Statistica Sinica Preprint No: SS-2022-0336	
Title	Empirical Likelihood Inference of Variance Components in Linear Mixed-Effects Models
Manuscript ID	SS-2022-0336
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0336
Complete List of Authors	Jingru Zhang, Wei Guo, Joanne S. Carpenter, Andrew Leroux, Kathleen R. Merikangas, Nicholas G. Martin, Ian B. Hickie, Haochang Shou and Hongzhe Li
Corresponding Authors	Hongzhe Li
E-mails	hongzhe@upenn.edu
Notice: Accepted version subject to English editing.	

Empirical Likelihood Inference of Variance Components in Linear Mixed-Effects Models

Jingru Zhang[1], Wei Guo[2], Joanne S. Carpenter[3], Andrew Leroux[2,4], Kathleen R. Merikangas[2], Nicholas G. Martin[5], Ian B. Hickie[3], Haochang Shou[6], Hongzhe Li[6,*]

[1]*Fudan University,* [2]*National Institutes of Health,* [3]*University of Sydney,* [4]*University of Colorado,* [5]*QIMR Berghofer Medical Research Institute,* [6]*University of Pennsylvania*

Abstract: Linear mixed-effects models are widely used in analyzing repeated measures data, including clustered and longitudinal data, where inferences of both fixed effects and variance components are of interest. Unlike inference on fixed effect, which has been well studied, inference on the variance components is more challenging due to null value on the boundary and the unknown fixed effects as nuisance parameters. Existing methods require strong distributional assumptions

*Corresponding author: Hongzhe Li, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A. Email: hongzhe@pennmedicine.upenn.edu.

on the random effects and random errors. In this paper, we develop empirical likelihood-based methods for the inference of the variance components in the presence of fixed effects. We derive a nonparametric version of the Wilks' theorem for the proposed empirical likelihood ratio statistics for variance components. We also develop an empirical likelihood test for multiple variance components related to a sequence of correlated outcomes. Simulation studies demonstrate that the proposed methods exhibit better type 1 error control than the likelihood-based or score test when the Gaussian distributional assumptions of the random effects are violated. We apply the methods to investigate the heritability of physical activity measured by actigraph device in the Australian Twin study and observe that such activity is heritable only in the quantile range from 0.375 to 0.514.

Key words and phrases: Boundary value; Global test; Heritability; Nonparametric test; Wearable device data.

1. Introduction

Longitudinal and clustered data commonly arise from observational studies or clinical trials, where subjects are measured repeatedly over time or within a cluster. The repeated measures within a subject or a cluster are often correlated. To analyze such data, linear mixed-effects (LME) models that incorporate both fixed and random effects are widely used. Many statistical methods have been developed for such linear mixed-effects models, especially methods for inference of the fixed effects. However, inference on

the variance components is less studied and often requires strong distributional assumptions on the random effects and the error terms. When the underlying distributions are assumed to be multivariate normal, classical inference methods, such as the likelihood ratio test, the restricted likelihood ratio test, and the score test (Self and Liang, 1987; Zhang and Lin, 2003; Koh et al., 2019; Zhai et al., 2019), can be applied. However, these parametric methods are often restrictive and not robust if the model assumptions are violated. For example, we are interested in analyzing daily physical activity distributions as quantified by wearable device data such as actigraphy. Such data are typically measured at a high time resolution (e.g. minute-level) over several days for a given subject (Burton et al., 2013; Krane-Gartiser et al., 2014). The repeated measures enable us to account for day-to-day variability of the activity. Instead of focusing on the activity intensity levels at any given timestamp of the day, there is an increasing interest in summarizing daily activity time series into quantile functions and examining the activity distributions as a biologically meaningful measure of the activity traits (Ghosal et al., 2021; Zhang et al., 2022; Chang and McKeague, 2022). In analysis of such data, we will need to fit a linear mixed-effects model for each of the activity quantile $y_i(t)$ at t. Yet the typical distributional assumptions for random effects and errors terms in LME

might not hold.

Lin (1997) developed a global variance component test of all variance components equal to zero using an integrative quasi-likelihood and Laplacian expansion. This score test for the global variance component effectively uses the fact that under the global null, the observations are independent, and is robust in the special sense that the test does not require specifying the joint distribution of the random effects. However, the test cannot be applied for testing more general null hypothesis on the variance components. Lin (1997) also developed a score test for individual variance component. However, such a score test requires additional normality assumptions on the random effects, which allows one to use the existing maximum likelihood method or its approximations (Breslow and Lin, 1995) to estimate the unknown parameters under the null.

In this paper, we develop several tests of variance components based on empirical likelihood (EL) framework to avoid distributional assumptions on either the random effects or the error terms in LME. EL method, as an alternative to parametric likelihood-based methods, was first proposed by Owen (1988) and has been applied to many statistical inference problems (Owen, 2001). Without a prespecified distributional assumption on the data, EL methods incorporate side information through constraints or

prior distributions and have favorable statistical properties, including but not limited to Bartlett correctability, transformation invariance, better coverage accuracy of the corresponding confidence internals and greater power. Notable applications of the EL methods include mixture models (Zou et al., 2002) and censored survival data (Chang and McKeague, 2016).

Although the EL-based inference have been applied to longitudinal data modeling, majority of the methods only focus on the inference of the fixed effects term in LME. For example, You et al. (2006) proposed a block EL method for inference of the regression parameters assuming a working independence covariance, and Xue and Zhu (2007) considered a semiparametric regression model, where the repeated within-subject measures are summarized as a function over time in order to address the dependence issue. Wang et al. (2010) proposed a generalized EL method that takes into account the within-subject correlations. Li and Pan (2013) defined an empirical likelihood ratio (ELR) test by utilizing the extended score from quadratic inference functions for longitudinal data, which does not involve direct estimation of the correlation parameters.

In contrast, we consider a general setting of linear mixed-effects models and develop EL methods for the inference of the variance components. Specifically, suppose there are n subjects and denote by n_i the number of

repeated measures for the ith subject. For the ith subject, we observe a response vector $y_i \in R^{n_i}$, an $n_i \times p$ design matrix X_i for the fixed effects $\beta^* \in R^p$, and d $n_i \times n_i$ semi-positive design matrices Φ_{iq} $(q = 1, \cdots, d)$ for the variance components $\theta^* \in (R_+ \cup \{0\})^d$. The general linear mixed-effects model can be written as

$$y_i = X_i\beta^* + r_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where $r_i \in R^{n_i}$ is a zero-mean random variable with variance-covariance $H_i(\theta^*)$. We assume that $H_i(\theta^*)$ has a linear structure,

$$H_i(\theta^*) = \sum_{q=1}^{d} \theta_q^* \Phi_{iq}, \quad \theta^* = (\theta_1^*, \cdots, \theta_d^*)^T = (\theta_1^*, \theta_{(1)}^{*T})^T,$$

where $\theta^* = (\theta_1^*, \cdots, \theta_d^*)^T$ is the vector of the variance components. The linear structure of the variance components can be derived from $\mathrm{Cov}(r_i) = \mathrm{Cov}(Z_i b_i)$, where Z_i is the fixed design matrix, and b_i is the random effects that are generated from some distribution F with mean zero and covariance matrix $D(\theta^*)$. The linear structure holds when each components of $D(\theta^*)$ is a linear function of θ^* (Lin, 1997). This encompasses both nested, crossed and clustered designs (Michalski and Zmyślony, 1996; Zhai et al., 2019; Chen et al., 2019; Li et al., 2021). See Section 5.1 for a specific example of such a random-effect model for modeling the family data that includes additive genetic effect, common environment and unique subject-specific

environment effects. We emphasize that this general setting does not require any assumptions on the distributions of the data or the distributions of the random effects.

In many real applications, we are interested in making statistical inference on the variance components θ^* in model (1.1). For example, in the study of heritability based on twin data, each monozygotic twin or dizygotic twin is treated as one cluster, and the linear variance structure can be constructed based on the twin type (see details in Section 6). In the heritability analysis, a key question is whether there exists an genetic effect, which motivates us to study the inference of one of the variance components, say, θ_1^*. We propose to develop an EL based inference method for θ_1^* without any assumptions on the random components. The method can effectively account for the nuisance parameters, including the unknown fixed effects β^* and the variance components $\theta_{(1)}^{*T}$. The key difficulty when compared to the EL inference of the fixed effects is to deal with the boundary value problem when $\theta_1^* = 0$ in local testing problem $H_0 : \theta_1^* = \theta_1^0$. To solve the issues, we propose a new empirical likelihood ratio test by utilizing an unbiased estimator of β^* under very mild conditions, and prove that the asymptotic distribution of the test statistic is a mixture of χ^2 distributions.

