

Supplementary note one of Bivariate-LDER-GE

Zihan Dong, Wei Jaing, Andrew DeWan, Hongyu Zhao

1 Main derivation

We set up the model

$$Y_i = \sum_{j=1}^M G_{ji} \beta_j + \sum_{j=1}^M S_{ji} \gamma_j + \epsilon_{1i} E_i + \epsilon_{0i} \quad (1)$$

Where Y_i is the phenotype for subject i already adjusted for fixed effect including for the exposure covariate. E_i is the standardized exposure covariate for subject i . Suppose there are M variant. G_{ji} is the j th variant for subject i . $S_{ji} = G_{ji} * E_i$ is the GE interaction product term for variant j of subject i . ϵ_{1i} is the non-genetic residual that has exposure interaction effect. ϵ_{0i} is the residual independent from all other parts. β_j is the true additive effect for variant j . γ_j is the true interaction effect for variant j .

In matrix form, it's written as

$$\mathbf{Y} = \mathbf{G}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\epsilon}_1 \mathbf{E} + \boldsymbol{\epsilon}_0 \quad (2)$$

We model β_j and γ_j using random effect model:

$$\begin{pmatrix} \beta_j \\ \gamma_j \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} h_g^2/M & \rho_{gI}/M \\ \rho_{gI}/M & h_I^2/M \end{pmatrix} \right] \quad (3)$$

Where h_g^2 is the narrow-sense heritability and h_I^2 is the GE interaction contribution proportion that we are interested at. ρ_{gI} models the genome-level covariance between the additive effect and GE interaction effect.

We model ϵ_0 and ϵ_1 using random effect model:

$$\begin{pmatrix} \epsilon_{0i} \\ \epsilon_{1i} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_{0,1} \\ \rho_{0,1} & \sigma_1^2 \end{pmatrix} \right] \quad (4)$$

We assume that the genetic part and residual part are independent.

Suppose all variants and the exposure variable are standardized.

Then the main genetic effect estimate for variant j is:

$$\begin{aligned} \hat{\beta}_j &= (\mathbf{G}_{j\cdot}^{(G)T} \mathbf{G}_{j\cdot}^{(G)})^{-1} \mathbf{G}_{j\cdot}^{(G)T} \mathbf{Y} \\ &= \mathbf{G}_{j\cdot}^{(G)T} \mathbf{Y}^{(G)} / N_G \end{aligned} \quad (5)$$

Where the superscript (G) represents the sample of the main genetic association with N_G being its sample size. The

first equality holds because after standardization $\text{var}(G_{j.}) = 1$
And the Z score of the main genetic effect estimate for variant j is:

$$Z_{jG} = \mathbf{G}_{j.}^{(G)T} \mathbf{Y}^{(G)} / \sqrt{N_G} \quad (6)$$

The interaction effect estimate for variant k is:

$$\begin{aligned} \hat{\gamma}_j &= (\mathbf{S}_{k.}^{(I)T} \mathbf{S}_{k.}^{(I)})^{-1} \mathbf{S}_{k.}^{(I)T} \mathbf{Y}^{(I)} \\ &= \mathbf{S}_{k.}^{(I)T} \mathbf{Y}^{(I)} / N_I \end{aligned} \quad (7)$$

Where the superscript (I) represents the sample of the GE interaction association with N_I being its sample size, which maybe different from (G). Denote the overlap sample size as N_S . The first equality holds because after standardization $\text{var}(S_{k.}) = 1$.

And the Z score of the interaction effect estimate for variant k is:

$$Z_{kI} = \mathbf{S}_{k.}^{(I)T} \mathbf{Y}^{(I)} / \sqrt{N_I} \quad (8)$$

Then for two variant j and k, the expected corss-variant Z score product is:

