Technical Report RM 706, Department of Statistics and Probability, Michigan State University

Kenward-Roger approximation for linear mixed models with missing covariates

Akshita Chawla[1,*], Tapabrata Maiti[1] and Samiran Sinha[2]

[1]Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA
[2]Department of Statistics, Texas A&M University, College Station, TX, USA
[*]*email:* chawlaak@stt.msu.edu

Abstract

Partially observed variables are common in scientific research. Ignoring the subjects with partial information may lead to a biased and or inefficient estimators, and consequently any test based only on the completely observed subjects may inflate the error probabilities. Missing data issue has been extensively considered in the regression model, especially in the independently identically (IID) data setup. Relatively less attention has been paid for handling missing covariate data in the linear mixed effect model– a dependent data scenario. In case of complete data, Kenward-Roger's F test is a well established method for testing of fixed effects in a linear mixed model. In this paper, we present a modified Kenward-Roger type test for testing fixed effects in a linear mixed model when the covariates are missing at random. In the proposed method, we attempt to reduce bias from three sources, the small sample bias, the bias due to missing values, and the bias due to estimation of variance components. The operating characteristics of the method is judged and compared with two existing approaches, listwise deletion and mean imputation, via simulation studies.

Keywords: Fixed effects; Kenward-Roger test; Likelihood ratio test; Linear mixed models; Missing covariates; REML; Small sample.

1 Introduction

Linear mixed model (Laird & Ware, 1982) is a popular technique to deal with correlated data such as longitudinal data. The model parameters are typically estimated using ML or REML estimates (Patterson & Thompson, 1971 and Harville, 1977). For complete data, to draw inferences on fixed effects of the model, many testing procedures have been proposed among which likelihood ratio test and Wald tests are the most commonly used methods. However, it is established (Leeper & Chang, 1992; Zucker et al. 2000) that these methods do not yield accurate results under small sample conditions. Any method that ignores the uncertainty of the estimated variance component in the inference of the fixed effect parameter $\boldsymbol{\beta}$ will result in deflated confidence interval, specially when sample size is small. The first problem of approximating the small sample precision of $\widehat{\boldsymbol{\beta}}$ has been studied by Kackar & Harville (1984), Harville (1985) and Harville & Jeske (1992). In order to address the second problem of underestimation of variance of $\widehat{\boldsymbol{\beta}}$, the estimator of $\boldsymbol{\beta}$, an approximate t or an F-statistic is used. In general, to test the hypothesis $H_0 : L^T\boldsymbol{\beta} = 0$, a Wald-type test statistic is used which is approximated as an F distribution with the numerator degrees of freedom as rank(L) and the denominator degrees of freedom being estimated from the data. The degrees of freedoms were used to be estimated by the residual method, containment method or the Satterthwaite-type approximation. Later Kenward & Roger (1997) addressed both the above mentioned issues. They proposed a scaled Wald statistic which involved an approximate covariance matrix, hence accounting for the variability introduced due to estimation of the variance components. Further, they showed that its small sample distribution can be approximated by an F-distribution with the denominator degrees of freedom obtained from the data using similar approximations as in Satterthwaite (1941). This approach has been established to perform exceptionally well under small sample conditions as compared to the other testing procedures. Alnosaier (2007) and Gregory (2011) illustrated the superiority of the Kenward Roger method as compared to the Satterthwaite and containment methods.

Partially missing variable is common in clinical studies. The causes of missingness could vary; for example, subjects may refuse to provide some information, some observations may not be recorded due to drop out or it could simply happen due to manual error. Ignoring the subjects with partially missing observation may distort statistical decision, especially when the sample size is small and or the proportion of observations with partially missing values is high. Depending on the missingness mechanishm missing data can be classified into three main categories, missing completely at random (MCAR) where the missingness mechanism is completely random and does not depend on any variables, missing at random (MAR) where missingness mechanism depends only on the observed values of the variables, and missing not at random (MNAR) where the missingness mechanism may depend on the unobserved values of the variables along with the completely observed variables. The first two mechanisms are ignorable in a likelihood framework while the third is not. Missing values may occur in a response variable, in a covariate or in both. There are several methods of dealing with missing responses depending on the missing mechanism. Among the most common methods are listwise deletion where the subject with any missing value is deleted from the study, mean or multiple imputation and the maximum likelihood method that do not involve modelling the missing mechanism when the missing mechanism is either MCAR or MAR. Padilla and Algina (2004) showed that when the responses are missing (MCAR or MAR) in a small sample setup, Kenward-Roger test preserves the Type I error rates for a single factor within-subject ANOVA as compared to the Hotelling-Lawley-McKeon procedure (McKeon, 1974). Therefore, Kenward Roger F test is proven to be superior even in case of missing responses. In several aricles, Ibrahim and coworkers considered partially missing response or covariate in regression models (Ibrahim, 1990; Ibrahim & Chen, 2001; Stubbendick & Ibrahim, 2003; Ibrahim et al., 2005; Chen et al., 2008). However, little has been done in case of missing covariates in a linear mixed model.

