Supplementary Document for "FSEM: Functional Structural Equation Models for Twin Functional Data"

September 24, 2017

1   Notations

We consider a functional ACE model as follows:

$$y_{ij}(v) = x_{ij}^T \boldsymbol{\beta}(v) + r_{ij}(v), \tag{1}$$

$$r_{ij}(v) = \sqrt{0.5}\mathbf{1}(\text{DZ})a_{ij}(v) + [\mathbf{1}(\text{MZ}) + \sqrt{0.5}\mathbf{1}(\text{DZ})]a_i(v) + c_i(v) + e_{ij}(v),$$

where $\boldsymbol{\beta}(v) = (\beta_1(v), \ldots, \beta_p(v))^T$ is a $p \times 1$ vector of coefficient functions and v is a grid point in $V_0 = \{v_1, v_2, \ldots, v_{N_G}\}$, which is a set of grid points in a common compact space, denoted by V. Furthermore, $a_{ij}(v)$ and $a_i(v)$ are introduced to represent within-curve and between-curve functional additive genetic effects on the i-th twin pair. $c_i(v)$ represents the common environmental effects. It is also assumed that

$$e_{ij}(v) = e_{ij,G}(v) + e_{ij,L}(v), \tag{2}$$

where $e_{ij,L}(v)$ are measurement errors representing local variability and $e_{ij,G}(s)$ are stochastic processes representing unique functional environmental effects.

The functional ACE model reduces to standard ACE model at each grid point. We denote the density of observations at tract $v_k \in V_0$ as

$$f^{(T)}(\boldsymbol{y}_{ik}; \boldsymbol{\sigma}^2(v_k), \boldsymbol{\beta}(v_k)) = f^{(T)}(\boldsymbol{\sigma}^2(v_k), \boldsymbol{\beta}(v_k); \boldsymbol{y}_{ik}),$$

1

where $\boldsymbol{y}_{ik} = (y_{i1}(v_k), y_{i2}(v_k))^T$ for MZ and DZ twin pairs, $\boldsymbol{y}_{ik} = y_{i1}(v_k)$ for singletons and $\boldsymbol{\sigma}^2(v) = (\sigma_a^2(v), \sigma_c^2(v), \sigma_e^2(v))$. Furthermore, we denote the log-likelihood function and its corresponding first order and second order derivatives with respect to θ as $\ell^{(T)}$, $\dot{\ell}_\theta^{(T)}$ and $\ddot{\ell}_{\theta\theta}^{(T)}$, respectively. Superscript $T \in \{M, D, I\}$ denotes a particular type of twin pairs, including MZ, DZ, and singleton.

Let $\boldsymbol{x}_i = (x_{i1}, x_{i2})^T$ and $\boldsymbol{\eta}_i(v) = (\eta_{i1}(v), \eta_{i2}(v))^T$, where

$$\eta_{ij}(v) = \sqrt{0.5}\mathbf{1}(\mathrm{DZ})a_{ij}(v) + [\mathbf{1}(\mathrm{MZ}) + \sqrt{0.5}\mathbf{1}(\mathrm{DZ})]a_i(v) + c_i(v)$$

for $1 \le i \le n_1 + n_2$. Let $\boldsymbol{x}_i = (x_{i1})$ and $\boldsymbol{\eta}_i(v) = (\eta_{i1}(v))$ for $n_1 + n_2 + 1 \le i \le n$. Denote $\sigma_1^2(v) = \sigma_a^2(v) + \sigma_c^2(v) + \sigma_e^2(v)$, $\sigma_2^2(v) = \sigma_a^2(v) + \sigma_c^2(v)$ and $\sigma_3^2(v) = 0.5\sigma_a^2(v) + \sigma_c^2(v)$.

We first introduce the following notation related to the log-likelihood function:

$$\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{Y}_k) = \mathcal{L}_{n_1}^{(M)}(\boldsymbol{\theta}; \boldsymbol{Y}_k) + \mathcal{L}_{n_2}^{(D)}(\boldsymbol{\theta}; \boldsymbol{Y}_k) + \mathcal{L}_{n_3}^{(I)}(\boldsymbol{\theta}; \boldsymbol{Y}_k), \tag{3}$$

$$\mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) = \frac{1}{\sqrt{n}} \left[\sqrt{n_1}\mathcal{J}_{n_1}^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) + \sqrt{n_2}\mathcal{J}_{n_2}^{(D)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \right.$$
$$\left. + \sqrt{n_3}\mathcal{J}_{n_3}^{(I)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \right],$$

$$\mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) = \frac{1}{n} \left[n_1\mathcal{I}_{n_1}^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) + n_2\mathcal{I}_{n_2}^{(D)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \right.$$
$$\left. + n_3\mathcal{I}_{n_3}^{(I)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \right],$$

where $\mathcal{L}_{n_1}^{(M)}(\boldsymbol{\theta}; \boldsymbol{Y}_k) = \sum_{i=1}^{n_1} \ell^{(M)}(\boldsymbol{\theta}; \boldsymbol{y}_{ik})$ and

$$\mathcal{J}_{n_1}^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \dot{\ell}_\theta^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{y}_{ik}),$$

$$\mathcal{I}_{n_1}^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) = \frac{1}{n_1} \sum_{i=1}^{n_1} \dot{\ell}_\theta^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{y}_{ik})^{\otimes 2}.$$

Similarly, we can define the corresponding terms for (D) and (I).

Then, we consider a quadratic expansion of the log-likelihood function as follows:

$$\mathcal{L}_n(\boldsymbol{\theta}(v_k); \boldsymbol{Y}_k) = \mathcal{L}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) - 0.5Q_n(\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)))$$
$$+ 0.5Z_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)^T \mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) Z_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)$$
$$+ R_n(\boldsymbol{\theta}(v_k), \boldsymbol{\theta}_*(v_k)), \tag{4}$$

where

$$Z_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) = \mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)^{-1} \mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k),$$

$$Q_n(\boldsymbol{\lambda}) = (\boldsymbol{\lambda} - Z_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k))^T \mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)(\boldsymbol{\lambda} - Z_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)),$$

and $R_n(\boldsymbol{\theta}(v_k), \boldsymbol{\theta}_*(v_k))$ will be shown to be uniformly negligible in Lemma 2.8.

We introduce more notation related to the weighted log-likelihood function as follows:

$$\mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \boldsymbol{R}) = \mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}^2(v); \boldsymbol{R}) + \mathcal{L}_{n_2,K}^{(D)}(\boldsymbol{\sigma}^2(v); \boldsymbol{R}) + \mathcal{L}_{n_3,K}^{(I)}(\boldsymbol{\sigma}^2(v); \boldsymbol{R}), \qquad (5)$$

$$\mathcal{J}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) = \frac{1}{\sqrt{n}} \left[\sqrt{n_1} \mathcal{J}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) + \sqrt{n_2} \mathcal{J}_{n_2,K}^{(D)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) \right.$$
$$\left. + \sqrt{n_3} \mathcal{J}_{n_3,K}^{(I)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) \right],$$

$$\mathcal{I}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) = \frac{1}{n} \left[n_1 \mathcal{I}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) + n_2 \mathcal{I}_{n_2,K}^{(D)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) \right.$$
$$\left. + n_3 \mathcal{I}_{n_3,K}^{(I)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) \right],$$

where

$$\mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}^2(v); \boldsymbol{R}) = \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \ell_\sigma^{(M)}(\boldsymbol{\sigma}^2(v); \boldsymbol{r}_{ik}) K_{h_1}(v_k - v),$$

$$\mathcal{J}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{r}_{ik}) K_{h_1}(v_k - v),$$

$$\mathcal{I}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{r}_{ik})^{\otimes 2} K_{h_1}(v_k - v).$$

Similarly, we can define the corresponding terms for (D) and (I).

We consider a quadratic expansion of the weighted log-likelihood function given by

$$\mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}}) = \mathcal{L}_{n,K}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}}) - 0.5 Q_{n,K}(\sqrt{n}(\boldsymbol{\sigma}^2(v) - \boldsymbol{\sigma}_*^2(v))) \qquad (6)$$
$$+ 0.5 \widehat{Z}_n(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y})^T \widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}) \widehat{Z}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y})$$
$$+ R_{n,K}(\boldsymbol{\sigma}^2(v), \boldsymbol{\sigma}_*^2(v)),$$

where

$$\widehat{Z}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}) = \widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y})^{-1} \widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}), \qquad (7)$$

$$Q_{n,K}(\boldsymbol{\lambda}) = (\boldsymbol{\lambda} - \widehat{Z}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}))^T \widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}) (\boldsymbol{\lambda} - \widehat{Z}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y})),$$

and the remainder term $R_{n,K}(\boldsymbol{\sigma}^2(v), \boldsymbol{\sigma}_*^2(v))$ will be shown to be uniformly negligible in Lemma 2.11.

We will use C and K to denote some universal constants and use K_α to denote constant depending only on α, which is a parameter. In both cases, the values of constants may change from line to line. We will also use the classical definition of Orlicz norms.

Definition 1.1. *For $\alpha > 0$, define the function $\psi_\alpha = \exp(x^\alpha) - 1$ with $x \geq 0$. For a*

random variable X, define the Orlicz norm

$$\|X\|_{\psi_\alpha} = \inf \left\{ \lambda > 0 : \mathbb{E}\psi_\alpha(|X|/\lambda) \le 1 \right\}.$$

REMARK. For $\alpha < 1$, it is widely accepted in literature to change the function ψ_α near zero $((0, [(1 - \alpha)/\alpha]^{1/\alpha}))$ via a linear interpolation to make it convex. We still use ψ_α to denote the interpolated function. It is easy to show that $\|X\|_{\psi_\alpha} < \infty$ if and only if $\mathbb{E}\{\psi_\alpha(|X|/C)\} < \infty$ for at least a finite constant C.

Definition 1.2. *For $x \in R^s$, define $\|x\|$ as the ℓ_2 norm of x, i.e., $\|x\| = \sqrt{\sum_{i=1}^s x_i^2}$. For matrix $X \in R^{s \times s}$, define $\|X\|$ as its element-wise sup norm, i.e. $\|X\| = \max\limits_{i,j} |X_{ij}|$. For a sequence of matrices $X_1, \ldots, X_m$, we define $\inf_{1 \le k \le m} X_k > 0$ if and only if $\inf_{1 \le k \le m} \lambda_{\min}(X_k) > 0$ where $\lambda_{\min}(X)$ denotes the minimum eigenvalue of X. Similarly, we define $\sup_{1 \le k \le m} X_k < 0$ if and only if $\sup_{1 \le k \le m} \lambda_{\max}(X_k) < 0$ where $\lambda_{\max}(X)$ denotes the maximum eigenvalue of X.*

Definition 1.3. *The set $\Omega \subset R^s$ is approximated at θ_0 by a cone with vertex at θ_0, C_Ω, if*

$$\inf_{x \in C_\Omega} \|x - y\| = o(\|y - \theta_0\|) \text{ for all } y \in \Omega$$

and

$$\inf_{y \in \Omega} \|x - y\| = o(\|x - \theta_0\|) \text{ for all } x \in \Omega.$$

Recall that a cone with vertex at θ_0, C, is a set of points such that if $x \in C$ then $a(x - \theta_0) + \theta_0 \in C$, where a is any real, non-negative number.

2 Proof of Theorems

We first introduce several lemmas that we are going to use in the sequel.

