Statistical Inference for Conditional Correlation in Neuroimaging Data
STA4000 Report

Amanda Ng

December 10, 2025

1 CEIDR (Pan et al. (2025))

1.1 Notation

For subject $i = 1, \ldots, N$ and vertex $v = 1, \ldots, V$, we denote subject-level brain measurements for two modalities (s and f) as $y_i^{(s)}(v)$ and $y_i^{(f)}(v)$, respectively. Let x_i be the covariate of interest and z_i as nuisance covariate vector being adjusted for subject i.

1.2 Model

For the m imaging modality, we consider the following mean and variance functions depending on covariates (i.e., $\mu_m(z_i, x_i, v)$ and $\sigma_m(z_i, x_i, v)$) for a model for $y_i^{(m)}(v)$:

$$y_i^{(m)}(v) = \mu_m(z_i, x_i, v) + \sigma_m(z_i, x_i, v) \times \epsilon_{mi}(v), \tag{1}$$

where the residual terms $\epsilon_{si}(v)$ and $\epsilon_{fi}(v)$ follow the joint distribution given by:

$$\begin{bmatrix} \epsilon_{si}(v) \\ \epsilon_{fi}(v) \end{bmatrix} \sim \mathcal{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho(z_i, x_i, v) \\ \rho(z_i, x_i, v) & 1 \end{bmatrix} \right). \tag{2}$$

The residual correlation $\rho(z_i, x_i, v)$ is the subject-level intermodal coupling of interest, modeled as follows:

$$\rho(z_i, x_i, v) = tanh(\beta_0 + \beta_1 \times z_i + \kappa(v) \times x_i) \tag{3}$$

The null hypothesis is $H_0\colon \kappa(v) = 0$. CEIDR tests H_0 by following these 6 steps: between-subject adjustment, within-subject adjustment, estimating subject-level intermodal coupling $\hat{\rho}_i(v)$, modeling $\hat{\rho}_i(v)$ in terms of covariates, adaptive cluster enhancement, and permutation-based test statistic.

1.3 Impact of Model Misspecification Simulation in Between-Subject Adjustment

False positives in intermodal correlation could be inflated when ”the observed covariate effects” on mean and variance were misspecified, either through an incorrect model or missing covariates.

1.3.1 Simulation Set-up

We considered $N = 100$ subjects and focus on a single vertex v. We first generated two covariates, $x_i \sim N(0, 1)$ (variable of interest) and $z_i \sim N(0, 1)$. Mean and variance were set as:

$$\mu_s(z_i, x_i, v) = 0.3 + 0.2z_i + 2x_i$$

$$\mu_f(z_i, x_i, v) = 0.5 + 0.1z_i + 3x_i$$

$$\sigma_s(z_i, x_i, v) = \sigma_f(z_i, x_i, v) = 1$$

The residual correlation was set as:

$$\rho(z_i, x_i, v) = tanh(0.3 + 0.1z_i + \kappa(v) \times x_i)$$

To investigate Type I error, we set $\kappa(v) = 0$.

1.3.2 Simulation Method

We focus on the effect of missing covariates in between-subject adjustment modeling, we fitted the true model shape, i.e. linear regression model, to estimate individual means $\mu_m(z_i, x_i, v)$ and variances $\sigma_m(z_i, x_i, v)$. For $m = s$, we on purpose misspecified the model as

$$y^{(s)}(z_i, x_i, v) = \alpha_0(v) + \alpha_1(v) \times z_i + \tau_i(v),$$

i.e. we left out the covariate effect of x_i. For $m = f$, we fitted the correct model

$$y^{(f)}(z_i, x_i, v) = \gamma_0(v) + \gamma_1(v) \times z_i + \gamma_2(v) \times x_i + \eta_i(v).$$

For simplicity, we did not conduct within subject adjustments on the residuals.

The standardized residuals were obtained as follows:

$$\hat{\epsilon}_{mi}(v) = \frac{y_i^{(m)}(v) - \hat{\mu}_m(z_i, x_i, v)}{\hat{\sigma}_m(z_i, x_i, v)}. \tag{4}$$

where $\hat{\mu}_m(z_i, x_i, v)$ is the fitted mean from the linear regression model and $\hat{\sigma}_m(z_i, x_i, v)$ is standard error of residuals. We then estimated subject-level intermodel coupling as $\hat{\rho}_i(v) = \hat{\epsilon}_{1i}(v) \times \hat{\epsilon}_{2i}(v)$. H_0 was tested by fitting the linear regression model

$$\hat{\rho}_i(v) = \beta_0(v) + z_i\beta_1(v) + x_i \times \kappa(v) \tag{5}$$

and extracting the corresponding p-value. The simulation was run 1000 times and the empirical Type I error rate was calculated as the average number of rejections with $\alpha = 0.05$.

1.3.3 Simulation Results

Under model misspecification, the Type I error rate is inflated to 0.128. Meanwhile, it is controlled at 0.042 under correct model specification.

1.4 Impact of Cluster Enhancement

1.4.1 Simulation Set-up

For a single vertex, the data generation process is the same as in Section 1.3.1, except that we set $\kappa(v) = 0$ when evaluating Type I error and $\kappa(v) = 0.2$ when evaluating power. We generated data for 5 vertices. All vertices data $y_{mi}(v^*)$ were generated using the same $z_i, x_i, \rho_i(v), \kappa(v)$, but independent to other vertices. Hence, the only difference occurred between the vertices are due to randomness in $\epsilon_{mi}(v^*)$ generation.

1.4.2 Simulation Method

Following the steps outlined in Section 1.3.2 with correct model specification, we estimated $\hat{\rho}_i(v)$ for each vertex. We consider two approaches: with and without cluster enhancement.

Without cluster enhancement For a single chosen vertex v, we computed the U score test statistic

$$U(v) = \sum_{i=1}^{N} x_i * (\hat{\rho}_i(v) - \tilde{\beta}_0(v) - z_i * \tilde{\beta}_1(v)), \tag{6}$$

where $\tilde{\beta}_0$ and $\tilde{\beta}_1$ were obtained from the null model. See Section A.1 for more details about the derivation of $U(v)$.

$B = 1000$ permutations were conducted by shuffling x_i in (6) to obtain $U_{\text{null}}(v)^{(b)}$. The empirical p-value was obtained as

$$p_{\text{perm}}(v) = \frac{\sum_{b=1}^{B}\{|U_{\text{null}}(v)^{(b)}| > |U(v)|\}}{B}. \tag{7}$$

With cluster enhancement We considered the remaining 4 vertices as neighbour of the original chosen vertex v, i.e. $|N_r(v)| = 5$. Denote these clustered vertices as v^*. Denote the between-subject adjusted residuals for each vertix as $\hat{\epsilon}_{\text{between } mi}(v)$ for each $v^* \in N_r(v)$. Then, we conducted a within-subject adjustment step on these $\hat{\epsilon}_{\text{between } mi}(v)$ as follow

$$\hat{\epsilon}_{mi}(v) = \frac{\hat{\epsilon}_{\text{between } mi}(v) - \bar{\epsilon}_{\text{between } mi}(v)}{\frac{1}{|N_r(v)|-1}\sum_{v^* \in N_r(v)}(\hat{\epsilon}_{\text{between } mi}(v^*) - \bar{\epsilon}_{\text{between}mi}(v))^2}, \tag{8}$$

where $\bar{\epsilon}_{\text{between}mi}(v) = \frac{1}{|N_r(v)|}\sum_{v^* \in N_r(v)} \hat{\epsilon}_{\text{between } mi}(v^*)$. Using these $\hat{\epsilon}_{mi}(v)$, we proceeded to obtain $U(v)$ for each $v^* \in N_r(v)$. The cluster-enhanced test statistic was calculated as follows:

$$T(v) = \frac{(\sum_{v* \in N_r(v)} U(v^*))^2}{\widehat{\text{Var}}_{H_0}(\sum_{v^* \in N_r(v)} U(v^*))}, \tag{9}$$

where $\widehat{\text{Var}}_{H_0}(\sum_{v^* \in N_r(v)} U(v^*))$ was the sample variance of the sum of $|N_r(v)|$ $U_{\text{null}}(v*)^{(b)}$ obtained by permutation done through shuffling x_i in (6) $B = 1000$ times. [*Note: We assumed that* $\kappa(v^*)$ *is the same* $\forall v^* \in N_r(v)$. *See Section A.2 for derivation of clustered score test statistic with a looser assumption.*] See Section A.3 for the derivation of theoretical $Var_{H_0}(\sum_{v^* \in N_r(v)} U(v^*))$.

Using the permuted $U_{\text{null}}(v^*)^{(b)}$, we computed

$$T_{\text{null}}(v)^{(b)} = \frac{(\sum_{v* \in N_r(v)} U_{\text{null}}(v^*)^{(b)})^2}{\widehat{\text{Var}}_{H_0}(\sum_{v* \in N_r(v)} U(v^*))}. \tag{10}$$

The empirical p-value was obtained as

$$p_{\text{perm}}(v) = \frac{\sum_{b=1}^{B}\{|T_{\text{null}}(v)^{(b)}| > |T(v)|\}}{B}. \tag{11}$$

The simulation was run 1000 times.

1.4.3 Simulation Results

Without cluster enhancement, the Type I error rate (when $\kappa = 0$) and power (when $\kappa = 0.2$) are 0.02 and 0.36 respectively. With cluster enhancement, the Type I error rate and power are 0.04 and 0.79 respectively. We observe an increase in power when cluster enhancement was applied.

1.5 Adjustment on ρ model

As shown in Section 1.3, even under correct model specification adjustment on mean and variance, we notice that the Type I error rate is still slightly below 0.05. We wish to allow the type I error to be closer to 0.05 and hence increase power. In this section, instead of fitting (5) for $\hat{\rho}_i(v)$), we consider

$$\mathbb{E}[\tanh^{-1}(\hat{\rho}_i(v))|x_i, z_i] = \beta_0(v) + z_i\beta_1(v) + x_i \times \kappa(v). \tag{12}$$

This is because while true $\rho(v)$ is strictly between $[-1, 1]$, such restriction may not hold for $\hat{\rho}_i(v)$. Therefore, we imposed the inverse tanh function on $\hat{\rho}_i(v)$ to rescale them.

We generated data for a single vertex and the data generation process is the same as in 1.3.1, except that we set $\kappa(v) = 0$ when evaluating Type I error and $\kappa(v) = 0.2$ when evaluating power. Following the "without cluster enhancement approach" outlined in Section 1.4.2, we tested H_0 by extracting the corresponding permutation-based p-value. With adjustment on the $\hat{\rho}$ model, Type 1 error rate increased to 0.044 (compared to 0.042 without adjustment) and the power when $\rho(v) = 0.2$ increased to 0.357 (compared to 0.35 without adjustment).

2 SKAT variance-component score statistic (Wu et al. (2011))

2.1 Notation

Assume n subjects are sequenced in a region with p variant sites observed. For the ith subject, y_i denotes the phenotype variable, $\mathbf{X_i} = (X_{i1}, \dots, X_{im})$ denotes the m dimensional nuisance covariates and $\mathbf{G_i} = (G_{i1}, \dots, G_{ip})$ denotes the genotypes for the p variants within the region (covariate of interest).