$$\begin{aligned} \mathbb{E}(Z_{jG} Z_{kI}) &= \mathbb{E}(\mathbb{E}(Z_{jG} Z_{kI} | \beta, \gamma, \epsilon_0, \epsilon_1)) \\ &= \frac{1}{\sqrt{N_G N_I}} \mathbb{E}(\mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{Y}^{(G)} \mathbf{S}_{k.}^{(I)T} \mathbf{Y}^{(I)} | \beta, \gamma, \epsilon_0, \epsilon_1)) \\ &= \frac{1}{\sqrt{N_G N_I}} \mathbb{E}(\mathbf{G}_{j.}^{(G)T} (\mathbf{G}^{(G)} \beta + \mathbf{S}^{(G)} \gamma + \epsilon_1^{(G)} \mathbf{E}^{(G)} + \epsilon_0^{(G)})) \\ &\quad (\mathbf{S}_{k.}^{(I)T} (\mathbf{G}^{(I)} \beta + \mathbf{S}^{(I)} \gamma + \epsilon_1^{(I)} \mathbf{E}^{(I)} + \epsilon_0^{(I)})) \\ &= \frac{1}{\sqrt{N_G N_I}} \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{G}^{(G)} \beta \mathbf{S}_{k.}^{(I)T} \mathbf{S}^{(I)} \gamma) \\ &\quad + \frac{1}{\sqrt{N_G N_I}} \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{S}^{(G)} \beta \mathbf{S}_{k.}^{(I)T} \mathbf{G}^{(I)} \gamma) \\ &\quad + \frac{1}{\sqrt{N_G N_I}} \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \epsilon_1^{(G)} \mathbf{E}^{(G)} \mathbf{S}_{k.}^{(I)T} \epsilon_0^{(I)}) \\ &\quad + \frac{1}{\sqrt{N_G N_I}} \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \epsilon_0^{(G)} \mathbf{S}_{k.}^{(I)T} \epsilon_1^{(I)} \mathbf{E}^{(I)}) \\ &= \frac{1}{\sqrt{N_G N_I}} \left(\frac{\rho_{g,I}}{M} ((N_G N_I + N_S) \ell_{j,k} + N_S M r_{j,k}) + \frac{\rho_{g,I}}{M} N_S (2\ell_{j,k} + M r_{j,k}) + 2\rho_{0,1} N_S r_{j,k} \right) \\ &\approx \frac{\sqrt{N_G N_I} \rho_{g,I}}{M} \ell_{j,k} + \frac{2N_S}{\sqrt{N_G N_I}} (\rho_{0,1} + \rho_{g,I}) r_{j,k} \end{aligned} \quad (9)$$

where r_{jk} is the true correlation between the SNP j and k, and $\ell_{jk} = \sum_{t=1}^M r_{jt} r_{kt}$ is the off diagonal element of the LD score matrix. The details of the forth equality are in supplementary.

In matrix form, the above quantity can be written as

$$\text{COV}(\mathbf{Z}_G, \mathbf{Z}_I) = \frac{\sqrt{N_G N_I} \rho_{g,I}}{M} \mathbf{L} + \frac{2N_S}{\sqrt{N_G N_I}} (\rho_{0,1} + \rho_{g,I}) \mathbf{R} \quad (10)$$

where $\mathbf{L} = \mathbf{R}^T \mathbf{R}$ is the LD score matrix.

2 Details

This section provides the derivation details.

For equation (9) equality 4 there are 4 parts omitting constant term $\frac{1}{\sqrt{N_G N_I}}$ for convenience:

$$\begin{aligned}
& \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{G}^{(G)} \beta \mathbf{S}_{k.}^{(I)T} \mathbf{S}^{(I)} \boldsymbol{\gamma}) + \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{S}^{(G)} \beta \mathbf{S}_{k.}^{(I)T} \mathbf{G}^{(I)} \boldsymbol{\gamma}) \\
& + \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \boldsymbol{\epsilon}_1^{(G)} \mathbf{E}^{(G)} \mathbf{S}_{k.}^{(I)T} \boldsymbol{\epsilon}_0^{(I)}) + \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \boldsymbol{\epsilon}_0^{(G)} \mathbf{S}_{k.}^{(I)T} \boldsymbol{\epsilon}_1^{(I)} \mathbf{E}^{(I)}) \\
& = \left(\frac{\rho_{g,I}}{M} (N_G N_I \ell_{j,k} + N_S M r_{j,k}) + \frac{\rho_{g,I}}{M} N_S (2\ell_{j,k} + M r_{j,k}) + 2\rho_{0,1} N_S r_{j,k} \right)
\end{aligned} \tag{11}$$