Lemma 2.1. *For a random variable X with $0 < \|X\|_{\psi_\alpha} < \infty$, we have, for $t \ge 0$,*
$$\mathbb{P}(|X| \ge t) \le 2 \exp \left(- \left(\frac{t}{\|X\|_{\psi_\alpha}} \right)^\alpha \right).$$

Proof. This is a simple result from Chebyshev's inequality, and thus we omit its proof. $\square$

Lemma 2.2. *Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and*

$$\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$$

for some constant c. Then, for any random variables $X_1, \ldots, X_m$,

$$\|\max_{1\le i\le m} X_i\|_\psi \le K\psi^{-1}(m)\max_i \|X_i\|_\psi,$$

for some constant K depending only on ψ.

Proof. This is the Lemma 2.2.2 of van der Vaart and Wellner (1996). $\qquad\square$

Lemma 2.3. *Let $X_1, \ldots, X_n$ be independent random variables, such that $\mathbb{E}(X_i) = 0$ and for some $\alpha \in (0,1]$, $\|X_i\|_{\psi_\alpha} < \infty$. Let $Z = |\sum_{i=1}^n X_i|$, then we have,*

$$\|Z\|_{\psi_\alpha} \le K_\alpha \left(\|Z\|_1 + \|\max_i |X_i|\|_{\psi_\alpha} \right)$$

Proof. This is a direct result of Theorem 5 of Adamczak (2007) when we take $\mathcal{F}$ as a class with only the identity function. $\qquad\square$

Lemma 2.4. *Suppose $Y_i, i = 1, 2, \ldots, n$ are independent and identically distributed random variables such that $\mathbb{E}(Y_i) = 0$, $\mathbb{E}(Y_i^2) = 1$ and $\|Y_i\|_{\psi_\alpha} \le c$ where $0 < \alpha \le 1$, and $\|\cdot\|_{\psi_\alpha}$ denotes the Orlicz norm for $\psi_\alpha(x) = \exp(x^\alpha) - 1$. Define*

$$W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i,$$

then we have $\|W_n\|_{\psi_\alpha} \le C$ for all n where C only depends on c and α.

Proof. Let $X = \max_{1\le i\le n} |Y_i|$, $Y_{i1} = Y_i \mathbf{1}\{X \le \rho\}$ and $Y_{i2} = Y_i \mathbf{1}\{X > \rho\}$ where the truncation level $\rho = 8\mathbb{E}(X)$. We have, $W_n = W_{n1} + W_{n2}$ where

$$W_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_{i1} - \mathbb{E}(Y_{i1})], \quad W_{n2} = \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_{i2} - \mathbb{E}(Y_{i2})].$$

Next, we try to control the Orlicz norm of W_{n1} and W_{n2} separately. We first control W_{n1}. Since $|Y_{i1} - \mathbb{E}(Y_{i1})| \le 2\rho$ and $\mathbb{E}(|Y_{i1} - \mathbb{E}(Y_{i1})|^2) \le \mathbb{E}(Y_i^2) = 1$, Bernstein inequality indicates that

$$\mathbb{P}(|W_{n1}| > t) \le 2\exp\left(-c_1 \min\left(t^2, \frac{\sqrt{n}t}{2\rho}\right)\right)$$

where c_1 is a universal constant. Furthermore, Lemma 2.2 indicates that $\|X\|_{\psi_\alpha} \le c\log^{1/\alpha}(1+n)$. Therefore,

$$\mathbb{P}(|W_{n1}| > t) \le 2\exp\left(-c_1 \min\left(t^2, \frac{\sqrt{n}t}{2c\log^{1/\alpha}(1+n)}\right)\right).$$

Finally, by Fubini theorem,

$$\mathbb{E}\left[\psi_\alpha(|W_{n1}|/w)\right] = \int_0^\infty \mathbb{P}(|W_{n1}| > s^{1/\alpha}w)\exp(s)ds.$$

Thus, it is easy to show that as long as

$$w \geq \max\left(\frac{2}{\sqrt{c_1}}, \frac{4c\log^{1/\alpha}(1+n)}{c_1\sqrt{n}}\right),$$

we have $\mathbb{E}\left[\psi_\alpha(|W_{n1}|/w)\right] < 2e$ for all n. Therefore, for all n, $\|W_{n1}\|_{\psi_\alpha} \leq C_1$ for some constant C_1 where C_1 depends only on c and α.

Then, we control W_{n2}. By Chebyshev inequality,

$$\mathbb{P}(\max_{k\leq n}|\sum_{i=1}^k Y_{i2}| > 0) \leq \mathbb{P}(X > \rho) \leq 1/8$$

and thus by Hoffman-Jorgensen inequality (see e.g. Proposition 6.8, Chapter 6, Ledoux and Talagrand (2013)), we obtain

$$\mathbb{E}|\sum_{i=1}^n Y_{i2}| \leq 8\mathbb{E}\left[X\right].$$

In consequence

$$\mathbb{E}|\sum_{i=1}^n \left[Y_{i2} - \mathbb{E}(Y_{i2})\right]| \leq 16\mathbb{E}\left[X\right] \leq K_\alpha\|X\|_{\psi_\alpha}.$$

We then have, by Lemma 2.3,

$$\|\sum_{i=1}^n \left[Y_{i2} - \mathbb{E}(Y_{i2})\right]\|_{\psi_\alpha} \leq K_\alpha\left(\mathbb{E}|\sum_{i=1}^n \left[Y_{i2} - \mathbb{E}(Y_{i2})\right]| + \|X\|_{\psi_\alpha}\right)$$

$$\leq K_\alpha\|X\|_{\psi_\alpha} \leq K_\alpha c\log^{1/\alpha}(1+n).$$

Therefore, $\|W_{n2}\|_{\psi_\alpha} \leq K_\alpha c\log^{1/\alpha}(1+n)/\sqrt{n} \leq C_2$ for some constant C_2 where C_2 only depends on c and α. The result now follows by triangle inequality.

Lemma 2.5. *For $1 \leq k \leq m$, $\{Z_{ik}, 1 \leq i \leq n\}$ are independent random variables such that for some constants $c > 0$, $\|Z_{ik}\|_{\psi_2} \leq c$. Let*

$$W_k = \left[\frac{1}{n}\sum_{i=1}^n |Z_{ik}|^s\right]^{1/s}, s > 0$$

we then have the following result

$$\sup_{1\leq k\leq m}|W_k| = O_p\left(\sqrt{\log(1+m)}\right).$$

Proof. For a given $w > 0$,

$$\mathbb{E}\left[\psi_2(W_k/w)\right] = \int_0^\infty \mathbb{P}(W_k > wt^{1/2})\exp(t)dt, \tag{8}$$

where

$$\mathbb{P}(W_k > wt^{1/2}) = \mathbb{P}(\sum_{i=1}^n |Z_{ik}|^s > nw^s t^{s/2}). \tag{9}$$

Since $\|Z_{ik}\|_{\psi_2} \leq c$, we have $\||Z_{ik}|^s\|_{\psi_{\frac{2}{s}}} \leq c^s$, as a result, $\|\sum_{i=1}^n |Z_{ik}|^s\|_{\psi_{\frac{2}{s}}} \leq nc^s$. Then, by Lemma 2.1 and (9)

$$\mathbb{P}(W_k > wt^{1/2}) \leq 2\exp\left(-\left[\frac{nw^s t^{s/2}}{nc^s}\right]^{2/s}\right) = 2\exp\left(-tw^2/c^2\right). \tag{10}$$

Thus, as long as $w \geq \sqrt{3}c$, $\mathbb{E}\left[\psi_2(W_k/w)\right] \leq 1$, which indicates $\|W_k\|_{\psi_2} \leq \sqrt{3}c$. We then have, by Lemma 2.2,

$$\sup_{1\leq k\leq m} |W_k| = O_p(\sqrt{\log(1+m)}).$$

2.1 Proof of Theorem 1

We first show several lemmas used to prove Theorem 1.

Lemma 2.6. *Under Assumptions C1-C3 and C7a, we have*

$$\sup_{1\leq k\leq N_G}\left\|\frac{1}{n_1}\sum_{i=1}^{n_1}\left[\dot{\ell}_\theta^{(M)\otimes 2} + \ddot{\ell}_{\theta\theta}^{(M)}\right](\boldsymbol{\theta}_*(v_k);\boldsymbol{y}_{ik})\right\| = O_p(\log^2(1+N_G)/\sqrt{n}), \tag{11}$$

$$\sup_{1\leq k\leq N_G}\left\|\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\left[\dot{\ell}_\theta^{(D)\otimes 2} + \ddot{\ell}_{\theta\theta}^{(D)}\right](\boldsymbol{\theta}_*(v_k);\boldsymbol{y}_{ik})\right\| = O_p(\log^2(1+N_G)/\sqrt{n}), \tag{12}$$

$$\sup_{1\leq k\leq N_G}\left\|\frac{1}{n_3}\sum_{i=n_1+n_2+1}^{n}\left[\dot{\ell}_\theta^{(I)\otimes 2} + \ddot{\ell}_{\theta\theta}^{(I)}\right](\boldsymbol{\theta}_*(v_k);\boldsymbol{y}_{ik})\right\| = O_p(\log^2(1+N_G)/\sqrt{n}), \tag{13}$$

$$\sup_{1\leq k\leq N_G}\|\mathcal{J}_n(\boldsymbol{\theta}_*(v_k);\boldsymbol{Y}_k)\| = O_p(\log(1+N_G)), \tag{14}$$

$$\sup_{1\leq k\leq N_G}\|\mathcal{I}_n(\boldsymbol{\theta}_*(v_k);\boldsymbol{Y}_k) - \mathbb{E}\left[\mathcal{I}_n(\boldsymbol{\theta}_*(v_k);\boldsymbol{Y}_k)\right]\| = O_p(\log^2(1+N_G)/\sqrt{n}). \tag{15}$$

Proof. The proofs of (11)-(15) are quite similar to each other. Let's consider (11) as an example. Based on formulas (44)-(46) and (50)-(52) in Section (4), the expression in the norm can be written as the sum of several empirical processes. Lemma 2.4 can be used to control the Orlicz norm of each empirical process for each k. Then we can use Lemma 2.2 to control the Orlicz norm of the supreme of each empirical process over $1 \leq k \leq N_G$. Thus, (11) follows directly. We omit the trivial details.