2.2 Model

Consider the linear model

$$y_i = \alpha_0 + \boldsymbol{\alpha}\mathbf{X_i} + \boldsymbol{\beta}\mathbf{G_i} + \epsilon_i \tag{13}$$

α_0 is the intercept term. $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_m]$ is the vector of regression coefficient for the m nuisance covariates. $\boldsymbol{\beta} = [\beta_1, \dots, \beta_m]$ is the vector of regression coefficient for the p observed gene variants in the region. ϵ_i is error term with mean zero and variance σ^2.

The goal is to evaluate whether the gene variants $\mathbf{G_i}$ influence genotypes y_i. The null hypothesis is $H_0\colon \boldsymbol{\beta} = 0$.

2.3 SKAT Test Statistic

SKAT test H_0 by assuming that each β_j follows an arbitrary distribution with mean zero and variance $w_j\tau$, where τ is the variance component and w_j is a prespecified weight for variant j. Hence, testing $H_0\colon \boldsymbol{\beta} = 0$ is equivalent to testing $H_0\colon \tau = 0$.

The variance-component score test in the corresponding mixed model is defined as

$$Q = (\mathbf{y} - \hat{\boldsymbol{\mu}})^T \mathbf{K} (\mathbf{y} - \hat{\boldsymbol{\mu}}) \tag{14}$$

where $\mathbf{K} = \mathbf{G}^T\mathbf{W}\mathbf{G}$, and $\hat{\boldsymbol{\mu}}$ is predicted mean of y under H_0, that is $\hat{\boldsymbol{\mu}} = \hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}\mathbf{X_i}$. $\hat{\alpha}_0$ and $\hat{\boldsymbol{\alpha}}$ can be estimated under the null model by regressing on only $\mathbf{X_i}$. $\mathbf{G}$ is a $n \times p$ matrix where the (i, j) entry is genotype of variant j for subject i. $\mathbf{W} = diag(w_1, \dots, w_p)$ contains the weights of the p variants.

SKAT test statistic only requires fitting the null model once to obtain $\hat{\alpha}_0$ and $\hat{\boldsymbol{\alpha}}$ and good choices of weights have been shown to improve power. It is used to define test statistic in Section 3.3 and Section 4.3.

3 CLEAN-V (Pan et al. (2024))

3.1 Notation

For images $n = 1, \ldots, N$ and vertex $v = 1, \ldots, V$, we denote the observed imaging data as $\mathbf{y}(v) = (y_1(v), \ldots, y_N(v))^T$ at vertex v. For test-retest studies, we consider images from the same person as a group, denoted as j. For twin studies, we consider images from the same family as a group. We only consider 2 images ($i = 1, 2$) per family, so there are $N/2$ families. We denote the ith image within family j as image ij. $\mathbf{X}$ is an $N \times p$ nuisance covariates matrix, each row of the matrix is a p-dimensional nuisance covariates vector for the ith image. $\mathbf{r}(v) = (r_1(v), \ldots, r_N(v))^T$ characterizes the dependence among all N images at vertex v through K.

$$K = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{N/2} \end{bmatrix}$$

where J_j is a 2 × 2 block matrices

$$J_j = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

for test-retest images in reliability studies or monozygotic twins in heritability studies, or

$$J_j = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

for dizygotic twins in heritability studies.

$\epsilon_i = (\epsilon_i(1), \ldots, \epsilon_i(V))^T$ characterizes both spatial and non-spatial variations within images. In this report, we only consider the non-spatial variations, controlled by σ^2.

3.2 Model

$$\mathbf{y}(v) = \mathbf{X}\boldsymbol{\beta}(v) + \mathbf{r}(v) + \epsilon(v) \tag{15}$$

$$\mathbf{r}(v) \sim MVN(\mathbf{0}_N, \theta^2(v) \cdot K) \qquad v = 1, \ldots, V \tag{16}$$

$$\epsilon_i = (\epsilon_i(1), \ldots, \epsilon_i(V))^T \sim^{iid} MVN(\mathbf{0}_V, \sigma^2 I_v) \quad n = 1, \ldots, N \tag{17}$$

$\mathbf{r}(v)$ and $\epsilon(v)$ are assumed to be independent of each others. Assuming $\theta^2(v)$ is a constant θ^2, the null hypothesis is $H_0 \colon \theta^2 = 0$.

3.3 CLEAN-V Test Statistic Distribution

The CLEAN-V variance component score-based test statistic based on (14) for a single vertex v is

$$U(v) = (\mathbf{y}(v) - \mathbf{X}\hat{\boldsymbol{\beta}}(v))^T K (\mathbf{y}(v) - \mathbf{X}\hat{\boldsymbol{\beta}}(v)) \quad \text{for } v = 1, \ldots, V \tag{18}$$

where $\hat{\boldsymbol{\beta}}(v)$ is to be estimated under the null hypothesis. Define $N_r(v)$ as the collection of all vertices within the radius r neighborhood of some central vertex v. The CLEAN-V clustered test statistic is

$$U_{N_r(v)}(v) = \sum_{v^* \in N_r(v)} U(v^*) \tag{19}$$

We can show that $U_{N_r(v)}(v)$ follows mixture of χ_1^2 distribution.

$$\sum_{v^* \in N_r(v)} U(v^*) = \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} \lambda_{ij} \sigma^2 H_{ij}(v^*)$$

where $H_{ij}(v^*)$ are independent χ^2_1 random variables.

Under H_0,

$$\mathbb{E}_{H_0}(U_{N_r(v)}(v)) = |N_r(v)|N\sigma^2$$

$$Var_{H_0}(U_{N_r(v)}(v)) = \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} 2\sigma^4 \lambda_{ij}^2$$

where λ_{ij} are eigenvalues of J_j. See Section A.4 for full derivation.

Using Lyapunov Central Limit Theorem, we can also show that, under H_0,

$$U_{N_r(v)}(v) \xrightarrow{d} N(|N_r(v)|N\sigma^2, \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} 2\sigma^4 \lambda_{ij}^2)$$

See Section A.5 for full derivation.

The standardized test statistic is defined as

$$T_r(v) = \frac{\sum_{v^* \in N_r(v)} U(v^*) - \mathbb{E}_{H_0}(\sum_{v^* \in N_r(v)} U(v^*))}{\sqrt{\text{VAR}_{H_0}(\sum_{v^* \in N_r(v)} U(v^*))}} \tag{20}$$

So, $T_r(v)$ asymptotically follows the standard normal distribution under H_0. In practice, $\widehat{\mathbb{E}}_{H_0}$ and $\widehat{\text{VAR}}_{H_0}$ are estimated by permutation.

3.4 Behavior of $T_r(v)$ as N increases

In traditional asymptotical setting, we usually only consider increasing N while the dimension of the design matrix is fixed. However, in CLEAN-V clustered test statistic, the dimension is $N/2$. This section is dedicated to investigate how well does the normal approximation perform under H_0 when N increases.

3.4.1 Simulation Set-up

Data is generated by setting $N = 4, 10, 50, 100, 200$, $\sigma^2 = 3$, $|N_r(v)| = 10$, $\theta^2(v) = 0$. For a given N, we first constructed the matrix K, which is a block-diagonal matrix with $N/2$ identical 2×2 all-ones blocks. For each $v^* \in N_r(v)$, we then generated $\epsilon(v^*) \sim N(0, \sigma^2 I_N)$. In this simulation, we did not generate $\mathbf{y}(v)$ or $\mathbf{X}$ to obtain residuals $\hat{\epsilon}(v^*)$, we used the actual noise $\epsilon(v^*)$ directly.

3.4.2 Simulation Method

We computed the single vertex statistic

$$U(v^*) = \epsilon(v^*)^T K \epsilon(v^*).$$

The clustered test statistic $U_{N_r(v)}(v)$ was obtained by summing $|N_r(v)|$ independent copies of $U(v^*)$. That is $U_{N_r(v)}(v) = \sum_{v^* \in N_r(v)} U(v^*)$. This data-generation and clustered test statistic calculation procedure was repeated 1000 times.

For each N, we computed the empirical mean and empirical variance of these $U_{N_r(v)}(v)$. As shown in Section 3.3, in test-retest scenario, $U_{N_r(v)}(v)$ should asymptotically follow $N(|N_r(v)|N\sigma^2, 4\sigma^4|N_r(v)|N)$.

3.4.3 Simulation Results

Below are the the empirical distribution and compared to their exact mixture χ^2 distribution as well as its normal approximation. As N increases, the normal approximation distribution is closer to the actual mixture χ^2 distribution.

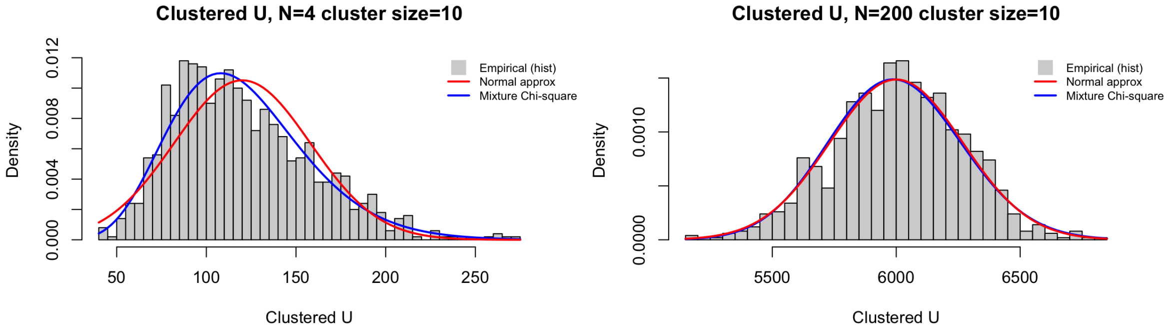


Figure 1: Comparison of distributions of $U_{N_r(v)}(v)$ for $N = 4$ (left) and $N = 200$ (right). Grey = Empirical distribution; Red = Normal approximation; Blue = exact mixture χ^2 distribution

3.5 Behavior of $T_r(v)$ as neighborhood radius r increases

3.5.1 Simulation Set-up

Data is generated by setting $N = 100$, $\sigma^2 = 0.2$, $\beta(v) = (2, 1.5)$, X consists of a column of ones (intercept) and a single covariate sampled from iid standard normal. We constructed the matrix K as a block-diagonal matrix with $N/2$ identical 2×2 all-ones blocks. We first set up a 20×20 grid, and sampled $V = 50$ sets of vertex coordinates uniformly from it. Fixing the vertex of interest as v_{25}, we set $\theta(v)^2 = 0.5$ for vertices within signal radius $= 5$ from v_{25}, 0 otherwise.

