Blockwise Resampling for Robust Fixed Effects Inference in Linear Mixed Models

Andreella Angela[1] and Finos Livio[2]

[1] University of Trento, Department of Economics and Management, Trento, Italy
angela.andreella@unitn.it
[2] University of Padova, Department of Statistical Sciences, Padova, Italy

Abstract. Linear mixed models (LMM) are widely used for analyzing clustered data but face challenges with model misspecification, convergence issues, and variance heterogeneity. Alternatively, semi-parametric approaches like generalized estimating equation (GEE) are used to estimate the parameters of a generalized LM in the case of dependent observations. While GEE is robust to covariance misspecification, it is inefficient in handling unbalanced designs and underestimates the true standard errors unless a large sample size is utilized. To address these challenges, we propose a robust extension of the score-based statistical test using sign-flipping transformations. Our approach handles within-variance structure and heteroscedasticity nonparametrically by leveraging whole-block exchangeability. The proposed method provides robust and efficient inference for fixed effects, overcoming the limitations of traditional methodologies, e.g., the specification of the random structure.

Keywords: Longitudinal data, robustness, sign flipping, score test

1 Introduction

Linear mixed models (LMMs) are widely used for modeling clustered data, e.g., [5]. However, they rely on strong parametric assumptions and are sensitive to misspecification due to heteroscedasticity or omitted random and/or fixed effects, leading to inflated type I errors. Another common approach is using the *sandwich estimator* to estimate the fixed effects covariance matrix, as in Generalized Estimating Equations (GEE). While GEE is robust to covariance misspecification and computationally simpler than LMMs, it struggles with unbalanced designs, underestimating standard errors in small samples. LMMs also face convergence issues, particularly in complex random effect structures, high dimensionality, or sparse data, leading to computational challenges and potential inferential biases.

To address these challenges, nonparametric methods such as permutation tests provide a viable alternative for assessing covariate effects. [7] [8] introduced nonparametric tests for LMMs based on permuted residuals. However, these methods have limitations, including the inability to accommodate random slopes and computational inefficiency. As an alternative, [2] proposed an exact permutation test for generalized LMMs. However, its statistical power remains dependent on the quality of the covariance estimation.

To overcome these limitations, we propose a resampling-based statistical test for fixed effects using the score-flipping method [6], extending its applicability to within-subject dependence structures. Building on the sign-flipped score statistical test theory [6] [3], our approach incorporates whole-block exchangeability [13] into LMMs. The method proposed effectively accounts for within-subject correlations, enabling accurate inference in a repeated measurement framework without imposing a random correlation structure as LMMs.

2 Sign-flipping Score Test with Independent Observations

Let $y = \{y_1, \ldots, y_n\}^\top$ be a vector of n independent observations. We focus on:

$$y = X\beta + Z\gamma + \epsilon \tag{1}$$

where $X \in \mathbb{R}^{n\times 1}$ is a design vector of covariates of interest and $\beta \in \mathbb{R}$ the related parameter of interest. $Z \in \mathbb{R}^{n\times q}$ is a design matrix of nuisance covariates and $\gamma \in \mathbb{R}^q$ the related nuisance parameters. Finally, the error term is $\epsilon \sim (0, \Sigma)$.

We report the following assumption from [3]:

Assumption 1 ([3]) *Let* (1) *be the true model that generates the data. We assume that the mean is correctly specified.*

If assumption 1 holds, we have a consistent estimator for β and γ even if the assumed variance function is incorrect [1]. Related to the estimator $\hat{\Sigma}_n$ of Σ, [3] assumes that even if not consistent, it converges in probability as $n \to \infty$ to a strictly positive constant. We are interested in testing:

$$H_0 : \beta = \beta_0 \mid (\gamma, \Sigma) \in \Gamma \times \Phi \quad \text{with} \quad \Gamma \subseteq \mathbb{R}, \Phi \subseteq \mathbb{R}^{>0}. \tag{2}$$

[3] proposed the following asymptotic α level test:

Definition 1 ([3]).

$$S = \frac{\sigma X^\top (I - H)(y - \hat{\mu})}{[X^\top (I - H) X]^{1/2}} \tag{3}$$

where $\sigma^2 = Var(y_i)$, H is the projection matrix and $\hat{\mu}$ the estimated mean.