Lemma 2.7. *Under Assumptions C1-C4 and C7a, the maximum likelihood estimator* $\widehat{\boldsymbol{\theta}}(v_k)$ *satisfies*

$$\sup_{1 \leq k \leq N_G} \|\widehat{\boldsymbol{\theta}}(v_k) - \boldsymbol{\theta}_*(v_k)\| = o_p(1).$$

Proof. The proof of Lemma 2.7 consists of two steps. The first step is to show that log-likelihood function under model (1) converges uniformly to its expectation over V_0. The second step is to show that its expecation has well-separated maximum uniformly over V_0. We first consider the log-likelihood corresponding to MZ twin pairs given by

$$\mathcal{L}_{n_1}^{(M)}(\boldsymbol{\theta}; \boldsymbol{Y}_k) = -\frac{\sigma_1^2}{2(\sigma_1^4 - \sigma_2^4)} \sum_{i=1}^{n_1} \left[(y_{i1k} - x_{i1}^T \boldsymbol{\beta})^2 + (y_{i2k} - x_{i2}^T \boldsymbol{\beta})^2 \right]$$

$$+ \frac{\sigma_2^2}{\sigma_1^4 - \sigma_2^4} \sum_{i=1}^{n_1} \left[(y_{i1k} - x_{i1}^T \boldsymbol{\beta})(y_{i2k} - x_{i2}^T \boldsymbol{\beta}) \right] - \frac{n_1}{2} \log(\sigma_1^4 - \sigma_2^4).$$

The uniform convergence property can be similarly shown for every term in $\mathcal{L}_{n_1}^{(M)}(\boldsymbol{\theta}; \boldsymbol{Y}_k)$, therefore we only present the proof of the following one:

$$\frac{1}{n_1} \sum_{i=1}^{n_1} (y_{i1k} - x_{i1}^T \boldsymbol{\beta})^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} r_{i1}^2(v_k) + [\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}]^T \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} r_{i1}(v_k) \right\}$$

$$+ [\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}]^T \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} x_{i1}^T \right\} [\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}].$$

Assumptions C2-C3 indicates that

$$\|r_{i1}^2(v_k) - \mathbb{E}(r_{i1}^2(v_k))\|_{\psi_1} \leq C, \|x_{i1} r_{i1}(v_k)\|_{\psi_2} \leq C, \text{ for some universal constant } C.$$

From Lemmas 2.2 and 2.4, and Assumption C7a, it follows that we have

$$\sup_{1 \leq k \leq N_G} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \left[r_{i1}^2(v_k) - \mathbb{E}(r_{i1}^2(v_k)) \right] \right| = O_p\left(\frac{\log(1 + N_G)}{\sqrt{n}} \right) = o_p(1), \tag{16}$$

$$\sup_{1 \leq k \leq N_G} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} r_{i1}(v_k) \right| = O_p\left(\frac{\log^{1/2}(1 + N_G)}{\sqrt{n}} \right) = o_p(1). \tag{17}$$

Thus, $\mathcal{L}_{n_1}^{(M)}(\boldsymbol{\theta}; \boldsymbol{Y}_k)/n_1$ converges uniformly over V_0 to its expectation given by

$$\mathcal{L}^{(M)}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k))$$
$$= -\frac{\sigma_1^2 \sigma_{1*}^2(v_k)}{\sigma_1^4 - \sigma_2^4} + \frac{\sigma_2^2 \sigma_{2*}^2(v_k)}{\sigma_1^4 - \sigma_2^4} - \frac{1}{2}\log(\sigma_1^4 - \sigma_2^4)$$
$$- \frac{\sigma_1^2}{2(\sigma_1^4 - \sigma_2^4)}[\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}]^T \left\{ \frac{1}{n_1}\sum_{i=1}^{n_1}(x_{i1}x_{i1}^T + x_{i2}x_{i2}^T) \right\} [\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}]$$
$$+ \frac{\sigma_2^2}{2(\sigma_1^4 - \sigma_2^4)}[\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}]^T \left\{ \frac{1}{n_1}\sum_{i=1}^{n_1}(x_{i1}x_{i2}^T + x_{i2}x_{i1}^T) \right\} [\boldsymbol{\beta}_*(v_k) - \boldsymbol{\beta}].$$

Since the log-likelihood can be written as the sum of log-likelihood corresponding to MZ, DZ and singleton twins, the uniform convergence property of the log-likelihood corresponding to DZ and singleton twins can be shown similarly as above. As a result, the log-likelihood $\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{Y}_k)/n$ converges uniformly to

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k)) = \alpha_1 \mathcal{L}^{(M)}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k)) + \alpha_2 \mathcal{L}^{(D)}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k)) + \alpha_3 \mathcal{L}^{(I)}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k)).$$

Some algebraic calculations and Assumption C4 indicate that

$$\dot{\mathcal{L}}(\boldsymbol{\theta}_*(v_k); \boldsymbol{\theta}_*(v_k)) = 0, \quad \sup_{1 \leq k \leq N_G} \ddot{\mathcal{L}}(\boldsymbol{\theta}_*(v_k); \boldsymbol{\theta}_*(v_k)) < 0.$$

Thus, $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k))$ has a well-separated maximum $\boldsymbol{\theta}_*(v_k)$, i.e.

$$\sup_{1 \leq k \leq N_G} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \cap B(\boldsymbol{\theta}_*(v_k), \epsilon)^c} \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{\theta}_*(v_k)) < 0 \text{ for all } \epsilon > 0.$$

Finally, the uniform consisteny of $\widehat{\boldsymbol{\theta}}(v_k)$ is a direct result of argmax theorem (van der Vaart, 2000; van der Vaart and Wellner, 1996).

Lemma 2.8. *Let $\mathcal{I}(v_k)$ be the limit of $\mathbb{E}\left[\mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)\right]$, then for some random variable $\mathcal{J}(v_k) \sim N(\boldsymbol{0}, \mathcal{I}(v_k))$, we have, under Assumptions C1-C4 and C7a,*

$$\mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \to_d \mathcal{J}(v_k), \tag{18}$$

$$\mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \to_p \mathcal{I}(v_k), \tag{19}$$

and as $\gamma_n \to 0$,

$$\sup_{1 \leq k \leq N_G} \sup_{\boldsymbol{\theta}(v_k) \in \boldsymbol{\Theta}: \|\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)\| \leq \gamma_n} |R_n(\boldsymbol{\theta}(v_k), \boldsymbol{\theta}_*(v_k))| = o_p(1). \tag{20}$$

Proof. The (18) and (19) can be shown by using central limit theorem and law of large numbers. For (20), we first consider Taylor expansion on the log-likelihood function of

all MZ twin pairs

$$\mathcal{L}_n^{(M)}(\boldsymbol{\theta}(v_k); \boldsymbol{Y}_k) = \mathcal{L}_n^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) + [\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)]^T \sum_{i=1}^{n_1} \dot{\ell}_\theta^{(M)}(\boldsymbol{\theta}_*(v_k); \boldsymbol{y}_{ik})$$

$$+ \frac{1}{2} [\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)]^T \sum_{i=1}^{n_1} \ddot{\ell}_{\theta\theta}^{(M)}(\widetilde{\boldsymbol{\theta}}(v_k); \boldsymbol{y}_{ik}) [\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)],$$

where $\widetilde{\boldsymbol{\theta}}(v_k)$ is between $\boldsymbol{\theta}(v_k)$ and $\boldsymbol{\theta}_*(v_k)$. From the explicit formulas (50-52) in Section 4, we know that $\sum_{i=1}^{n_1} \ddot{\ell}_{\theta\theta}^{(M)}(\widetilde{\boldsymbol{\theta}}(v_k); \boldsymbol{y}_{ik})$ can be written as the sum of several empirical processes, of which one is given as follows,

$$A = \frac{1}{n_1} \sum_{i=1}^{n_1} \widetilde{r}_{i1}^2(v_k) = \frac{1}{n_1} \sum_{i=1}^{n_1} r_{i1}^2(v_k) - 2(\widetilde{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k))^T \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} r_{i1}(v_k) \qquad (21)$$

$$+ (\widetilde{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k))^T \left[\frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} x_{i1}^T \right] (\widetilde{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k)),$$

where $\widetilde{r}_{i1}(v_k) = y_{i1}(v_k) - x_{i1}^T \widetilde{\boldsymbol{\beta}}(v_k)$. Since $\|\widetilde{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k)\| \leq \gamma_n$, from (16) and (17), it follows that

$$A = \mathbb{E}\left(\frac{1}{n_1} \sum_{i=1}^{n_1} r_{i1}^2(v_k) \right) + O_{p,V_0}\left(\frac{\log(1 + N_G)}{\sqrt{n_1}} \right)$$

$$+ \gamma_n O_{p,V_0}\left(\frac{\log^{1/2}(1 + N_G)}{\sqrt{n_1}} \right) + \gamma_n^2 O_{p,V_0}(1) \qquad (22)$$

$$= \mathbb{E}\left(\frac{1}{n_1} \sum_{i=1}^{n_1} r_{i1}^2(v_k) \right) + o_{p,V_0}(1). \qquad (23)$$

Other involved empirical processes have similar results, therefore we have, for any $\boldsymbol{\theta}(v_k) \in \boldsymbol{\Theta}$ such that $\sup_{1 \leq k \leq N_G} \|\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)\| \leq \gamma_n \to 0$,

$$\mathcal{L}_n(\boldsymbol{\theta}(v_k); \boldsymbol{Y}_k) = \mathcal{L}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) + \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)) \right]^T \mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)$$

$$- \frac{1}{2} \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)) \right]^T (\mathcal{I}(v_k) + o_{p,V_0}(1)) \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)) \right]. \qquad (24)$$

Thus, following the proof of Theorem 1 in Andrews (1999) and (14) in Lemma 2.6, we have

$$\sup_{1 \leq k \leq N_G} \|\sqrt{n} (\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k)) \| = O_p(\log(1 + N_G)). \qquad (25)$$

Then, we can safely change γ_n to $\log(1 + N_G)/\sqrt{n}$ in (22) to obtain a better rate, that is

$$A = \mathbb{E}\left(\frac{1}{n_1} \sum_{i=1}^{n_1} r_{i1}^2(v_k) \right) + O_{p,V_0}\left(\frac{\log(1 + N_G)}{\sqrt{n}} \right), \qquad (26)$$

and further

$$\mathcal{L}_n(\boldsymbol{\theta}(v_k); \boldsymbol{Y}_k) = \mathcal{L}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) + \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k))\right]^T \mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)$$
$$- \frac{1}{2} \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k))\right]^T \left(\mathcal{I}(v_k) + O_{p,V_0}\left(\frac{\log(1 + N_G)}{\sqrt{n}}\right)\right) \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k))\right].$$
$$(27)$$

Combining (15), (25), and (27), we have

$$\mathcal{L}_n(\boldsymbol{\theta}(v_k); \boldsymbol{Y}_k) = \mathcal{L}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) + \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k))\right]^T \mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)$$
$$- \frac{1}{2} \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k))\right]^T \mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) \left[\sqrt{n}(\boldsymbol{\theta}(v_k) - \boldsymbol{\theta}_*(v_k))\right]$$
$$+ O_{p,V_0}\left(\frac{\log^4(1 + N_G)}{\sqrt{n}}\right),$$

which indicates that $R_n(\boldsymbol{\theta}(v_k), \boldsymbol{\theta}_*(v_k)) = O_{p,V_0}\left(\log^4(1 + N_G)/\sqrt{n}\right) = o_{p,V_0}(1)$. Thus, the proof is now completed.

Proof of Theorem 1.