In particular, we will apply the proposed methods to estimate and test

for the heritability of daily physical activity distribution summarized as quantiles in a twin study. This is a setting when LME models are fitted to a sequence of dependent outcomes. Denote by $\theta^*(t)$ the variance components for activity quantile t. We are interested in testing the global null H_0 : $\theta_1^*(t) \equiv \theta_1^0,\ t \in [t_1, t_2]$. We develop a max-type statistic for this global testing problem. Since the numerator of the proposed empirical likelihood ratio (ELR) tests can be rewritten as the sum of approximately independent random variables over different subjects, a random perturbation method is developed to approximate the p-value of the proposed global test.

We first introduce some notation. Denote by $(A)_{-1}$ the submatrix of A without the first column of A. For two vectors or matrices A and B of compatible dimension, define the inner product $\langle A, B\rangle = \mathrm{tr}(A^T B)$. For a matrix $D_{m\times n} = (D_1, \cdots, D_n)$, where D_i is the ith column of D, the vectorized D is defined by $(D_1^T, \cdots, D_n^T)^T$. Let $E(x)$ and $\mathrm{var}(x)$ be the expectation and variance of a random vector x, and let $\mathrm{cov}(x, y)$ be the covariance of random vectors x and y. When x is a random matrix, $E(x)$ and $\mathrm{var}(x)$ represent the expectation and variance of the vectorized x. When x and y are random matrices, $\mathrm{cov}(x, y)$ denotes the covariance of the vectorized x and vectorized y. We use $a = O(b)$ to denote that a and b are of the same order, and $a = o(b)$ to denote that a is of a smaller order than b. We

use $x = O_p(y)$ to denote that x and y are of the same order in probability, and $x = o_p(y)$ to denote that x is of a smaller order than y in probability.

2. ELR test for the fixed effects β^*

Statistical tests for the fixed effects in the linear mixed-effects model (1.1), $H_0 : \beta^* = \beta_0$, has been well studied. We first briefly review the subject-wise EL method proposed in Wang et al. (2010), where the covariance structure for each subject is considered.

Let $\hat{H}_{in}$ be an estimator of H_i, and assume that $\hat{H}_{in}$ converges to some H_i^* in probability uniformly over all $i = 1, \cdots, n$. One such a nonparametric sample covariance matrix $\hat{H}_{in}$ can be obtained using a simple two-step procedure, including estimating the residuals $\hat{r}_i = y_i - X_i\hat{\beta}$, where $\hat{\beta}$ is the least-squares estimator using working independence correlation matrices, and solving the constrained optimization problem $\min_{\theta \geq 0} \sum_{i=1}^{n} \|H_i(\theta) - \hat{r}_i\hat{r}_i^T\|_F^2$. Let

$$\phi_i(\beta) = X_i^T \hat{H}_{in}^{-1}(y_i - X_i\beta),$$

which satisfies $E\{\phi_i(\beta)\} = 0$ when β is the true value. Denote by p_i the point mass at the ith subject. The nonparametric empirical likelihood is

defined as

$$L_0(\beta) = \sup_{p_i} \Big\{ \prod_{i=1}^{n} p_i : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \phi_i(\beta) = 0 \Big\}.$$

Since it can be proved that $\max_\beta L_0(\beta) = 1/n^n$ (Owen, 2001), Wang et al. (2010) proposed the ELR statistic

$$\text{ELR}_0(\beta_0) = \frac{L_0(\beta_0)}{\max_\beta L_0(\beta)} = n^n L_0(\beta_0).$$

To obtain the asymptotic distribution of the ELR statistic, the following regularity conditions are needed.

Condition 1. *As $n \to \infty$, $P(0 \in ch\{\phi_1(\beta_0), \cdots, \phi_n(\beta_0)\}) \to 1$, where $ch\{\}$ is the convex hull.*

Condition 2. *The limit $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} X_i^T H_i^{*-1} H_i H_i^{*-1} X_i$ exists and is positive definite.*

Condition 3. *The expectation $E\|\phi_i(\beta_0)\|_2^{2+\gamma_1}$ are upper bounded uniformly for some $\gamma_1 > 0$.*

Let $\hat{G}_{in} = \hat{H}_{in}^{-1}$ with element $\hat{g}_{ijk}$, x_{ij}^T be the jth row of X_i, and r_{ik} be the kth element of r_i.

Condition 4. *For $i, i' = 1, \cdots, n$ $(i \neq i')$, $\hat{g}_{ijk} - \hat{g}_{-(i,i')jk} = O_p(n^{-1})$, and sufficient moment conditions are satisfied so that $E(\hat{B}_{ii'}) = O(n^{-1})$ and*

$E(\hat{B}_{ii'}\hat{B}_{ii'}^T) = O(n^{-2})$, *where* $\hat{g}_{-(i,i')jk}$ *is* $\hat{g}_{ijk}$ *but computed with all the data except for subjects* i *and* i' *and* $\hat{B}_{ii'} = \sum_{j=1}^{n_i}\sum_{k=1}^{n_i}(\hat{g}_{ijk} - \hat{g}_{-(i,i')jk})x_{ij}r_{ik}$.

Conditions 1–3 are common conditions for the empirical likelihood methods (Owen, 1991). Condition 4 assumes mild constraints on $\hat{H}_{in}^{-1}$ to ensure that the difference between the statistic $\text{ELR}_0(\beta_0)$ defined with $\hat{H}_{in}$ and the one using H_i^* vanishes as $n \to \infty$. If $\hat{H}_{in} - H_i^* = O_p(n^{-1})$ and $E\|r_i\|_2^2$ has an upper bound, then Condition 4 could be satisfied. Under these regularity conditions, the following theorem provides the asymptotic distribution of the ELR test $\text{ELR}_0(\beta_0)$ (Wang et al., 2010) under the null.

Theorem 1. *Under the regularity Conditions 1–4, as* $n \to \infty$,

$$-2\log \text{ELR}_0(\beta_0) \to \chi_p^2,$$

in distribution under the null hypothesis $H_0 : \beta^* = \beta_0$.

The asymptotic result only requires that the $\hat{H}_{in}$ converge uniformly to some H_i^*, which may not be the true H_i (Wang et al., 2010). When the correlation structure is correctly specified, the estimator $\hat{H}_{in}$ is a consistent estimator of $H_i^* = H_i$. The statistic defined with the true H_i is asymptotically locally most powerful among all the choices of the weight matrices.

3. ELR test for the variance component θ_1^*

We consider the local test $H_0 : \theta_1^* = \theta_1^0$ in the framework of the empirical likelihood, including the null $H_0 : \theta_1^* = 0$, which is of the most interest. We define $r_i = y_i - X_i\beta^*$ and $R_i = r_i r_i^T$. Since $E(r_i) = 0$ and $\text{var}(r_i) = H_i(\theta^*)$, we have

$$R_i = H_i(\theta^*) + \delta_i = \sum_{q=1}^{d} \theta_q^* \Phi_{iq} + \delta_i,$$

where $E(\delta_i) = 0$ and $\text{var}(\delta_i)$ exists. Since β^* is unknown, we first need an estimator of β^*, denoted by $\hat{\beta}$. One simple choice is the least-squares estimator using the all data. Specifically, we stack the data from all subjects by denoting $X = (X_1^T, \cdots, X_n^T)^T$, $y = (y_1^T, \cdots, y_n^T)^T$, and $r = (r_1^T, \cdots, r_n^T)^T$. Model (1.1) can be rewritten as

$$y = X\beta^* + r.$$

Then the least-squares estimator is $\hat{\beta} = (X^T X)^{-1} X^T y$. For $i = 1, \cdots, n$, let $\hat{r}_i = y_i - X_i\hat{\beta} = r_i + X_i(\beta^* - \hat{\beta})$. We have

$$\begin{aligned}
\hat{R}_i &= \hat{r}_i \hat{r}_i^T \\
&= r_i r_i^T + r_i(\beta^* - \hat{\beta})^T X_i^T + X_i(\beta^* - \hat{\beta}) r_i^T + X_i(\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^T X_i^T \\
&= R_i + \hat{\epsilon}_i = H_i(\theta^*) + \delta_i + \hat{\epsilon}_i,
\end{aligned}$$

where $\hat{\epsilon}_i = r_i(\beta^* - \hat{\beta})^T X_i^T + X_i(\beta^* - \hat{\beta}) r_i^T + X_i(\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^T X_i^T$.

