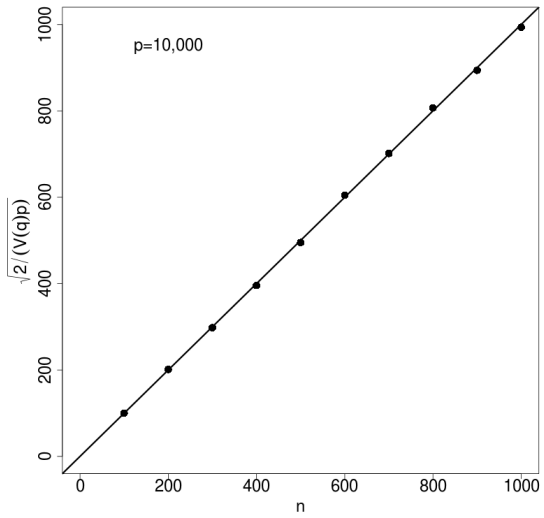
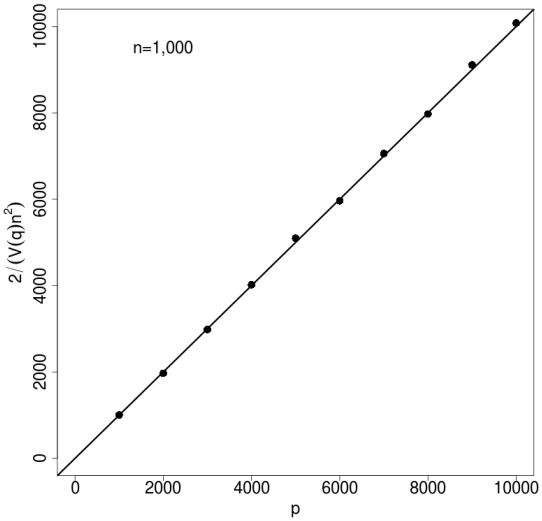
# Supplementary Material

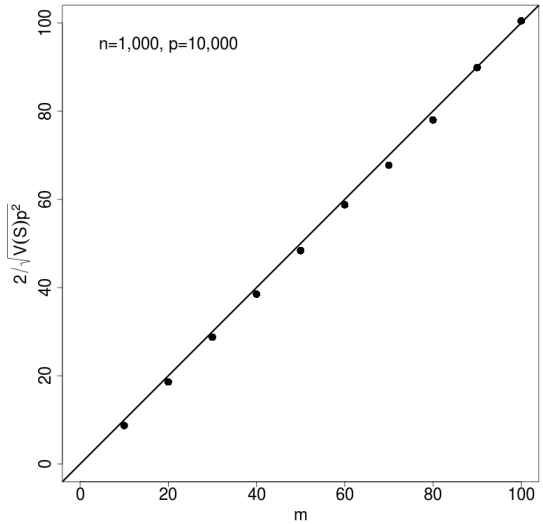
## Supplementary Figures



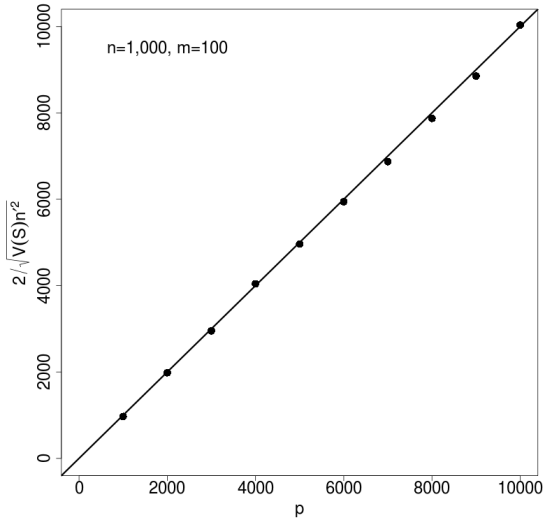

(a) V(q) as a function of n




(b) V(q) as a function of p

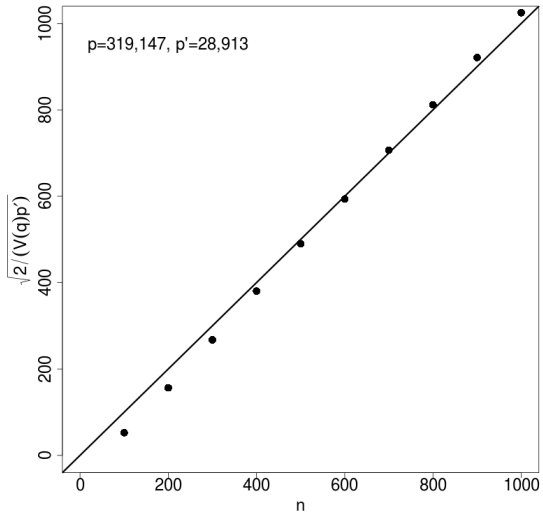



(c) V($\hat{S}$) as a function of m

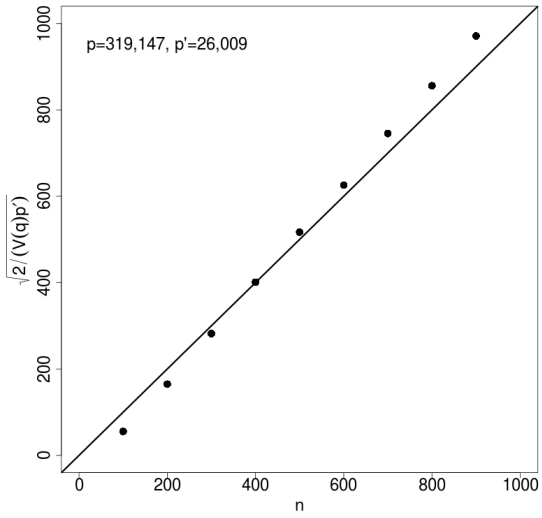



(d) V($\hat{S}$) as a function of p

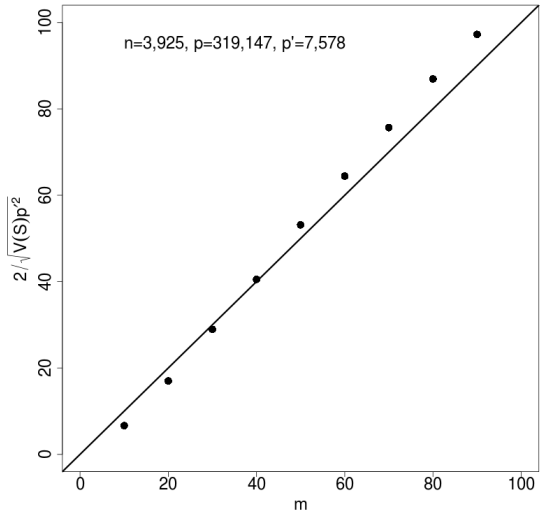
**Figure S1.** Simulations using normally distributed genotypes validate $E(V(q)) = O(2/(n^2p))$ and $E(V(\hat{S})) = O(4/(m^2p^2))$.



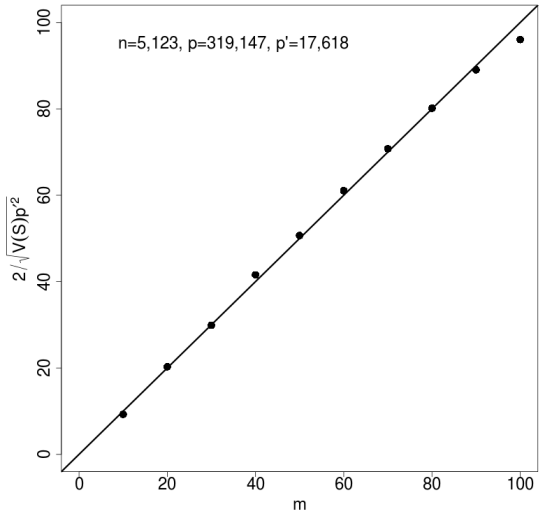

(a) V(q) vs n (Australia)




(b) V(q) vs n (Finland)

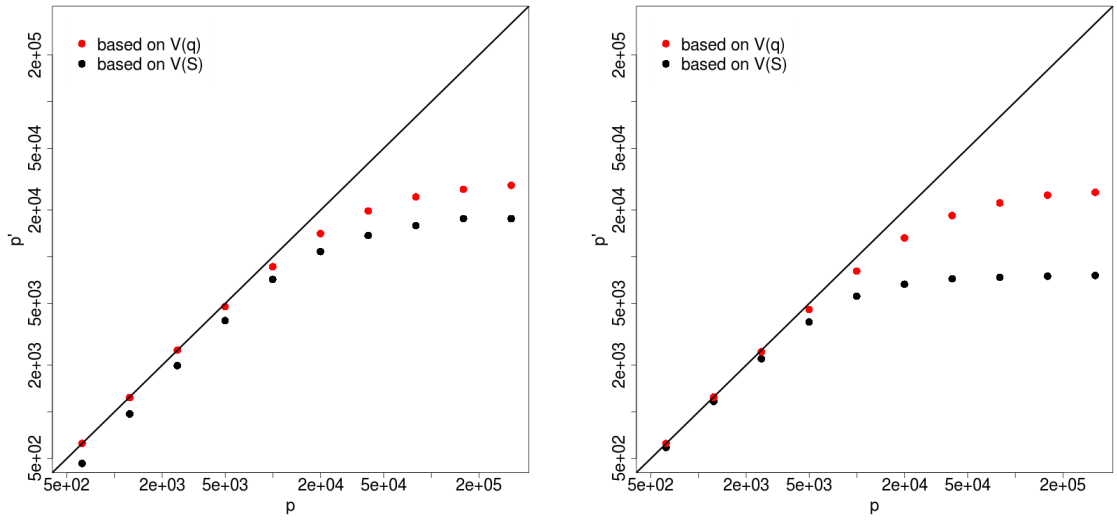



(c) V($\hat{S}$) vs m (Australia)




(d) V($\hat{S}$) vs m (Finland)

**Figure S2.** Simulations using real genotypes from either the Australia or the Finland data validate $V(q) = O(2/(n^2p'))$ and $V(\hat{S}) = O(4/(m^2p'^2))$, where $p'$ is the effective number of independent SNPs.




(a) Australia (b) Finland

**Figure S3.** Effective number of independent SNPs, $p'$, increases with the number of SNPs, $p$, in both the Australia and the Finland data. The effect number $p'$ is computed based on either $V(q) = O(2/(n^2 p'))$ or $V(S) = O(4/(m^2 p'^2))$. Note that both x-axis and y-axis are on log-scale.

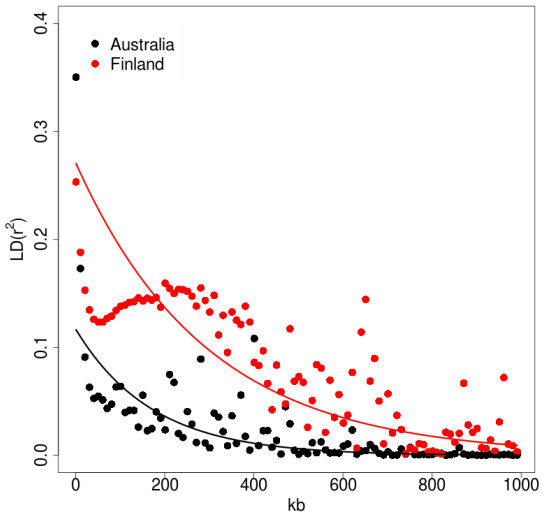
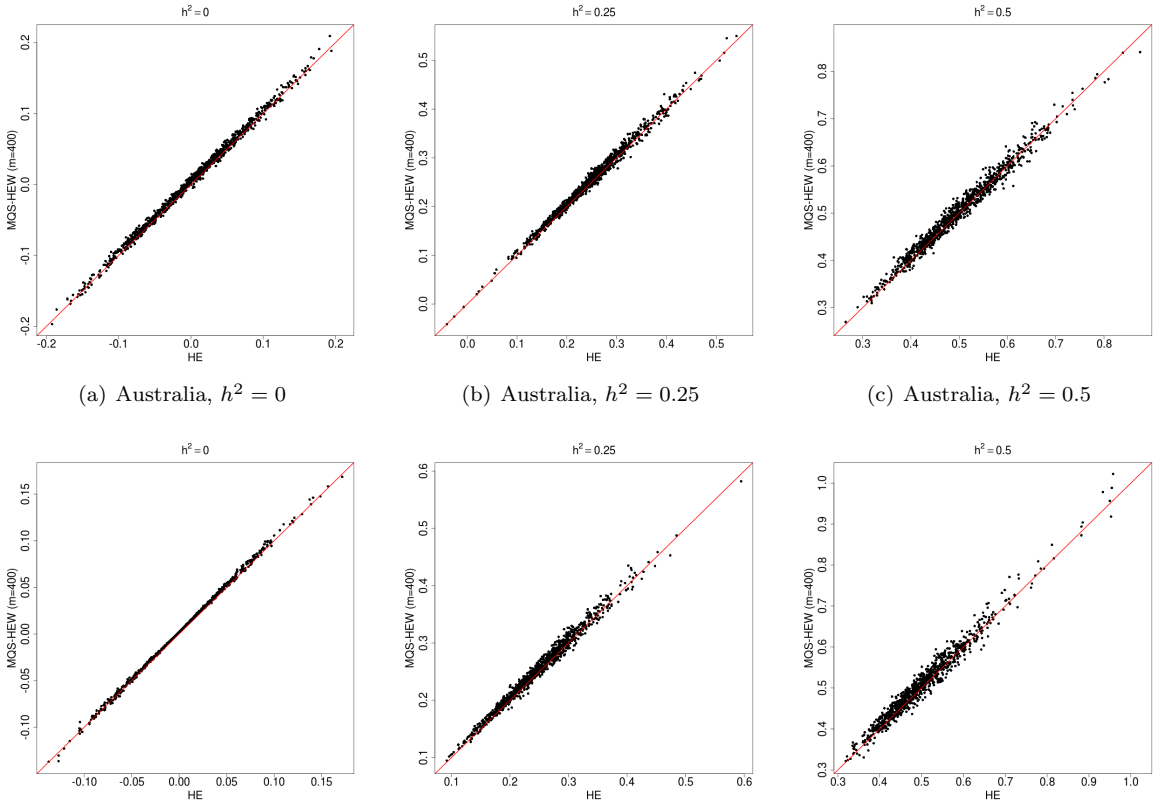


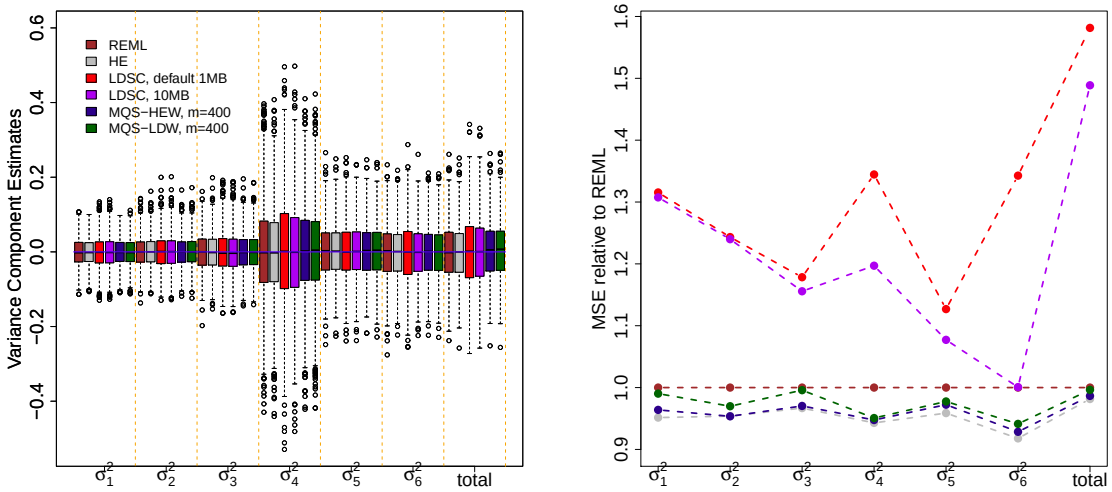

**Figure S4.** LD decay in the two data sets.




(a) Australia, $h^2 = 0$ (b) Australia, $h^2 = 0.25$ (c) Australia, $h^2 = 0.5$

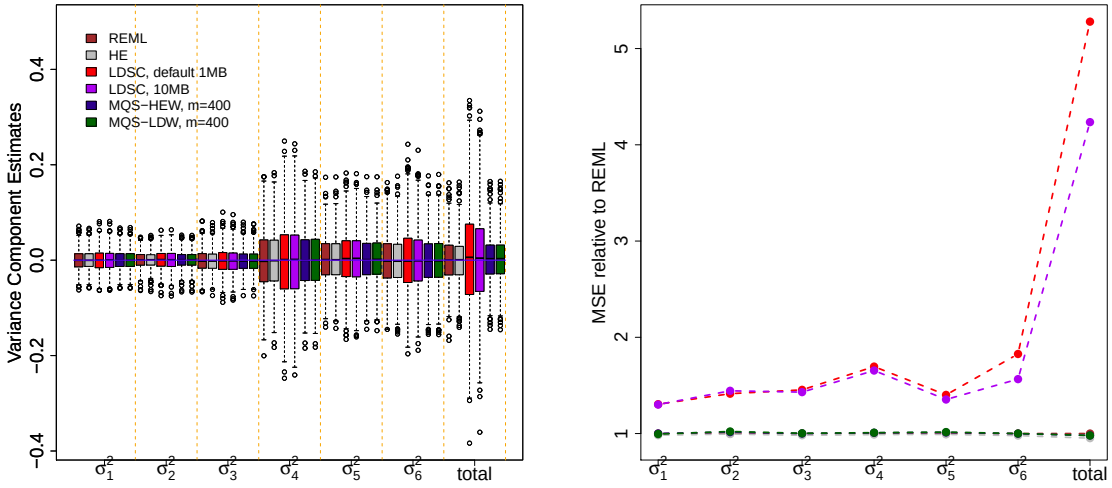
(d) Finland, $h^2 = 0$ (e) Finland, $h^2 = 0.25$ (f) Finland, $h^2 = 0.5$

**Figure S5.** Variance component estimates from MQS-HEW versus that from the HE regression in three simulations based on either the Australia data or Finland data. Note that large $h^2$ accompanies large estimation noise.