We have shown the uniform consistency and convergence rate of $\widehat{\boldsymbol{\theta}}(v_k)$ over $v_k \in V_0$ in Lemmas 2.7 and 2.8. Now, we show that asymptotic distribution of related estimators. We first introduce some notation as follows:

$$Z(v_k) = \mathcal{I}(v_k)^{-1} \mathcal{J}(v_k),$$
$$Q(\boldsymbol{\lambda}) = (\boldsymbol{\lambda} - Z(v_k))^T \mathcal{I}(\boldsymbol{\theta}_*(v_k))(\boldsymbol{\lambda} - Z(v_k)),$$
$$\sqrt{n}(\boldsymbol{\Theta} - \boldsymbol{\theta}_*(v_k)) := \{\boldsymbol{\lambda} \in R^q : \boldsymbol{\lambda} = \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_*(v_k)) \text{ for some } \boldsymbol{\theta} \in \boldsymbol{\Theta}\},$$
$$\boldsymbol{\Lambda}(v_k) := [0, \infty) \times R^{p+2} := \Lambda_1(v_k) \times \boldsymbol{\Lambda}_2(v_k) \text{ where}$$
$$\Lambda_1(v_k) = [0, \infty) \text{ and } \boldsymbol{\Lambda}_2(v_k) = R^{p+2},$$
$$Q_n(\widehat{\boldsymbol{\lambda}}_n(v_k)) = \inf_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}(v_k)} Q_n(\boldsymbol{\lambda}) \text{ and } Q(\widehat{\boldsymbol{\lambda}}(v_k)) = \inf_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}(v_k)} Q(\boldsymbol{\lambda}).$$

Now similar to arguments in Lemma 1 of Andrews (1999) or Theorem 4 of Zhu and Zhang (2004), we have

$$Q_n(\sqrt{n}(\widehat{\boldsymbol{\theta}}(v_k) - \boldsymbol{\theta}_*(v_k))) = \inf_{\boldsymbol{\lambda} \in \sqrt{n}(\boldsymbol{\Theta} - \boldsymbol{\theta}_*(v_k))} Q_n(\boldsymbol{\lambda}) + o_{p,V_0}(1)$$
$$= \inf_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}(v_k)} Q_n(\boldsymbol{\lambda}) + o_{p,V_0}(1).$$

By continuous mapping theorem, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}(v_k) - \boldsymbol{\theta}_*(v_k)) \to_d \widehat{\boldsymbol{\lambda}}(v_k).$$

Next, we provide the asymptotic distribution of $\sqrt{n}(\widehat{\sigma}_a^2(v_k) - \sigma_{a*}^2(v_k))$ by partitioning $\widehat{\boldsymbol{\theta}}(v_k)$, $\boldsymbol{\theta}_0(v_k)$, $\mathcal{J}(v_k)$, $\mathcal{I}(v_k)$, $Z(v_k)$ and $\widehat{\boldsymbol{\lambda}}(v_k)$ conformably with $\boldsymbol{\theta}(v_k) = (\theta_1(v_k), \boldsymbol{\theta}_2(v_k))$ where $\theta_1(v_k) = \sigma_a^2(v_k)$. We define some notation as follows:

$$\boldsymbol{\theta}(v_k) = \begin{pmatrix} \theta_1(v_k) \\ \boldsymbol{\theta}_2(v_k) \end{pmatrix}, \quad \widehat{\boldsymbol{\theta}}(v_k) = \begin{pmatrix} \widehat{\theta}_1(v_k) \\ \widehat{\boldsymbol{\theta}}_2(v_k) \end{pmatrix}, \quad \boldsymbol{\theta}_*(v_k) = \begin{pmatrix} \theta_{1*}(v_k) \\ \boldsymbol{\theta}_{2*}(v_k) \end{pmatrix},$$

$$\mathcal{J}(v_k) = \begin{pmatrix} \mathcal{J}_1(v_k) \\ \mathcal{J}_2(v_k) \end{pmatrix}, \quad \mathcal{I}(v_k) = \begin{bmatrix} \mathcal{I}_{11}(v_k) & \mathcal{I}_{12}(v_k) \\ \mathcal{I}_{21}(v_k) & \mathcal{I}_{22}(v_k) \end{bmatrix},$$

$$Z(v_k) = \begin{pmatrix} Z_1(v_k) \\ Z_2(v_k) \end{pmatrix}, \quad \boldsymbol{\lambda}(v_k) = \begin{pmatrix} \lambda_1(v_k) \\ \boldsymbol{\lambda}_2(v_k) \end{pmatrix}.$$

We also define the following two quadratic functions.

$$Q_1(\lambda_1) = (\lambda_1 - Z_1(v_k))^T \left[H\mathcal{I}^{-1}(v_k)H^T \right]^{-1} (\lambda_1 - Z_1(v_k)), \tag{28}$$

$$Q_2(\lambda_1, \boldsymbol{\lambda}_2) = (\boldsymbol{\lambda}_2 + \mathcal{I}_{22}^{-1}\mathcal{I}_{21}\lambda_1 - \mathcal{I}_{22}^{-1}\mathcal{J}_2(v_k))^T \mathcal{I}_{22}(\boldsymbol{\lambda}_2 + \mathcal{I}_{22}^{-1}\mathcal{I}_{21}\lambda_1 - \mathcal{I}_{22}^{-1}\mathcal{J}_2(v_k)).$$

where $H := [1 : \mathbf{0}] \in R^{1 \times (p+3)}$. Some algebra indicates that

$$Q(\lambda(v_k)) = Q_1(\lambda_1(v_k)) + Q_2(\lambda_1(v_k), \boldsymbol{\lambda}_2(v_k)).$$

When $\sigma_{c*}^2(v_k) > 0$ and $\sigma_{e*}^2(v_k) > 0$, $\boldsymbol{\Lambda}_2(v_k) = R^{p+2}$, thus for any $\lambda_1(v_k) \in R$, we have

$$\inf_{\boldsymbol{\lambda}_2(v_k) \in \boldsymbol{\Lambda}_2(v_k)} Q_2(\lambda_1(v_k), \boldsymbol{\lambda}_2(v_k)) = \inf_{\boldsymbol{\lambda}_2(v_k) \in R^{p+2}} Q_2(\lambda_1(v_k), \boldsymbol{\lambda}_2(v_k)) = 0.$$

Thus, we obtain

$$\inf_{\lambda(v_k) \in \boldsymbol{\Lambda}(v_k)} Q(\lambda(v_k)) = \inf_{\lambda_1(v_k) \in \Lambda_1(v_k)} Q_1(\lambda_1(v_k)). \tag{29}$$

Based on this, we have the following results:

$$Q_1(\widehat{\lambda}_1(v_k)) = \inf_{\lambda_1(v_k) \in \Lambda_1(v_k)} Q_1(\lambda_1(v_k)), \tag{30}$$

$$\widehat{\lambda}_1(v_k) = Z_1(v_k)\mathbf{1}(Z_1(v_k) \geq 0). \tag{31}$$

Finally, we have the following asymptotic distributions:

$$\sqrt{n}(\widehat{\sigma}_a^2(v_k) - \sigma_{a*}^2(v_k)) = \sqrt{n}(\widehat{\theta}_1(v_k) - \theta_{1*}(v_k)) \to_d \widehat{\lambda}_1(v_k), \tag{32}$$

$$2\left[\mathcal{L}_n(\widehat{\boldsymbol{\theta}}(v_k)) - \mathcal{L}_n(\boldsymbol{\theta}_*(v_k)) \right] \to_d \widehat{\lambda}_1(v_k)^T \left[H\mathcal{I}^{-1}(v_k)H^T \right]^{-1} \widehat{\lambda}_1(v_k)$$
$$+ \mathcal{J}_2^T(v_k)\mathcal{I}_{22}^{-1}(v_k)\mathcal{J}_2(v_k),$$

$$\mathrm{LRT}_n(v_k) \to_d \widehat{\lambda}_1(v_k)^T \left[H\mathcal{I}^{-1}(v_k)H^T \right]^{-1} \widehat{\lambda}_1(v_k).$$

Furthermore, since $Z(v_k)$ is normally distributed with mean $\mathbf{0}$ and covariance $\mathcal{I}(v_k)^{-1}$,

$Z_1(v_k) = HZ(v_k)$ is normally distributed with mean 0 and variance $H\mathcal{I}(v_k)^{-1}H^T$. Result (31) indicates that

$$\sqrt{n}(\widehat{\sigma}_a^2(v_k) - \sigma_{a*}^2(v_k)) \to_d (H\mathcal{I}(v_k)^{-1}H^T)^{1/2}N(0,1)\mathbf{1}(N(0,1) \geq 0),$$

$$\mathrm{LRT}_n(v_k) \to_d \frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_0^2.$$

Now, we consider local alternatives, $H_n : \sigma_a^2(v_k) = h(v_k)/\sqrt{n}$, we define $\boldsymbol{h}(v_k) = (h(v_k), \mathbf{0})^T \in R^{p+3}$, then similarly we have

$$\mathcal{L}_n(\boldsymbol{h}(v_k); \boldsymbol{Y}_k) = \boldsymbol{h}(v_k)^T \mathcal{J}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k) - \frac{1}{2}\boldsymbol{h}(v_k)^T \mathcal{I}_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k)\boldsymbol{h}(v_k) + o_{p,V_0}(1).$$

Therefore, under H_0, we have $(Z_n(\boldsymbol{\theta}_*(v_k); \boldsymbol{Y}_k), \mathcal{L}_n(\boldsymbol{h}(v_k); \boldsymbol{Y}_k))$ converges to normal distribution with mean $\left[\mathbf{0}^T, -\frac{1}{2}\boldsymbol{h}(v_k)^T \mathcal{I}(v_k)\boldsymbol{h}(v_k)\right]^T$ and covariance matrix

$$\begin{bmatrix} \mathcal{I}(v_k)^{-1} & \boldsymbol{h}(v_k) \\ \boldsymbol{h}(v_k)^T & \boldsymbol{h}(v_k)^T \mathcal{I}(v_k)\boldsymbol{h}(v_k) \end{bmatrix}.$$

By LeCam's third lemma, we have:

$$\sqrt{n}\{\widehat{\sigma}_a^2(v_k) - \sigma_{a*}^2(v_k)\} \xrightarrow[d]{H_n} \{H\mathcal{I}(v_k)^{-1}H^T\}^{1/2} \times N(\widetilde{h}(v_k), 1)\mathbf{1}(N(\widetilde{h}(v_k), 1) \geq 0),$$

$$\mathrm{LR}_n(v_k) \xrightarrow[d]{H_n} N(\widetilde{h}(v_k), 1)^2 \mathbf{1}\left(N(\widetilde{h}(v_k), 1) \geq 0\right),$$

where $\widetilde{h}(v_k) = \left[H\mathcal{I}^{-1}(v_k)H^T\right]^{-1/2}h(v_k)$.

2.2 Proof of Theorem 2

Lemma 2.9. *Under Assumptions C1-C6, C7b, we have*

$$\sup_{v \in V} \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \left[\dot{\ell}_\sigma^{(M)\otimes 2} + \ddot{\ell}_{\sigma\sigma}^{(M)} \right] (\boldsymbol{\sigma}_*^2(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik}) K_{h_1}(v_k - v) \right\|$$

$$= O_p(\log^2(1 + N_G)/\sqrt{n}) \tag{33}$$

$$\sup_{v \in V} \left\| \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} \frac{1}{N_G} \sum_{k=1}^{N_G} \left[\dot{\ell}_\sigma^{(D)\otimes 2} + \ddot{\ell}_{\sigma\sigma}^{(D)} \right] (\boldsymbol{\sigma}_*^2(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik}) K_{h_1}(v_k - v) \right\|$$

$$= O_p(\log^2(1 + N_G)/\sqrt{n}) \tag{34}$$

$$\sup_{v \in V} \left\| \frac{1}{n_3} \sum_{i=n_1+n_2+1}^{n} \frac{1}{N_G} \sum_{k=1}^{N_G} \left[\dot{\ell}_\sigma^{(I)\otimes 2} + \ddot{\ell}_{\sigma\sigma}^{(I)} \right] (\boldsymbol{\sigma}_*^2(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik}) K_{h_1}(v_k - v) \right\|$$

$$= O_p(\log^2(1 + N_G)/\sqrt{n}) \tag{35}$$

$$\sup_{v \in V} \| \mathcal{J}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y}) \| = O_p(1) \tag{36}$$

$$\sup_{v \in V} \| \widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y}) - \mathcal{I}_K(v) \| = O_p\left(\log^2(1 + N_G)/\sqrt{n}\right) \tag{37}$$

$$\sup_{v \in V} \| \widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y}) - \mathcal{J}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y}) \| = O_p\left(\log^2(1 + N_G)/\sqrt{n}\right) \tag{38}$$