To control the rates of $E(\hat{\epsilon}_i)$, $\mathrm{cov}(r_i r_i^T, \hat{\epsilon}_j)$, and $\mathrm{cov}(\hat{\epsilon}_i, \hat{\epsilon}_j)$, we need the following condition, which is also commonly used for empirical likelihood methods.

Condition 5. *The expectation* $E\|r_i\|_2^{4+\gamma_1}$ *are upper bounded uniformly for some* $\gamma_1 > 0$.

Under Condition 5, we see that the least-squares estimator $\hat{\beta}$ is good enough.

Proposition 1. *Assume that* $n^{-1}X^TX \to \Sigma$ *and* $n^{-1/2}X^Tr \xrightarrow{d} \eta$ *as* $n \to \infty$, *where* $0 < \|\Sigma\|_2, \|\Sigma^{-1}\|_2 < \infty$, $E\eta = 0$ *and* $E\|\eta\|_2^4 = O(1)$. *When Condition 5 holds and* $\hat{\beta} = (X^TX)^{-1}X^Ty$, *we have* $E(\hat{\epsilon}_i) = O(n^{-1})$, $i = 1, \cdots, n$, *and* $cov(r_i r_i^T, \hat{\epsilon}_j)$, $cov(\hat{\epsilon}_i, \hat{\epsilon}_j) = O(n^{-2})$, $i, j = 1, \cdots, n, i \neq j$.

Let $\Xi = (\Xi_{kl})_{d\times d}$ with $\Xi_{kl} = \sum_{i=1}^n \mathrm{tr}(\Phi_{ik}\Phi_{il})$, and $\hat{\Upsilon} = (\hat{\Upsilon}_1, \cdots, \hat{\Upsilon}_d)^T$ with $\hat{\Upsilon}_k = \sum_{i=1}^n \mathrm{tr}(\Phi_{ik}\hat{R}_i)$. We define

$$\hat{Z}_i(\theta_1) = \mathrm{tr}\Big\{\Phi_{i1}\Big(\hat{R}_i - \Phi_{i1}\theta_1 - \sum_{q=2}^{d}\hat{\theta}_q\Phi_{iq}\Big)\Big\},\ i = 1, \cdots, n,$$

where

$$\hat{\theta}_{(1)} = (\hat{\theta}_2, \cdots, \hat{\theta}_q)^T = (\Xi^{-1})_{-1}^T\hat{\Upsilon}. \tag{3.2}$$

Since Proposition 1 implies $E\hat{Z}_i(\theta_1) = O(n^{-1})$ if θ_1 is the true value (see (**??**) in the supplementary material), we define the nonparametric likelihood

as

$$L(\theta_1) = \max_{p_i} \left\{ \prod_{i=1}^{n} p_i | p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{Z}_i(\theta_1) = 0 \right\}$$

and the corresponding ELR statistic as

$$\text{ELR}(\theta_1^0) = \frac{L(\theta_1^0)}{\max_{\theta_1 \geq 0} L(\theta_1)}. \tag{3.3}$$

If the true value $\theta_1^* = 0$ (i.e., the null hypothesis under the case $\theta_1^0 = 0$), the denominator in (3.3) would not be $1/n^n$ as usual owing to the boundary value issue, and thus the existing results are inapplicable. To derive the asymptotic distribution of the proposed test $\text{ELR}(\theta_1^0)$, we assume the following condition similar to Condition 1.

Condition 6. *As $n \to \infty$, $P(0 \in ch\{Z_1(\theta_1^0), \cdots, Z_n(\theta_1^0)\}) \to 1$, where $Z_i(\theta_1^0)$ is defined as $\hat{Z}_i(\theta_1^0)$ with $\hat{R}_i$ replaced by R_i.*

Under Conditions 5 and 6, we have the following theorem on the asymptotic distribution of the ELR test under the null.

Theorem 2. *Let $\hat{c}_n(\theta_1^0) = \hat{\nu}_{2n}^2(\theta_1^0)/\hat{\nu}_{1n}^2(\theta_1^0)$, where $\hat{\nu}_{1n}^2(\theta_1^0)$ is a consistent estimator of the asymptotic variance of $n^{-1/2}\sum_{i=1}^{n} \hat{Z}_i(\theta_1^0)$ and $\hat{\nu}_{2n}^2(\theta_1^0) = n^{-1}\sum_{i=1}^{n} \hat{Z}_i^2(\theta_1^0)$. If $\theta_{(1)}^* \in R_+^{d-1}$, and Conditions 5 and 6 hold, as $n \to \infty$,*

$$\hat{c}_n(\theta_1^0) \left\{ -2 \log \text{ELR}(\theta_1^0) \right\} \to \chi_1^2$$

in distribution when $\theta_1^0 > 0$, *and*

$$\hat{c}_n(0)\left\{-2\log \text{ELR}(0)\right\} \to U_+^2$$

in distribution, where $U \sim N(0,1)$ *and* $U_+ = \max(U, 0)$.

Although the ELR statistic $\hat{c}_n(\theta_1^0)\,(-2\log \text{ELR}(\theta_1^0))$ in Theorem 2 involves optimizations in the numerator and denominator, the following lemma shows that the statistic has an asymptotically equivalent expression that can be used to calculate the statistic efficiently.

Lemma 1. *If* $\theta^*_{(1)} \in R_+^{d-1}$, *then under Conditions 5 and 6,*

$$\begin{aligned}
&\hat{c}_n(\theta_1^0)\left\{-2\log \text{ELR}(\theta_1^0)\right\} \\
&= \begin{cases} \frac{\left\{n^{-1/2}\sum_{i=1}^n \hat{Z}_i(\theta_1^0)\right\}^2}{\hat{\nu}_{1n}^2(\theta_1^0)} + o_p(1), & \text{if} \quad \theta_1^0 > 0, \\ \frac{\left\{n^{-1/2}\sum_{i=1}^n \hat{Z}_i(\theta_1^0)\right\}^2}{\hat{\nu}_{1n}^2(\theta_1^0)} I(\sum_{i=1}^n \hat{Z}_i(\theta_1^0) \geq 0) + o_p(1), & \text{if} \quad \theta_1^0 = 0. \end{cases}
\end{aligned}$$

We next provide an estimator of the asymptotic variance of $n^{-1/2}\sum_{i=1}^n \hat{Z}_i(\theta_1^0)$. We rewrite Ξ as

$$\Xi = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}$$

with E_{11} being a scalar. Let $F = E_{22}^{-1}E_{21} = (F_1, \cdots, F_{d-1})^T$ and $\alpha = 1 - E_{12}F/E_{11} \in (0, 1]$. It can be verified that

$$\sum_{i=1}^n \hat{Z}_i(\theta_1^0) = \sum_{i=1}^n \hat{D}_i(\theta_1^0) = \sum_{i=1}^n \hat{M}_i(\theta_1^0), \tag{3.4}$$

where

$$\hat{D}_i(\theta_1^0) = \alpha^{-1}\langle \Phi_{i1} - \sum_{q=1}^{d-1} F_q \Phi_{iq+1}, \hat{R}_i - \theta_1^0 \Phi_{i1} \rangle,$$
$$\hat{M}_i(\theta_1^0) = \alpha^{-1}\langle \Phi_{i1} - \sum_{q=1}^{d-1} F_q \Phi_{iq+1}, \hat{R}_i - H_i((\theta_1^0, \hat{\theta}_{(1)}^{T})^T) \rangle.$$