The first part:

$$\begin{aligned}
& \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{G}^{(G)} \beta \mathbf{S}_{k.}^{(I)T} \mathbf{S}^{(I)} \boldsymbol{\gamma}) \\
& = \frac{\rho_{g,I}}{M} \mathbb{E} \left(\sum_{t=1}^M \mathbf{S}_{j,.}^T \mathbf{S}_{t,.} \mathbf{G}_{k,.}^T \mathbf{G}_{t,.} \right) \\
& = \frac{\rho_{g,I}}{M} \mathbb{E} \left(\sum_{t=1}^M \left(\sum_{a=1}^{N_I} E_a^2 G_{j,a} G_{t,a} \right) \left(\sum_{b=1}^{N_G} G_{k,b} G_{t,b} \right) \right) \\
& = \frac{\rho_{g,I}}{M} \sum_{t=1}^M \sum_{a=1}^{N_I} \sum_{b=1}^{N_G} \mathbb{E}(G_{j,a} G_{t,a} G_{k,b} G_{t,b}) \\
& = \frac{\rho_{g,I}}{M} ((N_G N_I + N_S) \ell_{j,k} + N_S M r_{j,k})
\end{aligned} \tag{12}$$

The forth equality utilizes Isserlis' Theorem:

$$\begin{aligned}
& \mathbb{E}(G_{j,a} G_{t,a} G_{k,b} G_{t,b}) = \\
& \mathbb{E}(G_{j,a} G_{t,a}) \mathbb{E}(G_{k,b} G_{t,b}) + \mathbb{E}(G_{j,a} G_{k,b}) \mathbb{E}(G_{t,a} G_{t,b}) + \mathbb{E}(G_{j,a} G_{t,b}) \mathbb{E}(G_{t,a} G_{k,b}) \\
& = r_{j,t} r_{k,t} + \mathbf{I}(a=b) r_{j,k} + \mathbf{I}(a=b) r_{j,t} r_{k,t}
\end{aligned} \tag{13}$$

Where $\mathbf{I}$ is the indicator operation.

The second part:

$$\begin{aligned}
& \mathbb{E}(\mathbf{G}_{j.}^{(G)T} \mathbf{S}^{(G)} \beta \mathbf{S}_{k.}^{(I)T} \mathbf{G}^{(I)} \boldsymbol{\gamma}) \\
& = \frac{\rho_{g,I}}{M} \mathbb{E} \left(\sum_{t=1}^M \mathbf{G}_{j,.}^T \mathbf{S}_{t,.} \mathbf{S}_{k,.}^T \mathbf{G}_{t,.} \right) \\
& = \frac{\rho_{g,I}}{M} \mathbb{E} \left(\sum_{t=1}^M \left(\sum_{a=1}^{N_I} E_a G_{j,a} S_{t,a} \right) \left(\sum_{b=1}^{N_G} E_b G_{k,b} S_{t,b} \right) \right) \\
& = \frac{\rho_{g,I}}{M} \sum_{t=1}^M \sum_{a=1}^{N_I} \sum_{b=1}^{N_G} \mathbb{E}(E_a E_b G_{j,a} G_{t,a} G_{k,b} G_{t,b}) \\
& = \frac{\rho_{g,I}}{M} \sum_{t=1}^M \sum_{a=1}^{N_I} \sum_{b=1}^{N_G} \mathbf{I}(a=b) \mathbb{E}(G_{j,a} G_{t,a} G_{k,b} G_{t,b}) \\
& = \frac{\rho_{g,I}}{M} \sum_{t=1}^M (2\ell_{j,k} + M r_{j,k})
\end{aligned} \tag{14}$$

The derivation of the fifth equality can be found in equation 17 and 18.