Proof. We provide the proof of (36) and (38), the proof of other equations is similar and thus omitted. We first show (38). Since $\widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y})$ can be written as the sum of several similar terms of which one is given as,

$$\widehat{\mathcal{J}}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v), \widehat{\boldsymbol{\beta}}(v_k); y_{ik}) K_{h_1}(v_k - v)$$

$$= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v), \boldsymbol{\beta}_*(v_k); y_{ik}) K_{h_1}(v_k - v)$$

$$+ \frac{1}{N_G} \sum_{k=1}^{N_G} \left[\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \ddot{\ell}_{\sigma\beta}^{(M)}(\boldsymbol{\sigma}_*^2(v), \widetilde{\boldsymbol{\beta}}(v_k); y_{ik}) \right] \left[\widehat{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k) \right] K_{h_1}(v_k - v)$$

$$\leq \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v), \boldsymbol{\beta}_*(v_k); y_{ik}) K_{h_1}(v_k - v)$$

$$+ \sup_{1 \leq k \leq N_G} \| \widehat{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k) \| \frac{1}{N_G} \sum_{k=1}^{N_G} \left\| \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \ddot{\ell}_{\sigma\beta}^{(M)}(\boldsymbol{\sigma}_*^2(v), \widetilde{\boldsymbol{\beta}}(v_k); y_{ik}) \right\| K_{h_1}(v_k - v)$$

where $\widetilde{\boldsymbol{\beta}}(v_k)$ is between $\widehat{\boldsymbol{\beta}}(v_k)$ and $\boldsymbol{\beta}_*(v_k)$. From (48) and (49) in Section 4, it follows that $\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \ddot{\ell}_{\sigma\beta}^{(M)}(\boldsymbol{\sigma}_*^2(v), \widetilde{\boldsymbol{\beta}}(v_k); y_{ik})$ can be written as the sum of several empirical processes.

We only consider the following one, other processes can be dealt with similarly.

$$\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \left[x_{i1}(y_{i1} - x_{i1}^T \widetilde{\boldsymbol{\beta}}(v_k)) \right] = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} x_{i1} r_{i1}(v_k)$$

$$+ \left[\frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1} x_{i1}^T \right] \left[\sqrt{n_1}(\widetilde{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k)) \right]$$

$$= O_{p,V_0}\left(\log^{1/2}(1 + N_G) \right) + O_{p,V_0}(1) O_{p,V_0}\left(\log(1 + N_G) \right) = O_{p,V_0}\left(\log(1 + N_G) \right).$$

Thus $\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \ddot{\ell}_{\sigma\beta}^{(M)}(\boldsymbol{\sigma}_*^2(v), \widetilde{\boldsymbol{\beta}}(v_k); y_{ik}) = O_{p,V_0}\left(\log(1 + N_G) \right)$. Since $\left[\widehat{\boldsymbol{\beta}}(v_k) - \boldsymbol{\beta}_*(v_k) \right] = O_{p,V_0}\left(\log(1 + N_G)/\sqrt{n} \right)$, we have that

$$\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v), \widehat{\boldsymbol{\beta}}(v_k); y_{ik}) K_{h_1}(v_k - v)$$

$$= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v), \boldsymbol{\beta}_*(v_k); y_{ik}) K_{h_1}(v_k - v) + O_{p,V}\left(\frac{[\log(1 + N_G)]^2}{\sqrt{n}} \right).$$

The reason that we can change O_{p,V_0} to $O_{p,V}$ is that v appears only in bounded coefficient functions (v only involved in a, b of (48) and (49)). Other involved terms of $\widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y})$ can be similarly shown to have same result, thus we have (38).

Then, let's focus on (36). $\mathcal{J}_{n,K}(\boldsymbol{\sigma}_*(v); \boldsymbol{Y})$ is also the sum of several empirical processes and one of them is given as,

$$\mathcal{J}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R}) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v), \boldsymbol{\beta}_*(v_k); y_{ik}) K_{h_1}(v_k - v)$$

which is the sum of several terms from (45-46) in Section 4. Then, we have

$$\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} r_{i1}^2(v_k) K_{h_1}(v_k - v) = (A_1) + (A_2) + (A_3)$$

$$:= \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \eta_{i1}^2(v_k) K_{h_1}(v_k - v) + \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} e_{i1}^2(v_k) K_{h_1}(v_k - v)$$

$$+ 2\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \eta_{i1}(v_k) e_{i1}(v_k) K_{h_1}(v_k - v),$$

15

where

$$A_1 = \left[\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v)\right]\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]$$

$$+ \frac{1}{N_G}\sum_{k=1}^{N_G}\left[\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v_k)-\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v)\right]K_{h_1}(v_k-v)$$

$$\leq \left[\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v)\right]\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]$$

$$+ \frac{1}{N_G}\sum_{k=1}^{N_G}\sup_{|v_k-v|\leq h_1}\left|\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v_k)-\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v)\right|K_{h_1}(v_k-v)$$

$$= \left[\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\eta_{i1}^2(v)\right]\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]+o_{p,V}(1),$$

where the last equality comes from the Donsker property in Assumption C5. Following the arguments in Einmahl and Mason (2000), it follows that

$$A_2 = \frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\frac{1}{N_G}\sum_{k=1}^{N_G}\Sigma_e(v_k,v_k)K_{h_1}(v_k-v)+O_{p,V}\left(\sqrt{\frac{|\log(h_1)|}{N_Gh_1}}\right)$$

$$A_3 = O_{p,V}\left(\sqrt{\frac{|\log(h_1)|}{N_Gh_1}}\right),$$

Furthermore, since

$$\left|\frac{1}{N_G}\sum_{k=1}^{N_G}(\Sigma_e(v_k,v_k)-\Sigma_e(v,v))K_{h_1}(v_k-v)\right|\leq\frac{1}{N_G}\sum_{k=1}^{N_G}|\Sigma_e(v_k,v_k)-\Sigma_e(v,v)|K_{h_1}(v_k-v)$$

$$\leq C_eh_1\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)=O_p(h_1).$$

we have

$$A_2 = \sqrt{n_1}\Sigma_e(v,v)\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]$$

$$+ \frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\frac{1}{N_G}\sum_{k=1}^{N_G}(\Sigma_e(v_k,v_k)-\Sigma_e(v,v))K_{h_1}(v_k-v)+O_{p,V}\left(\sqrt{\frac{|\log(h_1)|}{N_Gh_1}}\right)$$

$$= \sqrt{n_1}\Sigma_e(v,v)\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]+O_{p,V}(\sqrt{n_1}h_1)+O_{p,V}\left(\sqrt{\frac{|\log(h_1)|}{N_Gh_1}}\right).$$

We can then write the score function corresponds to MZ twin pairs as

$$
\mathcal{J}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R})
$$

$$
= \frac{\sigma_{1*}^4(v) + \sigma_{2*}^4(v)}{2(\sigma_{1*}^4(v) - \sigma_{2*}^4(v))^2} \left[\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (\eta_{i1}^2(v) + \eta_{i2}^2(v) - 2\sigma_{2*}^2(v)) \right] \left[\frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v) \right]
$$

$$
- \frac{2\sigma_{1*}^2(v)\sigma_{2*}^2(v)}{(\sigma_{1*}^4(v) - \sigma_{2*}^4(v))^2} \left[\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (\eta_{i1}(v)\eta_{i2}(v) - \sigma_{2*}^2(v)) \right] \left[\frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v) \right]
$$

$$
+ O_{p,V}(\sqrt{n_1}h_1) + O_{p,V}\left(\sqrt{\frac{|\log(h_1)|}{N_G h_1}} \right). \tag{39}
$$

Thus we have the weak convergence of $\mathcal{J}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{R})$ given Assumptions C5 and C7b. Similar results hold for DZ twin pairs and singleton twins, (36) now follows.

Lemma 2.10. *Under Assumptions C1-C6, C7b, the weighted maximum likelihood estimator suppose $\widehat{\boldsymbol{\sigma}}_K^2(v)$ satisfies*

$$
\sup_{v \in V} \|\widehat{\boldsymbol{\sigma}}_K^2(v) - \boldsymbol{\sigma}_*^2(v)\| = o_p(1).
$$

Proof. Similar to the proof of Lemma 2.7, we first consider the uniform convergence of weighted likelihood function. The weighted likelihood function also consists of three parts, corresponding to MZ, DZ and singleton twin pairs. We take MZ part as an example. The proof of DZ and singleton parts are similar and are thus omitted.

$$
\mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}}) = \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \ell^{(M)}(\boldsymbol{\sigma}^2(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik}) K_{h_1}(v_k - v)
$$

$$
= -\frac{\sigma_1^2(v)}{2(\sigma_1^4(v) - \sigma_2^4(v))} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \left[(y_{i1k} - x_{i1}^T\widehat{\boldsymbol{\beta}}(v_k))^2 + (y_{i2k} - x_{i2}^T\widehat{\boldsymbol{\beta}}(v_k))^2 \right] K_{h_1}(v_k - v)
$$

$$
+ \frac{\sigma_2^2(v)}{2(\sigma_1^4(v) - \sigma_2^4(v))} \sum_{i=1}^{n_1} \frac{1}{N_G} \sum_{k=1}^{N_G} \left[(y_{i1k} - x_{i1}^T\widehat{\boldsymbol{\beta}}(v_k))(y_{i2k} - x_{i2}^T\widehat{\boldsymbol{\beta}}(v_k)) \right] K_{h_1}(v_k - v)
$$

$$
- \frac{n_1}{2} \log(\sigma_1^4(v) - \sigma_2^4(v)) \frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v).
$$

Following the arguments in the proof of Lemma 2.9, we have

$$
\frac{1}{n_1}\mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}^2(v);\widehat{\boldsymbol{R}}) = -\frac{\sigma_1^2(v)}{2(\sigma_1^4(v)-\sigma_2^4(v))}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}(\eta_{i1}^2(v)+\eta_{i2}^2(v))\right]\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]
$$

$$
+\frac{\sigma_2^2(v)}{2(\sigma_1^4(v)-\sigma_2^4(v))}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\eta_{i1}(v)\eta_{i2}(v)\right]\left[\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)\right]
$$

$$
-\frac{1}{2}\log(\sigma_1^4(v)-\sigma_2^4(v))\frac{1}{N_G}\sum_{k=1}^{N_G}K_{h_1}(v_k-v)
$$

$$
+O_p\left(\frac{\log^2(1+N_G)}{n}\right)+o_{p,V}\left(\frac{1}{\sqrt{n}}\right)+O_{p,V}\left(\sqrt{\frac{|\log(h_1)|}{nN_Gh_1}}\right)+O_{p,V}(h_1).
$$

By Assumptions C5-C6 and C7b, we have the uniform convergence of $\mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}^2(v);\widehat{\boldsymbol{R}})$ over $v\in V$. The rest of proof completely mirrors that of Lemma 2.7 and is thus omitted.

Lemma 2.11. *Let $\mathcal{I}(v)$ as the limit of $\mathbb{E}\left[\widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y})\right]$, then for some gaussian process $\mathcal{J}_K(v)$, we have, under Assumptions C1-C6, C7b,*

$$
\left(\widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y}),\widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y})\right)\Rightarrow_d(\mathcal{J}_K(v),\mathcal{I}_K(v)), \tag{40}
$$

and as $\gamma_n\to 0$,

$$
\sup_{v\in V}\sup_{\boldsymbol{\sigma}^2(v)\in\mathcal{E}:\|\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v)\|\leq\gamma_n}|R_{n,K}(\boldsymbol{\sigma}^2(v),\boldsymbol{\sigma}_*^2(v))=o_p(1). \tag{41}
$$

REMARK. At each fixed point $v\in V$, $\mathcal{J}_K(v)$ is normal with mean 0 and variance $\mathcal{I}_{1,K}(v)$ rather than variance $\mathcal{I}_K(v)$ where $\mathcal{I}_{1,K}(v)$ is the asymptotic variance of $\widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y})$ and $\mathcal{I}_K(v)$ is the asymptotic limit of $\widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y})$.