3.5.2 Simulation Method

In each simulation, regression coefficients $\hat{\beta}(v) = \hat{\beta}$ were estimated, and $U(v)$ was computed $\forall v \in 1, ..., V$. We considered these cluster radii r candidates: $\{3, 4, 5, 6, 7\}$. For each $r \in \{3, 4, 5, 6, 7\}$, we checked how many vertices are within the radius i.e., $|N_r(v)|$, the standardized test statistic $T_r(v)$ was calculated using theoretical mean and variance under H_0 instead of permutation. That is,

$$\mathbb{E}_{H_0}\left[\sum_{v^* \in N_r(v)} U(v^*)\right] = |N_r(v)| N \sigma^2$$

and

$$\mathrm{Var}_{H_0}\left(\sum_{v^* \in N_r(v)} U(v^*)\right) = 4\sigma^4 |N_r(v)| N.$$

1000 simulations were conducted to investigate the average change in the $|N_r(v)|$, $T_r(v)$, nominator of $T_r(v)$: $\sum_{v^* \in N_r(v)} U(v^*) - \mathbb{E}_{H_0}\left[\sum_{v^* \in N_r(v)} U(v^*)\right]$, and denominator (squared) of $T_r(v)$: $\mathrm{Var}_{H_0}\left(\sum_{v^* \in N_r(v)} U(v^*)\right)$ as r increases.

3.5.3 Simulation Results

As expected, when r increases, $|N_r(v)|$ increases. The optimal $T_r(v)$ is obtained at the true significant radius (5). In terms of the nominator of $T_r(v)$, it is smaller when $r < 5$, and becomes steady afterward. In terms of the squared denominator of $T_r(v)$, it continues to increase as r increases. This implies that if we choose a

1. r smaller than the true significant radius to calculate $T_r(v)$, we are missing some significant vertices information;
2. r greater than the true significant radius to calculate $T_r(v)$, we do not gain extra significant vertices information, but we sacrificed power by increasing the variance since we included too many vertices in the cluster.

This shows that if an optimal r is used, we can effectively increase statistical power.

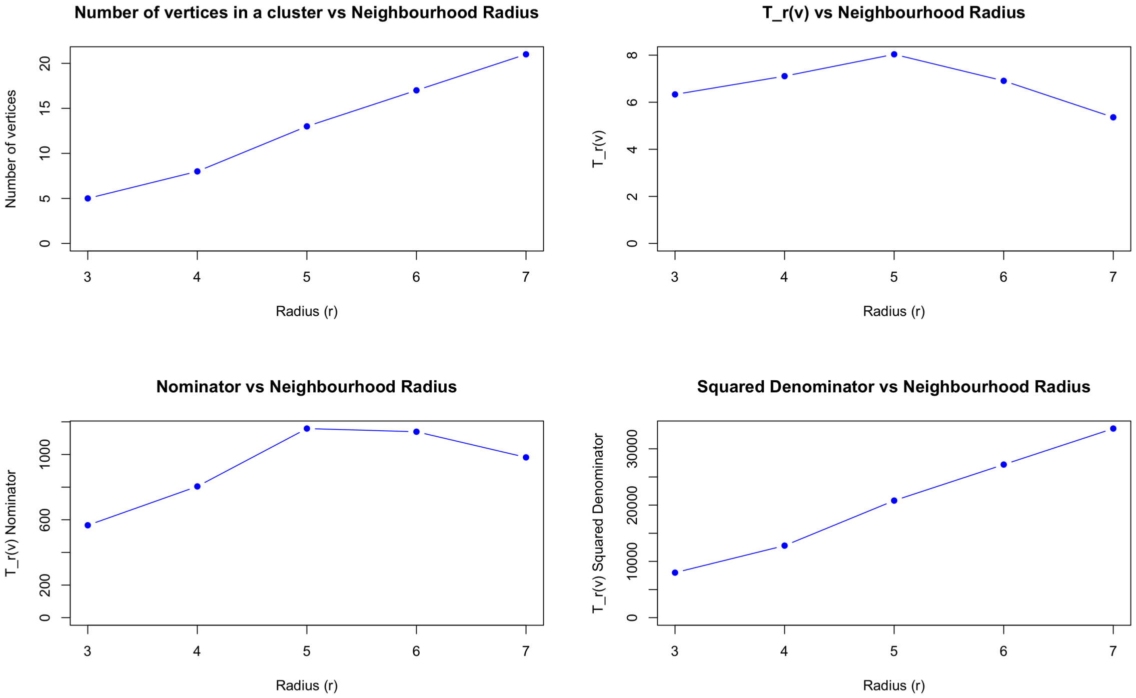


Figure 2: Comparison of $|N_r(v)|$ (top left), $T_r(v)$ (top right), nominator (bottom left) and squared denominator (bottom right) of $T_r(v)$, as a function of the neighborhood radius.

4 New Intermodel-coupling Model to Test Heritability Effect

4.1 Notation

We consider families $i = 1, \ldots, N$, each consisting of a twin pair indexed by $j = 1, 2$. Hence, we have $2N$ subjects in total. For each family $i = 1, \ldots, N$, we observe brain measurements for two modalities (s and f) as $\mathbf{y}^{(s)} = (y_{11}^{(s)}, y_{12}^{(s)}, \ldots, y_{N1}^{(s)}, y_{N2}^{(s)})^T$ and $\mathbf{y}^{(f)} = (y_{11}^{(f)}, y_{12}^{(f)}, \ldots, y_{N1}^{(f)}, y_{N2}^{(f)})^T$, respectively. We denote the ith image within family j as image ij. $\mathbf{X} \in \mathbb{R}^{2N \times p}$ is a nuisance covariates matrix, each row of the matrix is a p-dimensional nuisance covariates vector for the ijth subject. $\boldsymbol{\beta}^{(s)}, \boldsymbol{\beta}^{(f)} \in \mathbb{R}^p$ are fixed effect coefficients in first stage models (21), (22). Z is a design matrix for family level random effects. $b^{(s)}, b^{(f)}$ denote the family-specific random effects in the first stage models, each with variance σ_s^2 and σ_s^f respectively. $\epsilon_{ij}^{(s)}$ and $\epsilon_{ij}^{(f)}$ first stage model noise, each with variance τ_s^2 and τ_f^2 and have correlation ρ_{ij}. $\boldsymbol{\alpha} \in \mathbb{R}^p$ are fixed effect coefficients in second stage model (24). $b^{(c)}$ denote the family-specific random effects in the second stage models. θ^2 in (25) denote the true intermodel-coupling heritability signal of interest. In later sections, we will introduce pseudo signal η^2 and random noise in second stage model ω^2.

4.2 Model

The first stage models are:

$$\mathbf{y}^{(s)} = \mathbf{X}\boldsymbol{\beta}^{(s)} + Zb^{(s)} + \epsilon^{(s)} \tag{21}$$

$$\mathbf{y}^{(f)} = \mathbf{X}\boldsymbol{\beta}^{(f)} + Zb^{(f)} + \epsilon^{(f)} \tag{22}$$

where

$$b^{(s)} \sim MVN(0_N, \sigma_s^2 I_{N^2}),$$

$$b^{(f)} \sim MVN(0_N, \sigma_f^2 I_{N^2}),$$

$b^{(s)}$ and $b^{(f)}$ are independent.

The residuals for subject ij follow a bivariate normal distribution:

$$\begin{bmatrix}\epsilon_{ij}^{(s)}\\ \epsilon_{ij}^{(f)}\end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}\tau_s^2 & \rho_{ij}\tau_s\tau_f\\ \rho_{ij}\tau_s\tau_f & \tau_f^2\end{bmatrix}\right), \tag{23}$$

The second stage model is:

$$\rho = tanh(\mathbf{X}\boldsymbol{\alpha} + Zb^{(c)}) \tag{24}$$

$$b^{(c)} \sim MVN(0_N, \theta^2 I_{N^2}). \tag{25}$$

4.3 Hypothesis test θ^2 by Permutation

We wish to conduct hypothesis test on $H_0 : \theta^2 = 0$ using variance component score test statistic (SKAT), defined as follow

$$T = (\rho - \mathbf{X}\boldsymbol{\alpha})^T\Sigma^{-1}ZZ^T\Sigma^{-1}(\rho - \mathbf{X}\boldsymbol{\alpha}), \tag{26}$$

In practice, ρ is estimated by $\hat{\rho} = \hat{\epsilon}^{(s)} \times \hat{\epsilon}^{(f)}$. $\hat{\epsilon}^{(s)}$ and $\hat{\epsilon}^{(f)}$ can be obtained by fitting (21) and (22).

$\hat{\boldsymbol{\alpha}}$ is estimated under H_0, which is done by fitting the second stage model (24) without any random effect and assuming $tanh(t) \approx t$

$$\hat{\rho} = \mathbf{X}\boldsymbol{\alpha} + \epsilon^{(c)}.$$

Σ denote the variance of residuals after fitting the second stage null model, which is estimated by $\hat{\Sigma} = Cov(\hat{\epsilon}^{(c)})$. [*Note: In this section, we assume that the fitted residuals from data with* $\theta^2 = 0$ *do not contain pseudo signal, so* $\hat{\Sigma}$ *is a diagonal matrix.*]

In twin study, T can be rewritten as

$$T = \sum_{i=1}^{N} \frac{\{(\hat{\epsilon}_{i1}^{(s)} \times \hat{\epsilon}_{i1}^{(f)} - x_{i1}^T\hat{\alpha}) + (\hat{\epsilon}_{i2}^{(s)} \times \hat{\epsilon}_{i2}^{(f)} - x_{i2}^T\hat{\alpha})\}^2}{(\hat{\omega}_i^2)^2}, \tag{27}$$

where $\hat{\alpha}$ is estimated under null model: $\hat{\epsilon}_{ij}^{(s)} \times \hat{\epsilon}_{ij}^{(f)} = x_{ij}^T \times \alpha + \epsilon_{ij}^{(c)}$ and $\hat{\omega}_i^2$ is the corresponding variance of fitted residuals.

Using the original data, we can compute T_{obs}. Then, we proceed to conduct permutation test to determine of T_{obs} is significant. Permutation is performed by switching $\hat{\epsilon}_{i1}^{(s)}$ and $\hat{\epsilon}_{i2}^{(s)}$ in some families when calculating T_{perm}. That is, for some $i \in \{1, \ldots, N\}$, we computed

$$\frac{\{(\hat{\epsilon}_{i2}^{(s)} \times \hat{\epsilon}_{i1}^{(f)} - x_{i1}^T\hat{\alpha}) + (\hat{\epsilon}_{i1}^{(s)} \times \hat{\epsilon}_{i2}^{(f)} - x_{i2}^T\hat{\alpha})\}^2}{(\hat{\omega}_i^2)^2} \tag{28}$$

Note that $\hat{\alpha}$ should not be refitted in each permutation. Permutation based p-value is calculated as

$$p_{\text{perm}} = \frac{\sum_{b=1}^{B}\{|T_{\text{perm}}^{(b)}| > |T_{obs}|\}}{B}$$

We reject the hypothesis when p-value < 0.05.