In addition, for $i \neq j$,

$$\mathrm{cov}(\hat{R}_i, \hat{R}_j) = \mathrm{cov}(\delta_i + \hat{\epsilon}_i, \delta_j + \hat{\epsilon}_j) = \ \mathrm{cov}(\delta_i, \hat{\epsilon}_j) + \mathrm{cov}(\hat{\epsilon}_i, \delta_j) + \mathrm{cov}(\hat{\epsilon}_i, \hat{\epsilon}_j) = O(n^{-2})$$

based on Proposition 1. Therefore, $\hat{D}_i(\theta_1^0)$ $(i = 1, \cdots, n)$ are asymptotically independent with expectation

$$E(\hat{D}_i(\theta_1^0)) = \alpha^{-1}\langle \Phi_{i1} - \sum_{q=1}^{d-1} F_q \Phi_{iq+1}, \sum_{q=2}^{d} \theta_q^* \Phi_{iq} \rangle + O(n^{-1}),$$

while the expectations of $\hat{Z}_i(\theta_1^0)$ and $\hat{M}_i(\theta_1^0)$ are $O(n^{-1})$ (see (**??**) in the supplementary material). We have

$$\begin{aligned}\hat{D}_i(\theta_1^0) - E(\hat{D}_i(\theta_1^0)) &= \alpha^{-1}\langle \Phi_{i1} - \sum_{q=1}^{d-1} F_q \Phi_{iq+1}, \hat{R}_i - H_i((\theta_1^0, (\theta_{(1)}^*)^T)^T) \rangle + O(n^{-1}) \\ &= \hat{M}_i(\theta_1^0) + o_p(1).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{var}\{n^{-1/2} \sum_{i=1}^{n} \hat{Z}_i(\theta_1^0)\} &= \frac{1}{n} \sum_{i=1}^{n} \{\hat{D}_i(\theta_1^0) - E(\hat{D}_i(\theta_1^0))\}^2 + o_p(1) \\ &= \frac{1}{n} \sum_{i=1}^{n} \hat{M}_i(\theta_1^0)^2 + o_p(1), \qquad (3.5)\end{aligned}$$

which leads a consistent estimator of the variance of $n^{-1/2}\sum_{i=1}^{n}\hat{Z}_i(\theta_1^0)$ as

$$\hat{\nu}_{1n}^2(\theta_1^0) = n^{-1}\sum_{i=1}^{n}\hat{M}_i(\theta_1^0)^2.$$

4. Variance component analysis over a sequence of responses

In some applications, we are interested in testing whether the variance components are all zero over a set of possibly correlated outcomes. One example of such applications is to test the variance components for the activity distribution based on wearable device data where we are interested in testing the variance component at each of the quantiles t of the activity distribution. Extending model (1.1), we assume the following outcome model at level t,

$$y_i(t) = X_i\beta^*(t) + r_i(t), \quad i = 1, \cdots, n, \tag{4.6}$$

where $r_i(t) \in R^{n_i}$ is a zero-mean random variable with variance $H_i(\theta^*(t))$. We assume that $H_i(\theta^*(t))$ has the same linear structure for each t,

$$H_i(\theta^*(t)) = \sum_{q=1}^{d}\theta_q^*(t)\Phi_{iq}, \quad \theta^*(t) = (\theta_1^*(t), \cdots, \theta_d^*(t))^T = (\theta_1^*(t), \theta_{(1)}^{*T}(t))^T.$$

We are interested in testing the null $H_0 : \theta_1^*(t) \equiv \theta_1^0,\ t \in [t_1, t_2]$, where $[t_1, t_2]$ is a pre-defined interval. We propose the following maximally selected empirical likelihood ratio statistic (gELR),

$$\Gamma = \sup_{t\in[t_1,t_2]} \hat{c}_n(\theta_1^0, t)\left\{-2\log \text{ELR}(\theta_1^0, t)\right\}, \tag{4.7}$$

where $\hat{c}_n(\theta_1^0, t)\{-2\log \text{ELR}(\theta_1^0, t)\}$ is the ELR statistic for the outcome at t. It can be shown that $\Gamma = \sup_{t\in[t_1,t_2]} S(t) + o_p(1)$, with

$$S(t) = \begin{cases} \frac{\{n^{-1/2}\sum_{i=1}^n \hat{Z}_i(\theta_1^0,t)\}^2}{\hat{\nu}_{1n}^2(\theta_1^0,t)}, & \text{if } \theta_1^0 > 0, \\ \frac{\{n^{-1/2}\sum_{i=1}^n \hat{Z}_i(\theta_1^0,t)\}^2}{\hat{\nu}_{1n}^2(\theta_1^0,t)} I\{\sum_{i=1}^n \hat{Z}_i(\theta_1^0,t) \geq 0\}, & \text{if } \theta_1^0 = 0, \end{cases}$$

where

$$\hat{Z}_i(\theta_1^0, t) = \text{tr}\Big\{\Phi_{i1}\Big(\hat{R}_i(t) - \Phi_{i1}\theta_1^0 - \sum_{q=2}^{d}\hat{\theta}_q(t)\Phi_{iq}\Big)\Big\},$$

$$\hat{\nu}_{1n}^2(\theta_1^0, t) = n^{-1}\alpha^{-2}\sum_{i=1}^{n}\Big\langle \hat{R}_i(t) - H_i((\theta_1^0, \hat{\theta}_{(1)}(t)^T)^T), \Phi_{i1} - \sum_{q=1}^{d-1} F_q\Phi_{iq+1}\Big\rangle^2.$$

Assessment of the statistical significance of the statistic Γ defined in (4.7) is challenging because of the dependence of $\hat{Z}_i(\theta_1^0, t)$. We propose a simple way of evaluating its significance by perturbing the EL statistic. Specifically, we apply (3.4) to rewrite $\sum_{i=1}^n \hat{Z}_i(\theta_1^0, t)$ as $\sum_{i=1}^n \hat{M}_i(\theta_1^0, t)$, where $\hat{M}_i(\theta_1^0, t) = \alpha^{-1}\Big\langle \Phi_{i1} - \sum_{q=1}^{d-1} F_q\Phi_{iq+1}, \hat{R}_i(t) - H_i((\theta_1^0, \hat{\theta}_{(1)}^T(t))^T)\Big\rangle$. We can generate the null distribution of Γ by perturbing the test statistic $\Gamma^{(g)}$. Specifically, for each g $(g = 1, \cdots, G)$, we generate $\xi_i^{(g)}$ from i.i.d. standard normal distribution and define

$$S^{(g)}(t) = \begin{cases} \frac{\{n^{-1/2}\sum_{i=1}^n \hat{M}_i(\theta_1^0,t)\xi_i^{(g)}\}^2}{\hat{\nu}_{1n}^2(\theta_1^0,t)}, & \text{if } \theta_1^0 > 0, \\ \frac{\{n^{-1/2}\sum_{i=1}^n \hat{M}_i(\theta_1^0,t)\xi_i^{(g)}\}^2}{\hat{\nu}_{1n}^2(\theta_1^0,t)} I\{\sum_{i=1}^n \hat{M}_i(\theta_1^0,t)\xi_i^{(g)} \geq 0\}, & \text{if } \theta_1^0 = 0. \end{cases}$$

Define the corresponding perturbed test statistic as $\Gamma^{(g)} = \sup_{t\in[t_1,t_2]} S^{(g)}(t)$. The following Proportion 2 shows that the perturbed test statistics have the

same asymptotic distribution as the original test statistic under the null. Therefore, the p-value of Γ can be approximated by $\sum_{g=1}^{G} I(\Gamma^{(g)} > \Gamma)/G$.