(a) Estimates (Australia)

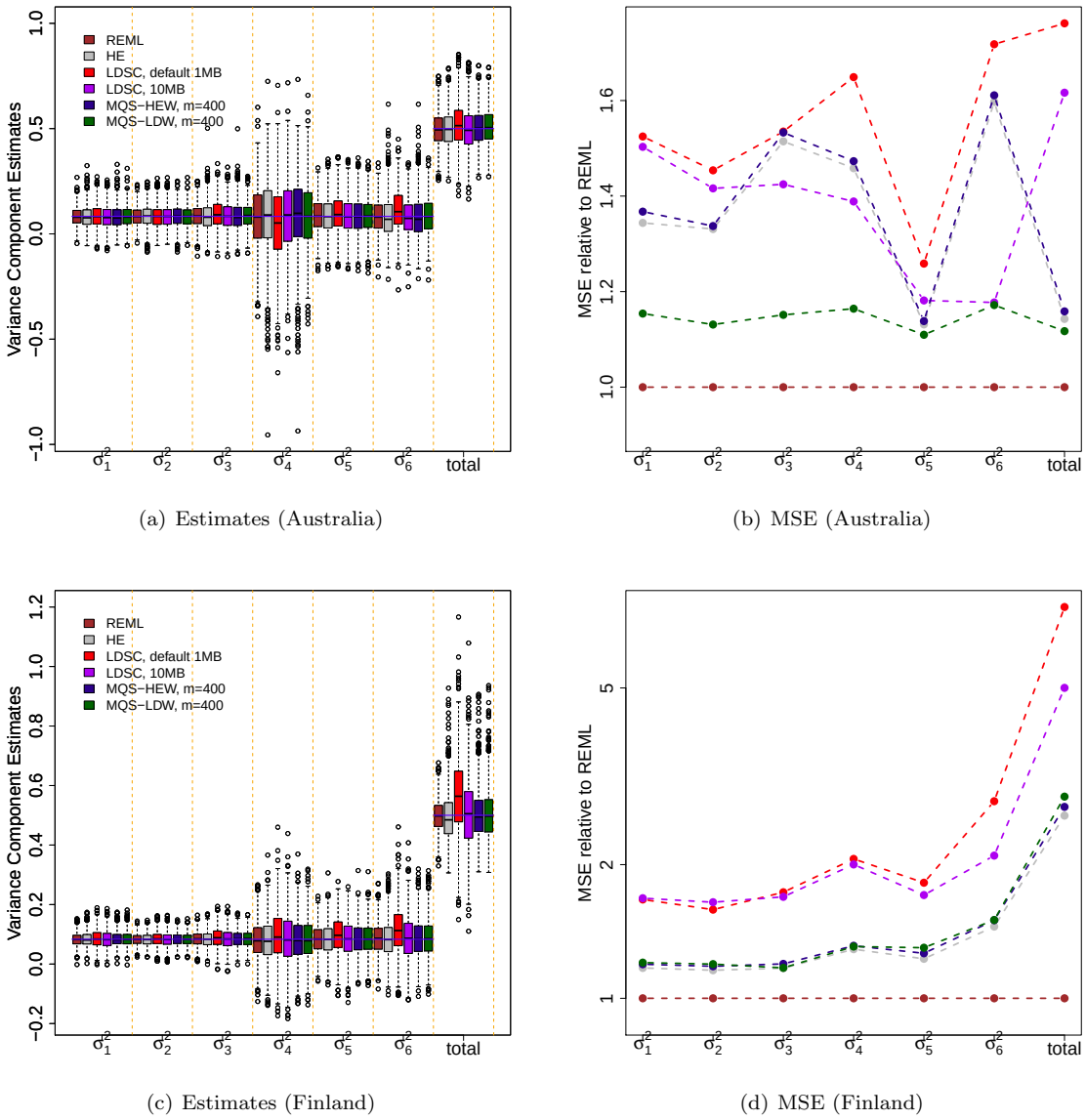
(b) MSE (Australia)




(c) Estimates (Finland)

(d) MSE (Finland)

**Figure S6.** Comparison of variance component estimates from REML (brown), HE (grey), LDSC (purple), MQS-HEW (blue), and MQS-LDW (green) for scenario I of the $k = 6$ simulations based on the Australian data (a, b) or the Finland data (c, d). LDSC estimates are obtained using either the default 1 MB window (red) or an extended 10 MB window (purple). (a) and (c) show boxplots. The true variance components are shown as blue horizontal lines. (b) and (d) show the mean squared error (MSE) relative to REML. MSE relative to REML measures the statistical efficiency of REML with respect to other methods; a higher relative MSE thus indicates low statistical efficiency. x-axis shows the six variance components and the total genetic variance.



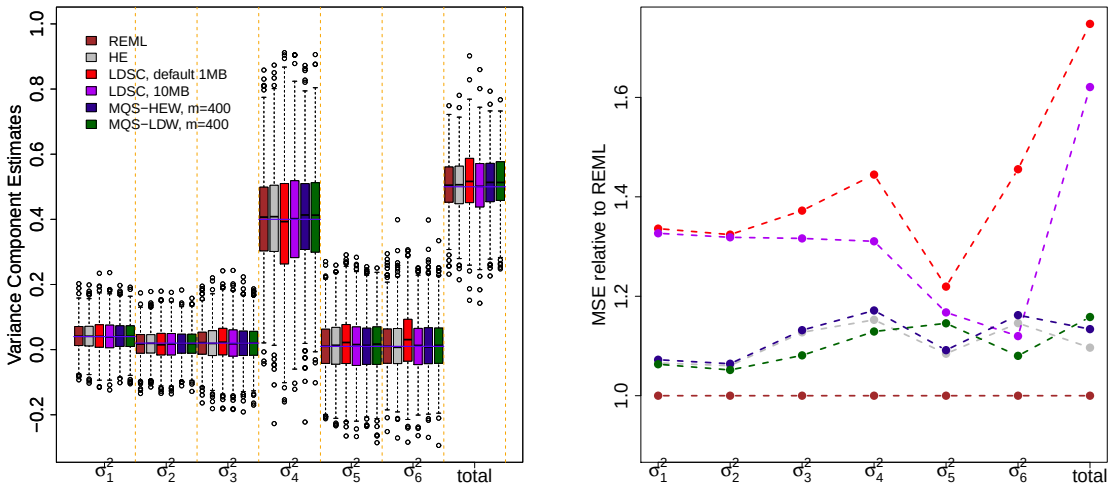

(a) Estimates (Australia)

(b) MSE (Australia)

(c) Estimates (Finland)

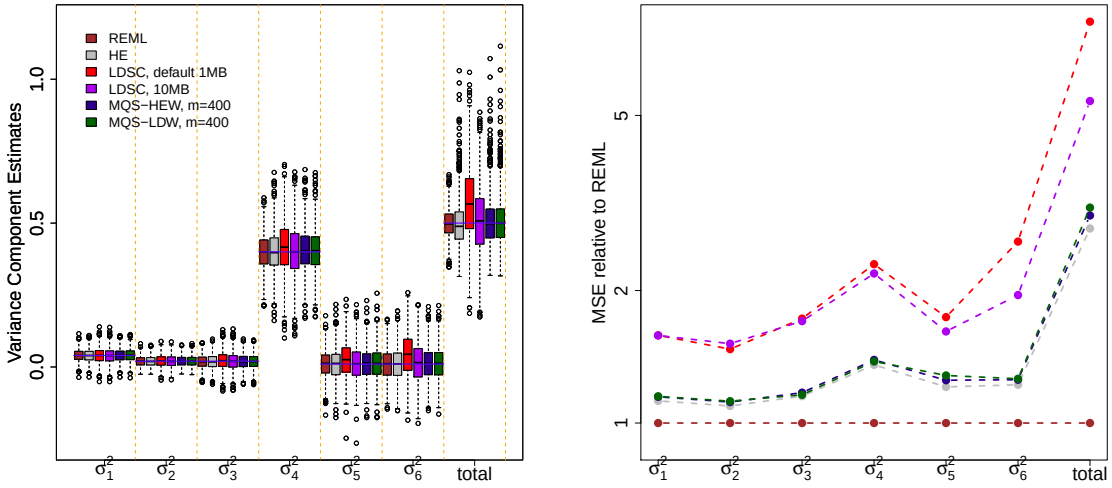
(d) MSE (Finland)

**Figure S7.** Comparison of variance component estimates from REML (brown), HE (grey), LDSC (purple), MQS-HEW (blue), and MQS-LDW (green) for scenario II of the $k = 6$ simulations based on the Australian data (a, b) or the Finland data (c, d). LDSC estimates are obtained using either the default 1 MB window (red) or an extended 10 MB window (purple). (a) and (c) show boxplots. The true variance components are shown as blue horizontal lines. (b) and (d) show the mean squared error (MSE) relative to REML. MSE relative to REML measures the statistical efficiency of REML with respect to other methods; a higher relative MSE thus indicates low statistical efficiency. x-axis shows the six variance components and the total genetic variance. y-axis in (d) is on log scale.




(a) Estimates (Australia)

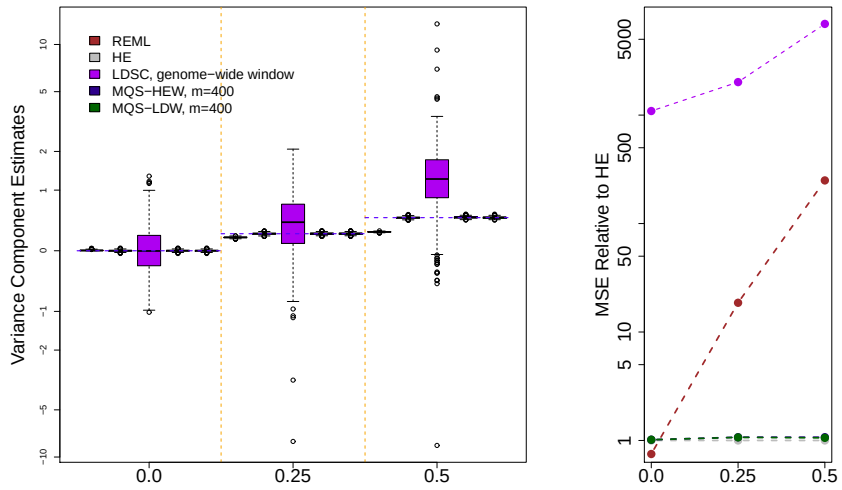
(b) MSE (Australia)



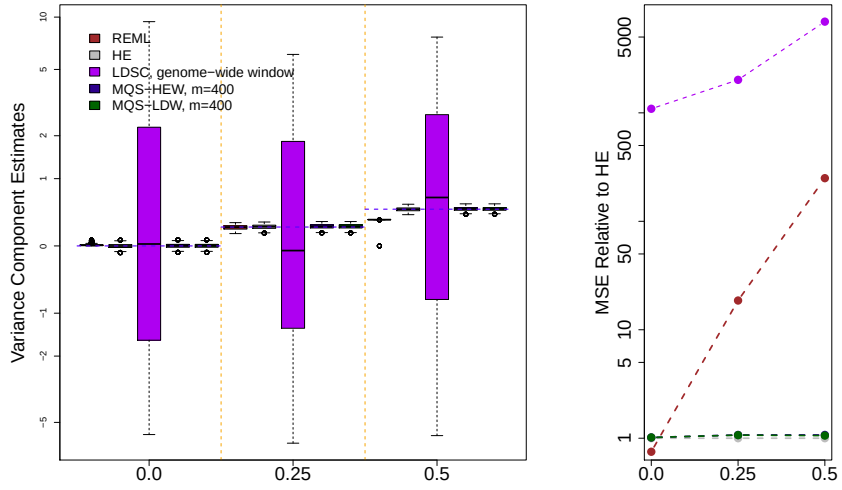

(c) Estimates (Finland)

(d) MSE (Finland)

**Figure S8.** Comparison of variance component estimates from REML (brown), HE (grey), LDSC (purple), MQS-HEW (blue), and MQS-LDW (green) for scenario III of the $k = 6$ simulations based on the Australian data (a, b) or the Finland data (c, d). LDSC estimates are obtained using either the default 1 MB window (red) or an extended 10 MB window (purple). (a) and (c) show boxplots. The true variance components are shown as blue horizontal lines. (b) and (d) show the mean squared error (MSE) relative to REML. MSE relative to REML measures the statistical efficiency of REML with respect to other methods; a higher relative MSE thus indicates low statistical efficiency. x-axis shows the six variance components and the total genetic variance. y-axis in (d) is on log scale.