To compute the null distribution of S illustrated in Definition 1, [3] rely on the permutation theory, i.e., we randomly flip the sign of the residual $(y - \hat{\mu})$ noting that S can be expressed as a sum of n independent components known as the effective score contributions [6]. S represents the inner product of the n vectors $\sigma X^\top (I - H)$ and $(y - \hat{\mu})$. So, we randomly draw W elements from the group of transformation $\mathcal{F}$ composed by all possible sign flipping matrix trasformations $\text{diag}(f_1, \ldots, f_n)$ with f_i takes values $\{-1, 1\}$ with equal probability. Denoting with F_1 the identity transformation and with $F_2, \ldots, F_W$ the realization of W

random sign-flipping transformations, the residuals in (3) are premultiplied by $F_1, \dots, F_W$ leading to $\{S(F_1), \dots, S(F_W)\}$ defined as:

$$S(F_w) = \frac{\sigma X^\top (I-H) F_w (y - \hat{\mu})}{[X^\top (I-H) F_w (I-H) F_w (I-H) X]^{1/2}} \quad w \in \{1, \dots, W\}. \tag{4}$$

At significance level $\alpha \in (0,1)$, we reject (2) against $H_1 : \beta > \beta_0$ if $S_1 > S_{(\lceil (1-\alpha) W \rceil)}$ where $S_{(1)} \leq S_{(2)} \leq \cdots \leq S_{(W)}$ are the sorted statistics and $\lceil \cdot \rceil$ is the ceiling function.

[3] proved that under certain mild assumptions, the standardized sign-flip score statistic described in (3) is as $n \to \infty$ asymptotically second-moment null-invariant [12]. The test is then asymptotically exact, i.e., as $n \to \infty$ under H_0, the probability of rejection converges to $\lfloor \alpha W \rfloor / W \leq \alpha$ [11].

3 Extension to the Non-independent Case

Consider n observations $y_1, \dots, y_n$ with a correlation structure dictated by the longitudinal data nature. We have then n_j observations in group j with $n = \sum_j^N n_j$, and N is the total number of clusters. We assume that observations within each cluster are dependent and introduce random effects for both the covariate of interest and the nuisance ones in (1). The model is then expressed as follows:

$$y_{ij} = X_{ij}\beta + Z_{ij}\gamma + X_{ij}u_j + Z_{ij}g_j + b_j + \epsilon_{ij} \tag{5}$$

where $\mathbb{E}(u_j) = \mathbb{E}(g_j) = \mathbb{E}(b_j) = 0$.

The approach proposed by [3] is based on Assumption 1. In our context, if the score function of the adopted model still leads to an unbiased estimating equation for β, then the maximum likelihood estimator remains consistent [10]. We rephrase Assumption 1 accordingly.

Assumption 2 *Let* (5) *be the true model which generates the data. We assume the estimating equation for β under the marginal model is unbiased.*

In other words, Assumption 2 allows us to focus solely on the variance specification of the misspecified model. Moreover, it implies that β retains the same interpretation in both the marginal and conditional models. This assumption generally does not hold when the link function differs from the identity.

Next, we demonstrate that using the estimator from the marginal model (1) instead of the conditional model (5) still results in a consistent estimator.

Theorem 1. *Let* (5) *be the true model that generates the data. The estimator of β, denoted as $\hat{\beta}_n$, coming from the misspecified (marginal) model defined in Equation* (1)*, is a consistent estimator for β defined in model* (5)*.*

Accounting for within-cluster dependence is essential for obtaining unbiased variance estimates and valid inference. Failure to properly address intra-cluster correlation, as commonly seen in traditional methods such as LMMs and GEEs, can lead to biased variance estimates—either inflated or deflated depending on the correlation structure. This bias propagates into inflated type I error rates, undermining the reliability of hypothesis testing [9]. The sign-flipping approach proposed by [3] is robust in this regard because it avoids direct variance estimation, bypassing reliance on the Fisher information matrix. This distinction frees the method from the limitations of traditional approaches that depend on accurate variance estimation. Consequently, the sign-flipping method ensures valid type I error rates and reliable inference without explicitly specifying the random structure of the model.