Proposition 2. *$\hat{M}_i(\theta_1^0, t)$ satisfies*

$$(i)\ E\{n^{-1/2}\sum_{i=1}^{n}\hat{M}_i(\theta_1^0,t)\xi_i^{(g)}\} - E\{n^{-1/2}\sum_{i=1}^{n}\hat{Z}_i(\theta_1^0,t)\} = o(1);$$

$$(ii)\ var\{n^{-1/2}\sum_{i=1}^{n}\hat{M}_i(\theta_1^0,t)\xi_i^{(g)}\} - var\{n^{-1/2}\sum_{i=1}^{n}\hat{Z}_i(\theta_1^0,t)\} = o(1);$$

$$(iii)\ cov\{n^{-1/2}\sum_{i=1}^{n}\hat{M}_i(\theta_1^0,s)\xi_i^{(g)}, n^{-1/2}\sum_{j=1}^{n}\hat{M}_j(\theta_1^0,t)\xi_j^{(g)}\}$$
$$- cov\{n^{-1/2}\sum_{i=1}^{n}\hat{Z}_i(\theta_1^0,s), n^{-1/2}\sum_{j=1}^{n}\hat{Z}_j(\theta_1^0,t)\} = o(1).$$

5. Simulation studies

5.1 Data generation

We examine the performance of the proposed empirical likelihood ratio tests for variance components and compare the results with the standard likelihood ratio (LR) test (Self and Liang, 1987) and the standard score test (Zhang and Lin, 2003) assuming Gaussian random effects and Gaussian errors. To mimic the twin design in the heritability analysis of wearable device data that we analyze next, we simulate data on a monozygotic or dizygotic twin pair. For the ith twin, let $n_i = n_{i1} + n_{i2}$, where n_{i1} and n_{i2} are the numbers of repeated measures for the twin. In wearable device data, $y_i(t)$

represents the tth quantile of the activity distributions over n_i days. The data are generated from the commonly-used model for heritability analysis (Ge et al., 2017):

$$y_i(t) = X_i\beta(t) + T_i a_i(t) + \tau_i(t), \quad i = 1, \cdots, n, \quad t = s_1, s_2, \cdots, s_m, \quad (5.8)$$

where $T_i = \text{blkdiag}\{\mathbf{1}_{n_{i1}}, \mathbf{1}_{n_{i2}}\}$, $a_i(t)$ is a random intercept, and $\tau_i(t)$ denotes zero-mean noise with variance $\sigma_M^2(t)\mathbf{I}_{n_i}$. Here, $a_i(t)$ is assumed as the sum of additive genetic effect $g_i(t)$, common environment $c_i(t)$, and unique subject-specific environment $e_i(t)$, i.e.,

$$a_i(t) = g_i(t) + c_i(t) + e_i(t),$$

where $g_i(t), c_i(t), e_i(t)$ are independent zero-mean random variables with variance-covariance $\sigma_A^2(t)K_i, \sigma_C^2(t)\Lambda_i$, and $\sigma_E^2(t)\mathbf{I}_2$, respectively. The variance components $\sigma_A^2(t), \sigma_C^2(t)$, and $\sigma_E^2(t)$ represent the additive genetic variance, common environmental variance, and unique environmental variance, respectively. For the ith twin, K_i is a genetic similarity matrix with $K_i = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ for monozygotic twin, $K_i = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$ for dizygotic twin, and Λ_i quantifies shared environment between the twin pair with $\Lambda_i = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Under this model, we have

$$H_i(\theta^*(t)) = \sigma_A^2(t)T_iK_iT_i^T + \sigma_C^2(t)T_i\Lambda_iT_i^T + \sigma_E^2(t)T_iT_i^T + \sigma_M^2(t)\mathbf{I}_{n_i},$$

$$\theta^*(t) = (\sigma_A^2(t), \sigma_C^2(t), \sigma_E^2(t), \sigma_M^2(t))^T.$$

We sample n_{ik} $(i = 1, \cdots, n; k = 1, 2)$ from $\{2, 3, 4\}$ with equal probability 1/3. We set $n = 300$, among which there are 150 monozygotic twin families and 150 dizygotic twin families, and $t = 0.01, 0.06, 0.11, \cdots, 0.96$. Let $X_i = (x_{i1}, \cdots, x_{in_i})^T$, where $x_{ij} \sim N(0, 1)$, and $\beta(t) = 3.5$. To evaluate the proposed tests for variance components in the case of correlated outcomes over an interval of t, we generate random effects and noises as follows. Let $g_i(t) = \sigma_A(t)\zeta_{ai}, c_i(t) = \sigma_C(t)\zeta_{ci}$, $e_i(t) = \sigma_E(t)\zeta_{ei}$, and $\tau_i(t) = \sigma_M(t)\zeta_{\tau i}$, where $\sigma_l^2(t)$ $(l = A, C, E, M)$ are set in a similar way as in Zhu et al. (2012)

$$\sigma_l^2(t) = C_{l1}(t)sin^2(2\pi t) + C_{l2}(t)cos^2(2\pi t), \quad l = A, C, E, M,$$

$(C_{A1}(t), C_{A2}(t)) = (1, 1.6)C_a I(t \in \{0.41, 0.46, 0.51, 0.56\})$, $(C_{C1}(t), C_{C2}(t)) = (0.2, 1)C_c$, $(C_{E1}(t), C_{E2}(t)) = (1, 0.2)C_e$, and $(C_{M1}(t), C_{M2}(t)) = (1.6, 1)C_m$, $I(\cdot)$ is an indicator function. We consider three types of distributions for $\zeta_{ai}, \zeta_{ci}, \zeta_{ei}, \zeta_{\tau i}$.

(i) normal random effects and noises: $\zeta_{ai} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, K_i)$, $\zeta_{ci} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Lambda_i)$, $\zeta_{ei} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, $\zeta_{\tau i} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_i})$.

(ii) t-distributed random effects and normal noises: $\zeta_{ai} \overset{iid}{\sim} t_3(\mathbf{0}, K_i/3)$, $\zeta_{ci} \overset{iid}{\sim} t_3(\mathbf{0}, \Lambda_i/3)$, $\zeta_{ei} \overset{iid}{\sim} t_3(\mathbf{0}, \mathbf{I}_2/3)$, $\zeta_{\tau i} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_i})$.

(iii) t-distributed random effects and noises: $\zeta_{ai} \overset{iid}{\sim} t_3(\mathbf{0}, K_i/3)$, $\zeta_{ci} \overset{iid}{\sim} t_3(\mathbf{0}, \Lambda_i/3)$, $\zeta_{ei} \overset{iid}{\sim} t_3(\mathbf{0}, \mathbf{I}_2/3)$, $\zeta_{\tau i} \overset{iid}{\sim} t_3(\mathbf{0}, \mathbf{I}_{n_i}/3)$.

5.2 ELR test for fixed effects

Since the focus of this paper is on ELR test for the various components, we focus our evaluation of the ELR test for the fixed effect (Wang et al., 2010) on its robustness to the distributional assumptions of the random effects and the errors. Let $C_a = 0.1$, $C_c = 0.05$, $C_e = 0.05$, and $C_m = 0.03$. For each $t \in T \equiv \{0.01, 0.06, 0.11, \cdots, 0.96\}$, we examine coverage probability for the ELR test reviewed in Section 2. Based on Theorem 1, the coverage probability for the ELR test is constructed using the asymptotic χ^2_1 distribution.

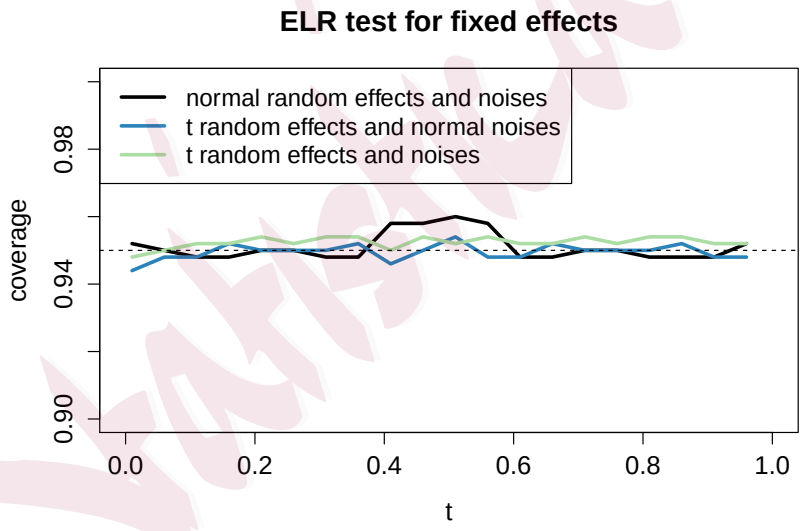


Figure 1: Coverage probability for the ELR test for fixed effects. The black dashed horizontal line is the 95% coverage probability.

Figure 1 shows the results of the simulation study. We observe that the ELR test produces coverage probabilities roughly at the nominal level under all types of distributions.