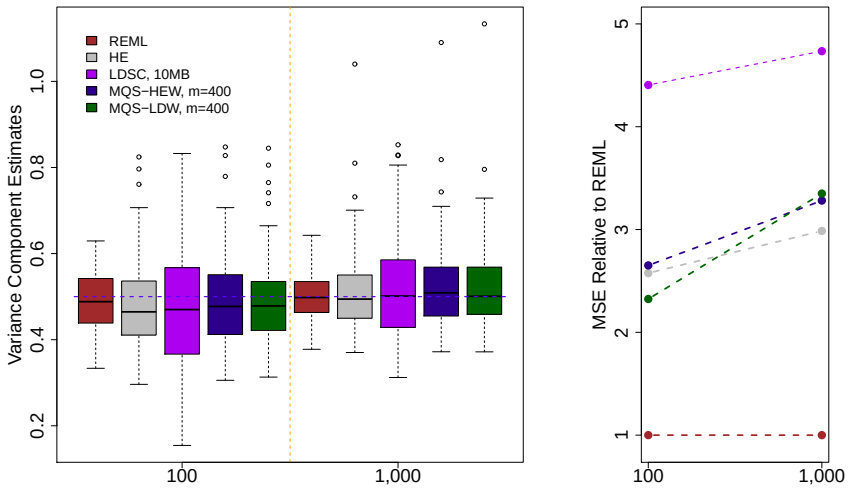



(a) Estimates (b) MSE



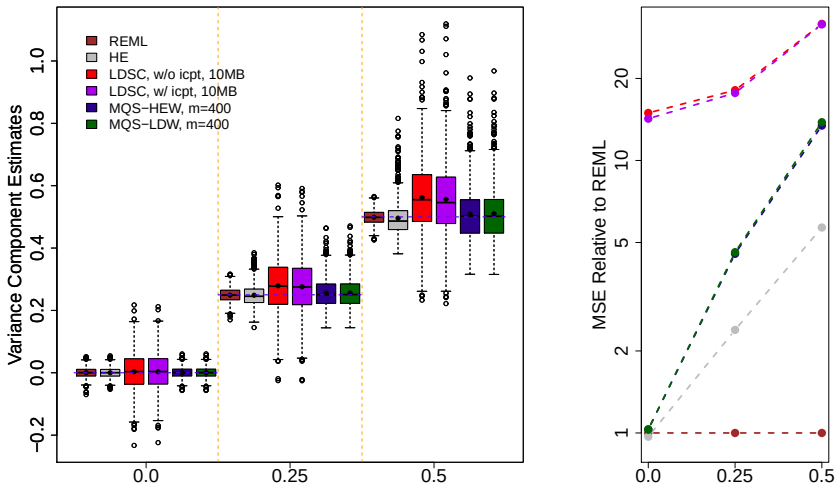

(c) Estimates (d) MSE

**Figure S9.** Comparison of variance component estimates from REML (brown), HE (grey), LDSC (purple), MQS-HEW (blue), and MQS-LDW (green) for the ascertained case control simulations with independent SNPs. LDSC estimates are obtained using a genome-wide window; thus LDSC is identical to MQS-LDW in this simulation. (a) and (c) show boxplots in the non-sparse scenario (with 10,000 SNPs total, all of which are causal) and the sparse scenario (with 100,000 SNPs total among which 10,000 are causal), where the methods are plotted in the same order as that in the legend. The true variance components (0, 0.25 and 0.5) are shown as blue horizontal lines. (b) shows the mean squared error (MSE) relative to REML. MSE relative to REML measures the statistical efficiency of REML with respect to other methods; a higher relative MSE thus indicates low statistical efficiency. x-axis shows the true variance components. y-axis is either on asinh scale (a) or log scale (b).




(a) Estimates (b) MSE

**Figure S10.** Comparison of variance component estimates from REML (brown), HE (grey), LDSC (purple), MQS-HEW (blue), and MQS-LDW (green) for the sparse simulations based on the Finland data. LDSC estimates are obtained using an extended 10 MB window. (a) shows boxplots, where the methods are plotted in the same order as that in the legend. The true variance component (0.5) is shown as a blue horizontal line. (b) shows the mean squared error (MSE) relative to REML. MSE relative to REML measures the statistical efficiency of REML with respect to other methods; a higher relative MSE thus indicates low statistical efficiency. x-axis shows the number of causal SNPs.




(a) Estimates (b) MSE

**Figure S11.** Comparison of variance component estimates from REML (brown), HE (grey), LDSC (red and purple), MQS-HEW (blue), and MQS-LDW (green) for $k = 1$ simulations based on the Framinghan heart study (FHS) data. LDSC estimates are obtained using either the default 1 MB window (red) or an extended 10 MB window (purple). Left panel shows the boxplot. The true variance components (0, 0.25 and 0.5) are shown as blue horizontal lines. Right panel shows the mean squared error (MSE) relative to REML. MSE relative to REML measures the statistical efficiency of REML with respect to other methods; a higher relative MSE thus indicates low statistical efficiency. Note that MQS-HEW and MQS-LDW yield almost identical results and thus blue and green curves overlap. x-axis shows the true variance components. y-axis in the right panel is on log scale.

# Supplementary Tables

**Table S1.** Comparison between $\hat{S}$ computed based on different values of $m$ and $S$ computed based on $n$. $\bar{S}$ is the sample mean of $\hat{S}$ across simulation replicates. We either simulate genotypes from a binomial distriution (simulated $\mathbf{X}$; $n = 1,000$ and $p = 10,000$), or use genotypes from the Australia data ($n = 3,925$, $p = 319,148$) or the Finland data ($n = 5,123$, $p = 319,148$). Both $\hat{S}$ and $S$ are scaled by the number of SNPs $p$. Because SNPs are not independent in real data, $\hat{S}$ may over-estimates $S$ for very small $m$ in both real data sets. However, the bias quickly diminishes for moderate $m$.

| $p\bar{S}$ | $m$ | | | | | | | | | | $n$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | |
| Simulated | 0.980 | 0.989 | 0.990 | 0.993 | 0.995 | 0.995 | 0.996 | 0.996 | 0.997 | 0.997 | 0.999 |
| Australia | 18.77 | 14.37 | 12.96 | 12.30 | 11.87 | 11.60 | 11.39 | 11.24 | 11.14 | 11.04 | 10.24 |
| Finland | 17.94 | 14.69 | 13.59 | 13.01 | 12.67 | 12.48 | 12.35 | 12.21 | 12.15 | 12.07 | 11.47 |

**Table S2.** Type I error control for different methods at different $\alpha$ levels in simulations with three different genotype data sets.

| Genotype Data | $\alpha$ | Methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | REML | HE | LDSC, 10 MB | MQS-HEW | | MQS-LDW | |
| | | | | | asymptotic | jacknife | asymptotic | jacknife |
| Australia | 0.10 | 0.127 | 0.119 | 0.139 | 0.117 | 0.125 | 0.118 | 0.126 |
| | 0.05 | 0.068 | 0.059 | 0.077 | 0.063 | 0.065 | 0.063 | 0.066 |
| | 0.01 | 0.029 | 0.011 | 0.021 | 0.012 | 0.014 | 0.013 | 0.012 |
| Finland | 0.10 | 0.124 | 0.102 | 0.121 | 0.103 | 0.161 | 0.101 | 0.162 |
| | 0.05 | 0.070 | 0.045 | 0.064 | 0.043 | 0.094 | 0.044 | 0.093 |
| | 0.01 | 0.026 | 0.008 | 0.018 | 0.008 | 0.021 | 0.008 | 0.022 |
| FHS | 0.10 | 0.107 | 0.084 | 0.105 | 0.079 | 0.469 | 0.079 | 0.467 |
| | 0.05 | 0.060 | 0.050 | 0.063 | 0.039 | 0.382 | 0.039 | 0.383 |
| | 0.01 | 0.022 | 0.012 | 0.011 | 0.007 | 0.240 | 0.007 | 0.243 |

**Table S3.** SNP heritability estimates from different methods for 11 quantitative traits and 7 binary phenotypes from three GWASs, after adjusting for the top 2 PCs. Values in parentheses are standard errors. For MQS-HEW and MQS-LDW, the standard errors are computed by the asymptotic form. The heritability estimates for the WTCCC data is presented at the observed scale. A small scaling factor is required to transform the estimates to liability scale. To ensure unbiasedness, we do not constrain variance components to be positive during estimation. Thus a low heritable trait may have its SNP heritability estimated to be below zero. $\lambda_{GC}$ is the genomic control factor while icpt is the intercept estimate from LDSC.

| | Methods | | | | | | |
|---|---|---|---|---|---|---|---|
| Trait | REML | HE | LDSC, 10 MB | MQS ($m = 400$; asymptotic) | | icpt | $\lambda_{GC}$ |
| | | | | HEW | LDW | | |
| Australia, $n = 3,925$, $p = 4,352,968$ | | | | | | | |
| Height | 0.27 (0.072) | 0.25 (0.071) | 0.21 (0.089) | 0.27 (0.071) | 0.28 (0.071) | 1.008 | 1.025 |
| Finland, $n = 5,123$, $p = 319,148$ | | | | | | | |
| BMI | 0.20 (0.052) | 0.17 (0.048) | 0.17 (0.10) | 0.17 (0.049) | 0.18 (0.049) | 1.007 | 1.022 |
| CRP | 0.037 (0.051) | 0.026 (0.043) | 0.085 (0.091) | 0.030 (0.045) | 0.030 (0.045) | 0.993 | 0.994 |
| DiaBP | 0.066 (0.048) | 0.064 (0.048) | 0.036 (0.072) | 0.069 (0.050) | 0.070 (0.050) | 1.007 | 1.010 |
| Glucose | 0.16 (0.050) | 0.16 (0.052) | 0.14 (0.087) | 0.17 (0.054) | 0.16 (0.052) | 1.008 | 1.027 |
| HDL | 0.34 (0.052) | 0.34 (0.062) | 0.51 (0.12) | 0.33 (0.061) | 0.31 (0.055) | 0.982 | 1.046 |
| Insulin | -0.098 (0.045) | -0.11 (0.042) | -0.043 (0.083) | -0.11 (0.043) | -0.11 (0.043) | 0.987 | 0.984 |
| LDL | 0.35 (0.052) | 0.41 (0.078) | 0.16 (0.10) | 0.42 (0.080) | 0.42 (0.079) | 1.051 | 1.078 |
| SysBP | 0.17 (0.051) | 0.16 (0.052) | 0.11 (0.090) | 0.17 (0.053) | 0.17 (0.053) | 1.015 | 1.034 |
| TC | 0.26 (0.051) | 0.29 (0.065) | 0.12 (0.11) | 0.30 (0.067) | 0.30 (0.066) | 1.036 | 1.052 |
| TG | 0.18 (0.052) | 0.15 (0.047) | 0.13 (0.11) | 0.15 (0.048) | 0.15 (0.048) | 1.008 | 1.026 |
| WTCCC, $n = \sim 5,000$, $p = 458,868$ | | | | | | | |
| BD | 0.83 (0.057) | 1.04 (0.094) | 0.43 (0.11) | 1.07 (0.097) | 1.15 (0.10) | 1.080 | 1.114 |
| CAD | 0.57 (0.061) | 0.57 (0.070) | 0.18 (0.11) | 0.59 (0.071) | 0.60 (0.071) | 1.050 | 1.071 |
| CD | 0.68 (0.061) | 0.76 (0.082) | 0.44 (0.14) | 0.77 (0.084) | 0.79 (0.081) | 1.043 | 1.078 |
| HT | 0.58 (0.060) | 0.61 (0.070) | 0.12 (0.097) | 0.62 (0.071) | 0.66 (0.072) | 1.061 | 1.066 |
| RA | 0.68 (0.059) | 0.80 (0.081) | 0.50 (0.33) | 0.82 (0.083) | 0.81 (0.082) | 1.039 | 1.044 |
| T1D | 0.97 (0.053) | 1.19 (0.091) | 0.91 (0.57) | 1.22 (0.092) | 1.19 (0.088) | 1.035 | 1.053 |
| T2D | 0.60 (0.061) | 0.64 (0.075) | 0.15 (0.10) | 0.66 (0.077) | 0.70 (0.078) | 1.061 | 1.073 |

**Table S4.** SNP heritability estimates from different methods for 11 quantitative traits and 7 binary phenotypes from three GWASs, after adjusting for the top 10 PCs. Values in parentheses are standard errors. For MQS-HEW and MQS-LDW, the standard errors are computed by the asymptotic form. The heritability estimates for the WTCCC data is presented at the observed scale. A small scaling factor is required to transform the estimates to liability scale. To ensure unbiasedness, we do not constrain variance components to be positive during estimation. Thus a low heritable trait may have its SNP heritability estimated to be below zero. $\lambda_{GC}$ is the genomic control factor while icpt is the intercept estimate from LDSC.

| | | | Methods | | | | |
|---|---|---|---|---|---|---|---|
| Trait | REML | HE | LDSC, 10 MB | MQS ($m = 400$; asymptotic) | | icpt | $\lambda_{GC}$ |
| | | | | HEW | LDW | | |
| Australia, $n = 3,925$, $p = 4,352,968$ | | | | | | | |
| Height | 0.25 (0.073) | 0.21 (0.066) | 0.16 (0.090) | 0.22 (0.067) | 0.24 (0.068) | 1.009 | 1.015 |
| Finland, $n = 5,123$, $p = 319,148$ | | | | | | | |
| BMI | 0.19 (0.052) | 0.14 (0.045) | 0.10 (0.12) | 0.15 (0.046) | 0.16 ( 0.046) | 1.012 | 1.013 |
| CRP | 0.019 (0.051) | 0.012 (0.042) | 0.060 (0.094) | 0.014 (0.044) | 0.015 (0.044) | 0.994 | 0.993 |
| DiaBP | 0.034 (0.049) | 0.028 (0.044) | -0.019 (0.074) | 0.031 (0.046) | 0.032 (0.046) | 1.008 | 1.004 |
| Glucose | 0.13 (0.052) | 0.097 (0.044) | 0.041 (0.079) | 0.10 (0.046) | 0.11 (0.046) | 1.012 | 1.018 |
| HDL | 0.30 (0.053) | 0.23 (0.047) | 0.42 (0.11) | 0.23 (0.046) | 0.23 (0.046) | 0.982 | 1.035 |
| Insulin | -0.16 (0.046) | -0.13 (0.041) | -0.067 (0.076) | -0.13 (0.042) | -0.13 (0.042) | 0.987 | 0.979 |
| LDL | 0.29 (0.053) | 0.22 (0.046) | 0.0049 (0.10) | 0.23 (0.047) | 0.24 (0.048) | 1.041 | 1.044 |
| SysBP | 0.13 (0.052) | 0.098 (0.044) | 0.042 (0.096) | 0.10 (0.045) | 0.11 (0.045) | 1.013 | 1.016 |
| TC | 0.23 (0.052) | 0.18 (0.047) | 0.046 (0.11) | 0.19 (0.048) | 0.20 (0.049) | 1.028 | 1.036 |
| TG | 0.16 (0.052) | 0.13 (0.045) | 0.091 (0.094) | 0.13 (0.046) | 0.13 (0.046) | 1.010 | 1.021 |
| WTCCC, $n = \sim 5,000$, $p = 458,868$ | | | | | | | |
| BD | 0.78 (0.059) | 0.83 (0.077) | 0.36 (0.12) | 0.85 (0.078) | 0.91 (0.081) | 1.062 | 1.088 |
| CAD | 0.54 (0.062) | 0.51 (0.066) | 0.13 (0.11) | 0.52 (0.067) | 0.54 (0.067) | 1.048 | 1.058 |
| CD | 0.65 (0.062) | 0.66 (0.072) | 0.37 (0.13) | 0.68 (0.075) | 0.70 (0.075) | 1.041 | 1.065 |
| HT | 0.57 (0.060) | 0.59 (0.069) | 0.090 (0.095) | 0.61 (0.071) | 0.64 (0.071) | 1.062 | 1.062 |
| RA | 0.66 (0.060) | 0.73 (0.076) | 0.49 (0.32) | 0.76 (0.077) | 0.75 (0.077) | 1.032 | 1.036 |
| T1D | 0.97 (0.053) | 1.15 (0.087) | 0.90 (0.57) | 1.21 (0.090) | 1.15 (0.086) | 1.034 | 1.044 |
| T2D | 0.55 (0.062) | 0.52 (0.066) | 0.12 (0.10) | 0.53 (0.068) | 0.55 (0.068) | 1.049 | 1.058 |

**Table S5.** SNP heritability or heritability estimates from MQS-HEW and MQS-LDW for 15 quantitative traits and 7 binary phenotypes from four data sets. FHS is a family study while the rest are population studies. Values in parentheses are standard errors. Different from the figure in the main text, the standard error are computed based on block-wise jackknife. $S$ is estimated using either a random subset of 400 individuals from the data (first two columns) or 503 individuals of European ancestry from the 1,000 Genomes Project (1KGP, last two columns).