Proof. (40) is a direct result of Lemma 2.9 and (39). The proof of (41) completely mirrors the proof of Lemma 2.8. We first consider an ordinary Taylor expansion on the weighted log-likelihood function correpondint to MZ twin pairs. The DZ and singleton part are quite similar, and we omit the proof without explictly making the statement in the sequel.

$$
\mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}^2(v);\widehat{\boldsymbol{R}}) = \mathcal{L}_{n_1,K}^{(M)}(\boldsymbol{\sigma}_*^2(v);\widehat{\boldsymbol{R}})
$$

$$
+\left[\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v)\right]^T\left[\sum_{i=1}^{n_1}\frac{1}{N_G}\sum_{k=1}^{N_G}\dot{\ell}_\sigma^{(M)}(\boldsymbol{\sigma}_*^2(v),\widehat{\boldsymbol{\beta}}(v_k);\boldsymbol{y}_{ik})K_{h_1}(v_k-v)\right]
$$

$$
-\frac{1}{2}\left[\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v)\right]^T\left[\sum_{i=1}^{n_1}\frac{1}{N_G}\sum_{k=1}^{N_G}\ddot{\ell}_{\sigma\sigma}^{(M)}(\widetilde{\boldsymbol{\sigma}}_*^2(v),\widehat{\boldsymbol{\beta}}(v_k);\boldsymbol{y}_{ik})K_{h_1}(v_k-v)\right]
$$

$$
\times\left[\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v)\right]
$$

where $\widetilde{\boldsymbol{\sigma}}_*^2(v)$ is between $\boldsymbol{\sigma}^2(v)$ and $\boldsymbol{\sigma}_*^2(v)$. From the explicit formulas (50-52) in Section 4 and the proof of Lemma 2.9, it follows that the difference between

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_G}\sum_{k=1}^{N_G}\ddot{\ell}_{\sigma\sigma}^{(M)}(\widetilde{\boldsymbol{\sigma}}_*^2(v),\widehat{\boldsymbol{\beta}}(v_k);\boldsymbol{y}_{ik})K_{h_1}(v_k-v)$$

and

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_G}\sum_{k=1}^{N_G}\ddot{\ell}_{\sigma\sigma}^{(M)}(\boldsymbol{\sigma}_*^2(v),\widehat{\boldsymbol{\beta}}(v_k);\boldsymbol{y}_{ik})K_{h_1}(v_k-v)$$

is $O_{p,V}\left(\gamma_n\log^2(1+N_G)/\sqrt{n}\right)=o_p(1)$. Combining this with Lemma 2.9, we have that for any $\boldsymbol{\sigma}(v)\in\mathcal{E}$ such that $\sup_{v\in V}\|\boldsymbol{\sigma}(v)-\boldsymbol{\sigma}_*(v)\|\leq\gamma_n\to 0$,

$$\mathcal{L}_{n,K}(\boldsymbol{\sigma}(v);\widehat{\boldsymbol{R}})=\mathcal{L}_{n,K}(\boldsymbol{\sigma}_*^2(v);\widehat{\boldsymbol{R}})+\left[\sqrt{n}(\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v))\right]^T\widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y}) \qquad (42)$$
$$-\frac{1}{2}\left[\sqrt{n}(\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v))\right]^T\left[\widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v);\boldsymbol{Y})+o_{p,V}(1)\right]\left[\sqrt{n}(\boldsymbol{\sigma}^2(v)-\boldsymbol{\sigma}_*^2(v))\right].$$

Similar to the proof of Lemma 2.8, but notice that the weighted score vector is now uniformly $O_p(1)$ ((36) of Lemma 2.9), we have $\sup_{v\in V}\|\sqrt{n}(\boldsymbol{\sigma}(v)-\boldsymbol{\sigma}_*(v))\|=O_p(1)$. Finally, we can take the $o_{p,V}(1)$ term out of (42) to conclude this Lemma.

Proof of Theorem 2.

We have shown the uniform consistency and uniform convergence rate of $\widehat{\boldsymbol{\sigma}}_K^2(v)$ in Lemmas 2.10 and 2.11. We are now ready to show the asymptotic distribution of related estimators. Let's introduce some notation as follows:

$$Z_K(v)=\mathcal{I}_K(v)^{-1}\mathcal{J}_K(v),$$

$$Q_K(\boldsymbol{\lambda})=(\boldsymbol{\lambda}-Z_K(v))^T\mathcal{I}_K(v)(\boldsymbol{\lambda}-Z_K(v)),$$

$$\widehat{\boldsymbol{\lambda}}_K(v)=\operatorname*{arg\,min}_{\boldsymbol{\lambda}\in[0,\infty)\times R^2}Q_K(\boldsymbol{\lambda}),\ H_K=[1,0,0]$$

$$Q_{1,K}(\lambda_1)=[\lambda_1-Z_{1,K}(v)]^T\left[H_K\mathcal{I}_K^{-1}H_K^T\right]^{-1}[\lambda_1-Z_{1,K}(v)],$$

$$\widehat{\lambda}_{1,K}(v)=\operatorname*{arg\,min}_{\lambda_1\in[0,\infty)}Q_{1,K}(\lambda_1)=Z_{1,K}(v)\mathbf{1}(Z_{1,K}(v)\geq 0),$$

$$\mathcal{J}_K(v)=\begin{pmatrix}\mathcal{J}_{1,K}(v)\\\mathcal{J}_{2,K}(v)\end{pmatrix},\quad\mathcal{I}_K(v)=\begin{bmatrix}\mathcal{I}_{11,K}(v)&\mathcal{I}_{12,K}(v)\\\mathcal{I}_{21,K}(v)&\mathcal{I}_{22,K}(v)\end{bmatrix},$$

$$Z_K(v)=\begin{pmatrix}Z_{1,K}(v)\\Z_{2,K}(v)\end{pmatrix},\quad\widehat{\boldsymbol{\lambda}}_K(v)=\begin{pmatrix}\widehat{\lambda}_{1,K}(v)\\\widehat{\boldsymbol{\lambda}}_{2,K}(v)\end{pmatrix}.$$

It follows from Lemma 2.11 that we have

$$(\widehat{\mathcal{J}}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}), \widehat{\mathcal{I}}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y})) \Rightarrow_d (\mathcal{J}_K(\cdot), \mathcal{I}_K(\cdot)).$$

Similar to the proof of Theorem 1, we have

$$\sqrt{n}(\widehat{\boldsymbol{\sigma}}_K^2(v) - \boldsymbol{\sigma}_*^2(v)) \Rightarrow_d \widehat{\boldsymbol{\lambda}}_K(v),$$

$$\sqrt{n}(\widehat{\sigma}_{a,K}^2(v) - \sigma_{a*}^2(v)) \Rightarrow_d \widehat{\lambda}_{1,K}(v),$$

$$\text{WLR}_n(v) \Rightarrow_d \widehat{\lambda}_{1,K}^T(v) \left[H_K \mathcal{I}_K^{-1} H_K^T \right]^{-1} \widehat{\lambda}_{1,K}(v).$$

However, since $Z_K(v) = \mathcal{I}_K(v)^{-1} \mathcal{J}_K(v)$ is normally distributed with mean $\boldsymbol{0}$ and covariance matrix $\mathcal{I}_K^{-1}(v) \mathcal{I}_{1,K}(v) \mathcal{I}_K^{-1}(v)$, $Z_{1,K}(v) = H_K Z_K(v)$ is normally distributed with mean 0 and variance

$$H_K \mathcal{I}_K^{-1}(v) \mathcal{I}_{1,K}(v) \mathcal{I}_K^{-1}(v) H_K^T.$$

We can simplify the distribution of $\widehat{\sigma}_{a,K}^2(v)$ and $\text{WLR}_n(v)$ at each location as:

$$\sqrt{n}(\widehat{\sigma}_{a,K}^2(v) - \sigma_{a*}^2(v)) \to_d \left[H_K \mathcal{I}_K^{-1}(v) \mathcal{I}_{1,K}(v) \mathcal{I}_K^{-1}(v) H_K^T \right]^{1/2} N(0,1) \mathbf{1}(N(0,1) \geq 0),$$

$$\text{WLR}_n(v) \to_d \frac{H_K \mathcal{I}_K^{-1}(v) \mathcal{I}_{1,K}(v) \mathcal{I}_K^{-1}(v) H_K^T}{H_K \mathcal{I}_K^{-1}(v) H_K^T} \left[\frac{1}{2} \chi_1^2 + \frac{1}{2} \chi_0^2 \right].$$

Next, we consider the asymptotic distribution of $\widehat{\sigma}_{a,K}^2(v)$ and $\text{WLR}_n(v)$ under local alternative $H_n : \sigma_a^2(v) = \sigma_{a*}^2(v) + h(v)/\sqrt{n}$. Following the argument of Section 8 in Andrews (2001), we have

$$\mathcal{J}_{n,K}(\boldsymbol{\sigma}_*^2(v) + \boldsymbol{h}_K(v)/\sqrt{n}; \boldsymbol{Y}) \overset{H_n}{\Longrightarrow}_d \mathcal{J}_K(v)$$

where $\boldsymbol{h}_K(v) = [h(v), 0, 0]^T$. Furthermore, we have

$$\mathcal{J}_{n,K}(\boldsymbol{\sigma}_*^2(v) + \boldsymbol{h}_K(v)/\sqrt{n}; \boldsymbol{Y}) = \mathcal{J}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}) + \frac{1}{\sqrt{n}} \frac{\partial \mathcal{J}_{n,K}}{\partial \boldsymbol{\sigma}_*^2(v)} H_K^T h(v) + o_p(1),$$

and it is easy to show that

$$\frac{1}{\sqrt{n}} \frac{\partial \mathcal{J}_{n,K}}{\partial \boldsymbol{\sigma}_*^2(v)} \overset{H_n}{\Longrightarrow}_d -\mathcal{I}_K(v), \quad \mathcal{I}_{n,K}(\boldsymbol{\sigma}_*^2(v); \boldsymbol{Y}) \overset{H_n}{\Longrightarrow}_d \mathcal{I}_K(v).$$

Then we have, under local alternative $H_n : \sigma_a^2(v) = \sigma_{a*}^2(v) + h(v)/\sqrt{n}$

$$\sqrt{n}\{\widehat{\sigma}_{a,K}^2(v) - \sigma_{a*}^2(v)\} \overset{H_n}{\Longrightarrow}_d \left\{ H_K \mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T \right\}^{1/2}$$
$$\times N(\widetilde{h}_K(v), 1)\mathbf{1}\left(N(\widetilde{h}_K(v), 1) \geq 0 \right),$$

$$\mathrm{WLR}_n(v) \overset{H_n}{\Longrightarrow} \frac{H_K \mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T}{H_K \mathcal{I}_K^{-1}(v)H_K^T}$$
$$\times N(\widetilde{h}_K(v), 1)^2\mathbf{1}\left(N(\widetilde{h}_K(v), 1) \geq 0 \right),$$

where $\widetilde{h}_K(v) = \left[H_K \mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T \right]^{-1/2} h(v)$.