Simulation

We considered $N = 50$ families, $J = 2$ for each family, $p = 1$, $\beta^{(s)} = 3$, $\beta^{(f)} = -2$, $\alpha = 1$,$\sigma_s^2 = 0.5$, $\sigma_f^2 = 0.7$, $\tau_s^2 = 1.0$, $\tau_f^2 = 1.2$, $\theta^2 = 0$, $\mathbf{X} \sim^{iid} N(0, 1)$. We set the number of permutations as 500 and conducted 1000 simulations.

Type I error rate is 0.004 when $\hat{\alpha}$ is not refitted after each permutation, which is too conservative. If $\hat{\alpha}$ is refitted for each permutation, Type I error rate is inflated to 0.076.

4.4 Estimation and Hypothesis testing on θ^2 by Bootstrapping

In Section 4.3, the fitted residuals using data with no true signal do not contain pseudo signal and are uncorrelated. Given the complexity of the model, it is reasonable to assume that with null data, the second stage model fitted residuals may still contain some signal. That is, with null data

$$\hat{\rho} = \mathbf{X}\boldsymbol{\alpha} + Z(b_{pseudo}^{(c)}) + \epsilon^{(c)} = \mathbf{X}\boldsymbol{\alpha} + \delta$$

and $Cov(\delta)$ is not a diagonal matrix. If we wish to estimate θ^2, the traditional approach would be estimating the random effect variance component in (24) from the original data. However, with presence of pseudo signal, the model is

$$\begin{aligned}\hat{\rho} &= \mathbf{X}\boldsymbol{\alpha} + Z(b_{pseudo}^{(c)}) + Z(b_{true}^{(c)}) + \epsilon^{(c)} \\ &= \mathbf{X}\boldsymbol{\alpha} + Zb^{(c)} + \epsilon^{(c)},\end{aligned}$$

but we are only able to compute $Var(b^{(c)})$, which is an estimate of the sum of true and pseudo signal, i.e., $\widehat{\theta^2 + \eta^2}$. Our goal is to estimate pseudo signal η^2 using bootstrap samples.

4.4.1 Bootstrap sample generation method

1. Using the original data, fit (21) and (22) to obtain $\hat{\boldsymbol{\beta}}^{(s)}, \hat{\boldsymbol{\beta}}^{(f)}, \hat{b}^{(s)}, \hat{b}^{(f)}, \hat{\epsilon}^{(s)}$, $\hat{\epsilon}^{(f)}$, $\hat{\tau}^{(s)} = Var(\hat{\epsilon}^{(s)})$, and $\hat{\tau}^{(f)} = Var(\hat{\epsilon}^{(f)})$
2. Estimate $\hat{\boldsymbol{\alpha}}$
3. Compute $\hat{\rho}^{(b)} = \mathbf{X}\hat{\boldsymbol{\alpha}}$
4. Generate bootstrapped $\hat{\epsilon}^{(s)(b)}$, and $\hat{\epsilon}^{(f)(b)}$ by

$$\begin{pmatrix} \hat{\epsilon}_{ij}^{(s)(b)} \\ \hat{\epsilon}_{ij}^{(f)(b)} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\tau}_s^2 & \hat{\rho}_{ij}^{(b)} \hat{\tau}_s \hat{\tau}_f \\ \hat{\rho}_{ij}^{(b)} \hat{\tau}_s \hat{\tau}_f & \hat{\tau}_f^2 \end{pmatrix} \right)$$

5. Generate $y^{(s)(b)}$ and $y^{(f)(b)}$ by

$$y^{(s)(b)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(s)} + \hat{b}^{(s)} + \hat{\epsilon}^{(s)(b)} \tag{29}$$

$$y^{(f)(b)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(f)} + \hat{b}^{(f)} + \hat{\epsilon}^{(f)(b)} \tag{30}$$

In Step 2, there are two ways to estimate $\hat{\boldsymbol{\alpha}}$:

1. Simple linear regression: $\text{lm}(\hat{\epsilon}^{(s)}\hat{\epsilon}^{(f)} \sim \mathbf{X}\boldsymbol{\alpha})$
2. MLE using $\hat{\epsilon}^{(s)}, \hat{\epsilon}^{(f)}, \hat{\tau}^{(s)}, \hat{\tau}^{(f)}$

The MVN model that MLE is obtained from is

$$\begin{pmatrix} \hat{\epsilon}_{ij}^{(s)} \\ \hat{\epsilon}_{ij}^{(f)} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\tau}_s^2 & \rho_{ij} \hat{\tau}_s \hat{\tau}_f \\ \rho_{ij} \hat{\tau}_s \hat{\tau}_f & \hat{\tau}_f^2 \end{pmatrix} \right)$$

where $\rho_{ij} = tanh(x_{ij}^T \alpha)$.

In Step 5, an alternative approach is to generate $\hat{b}^{(b)(s)} \sim N(0, \hat{\sigma}^{(s)})$ and $\hat{b}^{(b)(f)} \sim N(0, \hat{\sigma}^{(f)})$ and use them in (29) and (30).

4.4.2 θ^2 Estimation

Using $y^{(s)(b)}$ and $y^{(f)(b)}$, fit (21), (22) and (24), the variance of bootstrapped random intercept $b^{(c)(b)}$ gives the pseudo signal estimate $(\widehat{\eta^{2(b)}})$. With the original data $y^{(s)}$ and $y^{(f)}$, also fit (21), (22) and (24), the variance of fitted random intercept $b^{(c)}$ gives an estimate of true plus pseudo signal $(\widehat{\eta^2 + \theta^2})$. Hence, our interested signal estimator can be calculated as the signal from original data minus signal from bootstrapped data, i.e. $\hat{\theta}^2 = \widehat{\eta^2 + \theta^2} - \widehat{\eta^{2(b)}}$.

4.4.3 θ^2 Hypothesis Testing by Bootstrapping

Following (26), the score test statistic is

$$T = (\rho - \mathbf{X}\boldsymbol{\alpha})^T \Sigma^{-1} Z Z^T \Sigma^{-1} (\rho - \mathbf{X}\boldsymbol{\alpha}),$$

where $\Sigma = Cov(\epsilon^{(c)})$. ρ is estimated with $\hat{\rho} = \hat{\epsilon}^{(s)}\hat{\epsilon}^{(f)}$. As shown above, $\epsilon^{(c)}$ contains pseudo signal and noise. So, $\Sigma = \eta^2 K + \omega^2 I$ and it is estimated by $\hat{\Sigma} = \hat{\eta}^2 K + \hat{\omega}^2 I$

For each bootstrap sample, we fit

$$\hat{\rho}^{(b)} = \hat{\epsilon}^{(s)(b)}\hat{\epsilon}^{(f)(b)} = \mathbf{X}\boldsymbol{\alpha}^{\mathbf{(b)}} + b^{(c)(b)} + \epsilon^{(c)}. \tag{31}$$

to obtain an estimate for pseudo signal variance ($\hat{\eta}^{2(b)} = Var(\hat{b}^{(c)(b)})$) and noise variance ($\hat{\omega}^{2(b)} = Var(\hat{\epsilon}^{(c)})$). Then, $\hat{\Sigma}^{(b)} = \hat{\eta}^{2(b)} K + \hat{\omega}^{2(b)} I$. $\boldsymbol{\alpha}^{\mathbf{(b)}}$ is obtained from fitting a simple linear regression: lm($\hat{\epsilon}^{(s)(b)}\hat{\epsilon}^{(f)(b)} \sim \mathbf{X}\boldsymbol{\alpha}$).

Bootstrap based p-value was calculated as

$$p_{\text{boot}} = \frac{\sum_{b=1}^{B} \{|T_{\text{boot}}^{(b)}| > |T_{obs}|\}}{B}$$

We reject the hypothesis when p-value < 0.05.

4.4.4 Simulation

Data was generated by setting $N = 30, 50, 100$, $p = 1$, $\hat{\beta}^{(s)} = 3$, $\hat{\beta}^{(s)} = -2$, $\alpha = 0.1, 0.5, 1$, $\theta^2 = 0$, $\tau_s^2 = \tau_f^2 = 1$, $\sigma_s^2 = \sigma_f^2 = 1$, X consists of a single covariate sampled from iid standard normal. For each simulation, we generated 500 bootstrap samples. We conducted 1000 simulations. α is estimated by Lm, MLE or a benchmark approach through MLE using true $\epsilon^{(s)}, \epsilon^{(f)}, \tau^{(s)}, \tau^{(f)}$.

Simulation Results

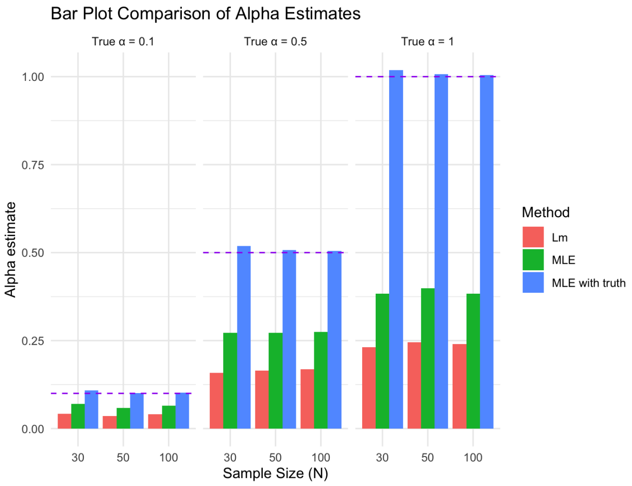
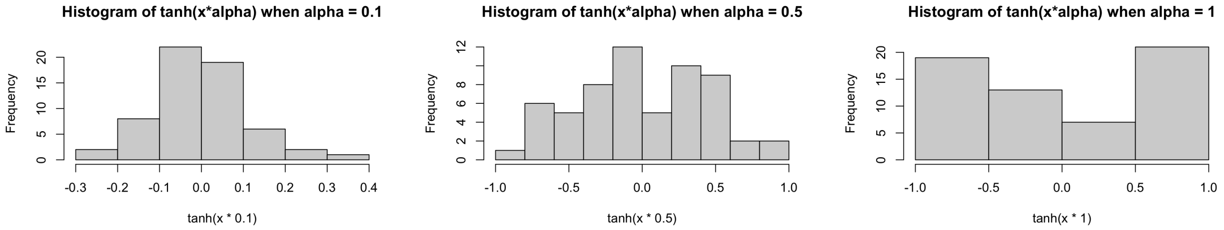


Figure 3: Estimates of alpha using different estimation approaches and N

As shown in Figure (3), MLE with truth performs the best as expected. MLE approach performs better than Lm approach in estimating α regardless of N. Notice that the performance of Lm and MLE worsen as α increases. This is because $tanh(\mathbf{X}\alpha)$ tends to land on the boundary of -1 and 1. See Figure (4).