| Trait | MQS ($m = 400$, jackknife) | | MQS ($m = 503$, 1KGP, jackknife) | |
|---|---|---|---|---|
| | HEW | LDW | HEW | LDW |
| Australia, $n = 3,925$, $p = 4,352,968$ | | | | |
| Height | 0.27(0.072) | 0.28(0.073) | 0.25(0.064) | 0.27(0.065) |
| Finland, $n = 5,123$, $p = 319,148$ | | | | |
| BMI | 0.19(0.042) | 0.19(0.043) | 0.33(0.079) | 0.35(0.080) |
| CRP | 0.037(0.046) | 0.037(0.046) | 0.059(0.080) | 0.060(0.080) |
| DiaBP | 0.075(0.042) | 0.076(0.042) | 0.13(0.073) | 0.13(0.074) |
| Glucose | 0.21(0.050) | 0.21(0.048) | 0.36(0.088) | 0.37(0.087) |
| HDL | 0.36(0.066) | 0.35(0.059) | 0.63(0.11) | 0.61(0.10) |
| Insulin | -0.082(0.039) | -0.081(0.039) | -0.15(0.068) | -0.15(0.068) |
| LDL | 0.61(0.058) | 0.61(0.057) | 1.04(0.10) | 1.10(0.10) |
| SysBP | 0.27(0.042) | 0.28(0.043) | 0.47(0.073) | 0.50(0.076) |
| TC | 0.42(0.055) | 0.43(0.054) | 0.73(0.094) | 0.76(0.096) |
| TG | 0.18(0.052) | 0.17(0.050) | 0.30(0.085) | 0.30(0.085) |
| WTCCC, $n = \sim 5,000$, $p = 458,868$ | | | | |
| BD | 1.08(0.068) | 1.16(0.071) | 1.02(0.068) | 1.09(0.071) |
| CAD | 0.60(0.065) | 0.63(0.066) | 0.57(0.064) | 0.60(0.065) |
| CD | 1.08(0.097) | 1.12(0.090) | 1.03(0.092) | 1.08(0.088) |
| HT | 0.63(0.064) | 0.66(0.064) | 0.59(0.060) | 0.62(0.061) |
| RA | 0.84(0.29) | 0.83(0.25) | 0.78(0.36) | 0.80(0.33) |
| T1D | 1.60(0.73) | 1.41(0.59) | 1.48(0.96) | 1.38(0.84) |
| T2D | 0.71(0.071) | 0.76(0.070) | 0.66(0.069) | 0.69(0.069) |
| FHS, $n = 3,806 - 6,855$, $p = 372,131$ | | | | |
| HDL | 0.40 (0.071) | 0.40 (0.073) | 2.38 (0.072) | 2.71 (0.084) |
| LDL | 0.58 (0.10) | 0.58 (0.10) | 2.91 (0.086) | 3.32 (0.099) |
| TC | 0.54 (0.091) | 0.55 (0.092) | 2.79 (0.10) | 3.04 (0.11) |
| TG | 0.47 (0.074) | 0.47 (0.075) | 2.46 (0.099) | 2.67 (0.10) |

# Supplementary Text

## 1 The Model

Although our method can be reasonably general, we introduce it by considering a particular application: partitioning heritability by different SNP functional categories. To do so, we assume that SNPs have been pre-classified into $k$ different, non-overlapping functional categories. Our goal is to estimate the proportion of phenotypic variance explained (PVE) by all SNPs in each category. To do so, we consider the following model

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_{k+1}^2 \mathbf{M}), \tag{1}$$

where $\mathbf{y}$ is an $n$-vector of phenotypes for $n$ individuals; $\mathbf{X}_i$ is an $n$ by $p_i$ genotype matrix for $p_i$ variants in $i$th category; $\boldsymbol{\beta}_i$ is a $p_i$-vector of corresponding effect sizes; $\boldsymbol{\epsilon}$ is a $n$-vector of residual errors; $\sigma_{k+1}^2$ is the residual error variance; $\mathbf{M} = \mathbf{I} - \mathbf{1}_n \mathbf{1}_n^T / n$ is the projection matrix onto the null space of the intercept; and MVN denotes a multivariate normal distribution. Both $\mathbf{y}$ and every column of $\mathbf{X}$ have been centered to have mean zero. Because of the centering, the residual errors follow a degenerate multivariate normal distribution with a low-rank covariance matrix $\mathbf{M}$ that constrains the errors to sum to zero. Centering does not affect results but simplifies the algebra. Extension of the model to incorporate other covariates in addition to the intercept is provided later in section 7.

Following previous approaches [1, 2], for each effect size vector $\boldsymbol{\beta}_i$, we specify a normal prior with mean zero and variance $\sigma_i^2/p_i$, or $\boldsymbol{\beta}_i \sim \text{MVN}(0, \sigma_i^2/p_i \mathbf{I})$. This normality assumption of effect size leads to an alternative but equivalent form – a LMM with $k$ variance components [2]

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{g}_i + \boldsymbol{\epsilon}, \quad \mathbf{g}_i \sim \text{MVN}(0, \sigma_i^2 \mathbf{K}_i), \quad \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_{k+1}^2 \mathbf{M}), \tag{2}$$

where $\mathbf{g}_i = \mathbf{X}_i \boldsymbol{\beta}_i$ is the combined genetic effects of $i$th category; $\mathbf{K}_i = \mathbf{X}_i \mathbf{X}_i^T / p_i$ is an $n$ by $n$ genetic relatedness matrix computed from SNPs in $i$th category; $\boldsymbol{\sigma}^2 = (\sigma_1^2, \cdots, \sigma_k^2)$ are the variance components.

As usual [1, 3, 4], we standardize every column of $\mathbf{X}$ to have variance one. Unlike centering, standardizing the $\mathbf{X}$ columns will affect the results because it changes our prior assumption [2]. Specifically, standardizing the $\mathbf{X}$ columns corresponds to making an assumption that rarer variants tend to have larger effects than common variants, and that marker effect sizes depend on the minor allele frequencies (MAFs) in a particular mathematical form (see [2] for relevant discussion). We do not, however, standardize $\mathbf{y}$, and the phenotype variance is $s_y^2 = \mathbf{y}^T \mathbf{y}/(n-1)$.

At this point, it is useful to define the scaled version of the variance components: $\mathbf{h}^2 = (h_1^2, \cdots, h_k^2) = (\sigma_1^2/s_y^2, \cdots, \sigma_k^2/s_y^2) = \boldsymbol{\sigma}^2/s_y^2$ and $h_{k+1}^2 = \sigma_{k+1}^2/s_y^2$. As we will show below, when the phenotype variance $s_y^2$ is unknown, $\boldsymbol{\sigma}^2$ and $\sigma_{k+1}^2$ are not estimable from summary statistics alone, but $\mathbf{h}^2$ and $h_{k+1}^2$ are. With these modeling assumptions, PVE by $i$th category is $\text{PVE}_i = \sigma_i^2/(\sum_{j=1}^{k+1} \sigma_j^2) = h_i^2/(\sum_{j=1}^{k+1} h_j^2),\ i \in \{1, \cdots, k\}$ [2].

## 2 MoM and MINQUE

Our goal is to estimate the variance components $\boldsymbol{\sigma}^2$ and the residual error variance $\sigma_{k+1}^2$, or correspondingly the scaled version $\mathbf{h}^2$ and $h_{k+1}^2$. The estimated parameters can then be used to compute PVE. To estimate the variance components, we consider the method of moments (MoM). Because variance components are parameters that define the second moments of LMMs, MoM for variance component

estimate is naturally based on a set of second moment matching equations (e.g. [5])

$$E(\mathbf{y}^T\mathbf{A}_j\mathbf{y}) = \sum_{i=1}^{k} tr(\mathbf{A}_j\mathbf{K}_i)\sigma_i^2 + tr(\mathbf{A}_j)\sigma_{k+1}^2, \tag{3}$$

where each $\mathbf{A}_j$ is an arbitrary symmetric non-negative definite matrix used to create weighted second moments; $tr$ denotes matrix trace. For estimation with MoM, we replace the expectation on the left hand side (LHS) of equation 3 with the realized value $\mathbf{y}^T\mathbf{A}_j\mathbf{y}$. We then solve the equations to obtain estimates for $\boldsymbol{\sigma}^2$ and $\sigma_{k+1}^2$. Because there are $k+1$ parameters in our model, we only need $k+1$ different $\mathbf{A}_j$ to obtain the estimates. Though unnecessarily (e.g. the MINQUE estimates described below are based on only k+1 different $\mathbf{A}_j$), we can also use more than $k+1$ $\mathbf{A}_j$ to set up an over-determined linear system, and use the ordinary least squares (OLS) to obtain unbiased estimates. For example, the naive LDSC equation [6] is based on using a set of $p$ different $\mathbf{A}_j$, where $\mathbf{A}_j = \mathbf{X}\boldsymbol{\Lambda}_j\mathbf{X}^T$ and $\boldsymbol{\Lambda}_j$ is a rank one matrix with $jj$th diagonal element being one and all other elements being zero. (Note that we call this naive LDSC to distinguish it from LDSC. LDSC does not use OLS estimates based on the naive LDSC equation because of low statistical efficiency. Rather, LDSC introduces two extra weights to improve statistical efficiency.)

Any choice of $\mathbf{A}_j$ will yield unbiased estimates for $\boldsymbol{\sigma}^2$ and $\sigma_{k+1}^2$, but different choices of $\mathbf{A}_j$ can affect statistical efficiency of the estimates. One particular choice of $\mathbf{A}_j$ can be found based on the Minimal Norm Quadratic Unbiased Estimation (MINQUE) criterion [5,7–9]. To see how MINQUE works, it helps to view the right hand side (RHS) of the equation 3 as a linear combination of the variance components, and view the realization of the LHS, $\mathbf{y}\mathbf{A}_j\mathbf{y}$, as a way to provide a good estimate of this linear combination. We can quantify the accuracy of using $\mathbf{y}\mathbf{A}_j\mathbf{y}$ to estimate the RHS by the expected squared error, or

$$\mathrm{SE}(\boldsymbol{\sigma}^2, \sigma_{k+1}^2) = E((\mathbf{y}^T\mathbf{A}_j\mathbf{y} - \sum_{i=1}^{k} tr(\mathbf{A}_j\mathbf{K}_i)\sigma_i^2 - tr(\mathbf{A}_j)\sigma_{k+1}^2)^2), \tag{4}$$

where the expectation is with respect to distribution of $\mathbf{y}$ and SE is a function of the variance components. Our goal is then to find a set of $\mathbf{A}_j$ that minimizes the SE given $\boldsymbol{\sigma}^2, \sigma_{k+1}^2$ conditional on $tr(\mathbf{A}_j\mathbf{K}_i)$ being a constant. It has been shown [10] that this particular choice of $\mathbf{A}_j$ is of the following form

$$\hat{\mathbf{A}}_j = \mathbf{H}^{-1}\mathbf{K}_j\mathbf{H}^{-1}, \tag{5}$$

where $j = 1, \cdots, k+1$, $\mathbf{K}_{k+1} = \mathbf{M}$ and $\mathbf{H} = \sum_{i=1}^{k+1} \sigma_i^2\mathbf{K}_i$. Note that we have loosely used the matrix inverse notation to denote a generalized inverse, and we have also loosely used hat on top of $\mathbf{A}_j$ to not denote an estimate but to denote an optimal choice under the MINQUE criterion.

This set of $\hat{\mathbf{A}}_j$ is also known as the best quadratic unbiased estimator (BQUE) for the variance components, where "best" is in the sense that the estimates minimize the expected squared error given the parameters [10]. The BQUE for the variance components is analogous to the best unbiased estimator (BLUE) of fixed effects. BQUE is equivalent to the minimal variance quadratic unbiased estimation (MIVQUE) criteria [8, 9], and is a special case of the minimum-norm quadratic unbiased estimators (MINQUE) [5,7–9] that have been used in the animal breeding programs [11]. Because of the equivalence and the commonality of MINQUE, we refer to $\mathbf{A}_j$ in the form of equation 5 as the optimal choice based on the MINQUE criterion and the resulting variance component estimating equations of equation 3 as the MINQUE equations.

This optimal $\hat{\mathbf{A}}_j$ depends on the variance components $\boldsymbol{\sigma}^2, \sigma_{k+1}^2$ that are unknown *a priori* and are the target of estimation. Thus, we cannot use the optimal $\hat{\mathbf{A}}_j$ directly. However, two options are available to obtain MINQUE estimates in practice. The first option is to apply an iterative procedure on equation 3: in each iteration $t$ we plug in the current estimates $\hat{\boldsymbol{\sigma}}_t^2, \hat{\sigma}_{k+1,t}^2$ in $\mathbf{H}$ and $\mathbf{A}_j$ to obtained the updated

estimates $\hat{\boldsymbol{\sigma}}^2_{t+1}, \hat{\sigma}^2_{k+1,t+1}$. The resulting algorithm is often referred to as an iterative MINQUE, or I-MINQUE, which produces estimates that are identical to REML estimates [12]; thus I-MINQUE can be viewed as an algorithm to obtain REML estimates. The second option is to use a pre-determined set of parameters $\tilde{\boldsymbol{\sigma}}^2_i, \tilde{\sigma}^2_{k+1}$ to construct $\tilde{\mathbf{H}} = \sum_{i=1}^{k+1} \tilde{\sigma}^2_i \mathbf{K}_i$ and then use $\tilde{\mathbf{A}}_j = \tilde{\mathbf{H}}^{-1}\mathbf{K}_j\tilde{\mathbf{H}}^{-1}$. For example, MINQUE(0) sets $\mathbf{h}^2 = \mathbf{0}$, $h^2_{k+1} = 1$ to obtain estimates of reasonable statistical efficiency [13, 14]. However, both these two options require individual-level data.

# 3 First Approximation: SNP Weights and Approximation to MINQUE

We aim to develop an approximation form of the optimal $\hat{\mathbf{A}}_j$ that eventually allows the use of summary statistics, while retaining the statistical efficiency comes with the optimal $\hat{\mathbf{A}}_j$. Summary statistics often comes in the form of marginal z-scores. For $l$th SNP, the marginal z-score is $z_l = \mathbf{x}_l^T\mathbf{y}/(\hat{\tau}^{-1}\sqrt{n-1})$, where $\mathbf{x}_l$ is an $n$ vector of genotypes for the SNP and $\hat{\tau}^{-1}$ is the estimated residual error variance of the marginal linear model. Because individual SNP effect size is small for most complex traits, $\hat{\tau}^{-1} \approx s^2_y$ and $z_l \approx \mathbf{x}_l^T\mathbf{y}/(s^2_y\sqrt{n-1})$. Therefore, summary statistics z-scores are accurately approximated by the cross-product of $\mathbf{X}$ and $\mathbf{y}$. To use summary statistics, we choose the following $\mathbf{A}_j$ matrix with $\mathbf{X}$ flanking the two sides to form $\mathbf{X}^T\mathbf{y}$ in equation 3:

$$\tilde{\mathbf{A}}_j = \mathbf{X}_j\mathbf{W}_j\mathbf{X}_j^T/p_j \ (j = 1, \cdots, k), \quad \tilde{\mathbf{A}}_{k+1} = \mathbf{M} \tag{6}$$

where $\mathbf{W}_j = diag(w_{j1}, \cdots, w_{jp_j})$ is a pre-specified $p_j$ by $p_j$ diagonal matrix of SNP weights and is used to ensure $\tilde{\mathbf{A}}_j$ approximates $\hat{\mathbf{A}}_j$ well to retain the statistical efficiency comes with $\hat{\mathbf{A}}_j$; and $\mathbf{X}_i$ again is an $n$ by $p_i$ genotype matrix for $p_i$ variants in $i$th category. $\hat{\mathbf{A}}_j$ $(j \neq k+1)$ is effectively a weighted genetic relatedness matrix, while $\tilde{\mathbf{A}}_{k+1}$ approximates $\hat{\mathbf{A}}_{k+1}$ by assuming that $\mathbf{H}$ is approximately $\mathbf{M}$. Note that $\tilde{\mathbf{A}}_{k+1}$ ensures that $\sum_{i=1}^{k+1}\sigma^2_i = s^2_y$ and $\sum_{i=1}^{k+1} h^2_i = 1$. Because a scale transformation of $\mathbf{W}_j$ does not affect results, to simplify the algebra, we constrain $tr(\mathbf{W}_j) = p_j$, or equivalently, assume that the average SNP weight equals one.

With this set of $\tilde{\mathbf{A}}_j$, the $k+1$ estimating equations from equation 3 become

$$\mathbf{y}^T\tilde{\mathbf{A}}_j\mathbf{y} = \sum_{i=1}^{k} tr(\mathbf{K}_i\tilde{\mathbf{A}}_j)\sigma^2_i + tr(\tilde{\mathbf{A}}_j)\sigma^2_{k+1}, \tag{7}$$

$$\mathbf{y}^T\mathbf{y} = \sum_{i=1}^{k} tr(\mathbf{K}_i)\sigma^2_i + tr(\mathbf{M})\sigma^2_{k+1}. \tag{8}$$

where $tr(\mathbf{K}_i) = tr(\mathbf{M}) = tr(\tilde{\mathbf{A}}_j) = n-1$. We refer to the above equations as MQS estimating equations.