3 Proof of Theorem 3

Given $u, v \in (0, 1)$, we consider $w(u, v)$ first.

$$\mathbb{E}\left[w(u, v) \right] = \mathbb{E}\left[K_{h_2}(T_1 - u)K_{h_2}(T_2 - v) \right] = 1.$$

We then consider $\widehat{S}_{w0}(u, v)$, $\widehat{S}_{w1}(u, v)$ and $\widehat{S}_{w2}(u, v)$. Their expectations are given as,

$$\mathbb{E}\left[\widehat{S}_{w0}(u, v) \right] = \Sigma_a(u, v) + \Sigma_c(u, v) + \Sigma_{e,G}(u, v)$$
$$+ \frac{1}{2}u_2(K)\left[\frac{\partial^2 \Sigma_a}{\partial u^2} + \frac{\partial^2 \Sigma_a}{\partial v^2} + \frac{\partial^2 \Sigma_c}{\partial u^2} + \frac{\partial^2 \Sigma_c}{\partial v^2} + \frac{\partial^2 \Sigma_{e,G}}{\partial u^2} + \frac{\partial^2 \Sigma_{e,G}}{\partial v^2} \right] h_2^2 + o(h_2^2) + O(n^{-1}),$$

$$\mathbb{E}\left[\widehat{S}_{w1}(u, v) \right] = \Sigma_a(u, v) + \Sigma_c(u, v)$$
$$+ \frac{1}{2}u_2(K)\left[\frac{\partial^2 \Sigma_a}{\partial u^2} + \frac{\partial^2 \Sigma_a}{\partial v^2} + \frac{\partial^2 \Sigma_c}{\partial u^2} + \frac{\partial^2 \Sigma_c}{\partial v^2} \right] h_2^2 + o(h_2^2) + O(n^{-1}),$$

$$\mathbb{E}\left[\widehat{S}_{w2}(u, v) \right] = \frac{1}{2}\Sigma_a(u, v) + \Sigma_c(u, v)$$
$$+ \frac{1}{2}u_2(K)\left[\frac{1}{2}\frac{\partial^2 \Sigma_a}{\partial u^2} + \frac{1}{2}\frac{\partial^2 \Sigma_a}{\partial v^2} + \frac{\partial^2 \Sigma_c}{\partial u^2} + \frac{\partial^2 \Sigma_c}{\partial v^2} \right] h_2^2 + o(h_2^2) + O(n^{-1}).$$

and their variance and covariance structures are given as

$$\mathrm{Var}\left[\widehat{S}_{w0}(u,v)\right] = \frac{1}{n}\left[\frac{\alpha_1}{(2\alpha_1 + 2\alpha_2 + \alpha_3)^2}V_1(u,v,u,v) + \frac{\alpha_2}{(2\alpha_1 + 2\alpha_2 + \alpha_3)^2}V_2(u,v,u,v)\right.$$

$$\left. + \frac{\alpha_{31}}{(2\alpha_1 + 2\alpha_2 + \alpha_3)^2}V_3(u,v,u,v) + \frac{\alpha_{32}}{(2\alpha_1 + 2\alpha_2 + \alpha_3)^2}V_4(u,v,u,v)\right] + o(h_2^2) + O(\frac{1}{N_G h_2}),$$

$$\mathrm{Var}\left[\widehat{S}_{w1}(u,v)\right] = \frac{1}{n}\left[\frac{V_7(u,v,u,v)}{\alpha_1}\right] + o(h_2^2) + O(\frac{1}{N_G h_2}),$$

$$\mathrm{Var}\left[\widehat{S}_{w2}(u,v)\right] = \frac{1}{n}\left[\frac{V_8(u,v,u,v)}{\alpha_2}\right] + o(h_2^2) + O(\frac{1}{N_G h_2}),$$

$$\mathrm{Cov}\left[\widehat{S}_{w1}(u,v),\widehat{S}_{w2}(u,v)\right] = o(h_2^2) + O(\frac{1}{N_G h_2}),$$

$$\mathrm{Cov}\left[\widehat{S}_{w0}(u,v),\widehat{S}_{w1}(u,v)\right] = \frac{1}{n}\left[\frac{1}{2\alpha_1 + 2\alpha_2 + \alpha_3}V_5(u,v,u,v)\right] + o(h_2^2) + O(\frac{1}{N_G h_2}),$$

$$\mathrm{Cov}\left[\widehat{S}_{w0}(u,v),\widehat{S}_{w2}(u,v)\right] = \frac{1}{n}\left[\frac{1}{2\alpha_1 + 2\alpha_2 + \alpha_3}V_6(u,v,u,v)\right] + o(h_2^2) + O(\frac{1}{N_G h_2}),$$

where V_1, V_2, V_3, V_4 are

$$V_1(u,v,u,v) = 4V_a(u,v,u,v) + 4V_c(u,v,u,v) + 2V_{e,G}(u,v,u,v)$$

$$+ 8\Sigma_a(u,v)\Sigma_c(u,v) + 4\Sigma_a(u,v)\Sigma_{e,G}(u,v) + 4\Sigma_c(u,v)\Sigma_{e,G}(u,v)$$

$$+ 4\Sigma_a(u,u)\Sigma_c(v,v) + 4\Sigma_a(v,v)\Sigma_c(u,u)$$

$$+ 2\Sigma_a(u,u)\Sigma_{e,G}(v,v) + 2\Sigma_a(v,v)\Sigma_{e,G}(u,u) + 2\Sigma_c(u,u)\Sigma_{e,G}(v,v) + 2\Sigma_c(v,v)\Sigma_{e,G}(u,u)$$

$$- 4\Sigma_a^2(u,v) - 4\Sigma_c^2(u,v) - 2\Sigma_{e,G}^2(u,v)$$

$$V_2(u,v,u,v) = \frac{3}{2}V_a(u,v,u,v) + \Sigma_a(u,u)\Sigma_a(v,v) + 4V_c(u,v,u,v) + 2V_{e,G}(u,v,u,v)$$

$$+ 6\Sigma_a(u,v)\Sigma_c(u,v) + 4\Sigma_a(u,v)\Sigma_{e,G}(u,v) + 4\Sigma_c(u,v)\Sigma_{e,G}(u,v)$$

$$+ 3\Sigma_a(u,u)\Sigma_c(v,v) + 3\Sigma_a(v,v)\Sigma_c(u,u)$$

$$+ \Sigma_a(u,u)\Sigma_{e,G}(v,v) + \Sigma_a(v,v)\Sigma_{e,G}(u,u) + 2\Sigma_c(u,u)\Sigma_{e,G}(v,v) + 2\Sigma_c(v,v)\Sigma_{e,G}(u,u)$$

$$- \frac{1}{2}\Sigma_a^2(u,v) - 4\Sigma_c^2(u,v) - 2\Sigma_{e,G}^2(u,v)$$

$$V_3(u,v,u,v) = V_a(u,v,u,v) + V_c(u,v,u,v) + V_{e,G}(u,v,u,v)$$

$$+ \Sigma_a(u,u)\Sigma_c(v,v) + \Sigma_a(v,v)\Sigma_c(u,u)$$

$$+ \Sigma_a(u,u)\Sigma_{e,G}(v,v) + \Sigma_a(v,v)\Sigma_{e,G}(u,u) + \Sigma_c(u,u)\Sigma_{e,G}(v,v) + \Sigma_c(v,v)\Sigma_{e,G}(u,u)$$

$$+ 2\Sigma_a(u,v)\Sigma_c(u,v) - \Sigma_a^2(u,v) - \Sigma_c^2(u,v) - \Sigma_{e,G}^2(u,v)$$

$$V_4(u,v,u,v) = \frac{1}{2}V_a(u,v,u,v) + \frac{1}{2}\Sigma_a(u,u)\Sigma_a(v,v) + V_c(u,v,u,v) + V_{e,G}(u,v,u,v)$$

$$+ 2\Sigma_a(u,v)\Sigma_c(u,v) + 2\Sigma_a(u,v)\Sigma_{e,G}(u,v) + 2\Sigma_c(u,v)\Sigma_{e,G}(u,v)$$

$$+ \Sigma_a(u,u)\Sigma_c(v,v) + \Sigma_a(v,v)\Sigma_c(u,u)$$

$$+ \Sigma_a(u,u)\Sigma_{e,G}(v,v) + \Sigma_a(v,v)\Sigma_{e,G}(u,u) + \Sigma_c(u,u)\Sigma_{e,G}(v,v) + \Sigma_c(v,v)\Sigma_{e,G}(u,u)$$

$$- \frac{1}{2}\Sigma_a^2(u,v) - \Sigma_c^2(u,v) - \Sigma_{e,G}^2(u,v)$$

and V_5, V_6, V_7, V_8 are

$$V_5(u, v, u, v) = V_a(u, v, u, v) + V_c(u, v, u, v) + \Sigma_a(u, u)\Sigma_c(v, v) + \Sigma_a(v, v)\Sigma_c(u, u)$$

$$+ \Sigma_a(u, u)\Sigma_{e,G}(v, v) + \Sigma_a(v, v)\Sigma_{e,G}(u, u) + \Sigma_c(u, u)\Sigma_{e,G}(v, v) + \Sigma_c(v, v)\Sigma_{e,G}(u, u)$$

$$- 2\Sigma_a^2(u, v) - 2\Sigma_c^2(u, v)$$

$$V_6(u, v, u, v) = \frac{1}{4}V_a(u, v, u, v) + \frac{1}{2}\Sigma_a(u, u)\Sigma_a(v, v) + V_c(u, v, u, v)$$

$$+ \Sigma_a(u, u)\Sigma_c(v, v) + \Sigma_a(v, v)\Sigma_c(u, u)$$

$$+ \frac{1}{2}\Sigma_a(u, u)\Sigma_{e,G}(v, v) + \frac{1}{2}\Sigma_a(v, v)\Sigma_{e,G}(u, u) + \Sigma_c(u, u)\Sigma_{e,G}(v, v) + \Sigma_c(v, v)\Sigma_{e,G}(u, u)$$

$$- \frac{1}{2}\Sigma_a^2(u, v) - 2\Sigma_c^2(u, v)$$

$$V_7(u, v, u, v) = V_a(u, v, u, v) + V_c(u, v, u, v) + \Sigma_a(u, u)\Sigma_c(v, v) + \Sigma_a(v, v)\Sigma_c(u, u)$$

$$+ \Sigma_a(u, u)\Sigma_{e,G}(v, v) + \Sigma_a(v, v)\Sigma_{e,G}(u, u) + \Sigma_c(u, u)\Sigma_{e,G}(v, v) + \Sigma_c(v, v)\Sigma_{e,G}(u, u)$$

$$+ 2\Sigma_a(u, v)\Sigma_c(u, v) + 2\Sigma_a(u, v)\Sigma_{e,G}(u, v) + 2\Sigma_c(u, v)\Sigma_{e,G}(u, v)$$

$$+ \Sigma_{e,G}(u, u)\Sigma_{e,G}(v, v) - \Sigma_a^2(u, v) - \Sigma_c^2(u, v)$$

$$V_8(u, v, u, v) = \frac{1}{4}V_a(u, v, u, v) + \frac{3}{4}\Sigma_a(u, u)\Sigma_a(v, v) + V_c(u, v, u, v)$$

$$+ \Sigma_a(u, u)\Sigma_c(v, v) + \Sigma_a(v, v)\Sigma_c(u, u)$$

$$+ \Sigma_a(u, u)\Sigma_{e,G}(v, v) + \Sigma_a(v, v)\Sigma_{e,G}(u, u) + \Sigma_c(u, u)\Sigma_{e,G}(v, v) + \Sigma_c(v, v)\Sigma_{e,G}(u, u)$$

$$+ 2\Sigma_a(u, v)\Sigma_c(u, v) + \Sigma_a(u, v)\Sigma_{e,G}(u, v) + 2\Sigma_c(u, v)\Sigma_{e,G}(u, v)$$

$$+ \Sigma_{e,G}(u, u)\Sigma_{e,G}(v, v) + \frac{1}{4}\Sigma_a^2(u, v) - \Sigma_c^2(u, v)$$

$$V_a(u, v, u, v) = \mathbb{E}\left[a_i^2(u)a_i^2(v)\right], \quad V_c(u, v, u, v) = \mathbb{E}\left[c_i^2(u)c_i^2(v)\right], \quad V_{e,G}(u, v, u, v) = \mathbb{E}\left[e_{ij,G}^2(u)e_{ij,G}^2(v)\right].$$