The following theorem states that the test using appropriate sign-flipping matrices is an asymptotic α-level test for testing (2). For the assumption in this theorem to hold, we require $\hat{\gamma}$ to be $\sqrt{n}$-consistent, e.g., $\max_{j \in \{1,\dots,N\}} n_j = o(n)$ or more strictly $\sup_{j \in \{1,\dots,N\}} n_j < \infty$ [9].

Theorem 2. *Let* (5) *be the true model that generates the data and* (3) *the null hypothesis of interest. If the sign-flipping matrix F is defined such that $F_{ii} = F_{kk}$ whenever y_i and y_k are in the same block, S_1^n is an asymptotic α level test. Then, as $N \to \infty$, the test level converges to $\lfloor \alpha W \rfloor / W \leq \alpha$.*

Theorem 2 also accounts for unbalanced designs, i.e., $n_j \neq n_{j'}$ for $j \neq j'$.

This result leverages the exchangeability assumption to construct a valid null distribution. It involves restricting the set of all possible permutations to those that preserve the relationships between observations [11]. To properly address within-cluster variance, we define the sign-flipping matrix as a block matrix:

$$\tilde{F} = \begin{bmatrix} \tilde{F}_1 & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{F}_2 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \mathbf{0} & \tilde{F}_j & \mathbf{0} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} & \tilde{F}_{N-1} & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} & \tilde{F}_N \end{bmatrix}. \tag{6}$$

Here, each $\tilde{F}_j$ $(1 \leq j \leq N)$ is defined as $\tilde{F}_j = cI_{n_j}$, where I_{n_j} is the identity matrix of size $n_j \times n_j$, and $c \in \{1, -1\}$ is randomly sampled. If all clusters have the same size, i.e., $n_j = n_{j'}$ for all j, j', then $\tilde{F}$ simplifies to $F \otimes I_{n_1}$ where F is a sign-flipping matrix of size $N \times N$

The set $\tilde{\mathbf{F}}$, containing the block sign-flipping transformations $\tilde{F}$ defined in (6), replaces the full set of sign-flipping transformations $\mathcal{F}$ when computing the standardized sign-flip score statistics $S(\tilde{F})$. Since the cardinality of $\tilde{\mathbf{F}}$ depends on N, the attainable α values also depend on N rather than n.

4 Simulation

We compare the proposed method (defined as flipscores) with LMM and GEE in terms of type I error control and power. The dependent variable y is simulated

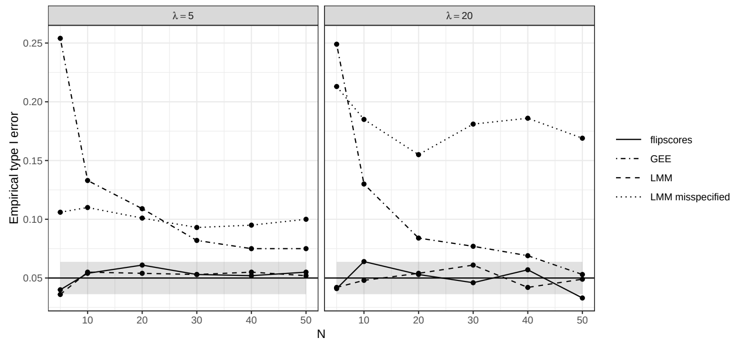

Fig. 1. Estimated type I error considering $N \in \{10, 20, 30, 40, 50\}$ number of clusters and $n_j \sim \text{Uniform}(3, \lambda)$ with $\lambda \in \{5, 20\}$ repeated measurements. Each line represents one model, and the grey area around the solid horizontal black line represents the 0.95 confidence bound for $\alpha = 0.05$.

as a normal random variable following model (5) with $q = 1$. The covariates X and Z are generated from a bivariate standard normal distribution with a correlation of 0.7. Within-subject random effects are drawn from a multivariate normal distribution with mean 0 and an equicorrelation structure with $\rho = 0.5$. The goal is to test $H_0 : \beta = 0$ under this correlation structure. We set $\beta = 0$ to evaluate type I error control and $\beta = 2$ to assess the power of the approaches. The nuisance parameter is fixed at $\gamma = 2$. In both scenarios, 1000 simulations are performed with $\alpha = 0.05$ and $W = 1000$.