The third part:

$$\begin{aligned}
& \mathbb{E}(\mathbf{G}_{j\cdot}^{(G)T} \boldsymbol{\epsilon}_1^{(G)} \mathbf{E}^{(G)} \mathbf{S}_{k\cdot}^{(I)T} \boldsymbol{\epsilon}_0^{(I)}) \\
&= \rho_{0,1} \mathbb{E}(\sum_{a=1}^{N_I} G_{j,a} E_a \sum_{b=1}^{N_G} E_b G_{k,b}) \\
&= \rho_{0,1} \mathbb{E}(\sum_{a=1}^{N_I} \sum_{b=1}^{N_G} \mathbf{I}(a=b) E_a E_b G_{j,a} G_{k,b}) \\
&= \rho_{0,1} N_{Sr_{j,k}}
\end{aligned} \tag{15}$$

The forth part:

$$\begin{aligned}
& \mathbb{E}(\mathbf{G}_{j\cdot}^{(G)T} \boldsymbol{\epsilon}_0^{(G)} \mathbf{S}_{k\cdot}^{(I)T} \boldsymbol{\epsilon}_1^{(I)} \mathbf{E}^{(I)}) \\
&= \rho_{0,1} \mathbb{E}(\sum_{a=1}^{N_I} G_{j,a} \sum_{b=1}^{N_G} E_b^2 G_{k,b}) \\
&= \rho_{0,1} \mathbb{E}(\sum_{a=1}^{N_I} \sum_{b=1}^{N_G} \mathbf{I}(a=b) E_b^2 G_{j,a} G_{k,b}) \\
&= \rho_{0,1} N_{Sr_{j,k}}
\end{aligned} \tag{16}$$

To show the sixth equality in equation 14, we model three standardized variants G_{jG}, G_{ki}, G_{ti} for one subject i as MVN distribution.

$$\begin{pmatrix} G_{jG} \\ G_{ki} \\ G_{ti} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & r_{jk} & r_{jt} \\ r_{jk} & 1 & r_{kt} \\ r_{jt} & r_{kt} & 1 \end{pmatrix} \right] \tag{17}$$

Then, by law of total expectation,

$$\begin{aligned}
& \mathbb{E}(G_{ji} G_{ki} G_{ti}^2) \\
&= \mathbb{E}(\mathbf{E}(G_{ji} G_{ki} G_{ti}^2 | G_{ti})) \\
&= 2r_{jt} r_{kt} + r_{jk}
\end{aligned} \tag{18}$$

3 Heritable exposure covariate $\mathbf{E}$

When the exposure $\mathbf{E}$ is heritable, where $\boldsymbol{\alpha}$ is the true polygenic effect modeled as a random effect, and $\boldsymbol{\epsilon}_e$ is the residual to make $\mathbf{E}$ standardized.

$$\mathbf{E} = \mathbf{G}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_e \tag{19}$$

$$\begin{pmatrix} \beta_j \\ \gamma_j \\ \alpha_j \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} h_g^2/M & \rho_{gI}/M & \rho_{ge}/M \\ \rho_{gI}/M & h_I^2/M & \rho_{eI}/M \\ \rho_{ge}/M & \rho_{eI}/M & h_e^2/M \end{pmatrix} \right] \tag{20}$$

$$\tag{21}$$

Miao et al in [1] models as follows:

$$\begin{aligned} \mathbf{Y} &= \mathbf{G}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\epsilon}_1\mathbf{E} + \boldsymbol{\epsilon}_0 \\ \mathbf{E} &= \mathbf{G}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_e \\ \mathbf{Y}_2 &= \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\epsilon}_2 \end{aligned} \quad (22)$$

Where $\mathbf{Z}$ is the genotype panel for $\mathbf{Y}_2$ (same or different from the sample of $\mathbf{Y}$), and $\boldsymbol{\theta}$ is its effect sizes. We denote $\rho_{G2I} = \text{Cov}(\theta_j, \gamma_j)$ and $\rho_{G2E} = \text{Cov}(\theta_j, \alpha_j)$. Miao et al derived the bivariate product of interaction analysis on $\mathbf{Y}$ and main genetic effect analysis on $\mathbf{Y}_2$ for the same variant, and we extend this to two possibly different variant j and k based on their previous work:

$$\begin{aligned} \mathbb{E}(Z_{jG_2}Z_{kI}|\mathbf{G}, \mathbf{Z}, \mathbf{E}) &= \frac{1}{\sqrt{N_{G2}N_I}} \mathbb{E}(\mathbf{Z}_j^T \mathbf{Y}_2 \mathbf{S}_k^T \mathbf{Y}) \\ &= \frac{1}{\sqrt{N_{G2}N_I}} (T1 + T2 - T3) \end{aligned} \quad (23)$$