5.3 ELR test for single variance component

We consider the model parameters $C_a = 0.3$, $C_c = 0.05$, $C_e = 0.05$, and $C_m = 0.03$. For each $t \in T$, we evaluate the performance of the tests for the null hypothesis $H_0 : \sigma_A^2(t) = 0$. We use ELR to denote the proposed empirical likelihood ratio test with unknown $\beta^*(t)$. Since $\sigma_A^2(t) > 0$ when $t \in T_1 \equiv \{0.41, 0.46, 0.51, 0.56\}$ and 0 otherwise, we use the simulations for $t \in T \setminus T_1$ to examine the type 1 errors, and use the simulations for $t \in T_1$ to evaluate the power of the proposed test. We compare the results with two other commonly used tests of variance components, LR test and score test. We repeat the simulations 500 times.

Figure 2 presents the results of empirical power and the type 1 errors for different values of t. For the type 1 errors, when $t \in T \setminus T_1$, we observe that all the methods perform well under the normal random effects and noises. However, LR and score tests show inflated type 1 errors when random effects follow a long-tailed t distribution.

For the comparisons, where $t \in T_1$, we observe that the proposed

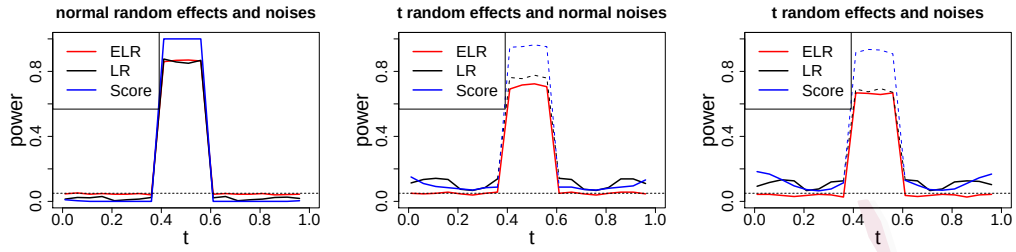


Figure 2: Empirical power for a given value of $t \in T_1 \equiv \{0.41, 0.46, 0.51, 0.56\}$, and type 1 errors for a given $t \in T \setminus T_1$. The black dashed horizontal line is the nominal threshold 0.05. Left: normal random effects and noises; middle: t-distributed random effects and normal noises; right: t-distributed random effects and noises. ELR: empirical likelihood ratio test with the least-square estimate of β^*; LR : likelihood ratio test under the normal assumption; Score : score test under the normal assumption. For the middle and right plots, dashed line segments for LR and Score tests indicate the t values where both tests have inflated type 1 errors.

method exhibits a similar power as the LR test, and the score test has the highest power under the normal distribution. However, when the random effects follow a t distribution, the ELR test does not lose much power. As a comparison, the empirical power of LR and score tests are also shown as the dashed lines in Figure 2. However, the power does not reflect the true power because of their inflated type 1 errors under the non-Gaussian

assumptions.

5.4 ELR test for variance components over an interval

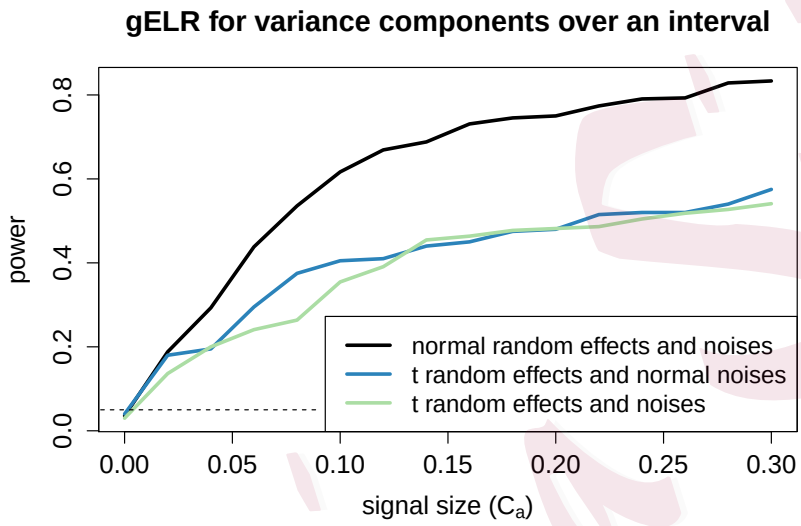


Figure 3: Empirical power for the global test $H_0 : \sigma_A^2(t) \equiv 0,\ t \in [0,1]$ at different values of C_a. The black dashed horizontal line is the nominal threshold 0.05.

We use gELR to denote the proposed global empirical likelihood ratio test with unknown $\beta^*(t)$. We consider the global test $H_0 : \sigma_A^2(t) \equiv 0,\ t \in [0,1]$, with different choices of the signal size $C_a = 0, 0.02, \cdots, 0.3$. Let $C_c = 0.05$, $C_e = 0.05$, and $C_m = 0.03$. We generate 500 datasets for each setting. Figure 3 presents the empirical power of gELR at 0.05 significance

level under different distributions of random errors and different C_a. As expected, the empirical power of rejecting the null hypothesis increases with the signal size. Compared to the results under the multivariate t distributed random effects, the proposed test gELR has higher power when data are normally distributed.

To further evaluate the type 1 error and the power, we consider models with $C_a = 0.3$, $C_c = 0.05$, $C_e = 0.05$, and $C_m = 0.03$. We consider to test each of the candidate intervals of lengths $\{3, 4, 5\}$ and denote them by scan3, scan4, and scan5, respectively. Let $\mathcal{J}_k$ be the set of candidate intervals under the scanning length k ($k = 3, 4, 5$) and let $\mathcal{J} = \cup_{k=3}^{5}\mathcal{J}_k$ be the set of all candidate intervals. For each candidate interval $L \in \mathcal{J}$, we test the null hypothesis $H_0 : \sigma_A^2(t) \equiv 0,\ t \in L$. The signal in the interval L is significant if

$$h(\Gamma_L) = \frac{\Gamma_L - \bar{\Gamma}_L}{\sqrt{\sum_{g=1}^{G}(\Gamma_L^{(g)} - \bar{\Gamma}_L)^2/(G-1)}} > \sqrt{2\log|\mathcal{J}|},$$

where $\bar{\Gamma}_L = (\sum_{g=1}^{G}\Gamma_L^{(g)})/G$ with $G = 1000$. The threshold $\sqrt{2\log|\mathcal{J}|}$ is selected based on the extreme value distribution of $|\mathcal{J}|$ normal random variables.

Under each type of error distributions, 500 datasets are generated. For the global test under the candidate interval $L = [t_1, t_2]$, we mark its em-

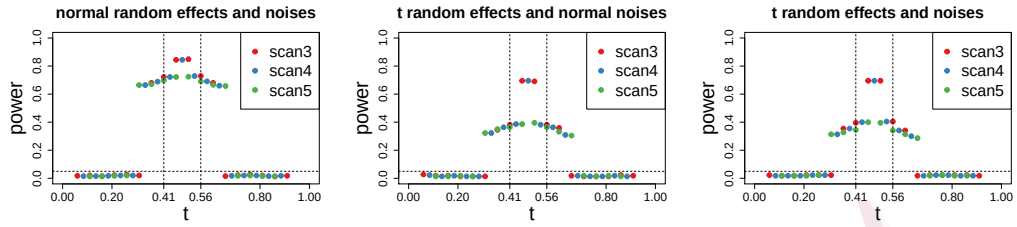


Figure 4: Empirical power of testing zero variance component in a given interval of length of 3, 4, and 5, where the true nonzero variance components are at $\{0.41,\ 0.46,\ 0.51,\ 0.56\}$. The black dashed horizontal line is the nominal threshold 0.05. Left: normal random effects and noises; middle: t-distributed random effects and normal noises; right: t-distributed random effects and noises. Fo the middle and left plots, dashed line segments for LR and Score indicate the

pirical power at $(t_1 + t_2)/2$. The results are shown in Figure 4. The proposed global test gELR exhibits high power if the interval involves at least one nonzero time points. When a candidate interval does not involve any nonzero time points, the empirical power is less than 0.05, and therefore the proposed gELR test controls type 1 error very well.