We are yet to specify the weighting matrices $\mathbf{W}_j$. In the present study, we consider two particular choices. The two choices represent different ways of approximating the optimal $\hat{\mathbf{A}}_j$ and are related to the HE regression [15–21] and the LDSC [6, 22, 23], respectively.

Both weights are derived based on approximations to ensure $\tilde{\mathbf{A}}_j \approx \hat{\mathbf{A}}_j$ and thus the statistical efficiency of the estimates (approximation details are provided in section 8). Both approximations become exact for unrelated individuals or a trait with zero heritability.

## 3.1 HE Weights

The first set, which we refer to as the HE weights, assigns equal weights for all SNPs, or

$$w_{jl} = 1. \tag{9}$$

We refer to the variation of MQS under the HE weights as MQS-HEW. Surprisingly, MQS-HEW is equivalent to both MINQUE(0) [13, 14] and the HE regression (section 8). Because of the equivalence between MQS-HEW and HE, MQS-HEW is expected to provide unbiased estimates for case control studies as shown before [24]. Compared with HE and MINQUE(0), however, our MQS formulation allows the use of summary statistics to compute both the point estimates and the standard errors (see below).

### 3.2 LD Weights

The second set, which we refer to as the LD weights, takes the following form

$$w_{jl} = (\sum_i (n-1)/p_i l_{(jl\sim i)} \tilde{h}_i^2 + 1)^{-2}/c_j, \tag{10}$$

where $\tilde{h}_i^2$ is a pre-specified estimate of $h_i^2$; $l_{(jl\sim i)}$ is the LD score of SNP $jl$ with respect to all SNPs (including itself if $j = i$) in $i$th category; and $c_j$ is a normalizing constant of $j$th category to ensure that $\sum_l w_{jl} = p_j$. In practice, we use the variance component estimates from MQS-HEW as $\tilde{h}_i^2$, but restricted them to be between 0 and 1 for algorithm stability. The LD score here is defined to be the summation of squared correlations between the $jl$th SNP and all SNPs in the $i$th category minus the expectation under the null, or $l_{(jl\sim i)} = \sum_{m=1}^{p_i} (r_{jl,im}^2 - 1/(n-1))$.

We refer to the variation of MQS under the LD weights as MQS-LDW. When $k = 1$ and when the LD scores are exact (i.e. computed based on its definition using all SNPs genome-wide), MQS-LDW is mathematically equivalent to LDSC without the intercept [6] (section 8). Due to computational reasons, however, LD scores cannot be computed exactly and can only be estimated using neighboring SNPs instead of all SNPs via a sliding window based approach [6]. Therefore, LD scores are always under-estimated in practice. Under-estimation of the LD scores have different impacts on MQS-LDW and LDSC. For MQS-LDW, because LD scores are only used to improve statistical efficiency of the estimates, the MQS-LDW estimates obtained using the estimated LD scores are still unbiased but are less statistically efficient than that based on exact LD scores. In contrast, LDSC uses LD scores to both increase statistical efficiency and to approximate $tr(\mathbf{K}_i\tilde{\mathbf{A}}_j)$ and account for SNP correlations. Approximating $tr(\mathbf{K}_i\tilde{\mathbf{A}}_j)$ with under-estimated LD scores will result in underestimation of SNP correlations and thus over-estimation of the variance components, which we have shown in the main text.

When $k > 1$, MQS-LDW is no longer equivalent to the stratified LDSC [23] even with exact LD scores. Unlike $k = 1$, it is no longer possible to convert MQS-LDW estimating equations to a set of linear equations that relate per-SNP z-scores to per-SNP LD scores as in the stratified LDSC.

## 4 Point Estimates and Confidence Intervals

With the MQS estimating equations 7 and 8, it is straightforward to obtain variance component estimates. To do so, we subtract equation 8 from each equation in 7, solve the resulting $k$ equations to estimate $\boldsymbol{\sigma}^2$, and plug in the estimated $\boldsymbol{\sigma}^2$ to equation 8 to estimate $\sigma_{k+1}^2$. The resulting MQS estimates are in a simple, closed-form solution

$$\hat{\boldsymbol{\sigma}}^2 = \mathbf{S}^{-1}\mathbf{q}, \tag{11}$$

$$\hat{\sigma}_{k+1}^2 = s_y^2 - \sum_{i=1}^{k} \hat{\sigma}_i^2, \tag{12}$$

where the elements in the $k$-vector $\mathbf{q}$ and the $k$ by $k$ matrix $\mathbf{S}$ are

$$\mathbf{q}_i = \mathbf{y}^T(\tilde{\mathbf{A}}_i - tr(\tilde{\mathbf{A}}_i)/tr(\mathbf{M}))\mathbf{y}/(n-1)^2 = \mathbf{y}^T(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{y}/(n-1)^2, \tag{13}$$

$$\mathbf{S}_{ij} = (tr(\tilde{\mathbf{A}}_j\mathbf{K}_i) - tr(\tilde{\mathbf{A}}_j)tr(\mathbf{K}_i)/tr(\mathbf{M}))/(n-1)^2 = tr(\tilde{\mathbf{A}}_j\mathbf{K}_i)/(n-1)^2 - 1/(n-1). \tag{14}$$

The variance for the estimates are

$$V(\hat{\boldsymbol{\sigma}}^2) = \mathbf{S}^{-1}V(\mathbf{q})\mathbf{S}^{-1}, \tag{15}$$

$$V(\hat{\sigma}^2_{k+1}) = \mathbf{1}_k^T\mathbf{S}^{-1}V(\mathbf{q})\mathbf{S}^{-1}\mathbf{1}_k, \tag{16}$$

where $\mathbf{1}_k$ is a $k$-vector of 1s and the $k$ by $k$ covariance matrix $V(\mathbf{q})$ is

$$V(\mathbf{q})_{ij} = V(\mathbf{y}^T(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{y}, \mathbf{y}^T(\tilde{\mathbf{A}}_j - \mathbf{I})\mathbf{y}) = 2tr(\mathbf{H}(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{H}(\tilde{\mathbf{A}}_j - \mathbf{I}))/(n-1)^4, \tag{17}$$

since $V(\mathbf{y}) = \mathbf{H}$.

The above form of $V(\mathbf{q})$ requires cubic operations to compute. To speed up computation, we instead consider a novel approximation to $V(\mathbf{q})$

$$V(\mathbf{q})_{ij} \approx 2\mathbf{y}^T(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{H}(\tilde{\mathbf{A}}_j - \mathbf{I})\mathbf{y}/(n-1)^4, \tag{18}$$

which is the asymptotic form we use in the main text. The above approximation is based on replacing the expectation $H = E(\mathbf{y}\mathbf{y}^T)$ with its realized value $\mathbf{y}\mathbf{y}^T$. The approximation is motivated by the average information (AI) algorithm [25] and effectively uses a realized information matrix in place of the expected information matrix. Our approximation not only makes computation $n$ times faster than the usual MoM, but also allows the use of summary statistics in computing standard errors (section 6). As we have shown in the main text, the standard errors constructed based on this approximation can be used to form a normal test for testing $H_0 : \sigma^2 = 0$ at the usual significance levels. However, when the sample size is small or when we want to obtain calibrated test statistics at higher significance levels (e.g. at $\alpha$ levels below 0.01), then we will need to use a mixture of chi-square distributions to obtain more calibrated $p$-values, as has been done in [26]. We do not explore this issue further here.

# 5 Second Approximation: Estimating S via Subsampling

Up to now, the total computational complexity of MQS scales quadratically with respect to the sample size and linearly with respect to the number of SNPs, or equivalently, on the order of $O(pn^2)$. Specifically, it takes $O(pn)$ time to compute $\mathbf{q}$, $O(pn^2)$ time to compute $\mathbf{S}$, and $O(n^2)$ time to compute $V(\mathbf{q})$. Because the most computationally expensive part is the computation of $\mathbf{S}$, we consider estimating $\mathbf{S}$ instead of computing it. Specifically, we consider using $m$ randomly selected individuals from the full sample to estimate $\mathbf{S}$, or

$$\hat{\mathbf{S}}_{ij} = (tr(\tilde{\mathbf{A}}'_j\mathbf{K}'_i) - tr(\tilde{\mathbf{A}}'_j)tr(\mathbf{K}'_i)/tr(\mathbf{M}'))/(m-1)^2, \quad \tilde{\mathbf{A}}'_j = \mathbf{X}'_j\mathbf{W}\mathbf{X}'^T_j/p_j, \quad \tilde{\mathbf{K}}'_i = \mathbf{X}'_i\mathbf{X}'^T_i/p_i. \tag{19}$$

where $\mathbf{X}'_i$ is the standardized genotype matrix for the subset of individuals and $\mathbf{M}' = \mathbf{I} - \mathbf{1}_m\mathbf{1}_m/m$ is the projecting matrix onto the null space of the intercept for this subset of individuals.

Using the estimated $\hat{\mathbf{S}}$ in place of $\mathbf{S}$ reduces computational complexity from $O(pn^2)$ to $O(pm^2)$, resulting in an overall computational complexity of $O(pn + pm^2)$ for MQS. In addition, using $\hat{\mathbf{S}}$ allows us to apply MQS to data from many consortium studies: we can pair $\mathbf{q}$ computed from the complete data with $\hat{\mathbf{S}}$ estimated from a random subsample of the study. When such a random subsample of the study is not available, we can also use a reference panel, such as the 1,000 genomes project [27], to estimate $\mathbf{S}$, so long as individuals in the reference panel can be viewed as a subsample of the study (e.g. of the same ethnic origin).

To account for the extra variance introduced by using $\hat{\mathbf{S}}$ instead of $\mathbf{S}$, we adjust the variance for the estimates by the Delta method

$$V(\hat{\boldsymbol{\sigma}}^2) = \hat{\mathbf{S}}^{-1}(V(\mathbf{q}) + \mathbf{D}(\hat{\boldsymbol{\sigma}}^2)V(\hat{\mathbf{S}})\mathbf{D}(\hat{\boldsymbol{\sigma}}^2))\hat{\mathbf{S}}^{-1}, \tag{20}$$

$$V(\hat{\sigma}^2_{k+1}) = \mathbf{1}^T\hat{\mathbf{S}}^{-1}(V(\mathbf{q}) + \mathbf{D}(\hat{\boldsymbol{\sigma}}^2)V(\hat{\mathbf{S}})\mathbf{D}(\hat{\boldsymbol{\sigma}}^2))\hat{\mathbf{S}}^{-1}\mathbf{1}, \tag{21}$$

where $\mathbf{D}(\hat{\boldsymbol{\sigma}}^2)$ denotes a $k$ by $k$ diagonal matrix with $i$th diagonal element $\hat{\sigma}^2_i$. We compute $V(\hat{\mathbf{S}})$ via Jackknife [28]. Specifically, we first compute $\mathbf{K}'_i$, $\tilde{\mathbf{A}}'_i$ and $\hat{\mathbf{S}}$. Then, we remove one individual at a time and use the corresponding submatrix $\mathbf{K}'_{m-1}$ and $\tilde{\mathbf{A}}'_{m-1}$ to compute $\hat{\mathbf{S}}_{m-1}$. Finally, we estimate the element-wise variance of $\hat{\mathbf{S}}$ as $V(\hat{\mathbf{S}})$. Because the second step of computing $\hat{\mathbf{S}}_{m-1}$ re-uses many quantities that are available from computing $\hat{\mathbf{S}}$, the overall complexity of estimating $V(\hat{\mathbf{S}})$ is only $O(m^3)$.

Our reasoning behind estimating $\mathbf{S}$ stems from the following alternative representation of the MQS solution

$$\mathbf{q}_i = \frac{1}{p_i}\sum_{l=1}^{p_i} w_{il} r^2_{il,y} - \frac{s^2_y}{n-1}, \tag{22}$$

$$\mathbf{S}_{ij} = \frac{1}{p_i p_j}\sum_{l=1}^{p_i}\sum_{l'=1}^{p_j} w_{il} r^2_{il,jl'} - \frac{1}{n-1}, \tag{23}$$

where $r_{il,y} = \mathbf{x}^T_{jl}\mathbf{y}/(n-1)$ is the correlation between phenotype and the genotype of $l$th SNP in $i$th category, and $r_{il,jl'} = \mathbf{x}^T_{il}\mathbf{x}_{jl'}/(n-1)$ is the genotype correlation between $l$th SNP in $i$th category and $l'$th SNP in $j$th category. Intuitively, each element of $\mathbf{S}$ is a weighted average of $p_i p_j$ different terms of $\hat{r}^2_{il,jl'}$ while each element of $\mathbf{q}$ is a weighted average of only $p_i$ different terms of $\hat{\beta}^2_{il}$. Therefore, $\mathbf{S}$ is much easier to be estimated accurately than $\mathbf{q}$.

Besides the intuitive explanation, we also provide two formal arguments to support the subsampling approach in MQS (details in section 10 with simulation details in section 11). We use MQS-HEW for illustration. Our first argument is that the variance of $\mathbf{S}$ estimated by using $m$ individuals, or $V(\mathbf{S})$, is often small compared with $\mathbf{S}$ itself. Because of this, the Delta method approximation will be accurate and $\mathbf{S}^{-1}$ can be estimated well by using $\hat{\mathbf{S}}^{-1}$. Specifically, when $k = 1$, for independent SNPs, $S$ is expected to be $1/p$ while $\sqrt{V(\hat{S})}$ is expected to be $2/(mp)$ under the null, where $p$ is the number of SNPs. Simulations with independently and binomially distributed SNPs confirm these relationship (Figure S1 and Table S1). For SNPs with LD, we can use the effective number of independent SNPs in place of $p$ to provide approximate forms $E(S) \approx 1/p'$ and $\sqrt{E(V(\hat{S}))} \approx 2/(mp')$. Simulations with real data confirm these approximate relationship, though the approximation can be less accurate with very small $m$ (Figures S2, S3 and Table S1).

Our second argument is that $V(\hat{\mathbf{S}})$ is also small compared with $V(\mathbf{q})$. Because of this, the variance of $\hat{\boldsymbol{\sigma}}^2$ and $\hat{\sigma}^2_{k+1}$ is dominated by $V(\mathbf{q})$ and estimating $\mathbf{S}$ does not introduce much extra variance. Specifically, when $k = 1$, for independent SNPs, $\sqrt{E(V(q))} = \sqrt{2}/(n\sqrt{p})$ under the null (Supplementary Text), which is $\sqrt{2p}m/n$ times larger than $\sqrt{V(\hat{S})}$. Simulations with independently and binomially distributed SNPs confirm this relationship (Figure S1). For SNPs with LD, we use the effective number of independent SNPs in place of $p$ to provide an approximate form $\sqrt{E(V(q))} \approx \sqrt{2}/(n\sqrt{p'})$. Simulations with real data again confirm the approximate form (Figures S2, S3).

The two arguments above represent the first attempt to understand the behavior of subsampling strategy and the reference panel idea that has been widely used in genetics [6, 29–33]. However, we note that the two arguments are by no means complete. For example, we have focused on $k = 1$ here. For $k > 1$, the number of variance components $k$ as well as the effective number of independent SNPs in each

category will both play a role in determining the size of $m$. A larger $m$ is likely required to ensure accurate estimation of $\mathbf{S}^{-1}$ as well as a small $V(\hat{\mathbf{S}})$ with respect to $V(\mathbf{q})$. In addition, we have focused only on MQS-HEW. MQS-LDW uses an iterative procedure that makes it harder for a thorough investigation. However, we have provided simulations as well as real data applications in the main text to show that estimating $\mathbf{S}$ does work well for either $k = 1$ or $k > 1$ and for both MQS-HEW and MQS-LDW.

# 6 Estimation with Summary Statistics

We are now ready to describe the details of MQS estimation when we only have summary statistics. In this case, we require marginal z-scores from the complete data and individual-level genotypes from a subset of individuals (or from a separate reference panel whose individuals can be thought of as a subset of the study sample). When we only have summary statistics and when the phenotype variance $s_y^2$ is unknown, $\boldsymbol{\sigma}^2$ and $\sigma_{k+1}^2$ are not estimable but their scaled versions $\mathbf{h}^2$ and $h_{k+1}^2$ are.