Where

$$\begin{aligned} T1 &= \frac{\rho_{\theta I}}{M} \mathbf{Z}_j^T \mathbf{Z} \mathbf{I}_M \mathbf{S}^T \mathbf{S}_k. \\ T2 &= (\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})_k^T (\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})\boldsymbol{\gamma}\boldsymbol{\theta}^T \mathbf{Z}^T \mathbf{Z}_j \\ T3 &= (\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})_k^T \mathbb{E}[(\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})\boldsymbol{\gamma}]\boldsymbol{\theta}^T \mathbf{Z}^T \mathbf{Z}_j \end{aligned} \quad (24)$$

From equation (9) we derive that

$$\begin{aligned} \mathbb{E}(T1) &= \mathbb{E}\left(\frac{\rho_{\theta I}}{M} \mathbf{Z}_j^T \mathbf{Z} \mathbf{I}_M \mathbf{S}^T \mathbf{S}_k\right) \\ &= (1 - h_e^2)(N_I N_{G2} \ell_{jk} \frac{\rho_{G2I}}{M} + N_S \rho_{G2I} r_{jk}) \end{aligned} \quad (25)$$

Following Miao et al's derivation, for T2 we generalize to two possibly different variants j and k :

$$\begin{aligned} \mathbb{E}(T2) &= \mathbb{E}[(\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})_k^T (\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})\boldsymbol{\gamma}\boldsymbol{\theta}^T \mathbf{Z}^T \mathbf{Z}_j] \\ &= \sum_{i=1}^{N_I} \sum_{q=1}^{N_{G2}} \sum_{a=1}^M \sum_{b=1}^M \sum_{c=1}^M \sum_{d=1}^M \mathbb{E}[G_{ik} G_{ia} \alpha_a G_{ib} \gamma_b G_{ic} \gamma_c Z_{qj} Z_{qd} \theta_d] \\ &= \sum_{i=1}^{N_I} \sum_{q=1}^{N_{G2}} \sum_{a=1}^M \sum_{b=1}^M \sum_{c=1}^M \sum_{d=1}^M \mathbb{E}[G_{ik} G_{ia} G_{ib} G_{ic} Z_{qj} Z_{qd}] \mathbb{E}[\alpha_a \gamma_b \gamma_c \theta_d] \\ &= N_I N_{G2} (h_e^2 \rho_{G2, I} + 2M \mathbb{E}(\alpha_a \gamma_a) \mathbb{E}(\alpha_a \theta_a) \ell_{jk} + N_S (h_e^2 \rho_{G2I} + 2M \mathbb{E}(\alpha_a \gamma_a) \mathbb{E}(\alpha_a \theta_a)) r_{jk}) \end{aligned} \quad (26)$$

Where the last equality is justified using Isserlis' theorem. For the third term, we did similar derivation:

$$\begin{aligned} \mathbb{E}(T3) &= \mathbb{E}[(\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})_k^T \mathbb{E}[(\text{diag}(\mathbf{G}\boldsymbol{\alpha})\mathbf{G})\boldsymbol{\gamma}]\boldsymbol{\theta}^T \mathbf{Z}^T \mathbf{Z}_j] \\ &= N_I N_{G2} M \mathbb{E}(\alpha_a \gamma_a) \mathbb{E}(\alpha_a \theta_a) \ell_{jk} + N_S \mathbb{E}(\alpha_a \gamma_a) \mathbb{E}(\alpha_a \theta_a) r_{jk} \end{aligned} \quad (27)$$