6. Application to genetic heritability analysis of physical activity distribution

6.1 Description of the data

We apply the methods to actigraphy data from the Australian twin study, which includes 189 healthy twin families (79 of them are monozygotic twin families and 110 are dizygotic twin families) and 6,103 observations (days) in total. The participants wore GENEActiv devices to track their physical activities for about 14 days. The minute-to-minute activity intensities derived from the device were collected in a 1440-dimensional vector per day. Since we are interested in the inference of heritability of the activity distributions, we obtain the empirical quantiles of activity counts at different quantiles, $t = 1/144,\ 2/144, \cdots, 144/144$. Specifically, for the kth person in the ith twin family, the raw data from the wearable device for the jth day is a minute-level vector, $\xi_{ikj} = (\xi_{ikj1}, \cdots, \xi_{ikj1440})^T$, where ξ_{ikjt} represents the activity (average Euclidean norm minus 1) measurement for the tth minute. Following the standard data processing step as implemented in the R package mMARCH.AC (Guo et al., 2022), we transform the data as

$$\tilde{\xi}_{ikj} = \log(9250\xi_{ikj} + 1), \quad i = 1, \cdots, n;\ k = 1, 2;\ j = 1, \cdots, n_{ik},$$

For the kth person in the ith twin family, the jth repeated measure of t-quantile of activity counts is obtained as

$$y_{ikj}(t) = \tilde{\xi}_{ikj}^{[1440\cdot t]}, \quad t = 1/144,\ 2/144, \cdots, 144/144,$$

where $\tilde{\xi}_{ikj}^{[s]}$ denotes the sth order statistic of $\tilde{\xi}_{ikj}$.

In our analysis, the covariate x_{ikj} includes gender, age, body mass index (BMI), and indicator of weekend, i.e., $x_{ikj} = (1, \text{Gender}, \text{Age}, \text{BMI}, \text{Weekend})^T$. Let $y_i(t) = (y_{i11}(t), \cdots, y_{i1n_{i1}}(t), y_{i21}(t), \cdots, y_{i2n_{i2}}(t))^T$ and $X_i = (x_{i11}, \cdots, x_{i1n_{i1}}, x_{i21}, \cdots, x_{i2n_{i2}})^T$. We consider the same model as (5.8):

$$y_i(t) = X_i\beta(t) + T_i(g_i(t) + c_i(t) + e_i(t)) + \tau_i(t), \quad i = 1, \cdots, n,$$

where $T_i = \text{blkdiag}\{\mathbf{1}_{n_{i1}}, \mathbf{1}_{n_{i2}}\}$, and $g_i(t)$, $c_i(t)$, and $e_i(t)$ represent the genetic effect $g_i(t)$, common environment $c_i(t)$, and unique subject-specific environment $e_i(t)$, respectively. The variance-covariance structure of these random effects are the same as those in Section 5.1.

Let $y(t) = (y_1^T(t), \cdots, y_n^T(t))^T$. The original data set includes 189 healthy twin families, among those 79 are monozygotic (MZ) twin families, and 110 dizygotic (DZ) twin families. The total number of days with activity data is 6,103. After removing the observations with missing covariate X or missing activity quantile Y, we have 172 healthy twin families, including 73 MZ twin families and 99 DZ twin families, with a total num-

ber of days with activity data being 4,698. We further remove the outliers under 1.5 IQR rule, which reduces to total number of observations to 3,741. Among them, we select the families with a twin pair, including 63 MZ twin families and 86 DZ twin families, and the total number of observations is 3,489. This is the final data set in our analysis.

6.2 Effects of gender, age, BMI, weekend on activity profiles

We first examine the associations between the covariates including gender, age, BMI, and weekend vs weekday and the overall activity distribution. For each of the four covariates and each of the t values, we obtain the EL estimator by solving the estimating equations $\sum_{i=1}^{n} \phi_i(\beta) = 0$, and apply Theorem 1 to construct the confidence interval $\{\beta_0 : -2\log \text{ELR}_0(\beta_0) \leq \chi_1^2(1-\alpha)\}$, where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)$ quantile of the χ_1^2 distribution. The first column of Figure 5 shows the estimated regression coefficient for each of the t values and its point-wise 95% confidence intervals using the EL method.

We then test whether there is any difference in activity profiles between individuals of different gender, age, BMI, and whether the activity profiles are different between weekdays and weekends. Specifically, we consider testing such differences at each of the quantile t. To test $H_0 : \beta_l(t) = 0,\ l \in$

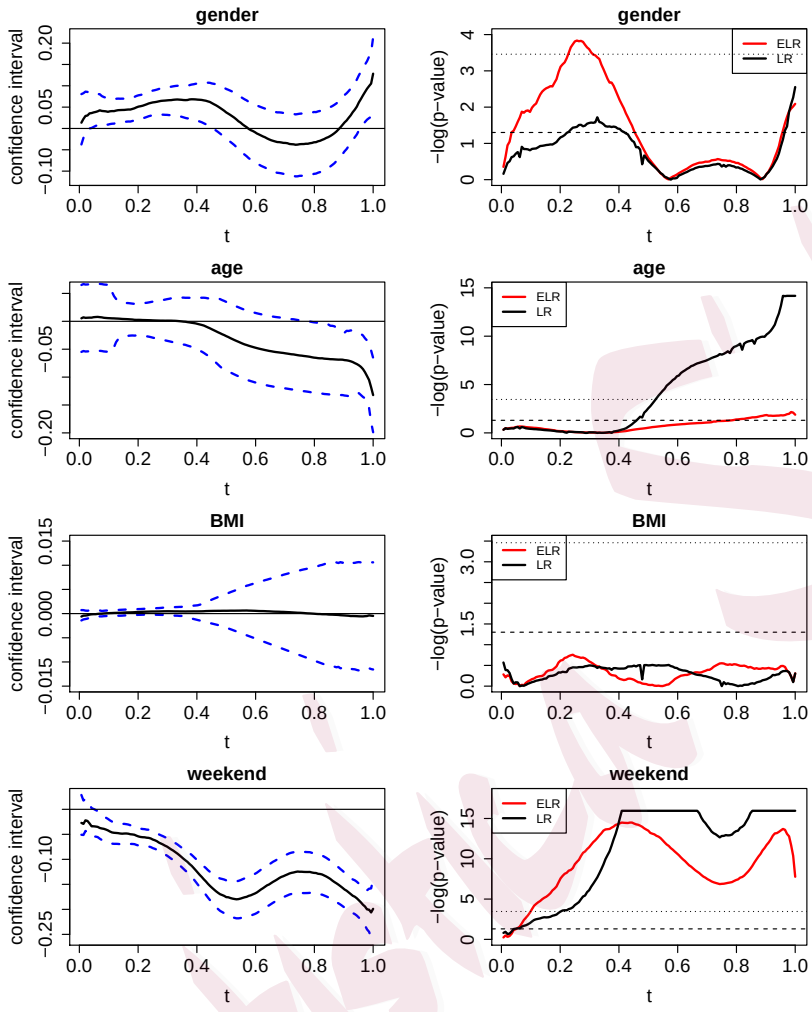


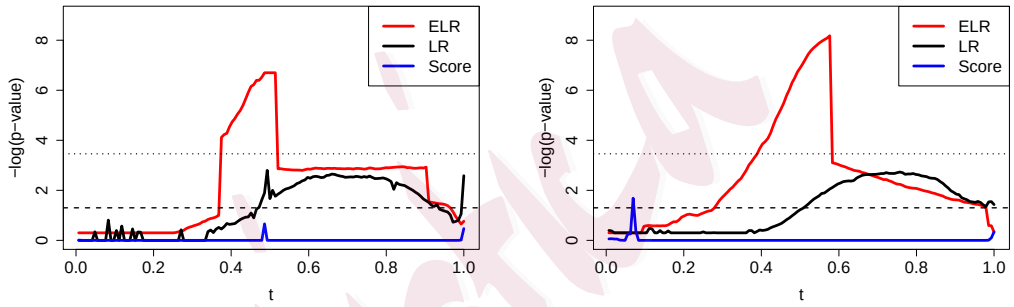
Figure 5: Estimate of $\beta^*(t)$, its 95% confidence interval (left panel) and the $-\log_{10}(p\text{-value})$ (right panel) for gender, age, BMI, and weekend for each of quantile t values. The black dashed horizontal line is the nominal threshold $-\log_{10}(0.05)$, and the black dotted horizontal line is the Bonferroni corrected threshold $-\log_{10}(0.05/144)$.