In this paper, we only deal with a missing covariate data, and assume that the covariate data are missing at random (MAR). We assume a parametric distribution on the partially missing covariate. Under this set up, we derive a Kenward-Roger type adjusted test for the fixed effects in a linear mixed model in case of small samples. To derive a robust test, we need to overcome three biases: the variance bias (such that the variability of the variance components is taken into account), the small sample bias and the bias due to missingness. We consider the bias in the estimation of the covariance matrix of $\widehat{\boldsymbol{\beta}}$ and then propose a new Wald statistic that uses this new adjusted covariance matrix. Further, this Wald statistic is approximated as an F distribution with the degrees of freedom calculated using the new covariance matrix. To demonstrate the accuracy of this method, we compare this method to the listwise deletion and imputation methods through simulation studies.

The rest of the paper is organized as follows. Models and assumptions are given in Section 2.

Section 3 contains the Kenward Roger type F test by approximating the new covariance matrix and the denominator degrees of freedom. Section 4 contains the simulation study that assesses the performance of the proposed method and compare it to other existing methods based on Type-I error rates. We have applied the proposed method to a real data and the details are collected in Section 5, followed by concluding remarks given in Section 6.

2 Model and assumptions

Consider a linear mixed model with m groups and n_i measurements in the i^{th} group. Denote $n = \sum_{i=1}^{m} n_i$ the total number of observations. For each group, we observe an $n_i \times 1$ vector of responses $\mathbf{Y}_i$, let X_i be an $n_i \times p$ fixed-effects matrix, Z_i be an $n_i \times q$ random-effects design matrix, $\boldsymbol{\beta}$ be a $p \times 1$ vector of fixed-effect coefficients and $\mathbf{b}_i$ be a group-specific vector of random regression coefficients. The model can be written as

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\epsilon}_i,\ i = 1, \cdots, m,$$

assuming that $\boldsymbol{\epsilon}_i \sim N_{n_i}(0, \sigma_e^2 I_{n_i})$, $\mathbf{b_i} \sim N_q(0, V)$ and $\boldsymbol{\epsilon}_i$ and $\mathbf{b_i}$ are independent. We can deduce that $\mathbf{Y}_1, \cdots, \mathbf{Y}_m$ are independent and $\mathbf{Y}_i \sim N_{n_i}(X_i\boldsymbol{\beta}, \Sigma_i)$ with $\Sigma_i = Z_i V Z_i^T + \sigma_e^2 I_{n_i}$. Denote the stacked vectors $\mathbf{Y} = (\mathbf{Y}_1^T, \cdots, \mathbf{Y}_m^T)^T$, $\mathbf{b} = (\mathbf{b}_1^T, \cdots, \mathbf{b}_m^T)^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \cdots, \boldsymbol{\epsilon}_m^T)^T$ and the stacked matrices $X_{n\times p} = (X_1^T, \cdots, X_n^T)^T$, $Z_{n\times mq} = diag(Z_1, \cdots, Z_m)$ and $\Sigma_{n\times n} = diag(\Sigma_1, \cdots, \Sigma_n)$. Then the model can be written as

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^{\star},$$

where $\boldsymbol{\epsilon}^{\star} = Z\mathbf{b} + \boldsymbol{\epsilon}$ such that $\boldsymbol{\epsilon}^{\star} \sim N_n(0, \Sigma(\boldsymbol{\sigma}))$. The covariance matrix Σ is a function of $1 + q(q+1)/2$ parameters, where the first parameter is the variance of the model errors i.e., σ_e^2 and the rest $q(q+1)/2$ parameters characterize the random effect covariance matrix V. These parameters are represented by the vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_r)^T$ where $r = 1 + q(q+1)/2$. Let $\mathbf{X_{(1)}} = (X_{1(1)}, X_{2(1)}, \cdots, X_{n(1)})^T$ be the first column of the covariate matrix X. Then $X_{n\times p}$ can be written as $X = (\mathbf{X_{(1)}} : X_{(-1)})$ where $\mathbf{X_{(1)}}$ is of dimension $n \times 1$ and $X_{(-1)}$ is a $n \times (p-1)$ matrix of the covariates other than the first. Without loss of generality, assume that $X_{(1)}$ contains partially missing covariates, and define the missing indicator as