Merging three terms T1, T2 and T3, we get

$$\begin{aligned} \mathbb{E}(Z_{jG_2}Z_{kI}) &= \mathbb{E}(\mathbb{E}(Z_{jG_2}Z_{kI}|\mathbf{G}, \mathbf{Z}, \mathbf{E})) \\ &= \mathbb{E}(T1) + \mathbb{E}(T2) + \mathbb{E}(T3) \\ &= \frac{\sqrt{N_I N_{G2}}(\rho_{G2I} + \rho_{G2E} \rho_{eI})}{M} \ell_{jk} + c_1 r_{jk} \end{aligned} \quad (28)$$

Where c_1 is the intercept which incorporates potential confounding effects. When Y_2 is E , we get

$$\mathbb{E}(Z_{jE}Z_{kI}) = \frac{\sqrt{N_I N_E}(\rho_{eI} + \rho_{eI}h_e^2)}{M}\ell_{jk} + c_1 r_{jk} \quad (29)$$

When Y_2 is Y , we get

$$\mathbb{E}(Z_{jG}Z_{kI}) = \frac{\sqrt{N_I N_G}(\rho_{gI} + \rho_{ge}\rho_{eI})}{M}\ell_{jk} + c_2 r_{jk} \quad (30)$$

We can utilize equation (29) to estimate ρ_{eI} and equation (30) to estimate ρ_{gI} .

$$\begin{aligned} \mathbb{E}(Z_{jI}Z_{kI}) &= \mathbb{E}(\mathbb{E}(Z_{jI}Z_{kI}|\beta, \gamma, \alpha, \epsilon_0, \epsilon_1, \epsilon_e)) \\ &= \mathbb{E}(\mathbb{E}(S_{j\cdot}^T Y S_{k\cdot}^T Y/n|\beta, \gamma, \alpha, \epsilon_0, \epsilon_1, \epsilon_e)) \\ &= \mathbb{E}(S_{j\cdot}^T (G\beta + S\gamma + \epsilon_1 E + \epsilon_0) S_{k\cdot}^T (G\beta + S\gamma + \epsilon_1 E + \epsilon_0))/n \\ &= \mathbb{E}(S_{j\cdot}^T (G\beta) S_{k\cdot}^T (G\beta))/n + \mathbb{E}(S_{j\cdot}^T (S\gamma) S_{k\cdot}^T (S\gamma))/n \\ &\quad + \mathbb{E}(S_{j\cdot}^T (\epsilon_1 E) S_{k\cdot}^T (\epsilon_1 E))/n + \mathbb{E}(S_{j\cdot}^T (\epsilon_0) S_{k\cdot}^T (\epsilon_0))/n \\ &= \mathbb{E}([diag(G\alpha + \epsilon_e)G]_{j\cdot}^T (G\beta) [diag(G\alpha + \epsilon_e)G]_{k\cdot}^T (G\beta))/n \\ &\quad + \mathbb{E}([diag(G\alpha + \epsilon_e)G]_{j\cdot}^T diag(G\alpha + \epsilon_e)G\gamma [diag(G\alpha + \epsilon_e)G]_{k\cdot}^T diag(G\alpha + \epsilon_e)G\gamma)/n \\ &\quad + \mathbb{E}([diag(G\alpha + \epsilon_e)G]_{j\cdot}^T (\epsilon_1 E) [diag(G\alpha + \epsilon_e)G]_{k\cdot}^T (\epsilon_1 E))/n \\ &\quad + \mathbb{E}([diag(G\alpha + \epsilon_e)G]_{j\cdot}^T (\epsilon_0) [diag(G\alpha + \epsilon_e)G]_{k\cdot}^T (\epsilon_0))/n \\ &= \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T (G\beta) [diag(G\alpha)G]_{k\cdot}^T (G\beta))/n + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T (G\beta) [diag(\epsilon_e)G]_{k\cdot}^T (G\beta))/n \\ &\quad + \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T diag(G\alpha)G\gamma [diag(G\alpha)G]_{k\cdot}^T diag(G\alpha)G\gamma)/n \\ &\quad + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T diag(\epsilon_e)G\gamma [diag(\epsilon_e)G]_{k\cdot}^T diag(\epsilon_e)G\gamma)/n \\ &\quad + \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T (\epsilon_1 E) [diag(G\alpha)G]_{k\cdot}^T (\epsilon_1 E))/n + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T (\epsilon_1 E) [diag(\epsilon_e)G]_{k\cdot}^T (\epsilon_1 E))/n \\ &\quad + \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T (\epsilon_0) [diag(G\alpha)G]_{k\cdot}^T (\epsilon_0))/n + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T (\epsilon_0) [diag(\epsilon_e)G]_{k\cdot}^T (\epsilon_0))/n \end{aligned} \quad (31)$$