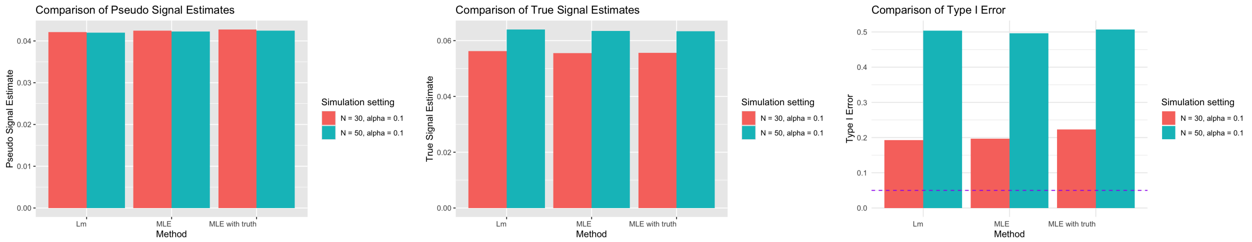
Since $\hat{\alpha}$ is no longer reliable when α is large, simulations on θ^2 estimation and hypothesis testing was only conducted on data with $\alpha = 0.1$, $N = 30, 50$. Results shown in Figure (5) have bootstrap samples generated using the fixed $\hat{b}^{(s)}, \hat{b}^{(f)}$. No obvious difference is observed in terms of $\hat{\eta}^2$ and $\hat{\theta}^2$ results between the 3 α estimation approaches given a simulation setting. In particular, there seems to be minimal difference in $\hat{\eta}^2$ as N changes. On the other hand, $\hat{\theta}^2$ is more inflated as N increases. In all methods and simulation setting, we observe inflated Type I error rates.

Figure (6) compares the performance between simulations using bootstrap samples generated with and without bootstrap first stage random intercepts. Using bootstrap samples generated with bootstrap first stage



(a) tanh 0.1 (b) tanh 0.5 (c) tanh 1.0

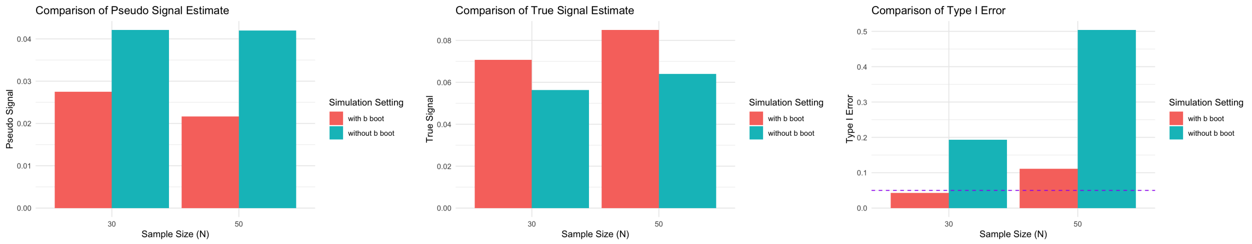
Figure 4: Comparison of tanh values at different α



(a) Pseudo signal estimate (b) True signal estimate (c) Type I Error

Figure 5: Comparison of pseudo signal, true signal, and T1E estimates.

random intercepts, both $\hat{\eta}^2$ and $\hat{\theta}^2$ are lower. When $N = 30$, Type I error rate is controlled. In $N = 50$, Type I error rate is still inflated but there is significant improvment compared to the previous approach.



(a) Pseudo signal estimate (b) True signal estimate (c) Type I Error

Figure 6: Comparison of Pseudo signal, True signal, and Type I Error with and without first stage bootstrap random intercepts.

4.5 Quantifying Pseudo Signal

To evaluate biasedness of $\hat{\eta}^2$, we need to quantify η^2 under H_0. As mentioned in Section 4.4, the second stage model of null data is

$$\hat{\rho} = \mathbf{X}\boldsymbol{\alpha} + Zb^{(c)}_{pseudo} + \epsilon^{(c)} = \mathbf{X}\boldsymbol{\alpha} + \delta$$

and $\hat{\rho} \sim (\mathbf{X}\boldsymbol{\alpha}, \eta^2 ZZ^T + \omega^2 I)$.
Since $\hat{\rho} = \hat{\epsilon}^{(s)} \times \hat{\epsilon}^{(f)}$,

$$\begin{aligned}
\eta^2 &= Cov(\hat{\rho}_{i1}, \hat{\rho}_{i2}) \\
&= Cov(\hat{\epsilon}^{(s)}_{i1}\hat{\epsilon}^{(f)}_{i1}, \hat{\epsilon}^{(s)}_{i2}\hat{\epsilon}^{(f)}_{i2}) \\
&= \mathbb{E}[\hat{\epsilon}^{(s)}_{i1}\hat{\epsilon}^{(f)}_{i1}\hat{\epsilon}^{(s)}_{i2}\hat{\epsilon}^{(f)}_{i2}] - \mathbb{E}[\hat{\epsilon}^{(s)}_{i1}\hat{\epsilon}^{(f)}_{i1}]\mathbb{E}[\hat{\epsilon}^{(s)}_{i2}\hat{\epsilon}^{(f)}_{i2}]
\end{aligned}$$

The joint distribution of the fitted first stage residuals is

$$\begin{bmatrix}\hat{\epsilon}^{(s)}\\ \hat{\epsilon}^{(f)}\end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix}P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\\ P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\end{bmatrix}, \begin{bmatrix}P^{(s)}\Sigma^{(s)}P^{(s)T} & P^{(s)}DP^{(f)T}\\ P^{(s)}DP^{(f)T} & P^{(f)}\Sigma^{(f)}P^{(f)T}\end{bmatrix}\right),$$

where $P^{(s)}, P^{(f)}, D, \Sigma^{(s)}, \Sigma^{(f)}$ can be expressed in terms of $\sigma_s^2, \sigma_f^2, \tau_s^2, \tau_f^2, \boldsymbol{\alpha}$, and $\mathbf{X}$.

By Isserlis' Theorem, we can derive that

$$\begin{align}
\eta^2 = &\{P^{(s)}\mathbf{X}\boldsymbol{\beta}_{i1}^{(s)}\}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(f)}\Sigma^{(f)}P^{(f)T}\}_{(i1,i2)} \tag{32}\\
&+ \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i1}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i2}\{P^{(f)}DP^{(s)T}\}_{(i2,i1)} \tag{33}\\
&+ \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(s)}DP^{(f)T}\}_{(i1,i2)} \tag{34}\\
&+ \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(f)}\mathbf{X}\beta^{(f)}\}_{i2}\{P^{(s)}\Sigma^{(s)}P^{(s)T}\}_{(i1,i2)} \tag{35}\\
&+ \{P^{(s)}\Sigma^{(s)}P^{(s)T}\}_{(i1,i2)}\{P^{(f)}\Sigma^{(f)}P^{(f)T}\}_{(i1,i2)} \tag{36}\\
&+ \{P^{(s)}DP^{(f)T}\}_{(i1,i2)}\{P^{(f)}DP^{(s)T}\}_{(i2,i1)} \tag{37}
\end{align}$$

See Section A.6 for full derivation.

Based (32), we can evaluate the biasedness of $\hat{\rho}$ in previous simulations. We may also consider plugging in estimated value of $\sigma_s^2, \sigma_f^2, \tau_s^2, \tau_f^2, \boldsymbol{\alpha}$ and use (32) to obtain $\hat{\Sigma}$ when computing test statistic.

A Appendix

A.1 Derivation of CEIDR Clustered score test statistic

According to (3) and treating $tanh(t) \approx t$, the loglikelihood is

$$\ell(\beta_0(v^*), \beta_1(v^*), \kappa(v^*)) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(\rho_i(v^*) - \beta_0(v) - z_i \times \beta_1(v^*) - x_i \times \kappa(v^*))^2.$$

The score function is

$$S(v^*) = \frac{\partial\ell(\beta_0(v^*), \beta_1(v^*), \kappa(v^*))}{\partial\kappa(v^*)} = \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i \times (\rho_i(v^*) - \beta_0(v^*) - z_i \times \beta_1(v^*) - x_i \times \kappa(v^*)).$$

So, the score test statistic for a single vertex v^* is

$$U(v^*) = \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i \times (\rho_i(v^*) - \tilde{\beta_0}(v^*) - z_i * \tilde{\beta_1}(v^*)),$$

where $\tilde{\beta_0}$ and $\tilde{\beta_1}$ are obtained from the null model.

For a fixed vertex v, consider a neighborhood of vertices denoted as $N_r(v)$. Assume $\kappa(v^*) = \kappa$ constant, $\forall v^* \in N_r(v)$. The corresponding loglikelihood is

$$\ell(\beta_0, \beta_1, \kappa) = \sum_{v^* \in N_r(v)} \ell(\beta_0(v^*), \beta_1(v^*), \kappa(v^*)),$$

and the corresponding score function is

$$\begin{aligned}\frac{\partial\ell(\beta_0, \beta_1, \kappa)}{\partial\kappa} &= \frac{\partial\sum_{v^* \in N_r(v)} \ell(\beta_0(v^*), \beta_1(v^*), \kappa(v^*))}{\partial\kappa} \\ &= \frac{\sum_{v^* \in N_r(v)} \partial\ell(\beta_0(v^*), \beta_1(v^*), \kappa(v^*))}{\partial\kappa} \\ &= \sum_{v^* \in N_r(v)} S(v^*)\end{aligned}$$

Hence, the cluster score test statistic is

$$\sum_{v^* \in N_r(v)} U(v^*) = \frac{1}{\sigma^2}\sum_{v^* \in N_r(v)}\sum_{i=1}^{N} x_i \times (\rho_i(v^*) - \beta_0(v^*) - z_i \times \beta_1(v^*)).$$

Notice that the cluster score test statistic obtained here is a sum of "sum of $x_i \times$ residuals under null model" across vertices.

A.2 Derivation of CEIDR Clustered score test statistic with looser assumption

Suppose that $\kappa(v^*)$ is some multiple of a fixed constant κ $\forall v^* \in N_r(v)$. That is, $\kappa(v^*) = w(v^*)\kappa$, where $1 \geq w(v^*)\kappa \geq -1$, $\forall v^* \in N_r(v)$.

According to (3) and treating $tanh(t) \approx t$, the score function is

$$\begin{aligned}S(v^*) &= \frac{\partial\ell(\beta_0(v^*), \beta_1(v^*), \kappa(v^*))}{\partial\kappa(v^*)} \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i \times w(v^*) \times (\rho_i(v^*) - \beta_0(v^*) - z_i \times \beta_1(v^*) - x_i \times w(v^*) \times \kappa),\end{aligned}$$

and the score test statistic is

$$U(v^*) = \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i \times w(v^*) \times (\rho_i(v^*) - \tilde{\beta_0}(v^*) - z_i * \tilde{\beta_1}(v^*)),$$

where $\tilde{\beta_0}$ and $\tilde{\beta_1}$ are obtained from the null model.
The corresponding score function for the clustered model is

$$\begin{aligned} S(N_r(v)) &= \sum_{v^* \in N_r(v)} S(v^*) \\ &= \frac{1}{\sigma^2} \sum_{v^* \in N_r(v)} \sum_{i=1}^{N} x_i \times w(v^*) \times (\rho_i(v^*) - \beta_0(v^*) - z_i \times \beta_1(v^*) - x_i \times w(v^*) \times \kappa), \end{aligned}$$

and the cluster score test statistic is

$$\begin{aligned} U(N_r(v)) &= \sum_{v^* \in N_r(v)} U(v^*) \\ &= \frac{1}{\sigma^2} \sum_{v^* \in N_r(v)} \sum_{i=1}^{N} x_i \times w(v^*) \times (\rho_i(v^*) - \beta_0(v^*) - z_i \times \beta_1(v^*)) \\ &= \frac{1}{\sigma^2} \sum_{v^* \in N_r(v)} w(v^*) \left[\sum_{i=1}^{N} x_i \times (\rho_i(v^*) - \beta_0(v^*) - z_i \times \beta_1(v^*)) \right]. \end{aligned}$$

Notice that the cluster score test statistic obtained here is a **weighted** sum of "sum of $x_i \times$ residuals under null model" across vertices.