Assumption (C7a) indicates that we have the asymptotic distributions of $\widehat{\Sigma}_a(u, v)$, $\widehat{\Sigma}_c(u, v)$ and $\widehat{\Sigma}_{e,G}(u, v)$:

$$\sqrt{n}\left\{\widehat{\Sigma}_a(u, v) - \Sigma_a(u, v)\right\} \to_d N(0, W_a(u, v, u, v)),$$

$$\sqrt{n}\left\{\widehat{\Sigma}_c(u, v) - \Sigma_c(u, v)\right\} \to_d N(0, W_c(u, v, u, v)),$$

$$\sqrt{n}\left\{\widehat{\Sigma}_{e,G}(u, v) - \Sigma_{e,G}(u, v)\right\} \to_d N(0, W_{e,G}(u, v, u, v)),$$

where $W_a(u, v, u, v) = \sum_{k=1}^{8} a_k V_i(u, v, u, v)$, $W_c(u, v, u, v) = \sum_{k=1}^{8} c_k V_i(u, v, u, v)$ and $W_{e,G}(u, v, u, v) = \sum_{k=1}^{8} e_k V_i(u, v, u, v)$ for some constants a_k's, c_k's and e_k's.

4 Explicit Forms

$$r_1 = y_1 - x_1^T \beta, r_2 = y_2 - x_2^T \beta$$

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a & b \\ b & a \end{bmatrix} \right)$$

Its log likelihood and corresponding derivative are given as

$$\ell = -\frac{a}{2(a^2 - b^2)}(r_1^2 + r_2^2) + \frac{b}{a^2 - b^2}r_1 r_2 - \frac{1}{2}\log(a^2 - b^2), \tag{43}$$

$$\dot{\ell}_\beta = \frac{a}{(a^2 - b^2)}(x_1 r_1 + x_2 r_2) - \frac{b}{(a^2 - b^2)}(x_1 r_2 + x_2 r_1), \tag{44}$$

$$\dot{\ell}_a = \frac{a^2 + b^2}{2(a^2 - b^2)^2}(r_1^2 + r_2^2) - \frac{2ab}{(a^2 - b^2)^2}r_1 r_2 - \frac{a}{a^2 - b^2}, \tag{45}$$

$$\dot{\ell}_b = -\frac{ab}{(a^2 - b^2)^2}(r_1^2 + r_2^2) + \frac{a^2 + b^2}{(a^2 - b^2)^2}r_1 r_2 + \frac{b}{a^2 - b^2}, \tag{46}$$

$$\ddot{\ell}_{\beta\beta} = -\frac{a}{a^2 - b^2}(x_1 x_1^T + x_2 x_2^T) + \frac{b}{a^2 - b^2}(x_1 x_2^T + x_2 x_1^T), \tag{47}$$

$$\ddot{\ell}_{\beta a} = -\frac{a^2 + b^2}{(a^2 - b^2)^2}(x_1 r_1 + x_2 r_2) + \frac{2ab}{(a^2 - b^2)^2}(x_1 r_2 + x_2 r_1), \tag{48}$$

$$\ddot{\ell}_{\beta b} = \frac{2ab}{(a^2 - b^2)^2}(x_1 r_1 + x_2 r_2) - \frac{a^2 + b^2}{(a^2 - b^2)^2}(x_1 r_2 + x_2 r_1), \tag{49}$$

$$\ddot{\ell}_{aa} = -\frac{a(a^2 + 3b^2)}{(a^2 - b^2)^3}(r_1^2 + r_2^2) + \frac{2b(3a^2 + b^2)}{(a^2 - b^2)^3}r_1 r_2 + \frac{a^2 + b^2}{(a^2 - b^2)^2}, \tag{50}$$

$$\ddot{\ell}_{bb} = -\frac{a(a^2 + 3b^2)}{(a^2 - b^2)^3}(r_1^2 + r_2^2) + \frac{2b(3a^2 + b^2)}{(a^2 - b^2)^3}r_1 r_2 + \frac{a^2 + b^2}{(a^2 - b^2)^2}, \tag{51}$$

$$\ddot{\ell}_{ab} = \frac{b(3a^2 + b^2)}{(a^2 - b^2)^3}(r_1^2 + r_2^2) - \frac{2a(a^2 + 3b^2)}{(a^2 - b^2)^3}r_1 r_2 - \frac{2ab}{(a^2 - b^2)^2}. \tag{52}$$

5 Simulations

5.1 Example 1 (continue) on Heritability

For better comparison, we rewrite Dr. Wang's (2011) functional mixed effects model for longitudinal family data as follows:

$$y_{ij}(v) = \mu(v) + \alpha_i + \mathbf{c}_{ij}(v)^T \beta + \eta_{ij}(v) + \epsilon_{ij}(v) \tag{53}$$

for $j = 1, \ldots, m_i$, where v is time in Wang (2011), $\mathbf{c}_{ij}(v)$ is a time-varying coefficient, α_i is a random family-specific shared environmental effect, $\eta_{ij}(v)$ is a random subject-

specific genetic effect, and $\epsilon_{ij}(v)$ is a residual measurement error. Furthermore, in Wang (2001), she modeled $\eta_{ij}(v)$ as $\eta_{ij}(v) = B(v)^T \boldsymbol{b}_{ij}$, where $B(v)$ is a vector of spline basis and $\boldsymbol{b}_{ij}$ is the corresponding vector of subject-specific polygenic coefficients. Moreover, let $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^T, \ldots, \boldsymbol{b}_{im_i}^T)^T$, it is assumed that $\text{Cov}(\boldsymbol{b}_i, \boldsymbol{b}_i) = K_i \otimes \Omega$, where Ω is the unknown covariance matrix of the polygenic effects basis, $\otimes$ denotes the Kronecker and K_i is the known kinship coefficient matrix of the $i-$th family. Compared with (1), model (53) assumes that α_i and β are independent of v. We directly applied model (53) to the simulated data sets generated in the first simulation study corresponding to $n = 300, c = 0.1$, and $\Sigma_e = 0.2$.

Figure 1 shows that the method in Wang's (2011) cannot correctly estimate the heritability curve when the common environment effect is not constant along the tract. As expected, for the case with the common environment effect, Wang's (2011) method works pretty well.
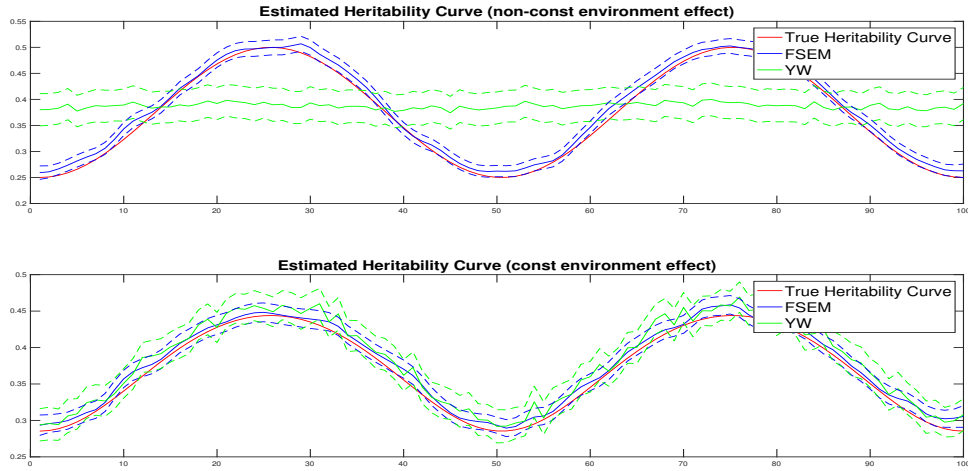

Figure 1: Estimated Heritability Curve when $n = 300, c = 0.1$, and $\Sigma_e = 0.2$.

5.2 Example 2 (continue)

Figure 2 presents the Type I and II error rates along the entire tract of both $\text{LR}_n(v)$ and $\text{WLR}_n(v)$ for $\Sigma_e = 2$.
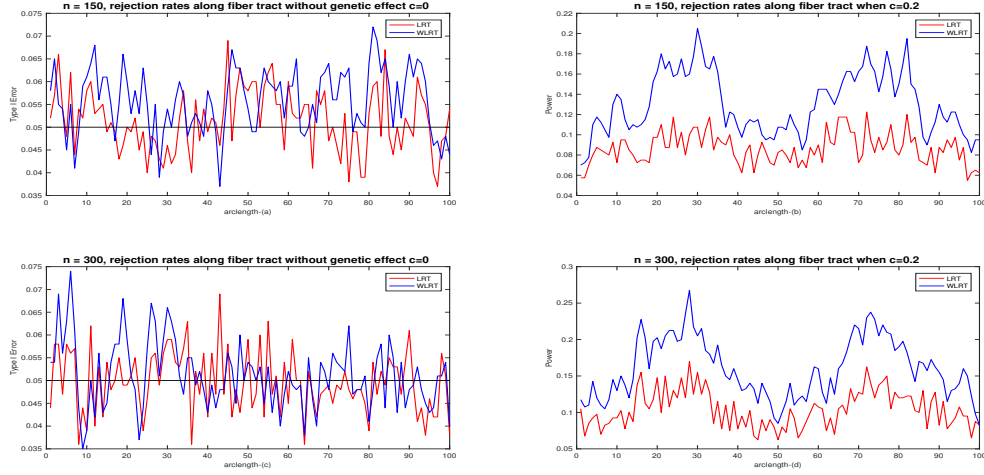
Figure 2: Inference Performance ($\Sigma_e = 2$): [(a),(c)] are rejection rates (type I error) of the two test statistics along fiber tract when $c = 0$ for $n = 150$ and $n = 300$; [(b),(d)] are rejection rates (power) of the two test statistics along fiber tract when $c = 0.2$ for $n = 150$ and $n = 300$.

References

Adamczak, R. (2007). A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electron. J. Probab 13*, 1000–1034.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica 67*(6), 1341–1383.

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*(3), 683–734.

Einmahl, U. and D. M. Mason (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability 13*, 1–37.

Fang, Y. and Y. Wang (2009). Testing for familial aggregation of functional traits. *Statistics in medicine 28*(29), 3611–3625.

Ledoux, M. and M. Talagrand (2013). *Probability in Banach Spaces: isoperimetry and processes*, Volume 23. Springer Science & Business Media.

van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3. Cambridge university press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

Wang, Y. (2011). Flexible estimation of covariance function by penalized spline with application to longitudinal family data. *Statistics in medicine 30*(15), 1883–1897.

Zhu, H. and H. Zhang (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B 66*(1), 3–16.