To estimate $\mathbf{h}^2$ and $h_{k+1}^2$, we first use the marginal z-scores to approximate $\mathbf{q}$

$$\mathbf{q}_i/s_y^2 \approx \frac{1}{n-1}\frac{1}{p_i}\sum_{l=1}^{p_i} w_{il}(\hat{z}_{il}^2 - 1). \tag{24}$$

The approximate assumes that the marginal variant effect size is small, which holds well for most GWASs. Next, we use the individual-level genotype data from a subsample of data or a separate reference panel to compute $\hat{\mathbf{S}}$ (equation 19). With $\mathbf{q}/s_y^2$ and $\hat{\mathbf{S}}$, we estimate $\mathbf{h}^2$ and $h_{k+1}^2$ with equations 11 and 12.

To compute the variance for the estimates, we consider two alternative strategies. The first strategy relies on the asymptotic form, is accurate but requires summary statistics in addition to the marginal $z$ scores. This strategy is based on an alternative expression of the equation 18

$$\begin{aligned} V(\mathbf{q})_{ij}/s_y^2 \approx 2\sum_{l=1}^{k} \hat{h}_l^2 (&\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_l\mathbf{X}_l^T\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/(p_ip_jp_l) - \mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_l\mathbf{z}_l/(p_ip_l) \\ &- \mathbf{z}_l^T\mathbf{X}_l\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/(p_jp_l) + \mathbf{z}_l^T\mathbf{z}_l/p_l)/(n-1)^3 \\ &+ 2\hat{h}_{k+1}^2(\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/p_ip_j - \mathbf{z}_i^T\mathbf{W}_i\mathbf{z}_i/p_i - \mathbf{z}_j^T\mathbf{W}_j\mathbf{z}_j/p_j + 1)/(n-1)^3 \end{aligned} \tag{25}$$

where $\mathbf{z}_i$ is a $p_i$ vector of marginal z-scores for SNPs in $i$th category. Thus, if we can compute additional summary statistics – specifically the $n$ by $k$ matrix of $\mathbf{X}_i\mathbf{z}_i$, the $n$ by $k$ matrix of $\mathbf{X}_i\mathbf{W}_i\mathbf{z}_i$, and the $p$ by $k$ matrix of $\mathbf{X}^T\mathbf{X}_i\mathbf{W}_i\mathbf{z}_i$ – then we can compute the standard errors. These additional summary statistics are easy to compute; in consortium studies, they can be computed within each substudy and then combined across studies without sharing individual level data. Importantly, for MQS-HEW, we only need to compute two of these three matrices, $\mathbf{X}_i\mathbf{z}_i$ and $\mathbf{X}^T\mathbf{X}_i\mathbf{z}_i$. These two matrices do not require $\mathbf{W}$, which is a function of variance components in MQS-LDW. Thus, MQS-HEW can be much more convenient than MQS-LDW. Finally, we note that, although it is tempting to use quantities computed from a subset of individuals to estimate the confidence interval, we find that $\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i'^T\mathbf{X}_l'\mathbf{X}_l'^T\mathbf{X}_j'\mathbf{W}_j\mathbf{z}_j/((m-1)^2)$ from $m$ individuals is not a good estimate of $\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_l\mathbf{X}_l^T\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/((n-1)^2)$.

The second strategy relies on jackknife but does not require extra summary statistics. It is based on the strategy used in LDSC [6]. It works well when SNPs are approximately block-wise independent but can work poorly otherwise. This second strategy is based on the observation that the covariance function $V(\mathbf{q})$ as a function of $\mathbf{y}$ can be written as a function of the marginal z-scores $\mathbf{z} = (\mathbf{z}_1, \cdots, \mathbf{z}_k)^T$. $\mathbf{z}_i \approx \mathbf{X}_i^T\mathbf{y}/(s_y\sqrt{n-1})$ follows approximately a degenerate multivariate normal distribution $\mathbf{z}_i \sim \text{MVN}(0, \mathbf{X}_i^T\mathbf{H}\mathbf{X}_i/(s_y^2(n-1)))$ with a low rank covariance matrix. When SNPs are independent (i.e. $\mathbf{X}_i^T\mathbf{H}\mathbf{X}_i$ is diagonal) or when $\mathbf{X}_i^T\mathbf{H}\mathbf{X}_i$ is block-diagonal, we can use block-wise permutation of

z-scores to estimate $V(\mathbf{q})$ as in LDSC [6]. However, as we show in the main text, when LD pattern is complicated, jackknife can yield untrustworthy confidence intervals.

Overall, we recommend the use of the asymptotic form. However, we recognize that the asymptotic form requires additional summary statistics that may not be readily available from many studies at the moment. Thus, we have also implemented jackknife in the software as a useful practical option.

# 7 Extensions of MQS

Although we have focused on the specific model outlined in equation 2, we note that it is straightforward to generalize our model to incorporate other modeling assumptions.

We can generalize the model to incorporate other prior assumptions on the dependence of effect sizes on the MAFs (and other variant information), as long as the dependency is linear in $\sigma_i^2$ (i.e. in the form of $g(f_l)\sigma_i^2/p_i$, where $f_l$ is the MAF for the $l$th variant in the $i$th category). Such generalization can be achieved by replacing the usual genetic relatedness matrix $\mathbf{K}_i$ with a corresponding weighted genetic relatedness matrix with variant weights depending on $g(f_l)$.

We can generalize our model to incorporate overlapping categories. If $l$th variant belongs to $k_l$ different categories, we can assume its effect size variance to be a weighted function of the variances of each category, or $\sum_{i=1}^{k_l}\sigma_i^2/(p_i k_l)$. With this assumption, we can again generalize our model by replacing the genetic relatedness matrix $\mathbf{K}_i$ with a corresponding weighted genetic relatedness matrix with variance weights $1_{l\in i}/k_l$, where the indicator function $1_{l\in i}$ equals one when the $l$th variant belongs to the $i$th category and equals zero otherwise.

Finally, we can also generalize our model to incorporate other covariates such as genotype PCs. Specifically, we can consider the following model

$$\mathbf{y} = \mathbf{W}\boldsymbol{\alpha} + \sum_{i=1}^{k}\mathbf{g}_i + \boldsymbol{\epsilon}, \quad \mathbf{g}_i \sim \mathrm{MVN}(0, \sigma_i^2\mathbf{K}_i), \quad \boldsymbol{\epsilon} \sim \mathrm{MVN}(0, \sigma_{k+1}^2\mathbf{I}), \tag{26}$$

where, in addition to the previous notations, $\mathbf{W}$ is an $n$ by $c$ matrix of covariates including the intercept, and $\boldsymbol{\alpha}$ is a $c$-vector of corresponding coefficients. To estimate variance components, we multiple on both sides of the equation the projection matrix $\mathbf{M} = \mathbf{I} - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$. This leads to

$$\mathbf{y}^* = \sum_{i=1}^{k}\mathbf{g}_i^* + \boldsymbol{\epsilon}^*, \quad \mathbf{g}_i^* \sim \mathrm{MVN}(0, \sigma_i^2\mathbf{M}\mathbf{K}_i\mathbf{M}), \quad \boldsymbol{\epsilon}^* \sim \mathrm{MVN}(0, \sigma_{k+1}^2\mathbf{M}), \tag{27}$$

where $\mathbf{y}^* = \mathbf{M}\mathbf{y}$, $\mathbf{g}_i^* = \mathbf{M}\mathbf{g}$ and $\boldsymbol{\epsilon}^* = \mathbf{M}\boldsymbol{\epsilon}$ are the transformed values. After the transformation, variance component estimation can be proceeded as usual. Note that $tr(\mathbf{M}) = n - c$ will replace $n - 1$ in all the estimation solutions.

# 8 Relationship to Other Methods

## MQS-HEW and the Haseman-Elston Regression

We first show that the HE weights are used to ensure $\tilde{\mathbf{A}}_j \approx \hat{\mathbf{A}}_j$. Specifically, when $\mathbf{H} \approx \mathbf{M}$ (or more precisely a scaled version of $\mathbf{M}$), $\tilde{\mathbf{A}}_j$ with HE weights is approximately $\hat{\mathbf{A}}_j$, as $\tilde{\mathbf{A}}_j = \mathbf{X}_j\mathbf{X}_j^T/p_j = \mathbf{K}_j \approx \mathbf{H}^{-1}\mathbf{K}_j\mathbf{H}^{-1} = \hat{\mathbf{A}}_j$. Clearly, the main approximation we used ($\mathbf{H} \approx \mathbf{M}$) is expected to be accurate either when the trait has a low heritability (i.e. $\mathbf{h}^2 = \mathbf{0}$ and $h_{k+1}^2 = 1$) or when individuals are unrelated (i.e. $\mathbf{K}_j \approx \mathbf{M}$). Thus, we expect the MQS-HEW estimates to be accurate for a data with unrelated individuals or a trait with low-heritability.

We next show that MQS-HEW is mathematically equivalent to the HE cross-product regression. This relationship is surprising because the HE regression is formulated as relating the phenotype correlation to the genotype correlation. Specifically, the HE regression is based on

$$E(y_l y_{l'}) = \sum_{i=1}^{k} (\mathbf{K}_i)_{ll'} \sigma_i^2 + \sigma_e^2, \tag{28}$$

for all pairs of $l$th and $l'$th individuals, where $(\mathbf{K}_i)_{ll'}$ denotes $ll'$th element of matrix $\mathbf{K}_i$. Plug in the realized value of the LHS leads to $n^2$ equations

$$\text{vec}(\mathbf{y}\mathbf{y}^T) \sim \sum_{i=1}^{k} \text{vec}(\mathbf{X}_i \mathbf{X}_i^T / p_i)\sigma_i^2 + \text{vec}(\mathbf{I})\sigma_e^2, \tag{29}$$

where *vec* denotes matrix vectorization by stacking columns. Above, we loosely use $\sim$ to denote a linear regression. The estimates from the above model are identical to that from MQS-HEW with computed $\mathbf{S}$, because of standard algebra properties

$$\text{vec}^T(\mathbf{X}_i^T \mathbf{X}_i)\text{vec}(\mathbf{X}_j \mathbf{X}_j^T) = \text{tr}(\mathbf{X}_i \mathbf{X}_i^T \mathbf{X}_j \mathbf{X}_j^T), \tag{30}$$

$$\text{vec}^T(\mathbf{X}_i^T \mathbf{X}_i)\text{vec}(\mathbf{y}\mathbf{y}^T) = \mathbf{y}^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{y}. \tag{31}$$

## 8.1 MQS-LDW and the LD Score Regression

We first show that the LD weights are used to ensure $\tilde{\mathbf{A}}_j \approx \hat{\mathbf{A}}_j$. To see this, let us denote $\mathbf{\Omega} = \sum_i \sigma_i^2 \mathbf{X}_j^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{X}_j / ((n-1)p_i) + \sigma_{k+1}^2 \mathbf{I}$. First, when $\mathbf{X}_j^T \mathbf{X}_j/(n-1) \approx \mathbf{I}$, we have $\mathbf{X}_j^T \mathbf{X}_j \mathbf{\Omega} \approx \mathbf{X}_j^T \mathbf{H} \mathbf{X}_j$. Thus $\mathbf{X}_j \mathbf{\Omega} \approx \mathbf{H} \mathbf{X}_j$, or equivalently $\mathbf{H}^{-1} \mathbf{X}_j \approx \mathbf{X}_j \mathbf{\Omega}^{-1}$. Next, we approximate $\mathbf{\Omega}$ with its diagonal elements, or $\mathbf{\Omega} \approx s_y^2 diag(w_{jl}^{-1/2})$. With these two approximations, we have $\tilde{\mathbf{A}}_j = \mathbf{X}_j \text{diag}(w_{jl})/(p_j s_y^4) \mathbf{X}_j^T \approx \mathbf{H}^{-1} \mathbf{K}_j \mathbf{H}^{-1} = \hat{\mathbf{A}}_j$. The first approximation is accurate when SNPs are independent while the second approximation is accurate either SNPs are independent or the trait has a low heritability. We thus expect the MQS-LDW estimates to be accurate for a data with independent SNPs or a trait with low-heritability. Moreover, the second approximation suggests that a non-diagonal form of $\mathbf{W}$ that uses a better approximation of $\mathbf{\Omega}^{-1}$ may provide more accurate estimates than the diagonal $\mathbf{W}$. Finally, compared with MQS-HEW, MQS-LDW uses LD information and relies on *a prior* set of $\tilde{h}_i^2$ from MQS-HEW. Thus, MQS-LDW may be more accurate than MQS-HEW in certain situations. However, these two differences also make it harder to investigate whether the MQS-LDW estimates are also unbiased for case control studies.

We next show that, when $k = 1$, MQS-LDW is mathematically equivalent to LDSC when the LD scores are exact. This relationship is also surprising because the LDSC is formulated as $p$ equations that relate per-SNP z-scores to per-SNP LD scores. For $k = 1$, the LDSC is based on

$$z_l^2 \sim (n-1)/p l_l h^2 + 1, \tag{32}$$

where $l_l$ is the LD score for $l$th SNP and we loosely use $\sim$ again to denote a linear regression. We refer to the equation above as the naive LDSC because the actual LDSC applies additional SNP weights and uses the weighted least squares to obtain these estimates. In particular, LDSC uses SNP weights

$$w_l c l_l^{-1} = ((n-1)/p l_l h^2 + 1)^{-2} l_l^{-1}. \tag{33}$$

Because $tr(\mathbf{K}\tilde{\mathbf{A}}) = (n-1)^2/p^2 \sum_l w_l l_l + (n-1)/p$, $\mathbf{y}^T \mathbf{X} \mathbf{W} \mathbf{X}^T \mathbf{y} = (n-1) s_y^2 \sum_l w_l z_l^2$, $\mathbf{y}^T \mathbf{y} = (n-1) s_y^2 \sum_l w_l / p$, the weighted least square estimates from LDSC are identical to that from MQS-LDW with computed $\mathbf{S}$. In practice, because LD scores cannot be computed exactly, $tr(\mathbf{K}\tilde{\mathbf{A}}) > (n-1)^2/p^2 \sum_l w_l \hat{l}_l +$

$(n-1)/p$ and the estimates from LDSC are always over-estimated. MQS-LDW uses $tr(\mathbf{K}\tilde{\mathbf{A}})$ directly and thus is an exact form of LDSC.

We note that there are two other minor differences between the LDSC introduced in [6] and our formulation above. First, LDSC uses $n$ instead of the correct $n-1$ in the two equations. Second, the LD scores defined in LDSC does not contain the correction term $-1/(n-1)$. The practice effects of these difference are expected to be small.

When $k > 1$, there is no equivalence between MQS-LDW and the stratified LDSC [23]. Unlike $k = 1$, it is no longer possible to convert MQS-LDW estimating equations into a set of linear equations that relate per-SNP z-scores to per-SNP LD scores. Indeed, the stratified LDSC has to rely on an *ad hoc* choice of SNP weights without clear justification.

# 9 Enrichment Score

Based on $\hat{\mathbf{h}}$, we can compute an enrichment score for each category, defined as

$$\boldsymbol{\rho} = (\rho_1, \cdots, \rho_k)^T = (\mathbf{1}^T\mathbf{D}_p\mathbf{1})(\mathbf{1}^T\mathbf{h})^{-1}\mathbf{D}_p^{-1}\mathbf{h} \tag{34}$$

where $\mathbf{D}_p$ is a diagonal matrix with $i$th diagonal element $p_i$. By the Delta method, we have

$$\begin{aligned}
\hat{\boldsymbol{\rho}} &= (\mathbf{1}^T\mathbf{D}_p\mathbf{1})(\mathbf{1}^T\hat{\mathbf{h}})^{-1}\mathbf{D}_p^{-1}\hat{\mathbf{h}} \\
\frac{\partial\boldsymbol{\rho}}{\partial\mathbf{h}} &= (\mathbf{1}^T\mathbf{D}_p\mathbf{1})((\mathbf{1}^T\mathbf{h})^{-1}\mathbf{D}_p^{-1} - \mathbf{D}_p^{-1}\hat{\mathbf{h}}(\mathbf{1}^T\hat{\mathbf{h}})^{-2}\mathbf{1}^T) \\
V(\hat{\boldsymbol{\rho}}) &= (\mathbf{1}^T\mathbf{D}_p\mathbf{1})^2(\mathbf{1}^T\hat{\mathbf{h}})^{-2}\mathbf{D}_p^{-1}(\mathbf{I} - (\mathbf{1}^T\hat{\mathbf{h}})^{-1}\hat{\mathbf{h}}\mathbf{1}^T)V(\hat{\mathbf{h}})(\mathbf{I} - (\mathbf{1}^T\hat{\mathbf{h}})^{-1}\mathbf{1}\hat{\mathbf{h}}^T)\mathbf{D}_p^{-1}
\end{aligned}$$

# 10 Expectations of $V(q)$ and $V(\hat{S})$ for MQS-HEW

Here, we derive the expectations for $V(q)$ and $V(\hat{S})$ in MQS-HEW when $k = 1$. To do so, we make a simplified assumption that SNP genotypes are independently and identically distributed from a standard normal distribution. This normality assumption allows us to use Isserlis' theorem (also known as Wick's theorem) to obtain a rather simple (but non-trivial) result. Though simplistic, the normality assumption is useful to understand the behavior of $V(q)$ and $V(\hat{S})$.