Recognizing that ϵ_e can be regarded the unstandardized non-heritable exposure covariate. Standardizing ϵ_e by multiplying $\frac{1}{\sqrt{1-h_e^2}}$ and utilizing the conclusion from the previous LDER-GE study [2], we have that

$$\begin{aligned} &\mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T (G\beta) [diag(\epsilon_e)G]_{k\cdot}^T (G\beta))/n + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T diag(\epsilon_e)G\gamma [diag(\epsilon_e)G]_{k\cdot}^T diag(\epsilon_e)G\gamma)/n \\ &\quad + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T (\epsilon_1 E) [diag(\epsilon_e)G]_{k\cdot}^T (\epsilon_1 E))/n + \mathbb{E}([diag(\epsilon_e)G]_{j\cdot}^T (\epsilon_0) [diag(\epsilon_e)G]_{k\cdot}^T (\epsilon_0))/n \\ &\approx (1 - h_e^2)h_g^2 M r_{jk}/M + (1 - h_e^2)^2 h_I^2/M * ((n^2 + 2nK(E) - 2n)\ell_{jk} + nMK(E)r_{jk})/n \\ &\quad + (1 - h_e^2)\sigma_1^2 K(E)r_{jk} + (1 - h_e^2)\sigma_0^2 r_{jk} \end{aligned} \quad (32)$$

Following the derivations for (26) and (27) and the Isserlis' theorem, we get

$$\begin{aligned} &\mathbb{E}([diag(G\alpha)G]_{j\cdot}^T (G\beta) [diag(G\alpha)G]_{k\cdot}^T (G\beta))/n + \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T diag(G\alpha)G\gamma [diag(G\alpha)G]_{k\cdot}^T diag(G\alpha)G\gamma)/n \\ &\quad + \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T (\epsilon_1 E) [diag(G\alpha)G]_{k\cdot}^T (\epsilon_1 E))/n + \mathbb{E}([diag(G\alpha)G]_{j\cdot}^T (\epsilon_0) [diag(G\alpha)G]_{k\cdot}^T (\epsilon_0))/n \\ &\approx (2h_e^2\rho_{eI}^2 + (h_e^2)^2 h_I^2 + 2(1 - h_e^2)(h_e^2 h_I^2 + \rho_{eI}^2))\ell_{jk}n/m + (h_e^2 h_g^2 + h_e^2 h_I^2)r_{jk} \end{aligned} \quad (33)$$

Merging the above two components, we get

$$\mathbb{E}(Z_{jI}Z_{kI}) = \frac{n(h_I^2 + 2\rho_{eI}^2)\ell_{jk}}{M} + c_3 r_{jk} \quad (34)$$

We can use the estimated $\hat{\rho}_{eI}$ from equation (29) and equation (34) to estimate the adjusted h_I^2 .

References

- [1] J. Miao, G. Song, Y. Wu, J. Hu, Y. Wu, S. Basu, J. S. Andrews, K. Schaumberg, J. M. Fletcher, L. L. Schmitz, *et al.*, “Reimagining gene-environment interaction analysis for human complex traits,” *bioRxiv*, pp. 2022–12, 2022.
- [2] Z. Dong, W. Jiang, H. Li, A. T. Dewan, and H. Zhao, “Lder-ge estimates phenotypic variance component of gene-environment interactions in human complex traits accurately with ge interaction summary statistics and full ld information,” *bioRxiv*, pp. 2023–11, 2023.