_i\tilde{\beta}_i}{s_i}. \tag{B.13}$$

- Given $(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, h, \rho)$, we update $\pi$ by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log((\pi-p^{-1})/(1-\pi))$.
- Given $(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \pi, \rho)$, we update $h$ by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log(h/(1-h))$.
- Given $(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \pi, h)$, we update $\rho$ by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log(\rho/(1-\rho))$.

The most computationally intensive step is drawing $\widetilde{\boldsymbol{\beta}}$ from a $p$-dimensional multivariate normal distribution (B.12). For each draw, one Cholesky decomposition of $BS^{-1}RS^{-1}B+I$ and two triangular linear systems are required. Since matrix $R$ is banded with some bandwidth $w$ (Wen and Stephens, 2010), the matrix $BS^{-1}RS^{-1}B+I$ also has the same bandwidth and therefore, the per-iteration cost of the algorithm above is at most $\mathcal{O}(pw^2+2p^2)$. For all the simulations, we set $P_{\text{g}} = 0.05$. For the analysis of adult height data, we set $P_{\text{g}} = 0.001$ (the default value in our software).

**B.4. Small world proposal.** To improve the convergence rate of the MCMC schemes, we use the "small-world" proposal (Guan and Krone, 2007) as an add-on for every Metropolis step in our main algorithms above. Specifically, with probability 0.3 in each iteration, a long-range move is made by compounding randomly many (from 2 to 20) local proposals.

## APPENDIX C: CONNECTION WITH LD SCORE REGRESSION

The LD score regression model (Bulik-Sullivan et al., 2015) is given by,

$$\mathsf{E}(\chi_j^2|\ell_j) = nh^2\ell_j/p + na + 1, \tag{C.1}$$

where $n$ is the sample size, $p$ is the number of SNPs, $h^2/p$ is the heritability per SNP, $a$ is the contribution of confounding biases per individual, $\chi_j^2 := (\widehat{\beta}_j/s_j)^2$ is the single-SNP association $\chi^2$ statistic and $\ell_j := \sum_{k=1}^p r_{jk}^2$ is the "LD score" of SNP $j$ ($r_{jk}$ is the pairwise LD between SNP $j$ and $k$).

To draw the connection between the LD score regression and RSS, we consider

$$\widehat{\boldsymbol{\beta}}|S,R,\boldsymbol{\beta} \sim \mathcal{N}(SRS^{-1}\boldsymbol{\beta}, SRS + na\cdot S^2), \tag{C.2}$$

which is a generalization of RSS accounting for possible over-dispersion in real data. When $a=0$, model (C.2) becomes the original RSS. Let $\mathbf{z} = (\mathsf{z}_1,\ldots,\mathsf{z}_p)^\intercal$, where $\mathsf{z}_j := \widehat{\beta}_j/s_j$ is the single-SNP $z$-score of SNP $j$ and $z_j^2 = \chi_j^2$. Noting that $\mathbf{z} = S^{-1}\boldsymbol{\beta}$, we rewrite (C.2) in terms of $z$-scores,

$$\mathbf{z}|S,R,\boldsymbol{\beta} \sim \mathcal{N}(RS^{-1}\boldsymbol{\beta}, R + na\cdot I_p). \tag{C.3}$$

Next, we specify the following prior on $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}|S,R) = \textstyle\prod_{j=1}^p p(\beta_j|S,R),\ \mathrm{E}(\beta_j|S,R) = 0,\ \mathrm{Var}(\beta_j|S,R) = nh^2 s_j^2/p. \tag{C.4}$$

Since $s_j := (\sqrt{n}\sigma_{x,j})^{-1}\sigma_y$, the the prior variance of $\beta_j$ is $(p\sigma_{x,j}^2)^{-1}(h^2\sigma_y^2)$ and thus prior (C.4) does not depend on the sample size $n$.