To simplify notation, we denote

$$U = tr(\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{X}^T), \quad U' = tr(\mathbf{X}'\mathbf{X}'^T\mathbf{X}'\mathbf{X}'^T), \quad V = tr(\mathbf{X}\mathbf{X}^T), \quad V' = tr(\mathbf{X}'\mathbf{X}'^T), \tag{35}$$

where $\mathbf{X}$ is an $n$ by $p$ matrix, $\mathbf{X}'$ is an $m$ by $p$ submatrix of $\mathbf{X}$. Under the normality assumption, each row of $\mathbf{X}$ follows $MVN(0, \boldsymbol{\Sigma})$ with an identity $p$ by $p$ SNP covariance matrix $\boldsymbol{\Sigma}$.

Both $V(q)$ and $V(\hat{S})$ are functions of $U$, $U'$, $V$ and $V'$. To obtain the expectations of $V(q)$ and $V(\hat{S})$, we first obtain the expectation and variance of $U$ and $V$.

First, we have

$$\begin{aligned}
E(U) &= E(\sum_{i=1}^{p}\sum_{j=1}^{p}(\sum_{l=1}^{n}x_{li}^2x_{lj}^2 + 2\sum_{l\neq l'}x_{li}x_{lj}x_{l'i}x_{l'j})) \\
&= \sum_{i=1}\sum_{j=1}(nE(x_{li}^2x_{lj}^2) + n(n-1)E(x_{li}x_{lj})^2) \\
&= \sum_{i=1}\sum_{j=1}(n(\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj} + 2\boldsymbol{\Sigma}_{ij}^2) + n(n-1)\boldsymbol{\Sigma}_{ij}^2) \\
&= np^2 + (n^2+n)p.
\end{aligned} \tag{36}$$

and

$$
\begin{aligned}
E(U^2) &= E(\sum_{i,j}\sum_{i',j'}(\sum_{l=1}^{n} x_{li}x_{lj})(\sum_{l=1}^{n} x_{li}x_{lj})(\sum_{l=1}^{n} x_{li'}x_{lj'})(\sum_{l=1}^{n} x_{li'}x_{lj'}) \\
&= \sum_{i,j}\sum_{i',j'}(nE(x_{li}^2x_{lj}^2x_{li'}^2x_{lj'}^2) + n(n-1)(2E(x_{li}^2x_{lj}^2x_{li'}x_{lj'})E(x_{li'}x_{lj'}) + 2E(x_{li}x_{lj}x_{li'}^2x_{lj'}^2)E(x_{li}x_{lj})) \\
&+ n(n-1)(E(x_{li}^2x_{lj}^2)E(x_{li'}^2x_{lj'}^2) + 2E(x_{li}x_{lj}x_{li'}x_{lj'})^2) \\
&+ n(n-1)(n-2)(E(x_{li}^2x_{lj}^2)E(x_{li'}x_{lj'})^2 + E(x_{li}x_{lj})^2E(x_{li'}^2x_{lj'}^2) + 4E(x_{li}x_{lj}x_{li'}x_{lj'})E(x_{li}x_{lj})E(x_{li'}x_{lj'})) \\
&+ n(n-1)(n-2)(n-3)E(x_{li}x_{lj})^2E(x_{li'}x_{lj'})^2) \\
&= \sum_{i,j}\sum_{i',j'}(nE(x_{li}^2x_{lj}^2x_{li'}^2x_{lj'}^2) + n(n-1)(2E(x_{li}^2x_{lj}^2x_{li'}x_{lj'})\mathbf{\Sigma}_{i'j'} + 2E(x_{li}x_{lj}x_{li'}^2x_{lj'}^2)\mathbf{\Sigma}_{ij}) \\
&+ n(n-1)((\mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj} + 2\mathbf{\Sigma}_{ij}^2)(\mathbf{\Sigma}_{i'i'}\mathbf{\Sigma}_{j'j'} + 2\mathbf{\Sigma}_{i'j'}^2) + 2(\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{i'j'} + \mathbf{\Sigma}_{ii'}\mathbf{\Sigma}_{jj'} + \mathbf{\Sigma}_{ij'}\mathbf{\Sigma}_{ji'})^2) \\
&+ n(n-1)(n-2)((\mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj} + 2\mathbf{\Sigma}_{ij}^2)\mathbf{\Sigma}_{i'j'}^2 + \mathbf{\Sigma}_{ij}^2(\mathbf{\Sigma}_{i'i'}\mathbf{\Sigma}_{j'j'} + 2\mathbf{\Sigma}_{i'j'}^2) + 4(\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{i'j'} + \mathbf{\Sigma}_{ii'}\mathbf{\Sigma}_{jj'} + \mathbf{\Sigma}_{ij'}\mathbf{\Sigma}_{ji'})\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{i'j'}) \\
&+ n(n-1)(n-2)(n-3)\mathbf{\Sigma}_{ij}^2\mathbf{\Sigma}_{i'j'}^2) \\
&= n(p^4 + 12p^3 + 44p^2 + 48p) + 4n(n-1)(p^3 + 3\times 2\times p^2 + 4\times 2\times p) + n(n-1)(p^4 + 4p^3 + 4p^2 + 2(3p^2 + 6p)) \\
&+ n(n-1)(n-2)(2(p^3 + 2p^2) + 4(p^2 + 2p)) + n(n-1)(n-2)(n-3)p^2 \\
&= n^2p^4 + (2n^3 + 2n^2 + 8n)p^3 + (n^4 + 2n^3 + 21n^2 + 20n)p^2 + (8n^3 + 20n^2 + 20n)p. \qquad (37)
\end{aligned}
$$

Next, we have

$$
E(V) = E(\sum_{i}\sum_{l} x_{li}^2) = n\sum_{i}\mathbf{\Sigma}_{ii} = np, \qquad (38)
$$

and

$$
\begin{aligned}
E(V^2) &= E((\sum_{i}\sum_{l} x_{li}^2)^2) = \sum_{i}\sum_{j} E((\sum_{l} x_{li}^2)(\sum_{l} x_{lj}^2)) \\
&= \sum_{i}\sum_{j}(nE(x_{li}^2x_{lj}^2) + n(n-1)E(x_{li}^2)E(x_{lj}^2)) \\
&= \sum_{i}\sum_{j}(n\mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj} + 2n\mathbf{\Sigma}_{ij}^2 + n(n-1)\mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj}) \\
&= n^2p^2 + 2np, \qquad (39)
\end{aligned}
$$

and

$$
\begin{aligned}
E(V^4) &= E((\sum_{i}\sum_{l} x_{li}^2)^4) = \sum_{i}\sum_{j}\sum_{i'}\sum_{j'} E((\sum_{l} x_{li}^2)(\sum_{l} x_{lj}^2)(\sum_{l} x_{li'}^2)(\sum_{l} x_{lj'}^2)) \\
&= \sum_{i}\sum_{j}\sum_{i'}\sum_{j'}(nE(x_{li}^2x_{lj}^2x_{li'}^2x_{lj'}^2) + 4n(n-1)E(x_{li}^2x_{lj}^2x_{li'}^2)E(x_{lj'}^2) \\
&+ 3n(n-1)E(x_{li}^2x_{lj}^2)E(x_{li'}^2x_{lj'}^2) + 6n(n-1)(n-2)E(x_{li}^2x_{lj}^2)E(x_{li'}^2)E(x_{lj'}^2) \\
&+ n(n-1)(n-2)(n-3)E(x_{li}^2)E(x_{lj}^2)E(x_{li'}^2)E(x_{lj'}^2) \\
&= n(p^4 + 12p^3 + 44p^2 + 48p) + 4n(n-1)p(p^3 + 6p^2 + 8p) + 3n(n-1)(p^2 + 2p)^2 \\
&+ 6n(n-1)(n-2)p^2(p^2 + 2p) + n(n-1)(n-2)(n-3)p^4 \\
&= n^4p^4 + 12n^3p^3 + 44n^2p^2 + 48np. \qquad (40)
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
E(UV^2) &= E((\sum_i \sum_j (\sum_l x_{li}x_{lj})^2)(\sum_i (\sum_l x_{li}^2))^2) = \sum_i \sum_j \sum_{i'} \sum_{j'} E((\sum_l x_{li}x_{lj})^2 (\sum_l x_{li'}^2)(\sum_l x_{lj'}^2)) \\
&= \sum_i \sum_j \sum_{i'} \sum_{j'} (nE(x_{li}^2 x_{lj}^2 x_{li'}^2 x_{lj'}^2) + n(n-1)(2E(x_{li}^2 x_{lj}^2 x_{li'}^2)E(x_{lj'}^2) + 2E(x_{li}x_{lj}x_{li'}^2 x_{lj'}^2)E(x_{li}x_{lj})) \\
&+ n(n-1)(E(x_{li}^2 x_{lj}^2)E(x_{li'}^2 x_{lj'}^2) + 2E(x_{li}x_{lj}x_{li'}^2)E(x_{li}x_{lj}x_{lj'}^2)) \\
&+ n(n-1)(n-2)(E(x_{li}^2 x_{lj}^2)E(x_{li'}^2)E(x_{lj'}^2) + 4E(x_{li}x_{lj}x_{li'}^2)E(x_{li}x_{lj})E(x_{lj'}^2) + E(x_{li}x_{lj})^2 E(x_{li'}^2 x_{lj'}^2)) \\
&+ n(n-1)(n-2)(n-3)E(x_{li}x_{lj})^2 E(x_{li'}^2)E(x_{lj'}^2) \\
&= n(p^4 + 12p^3 + 44p^2 + 48p) + n(n-1)(p^3 + 6p^2 + 8p)(2p+2) \\
&+ n(n-1)((p^2+2p)^2 + 2(p^3 + 4p^2 + 4p)) \\
&+ n(n-1)(n-2)(p^4 + 2p^3 + 5(p^2+2p)p) + n(n-1)(n-2)(n-3)p^3 \\
&= n^3p^4 + (n^4 + n^3 + 10n^2)p^3 + (10n^3 + 10n^2 + 24n)p^2 + (24n^2 + 24n)p. \qquad (41)
\end{aligned}
$$

Above, we have used the following properties based on the Isserlis' theorem

$$
\begin{aligned}
E(x_{li}x_{lj}) &= \mathbf{\Sigma}_{ij} = p, \qquad (42) \\
E(x_{li}^2 x_{lj}^2) &= \mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj} + 2\mathbf{\Sigma}_{ij}^2 = p^2 + 2p, \qquad (43) \\
E(x_{li}x_{lj}x_{li'}x_{lj'}) &= \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{i'j'} + \mathbf{\Sigma}_{ii'}\mathbf{\Sigma}_{jj'} + \mathbf{\Sigma}_{ij'}\mathbf{\Sigma}_{ji'} = 3p^2, \qquad (44) \\
E(x_{li}^2 x_{lj}^2 x_{li'}x_{lj'}) &= \mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj}\mathbf{\Sigma}_{i'j'} + 2\mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{ji'}\mathbf{\Sigma}_{jj'} + 2\mathbf{\Sigma}_{jj}\mathbf{\Sigma}_{ij'}\mathbf{\Sigma}_{ii'} \\
&\quad + 2\mathbf{\Sigma}_{ij}^2\mathbf{\Sigma}_{i'j'} + 4\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{ii'}\mathbf{\Sigma}_{jj'} + 4\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{ij'}\mathbf{\Sigma}_{ji'} \\
&= p^3 + 2\times 2\times p^2 + (2\times 2 + 2\times 3)\times p = p^3 + 4p^2 + 10p, \qquad (45) \\
\sum_{i,j,i',j'} E(x_{il}^2 x_{jl}^2 x_{i'l}^2 x_{j'l}^2) &= p^4 + \binom{4}{2}\times 2\times p^3 + \binom{4}{1}\times 4\times 2\times p^2 + \binom{4}{2}\times 2\times p^2 + 6\times 4\times 2p \\
&= p^4 + 12p^3 + 44p^2 + 48p. \qquad (46)
\end{aligned}
$$

Now we are ready to evaluate $E(V(q))$ and $E(V(\hat{S}))$. First, we have

$$
E(V(q)) = E(2tr((\mathbf{K}-\mathbf{I})(\mathbf{K}-\mathbf{I}))/n^4) = 2E(U/p^2 - 2V/p + n)/n^4 = O(\frac{2}{n^2 p}) \qquad (47)
$$

where we use $n$ instead of $n-1$ to reflect the fact that genotypes are not centered here.

Next, because $\hat{S} = U'/V'^2$, with Delta method, we have

$$
E(V(\hat{S})) = \frac{E(U')^2}{E(V'^2)^2}\frac{E(U'^2)E(V'^2)^2 + E(V'^4)E(U')^2 - E(U'V'^2)E(U')E(V'^2)}{E(U')^2 E(V'^2)^2} = O(\frac{4}{m^2 p^2}). \qquad (48)
$$

# 11 Simulations to Validate $E(V(q))$ and $E(V(\hat{S}))$

Because deriving the equations 47 and 48 involves complicated algebra, we verify the correctness of the above expectations using simulations.

First, we perform four sets of simulations using simulated genotypes. Specifically, for each SNP, we first simulate its minor allele frequency from a uniform distribution between 0.1 and 0.5. Based on the minor allele frequency, we simulate genotypes from a binomial distribution.

The first two sets of simulations are used to validate $E(V(q)) = O(\frac{2}{n^2 p})$. We first fix $p = 10,000$. For each $n \in \{100, 200, \cdots, 1,000\}$, we simulate one $\mathbf{X}$ and 10,000 $\mathbf{y}$ to obtain the sample variance $V(q)$. Each $\mathbf{y}$ is from a standard normal distribution. We use this set of simulations to show that $n = \sqrt{2/(V(q)p)}$ (Figure S1a). Next, we fix $n = 1,000$. For each $p \in \{1,000, 2,000, \cdots, 10,000\}$, we simulate one $\mathbf{X}$ and 10,000 $\mathbf{y}$ to obtain the sample variance $V(q)$. We use this set of simulations to show that $p = 2/(V(q)n^2)$ (Figure S1b).

The second two sets of simulations are used to validate $E(V(\hat{S})) = O(\frac{4}{m^2 p^2})$. We first fix $n = 1,000$, $p = 10,000$ and simulate one $\mathbf{X}$. For each value of $m \in \{10, 20, \cdots, 100\}$, we draw 10,000 $\mathbf{X}'$ from $\mathbf{X}$ to obtain the sample variance $V(\hat{S})$. We use this set of simulations to show that $n = 2/\sqrt{(V(\hat{S})p^2)}$ (Figure S1c). In addition, the sample mean of $\hat{S}$ computed with a small $m$ is a good estimate of $S$ (Table S1). Next, we fix $n = 1,000$ and $m = 100$. For each $p = \{1,000, 2,000, \cdots, 10,000\}$, we simulate one $\mathbf{X}$ and draw 10,000 $\mathbf{X}'$ from $\mathbf{X}$ to obtain the sample variance $V(\hat{S})$. We use this set of simulations to show that $p = 2/\sqrt{(V(\hat{S})n^2)}$ (Figure S1d).

Next, we perform three additional sets of simulations using real genotypes from either the Australia data or the Finland data. In the Australia data, we order all SNPs ($p = 4,352,968$) by their genomic location, and select from these SNPs 319,147 evenly spaced SNPs for simulations. Thus, we use $p = 319,147$ SNPs from both data sets for simulations.