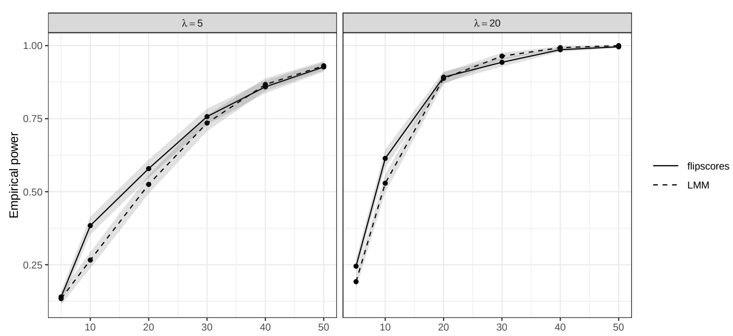

Fig. 2. Estimated power considering $N \in \{10, 20, 30, 40, 50\}$ number of clusters and $n_j \sim \text{Uniform}(3, \lambda)$ with $\lambda \in \{5, 20\}$ repeated measurements. Each line represents one model with corresponding shaded confidence bounds at level 0.95.

Figures 1 and 2 show the estimated error rate and power, respectively, considering $N \in \{10, 20, 30, 40, 50\}$ and $n_j \sim \text{Uniform}(3, \lambda)$ where $\lambda \in \{5, 20\}$. In the misspecified LMM, the random slopes are not included in the model. This highlights that LMM performance depends on correctly specifying the random structure, whereas GEE fails to control type I errors for small sample sizes. In contrast, the proposed approach successfully controls type I error while achieving empirical power levels comparable to those of LMM.

5 Further Research

The proposed method is highly flexible and can be extended to settings with multiple fixed effects and multiple dependent variables [4]. By leveraging the outcomes correlation structure through permutation theory, our approach provides greater statistical power than traditional parametric methods, e.g., LMMs with Bonferroni correction. Furthermore, this framework can be adapted to accommodate complex correlation structures of random effects, e.g., crossed designs.

References

1. Agresti, A. (2015). *Foundations of linear and generalized linear models.* John Wiley & Sons.
2. Basso, D. and Finos, L. (2012). Exact multivariate permutation tests for fixed effects in mixed-models. *Communications in Statistics-Theory and Methods*, 41(16-17):2991–3001.
3. De Santis, R., Goeman, J. J., Hemerik, J., and Finos, L. (2024). Inference in generalized linear models with robustness to misspecified variances. *arXiv:2209.13918 [stat.ME].*
4. De Santis, R., Goeman, J. J., Davenport, S., Hemerik, J., and Finos, L. (2024). Permutation-based multiple testing when fitting many generalized linear models. *arXiv:2403.02065 [math.ST].*
5. Epifania, O., Anselmi, P., and Robusto, R. (2024). A guided tutorial on linear mixed-effects models for the analysis of accuracies and response times in experiments with fully crossed design. *Psychological Methods.*
6. Hemerik, J., Goeman, J. J., and Finos, L. (2020). Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):841–864.
7. Kherad-Pajouh, S. and Renaud, O. (2010). An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova. *Computational Statistics & Data Analysis*, 54(7):1881–1893.
8. Lee, O. E. and Braun, T. M. (2012). Permutation tests for random effects in linear mixed models. *Biometrics*, 68(2):486–493.
9. Liang, K.-Y. and Zeger, S. L. (1993). Regression analysis for correlated data. *Annual review of public health*, 14(1):43–68.
10. Pace, L. and Salvan, A. (1996). *Teoria della Statistica: metodi, modelli, approssimazioni asintotiche.* Cedam.
11. Pesarin, F. (2001). *Multivariate permutation tests: with applications in biostatistics*, volume 240. Wiley Chichester.
12. Solari, A., Finos, L., and Goeman, J. J. (2014). Rotation-based multiple testing in the multivariate linear model. *Biometrics*, 70(4):954–961.
13. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, 92:381–397.