Integrating out $\boldsymbol{\beta}$ under prior (C.4), we obtain the LD score regression model:

$$\begin{aligned}
\mathrm{E}(\mathsf{z}_j^2|S,R) &= \mathrm{E}(\mathrm{Var}(\mathsf{z}_j|S,R,\boldsymbol{\beta})) + \mathrm{E}(\mathrm{E}^2(\mathsf{z}_j|S,R,\boldsymbol{\beta})) \\
&= 1 + na + \textstyle\sum_{k=1}^p r_{jk}^2 s_k^{-2}\mathrm{E}(\beta_k^2|S,R) + \sum_{k\neq\ell} r_{jk}r_{j\ell}s_k^{-1}s_\ell^{-1}\mathrm{E}(\beta_k\beta_\ell|S,R) \\
&= 1 + na + (nh^2/p)\textstyle\sum_{k=1}^p r_{jk}^2.
\end{aligned} \tag{C.5}$$

## APPENDIX D: GWAS META-ANALYSIS

The derivation of RSS assumes that $\{\hat{\beta}_j, \hat{\sigma}_j^2\}$ are calculated from a single study sample. When there are multiple studies, this assumption requires that $\{\hat{\beta}_j, \hat{\sigma}_j^2\}$ should be the summary results of a single sample where the individual-level data from each study are pooled. However, the summary data released by most large GWAS are not obtained from the pooled sample, and are actually synthesized results based on the single-SNP summary statistics from each research group via meta-analysis (Evangelou and Ioannidis, 2013). A series of results in literature have revealed the exact/asymptotic equivalence between meta-analysis estimates and pooled-sample estimates (Olkin and Sampson, 1998; Mathew and Nordstrom, 1999; Lin and Zeng, 2010a,b). Hence, RSS can be applied to the summary statistics from GWAS meta analysis.

## REFERENCES


Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S. et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467** 832–838.

Bulik-Sullivan, B., Loh, P.-R., Finucane, H., Ripke, S., Yang, J., Psychiatric Genomics Consortium, Schizophrenia Working Group, Patterson, N., Daly, M. J., Price, A. L. and Neale, B. M. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47** 291–295.

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

Evangelou, E. and Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics* **14** 379–389.

Guan, Y. and Krone, S. (2007). Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing. *The Annals of Applied Probability* **17** 284–304.

Guan, Y. and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *The Annals of Applied Statistics* **5** 1780–1815.

Lambert, J.-C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A. L., Bis, J. C., Beecham, G. W. et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45** 1452–1458.

Lin, D. Y. and Zeng, D. (2010a). Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genetic Epidemiology* **34** 60-66.

Lin, D. and Zeng, D. (2010b). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* 321–332.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J. et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518** 197–206.

Mathew, T. and Nordstrom, K. (1999). On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics* **55** 1221–1223.

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segre, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A. et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* **44** 981.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511** 421–427.

Olkin, I. and Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* **54** 317–322.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A. et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340** 1467–1471.

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss, M., Stewart, A. F., Barbalic, M., Gieger, C. et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics* **43** 333–338.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J. et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466** 707–713.

Tobacco and Genetics Consortium (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics* **42** 441–447.

Wen, X. and Stephens, M. (2010). Using linear predictors to impute allele frequencies from summary or pooled genotype data. *The Annals of Applied Statistics* **4** 1158–1182.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Luan, J., Kutalik, Z. et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46** 1173–1186.

| Phenotype | Reference |
|---|---|
| Adult human height | Allen et al. (2010) |
| Adult human height | Wood et al. (2014) |
| Body mass index (BMI) | Locke et al. (2015) |
| High-density lipoprotein (HDL) | Teslovich et al. (2010) |
| HDL | Global Lipids Genetics Consortium (2013) |
| Low-density lipoprotein (LDL) | Teslovich et al. (2010) |
| LDL | Global Lipids Genetics Consortium (2013) |
| Total cholesterol (TC) | Teslovich et al. (2010) |
| TC | Global Lipids Genetics Consortium (2013) |
| Triglycerides (TG) | Teslovich et al. (2010) |
| TG | Global Lipids Genetics Consortium (2013) |
| Cigarettes per day | Tobacco and Genetics Consortium (2010) |
| Smoking age of onset | Tobacco and Genetics Consortium (2010) |
| Ever versus never smoked | Tobacco and Genetics Consortium (2010) |
| Current versus former smoker | Tobacco and Genetics Consortium (2010) |
| Years of educational attainment | Rietveld et al. (2013) |
| College completion or not | Rietveld et al. (2013) |
| Schizophrenia | Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) |
| Alzheimer | Lambert et al. (2013) |
| Coronary artery disease (CAD) | Schunkert et al. (2011) |
| Type 2 diabetes (T2D) | Morris et al. (2012) |