A.3 Derivation of Variance of CEIDR score test statistic under H_0

For a single vertex and writing (3) in matrix form, we have

$$\vec{\rho}(v^*) = \beta_0(v^*) \times 1_n + \vec{Z} \times \beta_1(v^*) + \vec{X} \times \kappa(v^*) + \vec{\epsilon}(v^*),$$

where $\vec{\rho}(v^*) = (\rho_1(v^*), \ldots, \rho_N(v^*)) \in \mathbb{R}^N$, $\vec{Z} = (z_1, \ldots, z_N) \in \mathbb{R}^N$, $\vec{X} = (x_1, \ldots, x_N) \in \mathbb{R}^N$, $\vec{\epsilon}(v^*) = (\epsilon_1(v^*), \ldots, \epsilon_N(v^*)) \in \mathbb{R}^N$ with $\vec{\epsilon}(v^*) \sim MVN(0_n, \sigma^2 I_N)$.

Using Kronecker product, we can combine to get a full model for all vertices as follows, let

$$\rho = \begin{bmatrix} \rho_1(1) \\ \vdots \\ \rho_N(1) \\ \vdots \\ \rho_1(V) \\ \vdots \\ \rho_N(V) \end{bmatrix} \in \mathbb{R}^{NV \times 1} \qquad Z = \begin{bmatrix} 1 & z_1 \\ \vdots & \\ 1 & z_N \end{bmatrix} \in \mathbb{R}^{N \times 2} \qquad \vec{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^{N \times 1}$$

$$\alpha = \begin{bmatrix} \beta_0(1) \\ \beta_1(1) \\ \vdots \\ \beta_0(V) \\ \beta_1(V) \end{bmatrix} \in \mathbb{R}^{2V \times 1} \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1(1) \\ \vdots \\ \epsilon_N(1) \\ \vdots \\ \epsilon_1(V) \\ \vdots \\ \epsilon_N(V) \end{bmatrix} \in \mathbb{R}^{NV \times 1} \qquad \boldsymbol{\kappa} = \begin{bmatrix} \kappa(1) \\ \vdots \\ \kappa(V) \end{bmatrix} \in \mathbb{R}^{V \times 1}$$

$$X^* \;=\; I_V \otimes \vec{X}, \qquad Z^* \;=\; I_V \otimes Z \qquad \rho = Z^*\alpha + X^*\boldsymbol{\kappa} + \boldsymbol{\epsilon}.$$

Assuming that $\kappa(v^*) = \kappa \; \forall v^* \in N_r(v)$, the model reduces to

$$\rho = Z^*\alpha + X^*\kappa + \boldsymbol{\epsilon}.$$

Under the null hypothesis of $H_0\colon \kappa(v^*) = 0 \; \forall v^* \in N_r(v)$, we have $\rho \sim MVN(Z^*\alpha, I_V \otimes \sigma^2 I)$.

The score test statistic under H_0 is

$$\begin{aligned}
U(\kappa) &= \frac{1}{\sigma^2} X^{*T}(\rho - Z^*\tilde{\alpha}) \\
&= \frac{1}{\sigma^2} X^{*T}(I - Z^*(Z^{*T}Z^*)^{-1}Z^{*T})\rho \\
&= \sum_{v^* \in N_r(v)} \frac{1}{\sigma^2} \vec{X}^T(I - Z(Z^TZ)^{-1}Z^T)\vec{\rho}(v^*) \\
&= \frac{1}{\sigma^2} \vec{X}^T(I - Z(Z^TZ)^{-1}Z^T) \sum_{v^* \in N_r(v)} \vec{\rho}(v^*)
\end{aligned}$$

Let $Q = I - Z(Z^TZ)^{-1}Z^T$ and $A = \frac{1}{\sigma^2}\vec{X}^T(I - Z(Z^TZ)^{-1}Z^T) = \frac{1}{\sigma^2}\vec{X}^TQ$. Note that $Q^2 = Q$ and $\vec{\rho}(v^*)$ are pairwise uncorrelated as $\rho \sim MVN(Z^*\alpha, I_V \otimes \sigma^2 I)$. Then,

$$\begin{aligned}
Var_{H_0}\left[U(\kappa)\right] &= Var_{H_o}\left[A \sum_{v^* \in N_r(v)} \vec{\rho}(v^*)\right] \\
&= ACov_{H_o}\left[\sum_{v^* \in N_r(v)} \vec{\rho}(v^*)\right] A^T \\
&= \frac{1}{\sigma^2}\vec{X}^TQ \sum_{v^* \in N_r(v)} Var_{H_o}\left[\vec{\rho}(v^*)\right] Q\vec{X}\frac{1}{\sigma^2} \\
&= \frac{1}{\sigma^2}\vec{X}^TQ \left[\sum_{v^* \in N_r(v)} \sigma^2 I\right] Q\vec{X}\frac{1}{\sigma^2} \\
&= \frac{1}{\sigma^2}\vec{X}^TQ|N_r(v)|\sigma^2 IQ\vec{X}\frac{1}{\sigma^2} \\
&= \frac{|N_r(v)|}{\sigma^2}\vec{X}^TQQ\vec{X} \\
&= \frac{|N_r(v)|}{\sigma^2}\vec{X}^TQ\vec{X} \\
&= \frac{|N_r(v)|}{\sigma^2}\vec{X}^T(I - Z(Z^TZ)^{-1}Z^T)\vec{X}
\end{aligned}$$

A.4 Clustered U(v) Follows Mixture of χ^2 Distribution

For a single vertex, as shown in (18),

$$U(v) = \hat{\epsilon}(v)^T K \hat{\epsilon}(v),$$

where $\hat{\epsilon}(v)$ are the residuals after fitting the null model.

Note that K composes of 2x2 block matrices of either

$$J_j = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad J_j = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

along the diagonal. So,

$$K = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{N/2} \end{bmatrix}$$

Hence,

$$U(v) = \hat{\epsilon}(v)^T K \hat{\epsilon}(v) = \sum_{j=1}^{N/2} \hat{\epsilon}_j(v)^T J_j \hat{\epsilon}_j(v),$$

where $\epsilon_j(v)$ are 1x2 matrices with entries are residuals from the jth group (i.e., same subject in test-retest or twin pair). So, it is sufficient to consider the eigen-decomposition of block matrices J_j only.

$$J_j = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}},$$

which gives

$$Q_j = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \Lambda_j = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \qquad \lambda_{1j} = 2 \qquad \lambda_{2j} = 0,$$

or

$$J_j = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}}$$

which gives

$$Q_j = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \Lambda_j = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix} \qquad \lambda_{1j} = 1.5 \qquad \lambda_{2j} = 0.5$$

Then,

$$\begin{aligned} \hat{\epsilon}(v)^T K \hat{\epsilon}(v) &= \sum_{j=1}^{N/2} \hat{\epsilon}_j(v)^T J_j \hat{\epsilon}_j(v) \\ &= \sum_{j=1}^{N/2} \hat{\epsilon}_j(v)^T Q_j^T \Lambda_j Q_j \hat{\epsilon}_j(v) \\ &= \sum_{j=1}^{N/2} (Q_j \hat{\epsilon}_j(v))^T \Lambda_j (Q_j \hat{\epsilon}_j(v)) \\ &= \sum_{j=1}^{N/2} \sum_{i=1}^{2} \lambda_{ij} z_{ij}^2(v), \end{aligned}$$

where we denote z_{ij} as follows

$$Q_j \epsilon_j(v) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \epsilon_{1j}(v) \\ \epsilon_{2j}(v) \end{bmatrix} = \begin{bmatrix} \frac{\epsilon_{1j}(v)+\epsilon_{2j}(v)}{\sqrt{2}} \\ \frac{\epsilon_{1j}(v)-\epsilon_{2j}(v)}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} z_{1j}(v) \\ z_{2j}(v) \end{bmatrix}$$

Note that z_{ij} are linear combination of normal random variables $\epsilon_{1j}(v)$ and $\epsilon_{2j}(v)$, so it is normally distributed as well. In particular,

$$\mathbb{E}(z_{ij}(v)) = \mathbb{E}(\frac{\epsilon_{1j}(v) \pm \epsilon_{2j}(v)}{\sqrt{2}}) = \frac{1}{\sqrt{2}} \mathbb{E}(\epsilon_{1j}(v) \pm \epsilon_{2j}(v)) = \frac{1}{\sqrt{2}}(0 \pm 0) = 0,$$

$$Var(z_{ij}(v)) = Var(\frac{\epsilon_{1j}(v) \pm \epsilon_{2j}(v)}{\sqrt{2}}) = \frac{1}{2} Var(\epsilon_{1j}(v) \pm \epsilon_{2j}(v)) = \frac{1}{2}(2\sigma^2) = \sigma^2,$$

because $\epsilon_{1j}(v)$ and $\epsilon_{2j}(v)$ are pairwise uncorrelated and have common variance σ^2. Now, we will show that $z_{ij}(v)$ are pairwise uncorrelated. Let $k \neq j$ (i.e., non-test-retest pair; non-twin pair), then

$$Cov(z_{ij}(v), z_{lk}(v)) = \frac{1}{2}Cov(\epsilon_{1j}(v) \pm \epsilon_{2j}(v), \epsilon_{1k}(v) \pm \epsilon_{2k}(v)) = 0$$

because all ϵ terms are pairwise uncorrelated. Now, consider test-retest pair or twin pair, then

$$Cov(z_{1j}(v), z_{2j}(v)) = \frac{1}{2}Cov(\epsilon_{1j}(v) + \epsilon_{2j}(v), \epsilon_{1j}(v) - \epsilon_{2j}(v)) = \frac{1}{2}(\sigma^2 + 0 - 0 - \sigma^2) = 0$$

By the eigen-decomposition method above, we have the clustered test statistic as follows

$$\sum_{v^* \in N_r(v)} U(v^*) = \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} \lambda_{ij} z_{ij}^2(v^*) = \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} \lambda_{ij} \sigma^2 H_{ij}(v^*)$$

where $z_{ij} \sim^{iid} N(0, \sigma^2)$ and $z_{ij}^2 \sim^{iid} \sigma^2 \chi_1^2$. So, the clustered U(v) is a weighted sum of $N \times |N_r(v)|$ χ_1^2 random variables (i.e., $H_{ij}(v^*)$) with weights of $\lambda_{ij}\sigma^2$.