{Gender, Age, BMI, Weekend}, we apply the ELR test in Section 2 and the standard likelihood ratio (LR) test assuming normal random effects, and we obtain the p-value for each quantile t. The second column of Figure 5 shows the p-values for each t and for each of these four covariates. At the nominal p-value of 0.05, the ELR test shows that there is effect of gender when the activity counts are small (i.e., small t). In contrast, the standard LR test only shows such significance in a smaller interval from 0.23 to 0.42. For age, the ELR shows a significant effect for the large activity counts region (i.e., large t). Both the ELR and LR tests do not reject the null hypothesis that there is no effect of BMI, while the effects of weekend are statistically significant under almost the whole region of t.

6.3 Analysis of heritability of the activity distribution

We then address the question whether the activity distribution is heritable, where the distribution is summarized as the quantiles. This is equivalent to test the null hypothesis $H_0 : \sigma_A^2(t) = 0,\ t \in [0, 1]$. For each quantile t, we first estimate the fixed effects using the least-square estimate and then apply the proposed ELR to test the null hypothesis and to compare the results with the LR and score tests. Figure 6 (a) gives the p-values at different quantiles t. It shows that the test $H_0 : \sigma_A^2(t) = 0$ is rejected for $t \in [0.375, 0.958]$ based

on the ELR test and $t \in [0.472, 0.931]$ using LR, while the score test cannot reject the null hypothesis at the nominal 0.05 significance level. However, if we use the Bonferroni correction for multiple testing, only the proposed ELR test identifies significant heritability for the quantiles between 0.375 and 0.514. The p-value of global test $H_0 : \sigma_A^2(t) \equiv 0,\ t \in [0, 1]$ is 0 when applying the proposed gELR with 1000 permutations. Overall, our analysis shows that the activity distribution is heritable, especially in the quantile range from 0.375 to 0.514.



(a) Outlier removal #1 (b) Outlier removal #2

Figure 6: The $-\log_{10}(p\text{-value})$ of testing heritability $H_0 : \sigma_A^2(t) = 0$ for different quantile values t. The black dashed horizontal line is the nominal threshold $-\log_{10}(0.05)$, and the black dotted horizontal line is the Bonferroni corrected threshold $-\log_{10}(0.05/144)$.

To assess the sensitivity of heritability analysis of the activity distri-

bution, we also consider an alternative method of removing outliers under 3+/-SD rule, which result in a total of 172 healthy twin families, including 73 MZ twin families and 99 DZ twin families, and the total number of observations of 4,477. Among them, we select the twin families with a twin pair and obtain 152 twin families, including 64 MZ twin families and 88 DZ twin families, and the total number of observations of 4,190. The results are very similar (see Figure 6 (b)).

7. Discussion

In this paper, we have developed empirical likelihood methods for making inference of the variance components in general linear mixed-effects models. The proposed ELR test can be applied to a large set of related outcomes such as different quantiles of the activity distribution when we analyze the wearable device data sets. Simulation studies show that the proposed methods control type 1 error much better than likelihood-based or score test when the normality assumptions do not hold. Since the proposed ELR statistic has a closed form expression asymptotically (see Lemma 1), the ELR test can be implemented much more efficiently compared to the likelihood-based methods that require numerical optimizations.

We applied the proposed tests to investigate the heritability of physical

activity measured by actigraph device using the Australian twin study. We observed that such a physical activity is heritable in the quantile range from 37.5% to 51.4%, while the standard likelihood ratio test and the score test under the Gaussian assumption failed to show such a heritability. Our analysis showed that the physical activity is heritable, which warrants further genetic studies.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institute of Mental Health through grant ZIA MH002954-04 [Motor Activity Research Consortium for Health (mMARCH)]. We thank Dr. Hickie and Dr. Martin for sharing the Australian twin study data as part of the mMARCH network and the Genetic Epidemiology Research Branch at National Institute of Mental Health for processing the accelerometry data.

References

Breslow, N. E. and X. Lin (1995). Bias correction in generalised linear mixed models with a single component of dispersion.

Burton, C., B. McKinstry, A. S. Tătar, A. Serrano-Blanco, C. Pagliari, and

M. Wolters (2013). Activity monitoring in patients with depression: a systematic review. *Journal of Affective Disorders 145*(1), 21–28.

Chang, H.-w. and I. W. McKeague (2016). Empirical likelihood based tests for stochastic ordering under right censorship. *Electronic Journal of Statistics 10*(2), 2511.

Chang, H.-w. and I. W. McKeague (2022, 11). Empirical Likelihood-Based Inference for Functional Means with Application to Wearable Device Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(5), 1947–1968.

Chen, S. T., L. Xiao, and A.-M. Staicu (2019). An approximate restricted likelihood ratio test for variance components in generalized linear mixed models. *arXiv preprint arXiv:1906.03320*.

Ge, T., A. J. Holmes, R. L. Buckner, J. W. Smoller, and M. R. Sabuncu (2017). Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *Proceedings of the National Academy of Sciences 114*(21), 5521–5526.

Ghosal, R., V. R. Varma, D. Volfson, I. Hillel, J. Urbanek, J. M. Hausdorff, A. Watts, and V. Zipunnikov (2021, Nov). Distributional data analysis

via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's Disease. *Biostatistics*.

Guo, W., A. Leroux, H. Shou, L. Cui, S. J. Kang, M.-P. F. Strippoli, M. Preisig, V. Zipunnikov, and K. R. Merikangas (2022). Processing of accelerometry data with GGIR in motor activity research consortium for health. *Journal for the Measurement of Physical Behaviour 1*(aop), 1–8.

Koh, H., Y. Li, X. Zhan, J. Chen, and N. Zhao (2019). A distance-based kernel association test based on the generalized linear mixed model for correlated microbiome studies. *Frontiers in genetics 10*, 458.

Krane-Gartiser, K., T. E. G. Henriksen, G. Morken, A. Vaaler, and O. B. Fasmer (2014). Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder. *PloS One 9*(2), e89574.

Li, D. and J. Pan (2013). Empirical likelihood for generalized linear models with longitudinal data. *Journal of Multivariate Analysis 114*, 63–73.

Li, S., T. T. Cai, and H. Li (2021). Inference for high-dimensional linear mixed-effects models: A quasi-likelihood approach. *Journal of the American Statistical Association*, 1–12.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika 84*, 309–326.

Michalski, A. and R. Zmyślony (1996). Testing hypotheses for variance components in mixed linear models. *Statistics 27*(3-4), 297–310.

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 1725–1747.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC.

Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association 82*(398), 605–610.

Wang, S., L. Qian, and R. J. Carroll (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika 97*(1), 79–93.

Xue, L. and L. Zhu (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika 94*(4), 921–937.

You, J., G. Chen, and Y. Zhou (2006). Block empirical likelihood for lon-

gitudinal partially linear regression models. *Canadian Journal of Statistics 34*(1), 79–96.

Zhai, J., K. Knox, H. L. Twigg III, H. Zhou, and J. J. Zhou (2019). Exact variance component tests for longitudinal microbiome studies. *Genetic epidemiology 43*(3), 250–262.

Zhang, D. and X. Lin (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics 4*(1), 57–74.

Zhang, J., K. R. Merikangas, H. Li, and H. Shou (2022). Two-sample tests for multivariate repeated measurements of histogram objects with applications to wearable device data. *The Annals of Applied Statistics 16*(4), 2396 – 2416.

Zhu, H., R. Li, and L. Kong (2012). Multivariate varying coefficient model for functional responses. *The Annals of Statistics 40*(5), 2634–2666.

Zou, F., J. Fine, and B. Yandell (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika 89*(1), 61–75.

School of Data Science, Fudan University, Shanghai 200433, China.

E-mail: jr_zhang@fudan.edu.cn

Genetic Epidemiology Research Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

E-mail: wei.guo3@nih.gov

Brain and Mind Centre, University of Sydney, Camperdown, New South Wales, Australia.

E-mail: joanne.carpenter@sydney.edu.au

Genetic Epidemiology Research Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, U.S.A. and Department of Biostatistics & Informatics, University of Colorado, Aurora, Colorado 80045, U.S.A.

E-mail: andrew.leroux@cuanschutz.edu

Genetic Epidemiology Research Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

E-mail: merikank@mail.nih.gov

QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

E-mail: nick.martin@qimrberghofer.edu.au

Brain and Mind Centre, University of Sydney, Camperdown, New South Wales, Australia.

E-mail: ian.hickie@sydney.edu.au

Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

E-mail: hshou@pennmedicine.upenn.edu

Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

E-mail: hongzhe@pennmedicine.upenn.edu.