TABLE 1

*Full names of phenotypes and the corresponding references that are listed in the main text (Table 1)*

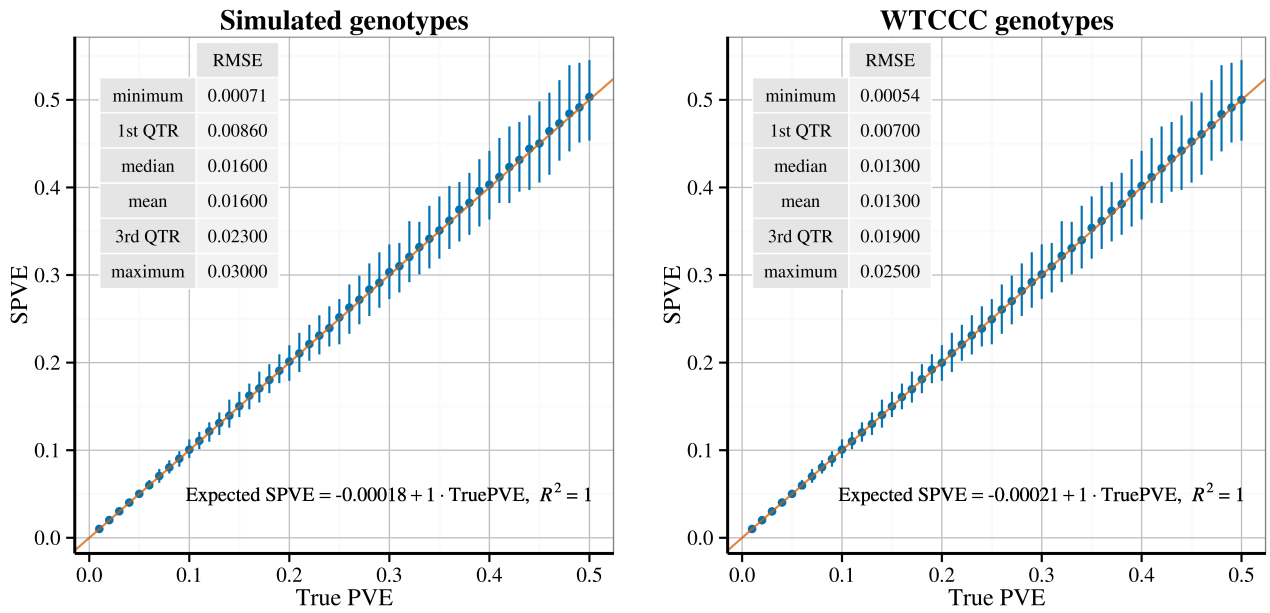


Fig 1: *Comparison of true PVE and SPVE given the true $\beta$. The simulated genotypes consist of 10,000 independent SNPs from 1000 individuals, so we set $\widehat{R}$ as identity matrix; The real genotypes are 10,000 correlated SNPs randomly drawn from Chromosome 16 (WTCCC UK Blood Service control group, 1458 individuals), and $\widehat{R}$ is estimated from WTCCC 1958 British Birth Cohort (1480 individuals) and HapMap CEU genetic maps using the shrinkage method in Wen and Stephens (2010). Solid dots indicate sample means of 200 replicates; vertical bars indicate symmetric 95% intervals; orange line indicates the reference line with intercept 0 and slope 1. The tables summarize the RMSEs between SPVE and true PVE.*