The mean and variance derived based on mixture χ^2 distribution are as follow

$$\begin{aligned}
\mathbb{E}_{H_0}(U_{N_r(v)}(v)) &= \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} \lambda_{ij} \sigma^2 \mathbb{E}(H_{ij}) \\
&= \sum_{i=1}^{2} \lambda_{ij} \sigma^2 \frac{|N_r(v)| N}{2} \\
&= |N_r(v)| N \sigma^2
\end{aligned}$$

since $\lambda_{i1} + \lambda_{i2} = 2$.

$$\begin{aligned}
Var_{H_0}(U_{N_r(v)}(v)) &= \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} \lambda_{ij}^2 \sigma^4 Var(H_{ij}(v^*)) \\
&= \sum_{v^* \in N_r(v)} \sum_{j=1}^{N/2} \sum_{i=1}^{2} 2\sigma^4 \lambda_{ij}^2
\end{aligned}$$

A.5 Clustered U(v) Normal Approximation

Lyapunov Central Limit Theorem states that for $\{X_1, \ldots, X_n, \ldots\}$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If for some $\delta > 0$,

$$\lim_{n \to \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\big[\, |X_i - \mu_i|^{2+\delta} \,\big] = 0,$$

then the sum of standardized variables

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{d} N(0, 1)$$

as $n \to \infty$.

In our case, we have independent χ_1^2 random variables $H_{ij}(v^*)$, each with expected value 1 and variance 2. We simplify the notation here to: we have χ_1^2 random variables $H_1, \ldots, H_N, \ldots$ and their corresponding weights $\lambda_1, \ldots, \lambda_N, \ldots$. Then, $\lambda_i \sigma^2 H_i$ are χ_1^2 random variables with expected value $\lambda_i \sigma^2$ and variance $2\lambda_i^2 \sigma^4$.

Let $X_i = \lambda_i \sigma^2 H_i$, and $\mu_i = \lambda_i \sigma^2$. Fix $\delta = 2$,

$$\sum_{i=1}^{n} \mathbb{E}\big[\, (\lambda_i \sigma^2 H_i - \lambda_i \sigma^2)^{2+\delta} \,\big] = \sum_{i=1}^{n} \mathbb{E}\big[\, (\lambda_i \sigma^2)^4 (H_i - 1)^4 \,\big] = 48\sigma^8 \sum_{i=1}^{n} \lambda_i^4$$

because the 4th moment of χ_1^2 variables is 48, i.e. $\mathbb{E}\big[(H_i-1)^4\big]=48$.

Also, $s_n^2=\sum_{i=1}^n Var(\lambda_i\sigma^2 H_i)=\sum_{i=1}^n \lambda_i^2\sigma^4 Var(H_i)=\sum_{i=1}^n 2\sigma^4\lambda_i^2$.

Then,

$$\begin{aligned}\frac{1}{s_n^4}\sum_{i=1}^n \mathbb{E}\big[(\lambda_i\sigma^2 H_i-\lambda_i\sigma^2)^4\big]=\frac{48\sigma^8\sum_{i=1}^n\lambda_i^4}{(\sum_{i=1}^n 2\sigma^4\lambda_i^2)^2}=\frac{48\sum_{i=1}^n\lambda_i^4}{4(\sum_{i=1}^n\lambda_i^2)^2}&\le\frac{12n\lambda_{max}^4}{(n\lambda_{min}^2)^2}\\&=\frac{12n\lambda_{max}^4}{n^2\lambda_{min}^4}\\&=\frac{12\lambda_{max}^4}{n\lambda_{min}^4}\end{aligned}$$

Then, as $n\to\infty$, it converges to 0. So, with Lyapunov Central Limit Theorem, we know that

$$\frac{1}{s_n}\sum_{i=1}^n(\lambda_i\sigma^2 H_i-\lambda_i\sigma^2)\xrightarrow{d}N(0,1)$$

as $n\to\infty$, which implies that

$$\sum_{i=1}^n\lambda_i\sigma^2 H_i\xrightarrow{d}N(\sum_{i=1}^n\lambda_i\sigma^2,s_n^2)$$

as $n\to\infty$.

So,

$$\sum_{v^*\in N_r(v)}U(v^*)\xrightarrow{d}N(\sum_{v^*\in N_r(v)}\sum_{j=1}^{N/2}\sum_{i=1}^2\lambda_{ij}\sigma^2,\sum_{v^*\in N_r(v)}\sum_{j=1}^{N/2}\sum_{i=1}^2 2\lambda_{ij}^2\sigma^4)$$

under H_0. Then, we have,

$$\mathbb{E}_{H_0}(U_{N_r(v)}(v))=\sum_{v\in N_r(v)}\sum_{j=1}^{N/2}\sum_{i=1}^2\lambda_{ij}\sigma^2=|N_r(v)|N\sigma^2$$

since $\lambda_{1j}+\lambda_{2j}=2\ \forall j$.

$$Var_{H_0}(U_{N_r(v)}(v))=\sum_{v^*\in N_r(v)}\sum_{j=1}^{N/2}\sum_{i=1}^2 2\sigma^4\lambda_{ij}^2$$

A.6 Derivation of Pseudo Signal as a function of parameters

Under H_0, data is generated with

$$\mathbf{y}^{(s)}=\mathbf{X}\boldsymbol{\beta}^{(s)}+Zb^{(s)}+\epsilon^{(s)}=\mathbf{X}\boldsymbol{\beta}^{(s)}+\delta^{(s)}$$

$$\mathbf{y}^{(f)}=\mathbf{X}\boldsymbol{\beta}^{(f)}+Zb^{(f)}+\epsilon^{(f)}=\mathbf{X}\boldsymbol{\beta}^{(f)}+\delta^{(f)}$$

$$\rho=tanh(\mathbf{X}\boldsymbol{\alpha})$$

where

$$b^{(s)}\sim MVN(0_N,\sigma_s^2 I_N),$$

$$b^{(f)}\sim MVN(0_N,\sigma_f^2 I_N),$$

$$\begin{bmatrix}\epsilon_{ij}^{(s)}\\ \epsilon_{ij}^{(f)}\end{bmatrix}\sim\text{MVN}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}\tau_s^2 & \rho_{ij}\tau_s\tau_f\\ \rho_{ij}\tau_s\tau_f & \tau_f^2\end{bmatrix}\right).$$

Then,

$$Cov(\mathbf{y}^{(s)})=Cov(\delta^{(s)})=\sigma_s^2 ZZ^T+\tau_s^2 I=\Sigma^{(s)},$$

$$Cov(\mathbf{y}^{(f)})=Cov(\delta^{(f)})=\sigma_f^2 ZZ^T+\tau_f^2 I=\Sigma^{(f)},$$

and

$$\begin{aligned}Cov(\mathbf{y}^{(s)}, \mathbf{y}^{(f)}) &= Cov(Zb^{(s)} + \epsilon^{(s)}, Zb^{(f)} + \epsilon^{(f)}) \\ &= Cov(Zb^{(s)}, Zb^{(f)}) + Cov(\epsilon^{(s)}, \epsilon^{(f)}) \\ &= Cov(\epsilon^{(s)}, \epsilon^{(f)})\end{aligned}$$

since $b^{(s)}$ is independent to $b^{(f)}$ and b terms indepndent to ϵ terms.
We have

$$Corr(\epsilon_{ij}^{(s)}, \epsilon_{ij}^{(f)}) = \rho_{ij}$$

and $Corr(\epsilon_{ij}^{(s)}, \epsilon_{kl}^{(f)}) = 0$ when $i \neq k$ or $j \neq l$.
So,

$$\text{Cov}\left(\begin{bmatrix}\mathbf{y}_{ij}^{(s)} \\ \mathbf{y}_{ij}^{(f)}\end{bmatrix}\right) = \begin{bmatrix}\Sigma_{(ij,ij)}^{(s)} & \rho_{ij}\sqrt{\Sigma_{(ij,ij)}^{(s)}\Sigma_{(ij,ij)}^{(f)}} \\ \rho_{ij}\sqrt{\Sigma_{(ij,ij)}^{(s)}\Sigma_{(ij,ij)}^{(f)}} & \Sigma_{(ij,ij)}^{(f)}\end{bmatrix}$$

and

$$\text{Cov}\left(\begin{bmatrix}\mathbf{y}^{(s)} \\ \mathbf{y}^{(f)}\end{bmatrix}\right) = \begin{bmatrix}\Sigma^{(s)} & D \\ D & \Sigma^{(f)}\end{bmatrix}$$

where $D = \text{diag}(\rho_{ij}\sqrt{\Sigma_{(ij,ij)}^{(s)}\Sigma_{(ij,ij)}^{(f)}})$ and $D^T = D$.
Now, consider the following estimators and predictors

- $\hat{\beta}_{GLS}^{(s)} = (\mathbf{X}^T\Sigma^{(s)}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{(s)-1}y^{(s)} = A^{(s)}y^{(s)}$
- $\hat{\beta}_{GLS}^{(f)} = (\mathbf{X}^T\Sigma^{(f)}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{(f)-1}y^{(f)} = A^{(f)}y^{(f)}$
- $\hat{b}_{BLUP}^{(s)} = \sigma_s^2 Z^T\Sigma^{(s)-1}(y^{(s)} - \mathbf{X}\hat{\beta}_{GLS}^{(s)}) = B^{(s)}y^{(s)}$
- $\hat{b}_{BLUP}^{(f)} = \sigma_f^2 Z^T\Sigma^{(f)-1}(y^{(f)} - \mathbf{X}\hat{\beta}_{GLS}^{(f)}) = B^{(f)}y^{(f)}$
- $\hat{\epsilon}^{(s)} = (I - \mathbf{X}A^{(s)} - ZB^{(s)})y^{(s)} = P^{(s)}y(s)$
- $\hat{\epsilon}^{(f)} = (I - \mathbf{X}A^{(f)} - ZB^{(f)})y^{(f)} = P^{(f)}y(f)$

Then the joint distribution of first stage model residuals is

$$\begin{bmatrix}\hat{\epsilon}^{(s)} \\ \hat{\epsilon}^{(f)}\end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix}P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)} \\ P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\end{bmatrix}, \begin{bmatrix}P^{(s)}\Sigma^{(s)}P^{(s)T} & P^{(s)}DP^{(f)T} \\ P^{(f)}DP^{(s)T} & P^{(f)}\Sigma^{(f)}P^{(f)T}\end{bmatrix}\right), \tag{38}$$

This allow us to compute the moments of first stage residuals.
For the second stage model, we fit

$$\hat{\rho} = \hat{\epsilon}^{(s)}\hat{\epsilon}^{(f)} = \mathbf{X}\boldsymbol{\alpha} + Zb_{pseudo}^{(c)} + \epsilon^{(s)}$$

with

$$b^{(c)} \sim MVN(0_N, \eta^2 I),$$

$$\epsilon^{(c)} \sim MVN(0_{2N}, \omega^2 I)$$

where η^2 denote pseudo signal and ω^2 denote noise.
So, $\hat{\rho} \sim (\mathbf{X}\boldsymbol{\alpha}, \eta^2 ZZ^T + \omega^2 I)$, that is

$$\mathrm{Cov}(\hat{\rho}) = \begin{bmatrix} \begin{bmatrix} \eta^2+\omega^2 & \eta^2 \\ \eta^2 & \eta^2+\omega^2 \end{bmatrix} & 0 & \\ 0 & \ddots & 0 \\ & 0 & \begin{bmatrix} \eta^2+\omega^2 & \eta^2 \\ \eta^2 & \eta^2+\omega^2 \end{bmatrix} \end{bmatrix}$$

Then,

$$\begin{aligned} \eta^2 &= Cov(\hat{\rho}_{i1}, \hat{\rho}_{i2}) \\ &= Cov(\hat{\epsilon}_{i1}^{(s)}\hat{\epsilon}_{i1}^{(f)}, \hat{\epsilon}_{i2}^{(s)}\hat{\epsilon}_{i2}^{(f)}) \\ &= \mathbb{E}[\hat{\epsilon}_{i1}^{(s)}\hat{\epsilon}_{i1}^{(f)}\hat{\epsilon}_{i2}^{(s)}\hat{\epsilon}_{i2}^{(f)}] - \mathbb{E}[\hat{\epsilon}_{i1}^{(s)}\hat{\epsilon}_{i1}^{(f)}]\mathbb{E}[\hat{\epsilon}_{i2}^{(s)}\hat{\epsilon}_{i2}^{(f)}] \end{aligned}$$

We need to compute the 2nd and 4th moment of the first stage residuals.