In the first set of simulations, we fix $p = 319,147$. For each $n \in \{100, 200, \cdots, 1,000\}$, we randomly draw $n$ individuals to obtain $\mathbf{X}$ and further simulate 10,000 $\mathbf{y}$ to obtain the sample variance $V(q)$. We use this set of simulations to show that $n = \sqrt{2/(V(q)p')}$ (Figure S1a and S1b), where $p'$ is estimated with a linear regression of $\sqrt{2/(V(q))}$ on $n$. In addition, the sample mean $\hat{S}$ across the 10,000 sets of $\mathbf{X}'$ based on small values of $m$ is a good estimate of $S$ computed using all $n$ samples. However, unlike the simulated genotype case, $\hat{S}$ often over-estimates $S$ for small $m$ though the bias quickly diminishes for moderate $m$ (Table S1).

In the second set of simulations, we fix both $n$ and $p = 319,147$ ($n = 3,925$ in the Australia data and $n = 5,123$ in the Finland data). For each $m \in \{10, 20, \cdots, 100\}$, we randomly draw 10,000 sets of $\mathbf{X}'$ to obtain $V(\hat{S})$. We use this set of simulations to show that $n = 2/\sqrt{(V(\hat{S})p'^2)}$ (Figure S1c and S1d), where $p'$ is estimated with a linear regression of $2/\sqrt{(V(\hat{S}))}$ on $n$.

Finally, we estimate the effective number of independent SNPs in the two data sets. Specifically, we order all 319,147 SNPs from the two data sets by their genomic location, and select from these SNPs a subset of $p$ evenly spaced SNPs for simulation. For each value of $p$, we estimate the effective number of independent SNPs, $p'$, in the two data sets. As expected, $p'$ increases with $p$ linearly at the beginning but eventually reaches a limiting value (Figure S3). In addition, $p'$ computed based on either $V(q)$ or $V(\hat{S})$ are close to each other (Figure S3). From the real data, we estimate $p'_q = 28,913, p'_s = 31,165, p'_v = 17,618$ in the Australia data, $p'_q = 28,913, p'_s = 27,842, p'_v = 7,578$ in the Finland data ($p'_s = 33,077$ for the full Australia data with $p = 4,352,968$). Here $p'_s$, $p'_v$ and $p'_q$ are the effective number of independent SNPs with respect to $S$, $V(S)$, and $V(q)$. $p'_q \approx p'_s$ because under the null $V(q) \approx 2S/n^2$.

# 12 GWAS Data Sets

We used five GWAS data sets in the present study.

The Australia data contain height measurements for 3925 Australian individuals with 4,352,968 imputed SNPs [1]. The phenotypes were already regressed out of age and sex effects, and were quantile normalized to a standard normal distribution afterwards [1]. A total of typed 294,831 SNPs were available after stringent quality control [1]. We imputed SNPs based on the 1000 genomes project reference panel [27] using Shapeit [34] and IMPUTE2 [32]. After imputation, we removed SNPs with either a low

minor allele frequency ($< 1\%$) or missing values.

The Finland data (NFBC1966) contain information for 5,123 individuals with multiple metabolic traits measured and 319,147 SNPs typed [35]. We selected ten phenotypes, including C-reactive protein (CRP), glucose, insulin, total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TG), body mass index (BMI), systolic bollod pressure (SysBP) and diastolic blood pressure (DiaBP). For each phenotype, we quantile-transformed the phenotypic values to a standard normal distribution, regressed out effects of sex, oral contraceptives and pregnancy status [35], and quantile-transformed the residuals to a standard normal distribution again. We selected individuals and SNPs following the previous study [36]. We replaced the missing genotypes for a given SNP with its mean genotype value. In addition, we excluded individuals with one or more phenotypes missing.

The WTCCC data set is from the Wellcome trust case control consortium (WTCCC) 1 study [37]. The data set consists of about 14,000 cases of seven common diseases, including 1868 cases of bipolar disorder (BD), 1926 cases of coronary artery disease (CAD), 1748 cases of Crohn's disease (CD), 1952 cases of hypertension (HT), 1860 cases rheumatoid arthritis (RA), 1963 cases of type 1 diabetes (T1D) and 1924 cases of type 2 diabetes (T2D), as well as 2938 shared controls. We selected a total of 458,868 shared SNPs following a previous study [2].

The Framingham heart study (FHS) data set contains genotype data on 6,950 individuals and 394,174 quality-controlled SNPs with MAF$> 0.01$ [38]. We performed analysis on four commonly used blood lipid traits including HDL ($n = 6,850$), LDL ($n = 6,855$), TC ($n = 3,806$), and TG ($n = 3,806$). Each trait was quantile normalized to a standard normal distribution before analysis.

Finally, we used eight phenotypes from four consortium studies. Only summary statistics (z-scores) are available for these phenotypes. These phenotypes include BMI ($n = 120,569$, $p =$), height (HT, $n = 129,945$) from the GINAT consortium [39, 40], HDL ($n = 88,754$), LDL ($n = 84,685$), TC ($n = 89,005$) and TG ($n = 85,691$) from the Global Lipids Genetics Consortium [41], fasting glucose (FG, $n = 58,074$) from the MAGIC consortium [42], and Crohn's disease (CD, $n = 21,447$) from the International Inflammatory Bowel Disease Genetics Consortium [43]. The data have been pre-processed by a previous study [44]. We further selected z-scores from a common set of $p = 5,014,740$ SNPs for the analysis. We paired the summary statistics with genotypes for the same set of SNPs from 503 individuals of European ancestry in the 1000 genomes project [27]. We emphasize that it is important to use the same set of SNPs for both the study sample and the reference panel; a mis-match could create severe bias. The 503 individuals are from five different populations and include Utah residents (CEPH) with northern and western European ancestry (CEU), British in England and Scotland (GBR), Iberian Population in Spain (IBS), Finnish in Finland (FIN) and Toscani in Italia (TSI). To control for the population stratification among the five populations, we used the population indicators as covariates and centered genotypes in each population separately before computing $\mathbf{S}$.

# 13 Methods for Comparison

We compare five different methods in the Results section. (1) REML is implemented GEMMA based on the AI algorithm [25]. We do not constrain the REML estimates to be non-negative. Unconstrained estimation ensures unbiasedness of the estimates and calibration of the standard errors. (2) HE is implemented in GEMMA. We use our asymptotic form to compute the standard errors. (3) LDSC is implemented in the LDSC software [6,22,23]. LD scores are computed using genotypes from all individuals based on either the default 1 MB sliding window. or an extended 10 MB sliding window. When fitting LDSC, we apply the flag "–not-M-5-50" to use all SNPs. (4) MQS-HEW is implemented in GEMMA. $\mathbf{S}$ is estimated using genotype data from $m = 400$ randomly selected individuals. (5) MQS-LDW is implemented in GEMMA. $\mathbf{S}$ is estimated using $m = 400$ randomly selected individuals. A different set of $m = 400$ individuals are chosen in each replicate. The LD scores are also computed from genotypes of these $m = 400$ individuals based on a 1 MB sliding window.

# 14 Statistical Efficiency and Its Empirical Estimation

Statistical efficiency is an important criterion to compare estimation methods and it effectively measures how close the estimates are to the ground truth. Statistical efficiency for the variance component is formally defined as $E((\hat{\sigma}^2 - \sigma^2)^2)$, which, in the case of unbiased estimate, is the variance $V(\hat{\sigma}^2)$. In simulations, we can use the mean squared error (MSE) of the estimates from M simulation replicates, or $MSE(\hat{\sigma}^2) = 1/M \sum_{m=1}^{M} (\hat{\sigma}_m^2 - \sigma^2)^2$, as an empirical estimate of the statistical efficiency. We can also contrast the MSE of estimates from one method with the MSE of REML estimates by computing the ratio between the two, or $MSE(\hat{\sigma}^2)/MSE(\hat{\sigma}^2_{REML})$, as an empirical measure of the relative statistical efficiency of REML with respect to the other method.

In addition, for MQS-HEW, we can formally quantify its loss of statistical efficiency due to the subsampling strategy as $E((\sigma^2_{MQS-HEW} - \sigma^2_{HE})^2)/E((\sigma^2_{HE} - \sigma^2)^2)$. This loss of statistical efficiency can again be empirically estimated by computing the ratio of the two corresponding MSEs across simulation replicates.

# 15 Simulations

We describe simulation details here.

For $k = 1$ LMM simulations, we first compute the relatedness matrix $\mathbf{K}$ using genotype matrix $\mathbf{X}$ with $\mathbf{K} = \mathbf{X}\mathbf{X}^T/p$. We then simulate the $n$-vector of genetic random effects $\mathbf{g}$ from a multivariate normal distribution with covariance $h^2\mathbf{K}$. We simulate the $n$-vector of residual errors $\boldsymbol{\epsilon}$ from independent normal distributions with variance $1 - h^2$. We obtain the phenotype as a summation of the genetic effects and the residual errors. We examine three scenarios with heritability $h^2$ set to be either 0, 0.25, or 0.5. For each scenario, we repeat the above steps to create 1,000 simulation replicates. Note that the above LMM simulation effectively simulates phenotypes from a polygenic model where all SNPs are causal and each SNP effect size comes from a normal distribution [2]. We use LMM as main simulations in order to examine the statistical properties of various methods for the variance component model. However, we also perform additional simulations to examine the robustness of these methods in ascertained case control settings as well as in sparse settings (see below).

For $k = 6$ LMM simulations, we first annotate the genome using six categories as in [45]. The six categories include coding, untranslated region (UTR), promoter, DNAse hyper-sensitivity regions (DHS), intronic and intergenic regions. SNPs are classified into these six categories based on genomic location. We then compute a genetic relatedness matrix $\mathbf{K}_i = \mathbf{X}_i\mathbf{X}_i^T/p_i$ for each category $i$ using the category specific genotype matrix $\mathbf{X}_i$. We then simulate the $n$-vector of genetic random effects $\mathbf{g}_i$ for $i$'th category from a multivariate normal distribution with covariance $h_i^2\mathbf{K}_i$. We simulate the $n$-vector of residual errors $\boldsymbol{\epsilon}$ from independent normal distributions with variance $1 - h^2$, where $h^2 = \sum_{i=1}^{6} h_i^2$. We obtain the phenotype as a summation of the genetic effects from all categories and the residual errors. We examine three different scenarios: (I) a null scenario where all variance components are zero ($h_i^2 = 0$ for $i = 1, \cdots, 6$); (II) an alternative scenario with total SNP heritability $h^2 = 0.5$, and with each category explaining an equal proportion ($h_i^2 = 0.5/6$ for $i = 1, \cdots, 6$); (III) a more realistic alternative scenario with $h^2 = 0.5$ and with DHS explaining a large proportion of heritability ($h_1^2 = 0.04, h_2^2 = 0.02, h_3^2 = 0.02, h_4^2 = 0.4, h_5^2 = 0.01, h_6^2 = 0.01$). For each scenario, we perform 1,000 replicates.

For ascertained case control simulations, we simulated 2,500 cases, 2,500 controls and 10,000 independent causal SNPs from a liability threshold model with $h^2 = 0.5$ and a disease prevalence of 0.1% using the software provided by [24]. We either use the 10,000 causal SNPs directly to form a non-sparse simulation scenario, or pair them with 90,000 non-causal independent SNPs to form a sparse simulation scenario. We then estimate heritability on the observed scale with different methods and transform the estimates back to the liability scale ( [2, 24, 46]). Because of the small number of simulated SNPs, we use genome-wide SNPs to compute the LD scores.

For sparse simulations, we randomly selected 100 or 1,000 SNPs from the Finland data to be causal. We simulated their effects independently from a standard normal distribution. We simulate the $n$-vector of residual errors $\boldsymbol{\epsilon}$ from independent normal distributions with variance $1-h^2$. We then scale the genetic effects by a constant so that the genetic effects in total explains a fixed heritability value of $h^2 = 0.5$ ($h^2 = V(\mathbf{g})/(V(\mathbf{g}) + V(\epsilon))$, where $V$ denotes sample variance and $\mathbf{g}$ is the summation of genetic effects from the causal SNPs).

# References


1. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42: 565-569.

2. Zhou X, Carbonetto P, Stephens M (2013) Polygenic modelling with Bayesian sparse linear mixed models. PLoS Genetics 9: e1003264.

3. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38: 904-909.

4. McVean G (2009) A genealogical interpretation of principal components analysis. PLoS Genetics 5: e1000686.

5. Rao CR (1970) Estimation of heteroscedastic variances in linear models. Journal of the American Statistical Association 65: 161-172.

6. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics 47: 291-295.

7. Rao CR (1971) Estimation of variance and covariance component – MINQUE theory. Journal of Multivariate Analysis 1: 257-275.

8. Rao CR (1971) Minimum variance quadratic unbiased estimation of variance components. Journal of Multivariate Analysis 1: 445-456.

9. Rao CR (1972) Estimation of variance and covariance components in linear models. Journal of the American Statistical Association 67: 112-115.

10. Rao CR (1973) Quadratic estimation of variance components. Biometrics 2: 311-330.

11. Zhu J, Weir B (1996) Mixed model approaches for diallele analysis based on a bio-model. Genetical Research 68: 233-240.

12. Brown KG (1976) Asymptotic behavior of minque-type estimators of variance components. Annals of Statistics 4: 746-754.

13. Hartley HO, Rao JNK, Lamotte LR (1978) A simple 'synthesis'-based method of variance component estimation. Biometrics 34: 233-242.

14. Swallow WH, Searle SR (1978) Minimum variance quadratic unbiased estimation (MIVQUE) of variance components. Technometrics 20: 265-272.

15. Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behavior Genetics 2: 3-19.

16. Drigalenko E (1998) How sib-pairs reveal linkage. American Journal of Human Genetics 63: 1242-1245.

17. Elston RC, Buxbaum S, Jacobs KB, Olson JM (2000) Haseman and Elston revisited. Genetic Epidemiology 19: 1-17.

18. Sham PC, Purcell S (2001) Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. American Journal of Human Genetics 68: 1527-1532.

19. Sham PC, Purcell S, Cherny SS, Abecasis GR (2002) Powerful regression-based quantitative-trait linkage analysis of general pedigrees. American Journal of Human Genetics 71: 238-253.

20. Chen WM, Broman KW, Liang KY (2004) Quantitative trait linkage analysis by generalized estimating equations: unification of variance components and Haseman-Elston regression. Genetic Epidemiology 26: 265-272.

21. Chen GB (2014) Estimating heritability of complex traits from genome-wide association studies using IBS-based HasemanElston regression. Frontiers in Genetics 5: 107.

22. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, et al. (2015) An atlas of genetic correlations across human diseases and traits. Nature Genetics 47: 1236-1241.

23. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, et al. (2015) Partitioning heritability by functional category using GWAS summary statistics. Nature Genetics 47: 1228-1235.

24. Golan D, Lander ES, Rosseta S (2014) Measuring missing heritability: Inferring the contribution of common variants. Proceedings of the National Academy of Sciences, USA 111: E5272E5281.

25. Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440-1450.

26. Wu MC, Lee S, Cai T, Li Y, Boehnke M, et al. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. American Journal of Human Genetics 89: 82-93.

27. Consortium TGP (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65.

28. Efron B, Stein C (1981) The Jackknife estimate of variance. Annals of Statistics 3: 586-596.

29. Browning SR (2006) Multilocus association mapping using variable-length Markov chains. American Journal of Human Genetics 78: 903-913.

30. Guan Y, Stephens M (2008) Practical issues in imputation-based association mapping. PLoS Genetics 4: e1000279.

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

32. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature Genetics 44: 955-959.

33. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature Genetics 44: 369-375.

34. Delaneau O, Zagury JF, Marchini J (2013) Improved whole chromosome phasing for disease and population genetic studies. Nature Methods 10: 5-6.

35. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, et al. (2008) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature Genetics 41: 35-46.

36. Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods : 407-409.

37. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.

38. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, et al. (2007) The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. American Journal of Epidemiology 165: 1328-1335.

39. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832-838.

40. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics 42: 937-948.

41. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707-713.

42. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature Genetics 44: 659-669.

43. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, et al. (2012) Hostmicrobe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491: 119-124.

44. Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. American Journal of Human Genetics 94: 559-573.

45. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjlmsson BJ, et al. (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. American Journal of Human Genetics 5: 535-552.

46. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. The American Journal of Human Genetics 88: 294-305.