Isserlis' Theorem states that for $X = (X_1, X_2, \ldots, X_p)^\top \sim \mathcal{N}_p(0, \Sigma)$ a zero-mean multivariate normal random vector,

- For any odd moments:

$$\mathbb{E}\big[X_{i_1}X_{i_2}\cdots X_{i_{2m+1}}\big] = 0$$

- For any even moments:

$$\mathbb{E}[X_{i_1}X_{i_2}\cdots X_{i_{2m}}] = \sum_{\text{all pairings of } \{i_1,\ldots,i_{2m}\}} \prod_{(a,b)} \Sigma_{i_a i_b}$$

So, for non-zero mean $X \sim \mathcal{N}_p(\mu, \Sigma)$. The second moment is

$$\mathbb{E}[X_iX_j] = \Sigma_{ij} + \mu_i\mu_j.$$

and the fourth moment is

$$\begin{aligned} \mathbb{E}[X_iX_jX_kX_l] = {} & \mu_i\mu_j\mu_k\mu_l \\ & + \mu_i\mu_j\Sigma_{kl} + \mu_i\mu_k\Sigma_{jl} + \mu_i\mu_l\Sigma_{jk} \\ & + \mu_j\mu_k\Sigma_{il} + \mu_j\mu_l\Sigma_{ik} + \mu_k\mu_l\Sigma_{ij} \\ & + \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}. \end{aligned}$$

According to (38) and Isserlis' Theorem, we have

$$\mathbb{E}[\hat{\epsilon}_{ij}^{(s)}] = \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{ij}$$

$$\mathbb{E}[\hat{\epsilon}_{ij}^{(f)}] = \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{ij}$$

$$\mathbb{E}[\hat{\epsilon}_{ij}^{(s)}\hat{\epsilon}_{ij}^{(f)}] = \{P^{(s)}DP^{(f)T}\}_{(ij,ij)} + \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{ij}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{ij}$$

$$
\begin{aligned}
\mathbb{E}[\hat{\epsilon}^{(s)}_{i1}\hat{\epsilon}^{(f)}_{i1}\hat{\epsilon}^{(s)}_{i2}\hat{\epsilon}^{(f)}_{i2}] = {} & \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i1}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i2} \\
& + \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i1}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(s)}DP^{(f)T}\}_{(i2,i2)} \\
& + \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i1}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(f)}\Sigma^{(f)}P^{(f)T}\}_{(i1,i2)} \\
& + \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i1}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i2}\{P^{(f)}DP^{(s)T}\}_{(i2,i1)} \\
& + \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(s)}DP^{(f)T}\}_{(i1,i2)} \\
& + \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i2}\{P^{(s)}\Sigma^{(s)}P^{(s)T}\}_{(i1,i2)} \\
& + \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i2}\{P^{(s)}DP^{(f)T}\}_{(i1,i1)} \\
& + \{P^{(s)}DP^{(f)T}\}_{(i1,i1)}\{P^{(s)}DP^{(f)T}\}_{(i2,i2)} \\
& + \{P^{(s)}\Sigma^{(s)}P^{(s)T}\}_{(i1,i2)}\{P^{(f)}\Sigma^{(f)}P^{(f)T}\}_{(i1,i2)} \\
& + \{P^{(s)}DP^{(f)T}\}_{(i1,i2)}\{P^{(f)}DP^{(s)T}\}_{(i2,i1)}
\end{aligned}
$$

Note that $P^{(s)}\Sigma^{(s)}P^{(s)T}, P^{(f)}\Sigma^{(f)}P^{(f)T}, P^{(s)}DP^{(f)T}, P^{(f)}DP^{(s)T}$ should be symmetric. So their $(i1, i2)$ entry should equal to $(i2, i1)$ entry.

Hence,

$$
\begin{aligned}
\eta^2 = {} & \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}_{i1}\}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(f)}\Sigma^{(f)}P^{(f)T}\}_{(i1,i2)} \\
& + \{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i1}\{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i2}\{P^{(f)}DP^{(s)T}\}_{(i2,i1)} \\
& + \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(s)}\mathbf{X}\boldsymbol{\beta}^{(s)}\}_{i2}\{P^{(s)}DP^{(f)T}\}_{(i1,i2)} \\
& + \{P^{(f)}\mathbf{X}\boldsymbol{\beta}^{(f)}\}_{i1}\{P^{(f)}\mathbf{X}\beta^{(f)}\}_{i2}\{P^{(s)}\Sigma^{(s)}P^{(s)T}\}_{(i1,i2)} \\
& + \{P^{(s)}\Sigma^{(s)}P^{(s)T}\}_{(i1,i2)}\{P^{(f)}\Sigma^{(f)}P^{(f)T}\}_{(i1,i2)} \\
& + \{P^{(s)}DP^{(f)T}\}_{(i1,i2)}\{P^{(f)}DP^{(s)T}\}_{(i2,i1)}
\end{aligned}
$$

where $P^{(s)}, P^{(f)}, D, \Sigma^{(s)}, \Sigma^{(f)}$ can be expressed in terms of $\sigma_s^2, \sigma_f^2, \tau_s^2, \tau_f^2, \boldsymbol{\alpha}, \mathbf{X}$.

A.7 Family level Correlation Hypothesis Test

The model in Section 4.2 assumes that the first stage family level random effects, $b^{(s)}$ and $b^{(f)}$, are independent.

In this section, we consider the case when $b^{(s)}$ and $b^{(f)}$ are correlated such that $\forall i = 1, \ldots, N$ family

$$
\begin{pmatrix} b_i^{(s)} \\ b_i^{(f)} \end{pmatrix} \sim^{iid} \text{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \omega\sigma_s\sigma_f \\ \omega\sigma_s\sigma_f & \sigma_f^2 \end{pmatrix} \right)
$$

Assume $\rho_{ij} = 0$, i.e. intermodel residuals are not correlated. The null hypothesis is $H_0\colon \omega = 0$

Data Generation Method

We generated data with $N = 1000$, $J = 2$,$p = 3$, $\beta^{(s)} = (1.2, 0.8, 0.5)$, $\beta^{(f)} = (0.6, 1.0, 0.3)$, $\sigma_s^2 = 0.5$, $\sigma_f^2 = 0.7$, $\tau_s^2 = 1.0$, $\tau_f^2 = 1.2$, $\rho = 0$, $\omega = 0$. We generated two covariates, age and height x_{ij}. For both covariates, a baseline value was first generated for each family. For age x_{1ij}, we simulated N baseline values

$$
x_{1i} \sim^{iid} \text{Unif}(18, 65),
$$

and assigned the same value to both twins within each family, so $x_{1i2} = x_{1i2}$ $\forall i = 1, \ldots, N$. For height x_{2ij}, we simulated N baseline values

$$
x_{2i} \sim^{iid} \text{Unif}(150, 190),
$$

then generated individual height differences from

$$
\Delta_{ij} \sim^{iid} N(0, 2).
$$

The height for each twin subject within a family was obtained by adding the individual-specific height difference to the family baseline value:

$$x_{2ij} = x_{2i} + \Delta_{ij}.$$

X consists of a column of ones (intercept) and the two covariates.

Appraoch 1: Likelihood Ratio Test

The full model fits

$$y_{ij}^{(t)} = \beta_0^{(t)} + \beta^{(t)} X_{ij} + b_i^{(t)} + \varepsilon_{ij}^{(t)}, \quad t \in \{s, f\},$$

$$\begin{pmatrix} b_i^{(s)} \\ b_i^{(f)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \omega\sigma_s\sigma_f \\ \omega\sigma_s\sigma_f & \sigma_f^2 \end{pmatrix} \right), \varepsilon_{ij}^{(t)} \overset{\text{iid}}{\sim} N(0, \tau^2)$$

The reduced model fits

$$\begin{pmatrix} b_i^{(s)} \\ b_i^{(f)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_f^2 \end{pmatrix} \right)$$

By conducting 1000 simulations, the empirical Type I error rate is 0.063, which is inflated.

Approach 2: Correlation Test on $\hat{b}_i^{(s)}$ and $\hat{b}_i^{(f)}$

By extracting $\hat{b}_i^{(s)}$ and $\hat{b}_i^{(f)}$ from the reduced model above, we calculated the observed correlation test statistic T_{obs}. Then we shuffled $\hat{b}_i^{(s)}$ across family i to obtain permuted correlation test statistics $T_{\text{perm}}^{(b)}$. The permutation p-value was calculated as

$$p_{\text{perm}} = \frac{\sum_{b=1}^{B} \{|T_{\text{perm}}^{(b)}| > |T_{obs}|\}}{B} \tag{39}$$

We reject the hypothesis when permuted p-value < 0.05. With 1000 permutations and 1000 simulations, it resulted in controlled Type I error (≈ 0.0512).

References

Pan, R., E. W. Dickie, C. Hawco, N. Reid, A. N. Voineskos, and J. Y. Park (2024, January). Spatial-extent inference for testing variance components in reliability and heritability studies. *Imaging Neurosci. (Camb.) 2*(imag-2-00058).

Pan, R., S. M. Weinstein, D. Tu, F. Hu, B. Tanrıverdi, R. Zhang, S. N. Vandekar, E. B. Baller, R. C. Gur, R. E. Gur, A. F. Alexander-Bloch, T. D. Satterthwaite, and J. Y. Park (2025, September). Mapping individual differences in intermodal coupling in neurodevelopment. *Imaging Neurosci. (Camb.) 3*(IMAG.a.156).

Wu, M., S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin (2011, Jul). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics 89*(1), 82–93.