$$B_j = \begin{cases} 0 & \text{if } X_{j(1)} \text{ is missing} \\ 1 & \text{otherwise,} \end{cases}$$

for $j = 1, \cdots, n$. Assume that the partially missing covariate $\mathbf{X_{(1)}}$ follows a parametric model $f(\mathbf{X_{(1)}}|\mathbf{X_{(-1)}}, \boldsymbol{\gamma})$ that is known upto a finite dimensional parameter $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_k)^T$. In the presence of missing data, we write

$$\mathbf{Y} = X^{\star}\boldsymbol{\beta} + \boldsymbol{\epsilon}^{\star},$$

where $X^{\star}_{n\times p} = (\mathbf{X^{\star}}_{(1)} : X_{(-1)})$, $\mathbf{X^{\star}_{(1)}} = (X_{1(1)}, \cdots, X_{n(1)})^T$ such that,

$$X^{\star}_{j(1)} = \begin{cases} E(\mathbf{X_{j(1)}}|\mathbf{X_{j(-1)}}, \boldsymbol{\gamma}) = \mu(\boldsymbol{\gamma}) & \text{if } B_j = 0 \\ X_{j(1)} & \text{if } B_j = 1, \end{cases}$$

for $j = 1, \cdots, n$.

Now we can estimate $\boldsymbol{\gamma}$ by maximizing the likelihood

$$L = \prod_{j=1}^{n} f^{B_j}(X_{i(1)}|X_{j(-1)}\boldsymbol{\gamma}). \tag{1}$$

The standard error of the parameter estimates can be determined from the Hessian matrix. We will be working with the following model setup in the rest of the paper. Now the n-dimensional $\mathbf{Y}$ follows a multivariate normal distribution,

$$\mathbf{Y} \sim \text{Normal}(X^{\star}\boldsymbol{\beta}, \Sigma),$$

where $X^{\star}$ is an $n \times p$ matrix of known/estimated covariates, assumed to be of full rank. $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and Σ is an unknown variance covariance matrix whose elements are functions of $\boldsymbol{\sigma}$=$(\sigma_1, \cdots, \sigma_r)^T$ which in turn is a function of $\boldsymbol{\gamma}$=$(\gamma_1, \cdots, \gamma_k)^T$.

The REML based estimated least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = X^{\star T}(\widehat{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\gamma}))X^{\star}(\widehat{\gamma})^{-1}X^{\star T}(\widehat{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\gamma}))\mathbf{Y},$$

where $\widehat{\boldsymbol{\sigma}}(\widehat{\gamma})$ is the REML estimator of $\boldsymbol{\sigma}$. In our study, $\boldsymbol{\beta}$ is the parameter of interest and $\Sigma(\boldsymbol{\sigma}(\boldsymbol{\gamma}))$ is the nuisance parameter. We will test the hypothesis $H_0 : L^T\boldsymbol{\beta} = 0$, where L is a $p \times l$ matrix.

3 Inference for the fixed effects

3.1 Estimation of the variance of the estimator $\widehat{\boldsymbol{\beta}}$

In order to make inferences about fixed effects such as confidence interval estimation and testing of hypothesis, we need to define a test statistic and compute its distribution under small sample situation. Under the linear mixed model framework, the variance of $\widehat{\boldsymbol{\beta}}$ is given by $\widehat{\Phi} = \Phi(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})) = X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X(\widehat{\boldsymbol{\gamma}})^{-1}$, where $\widehat{\boldsymbol{\gamma}}$ is the estimator obtained after maximizing the likelihood given in (1) and $\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}})$ is the REML estimate of $\boldsymbol{\sigma}$ obtained after using the imputed data set. There are three sources of bias when $\widehat{\Phi}$ is used as an estimator for the variance of $\widehat{\boldsymbol{\beta}}$ for small samples: $\widehat{\Phi}$ is a biased estimator of $\Phi(\boldsymbol{\sigma}(\boldsymbol{\gamma}))$, Φ does not take the variability of $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\gamma}}$ into account. We consider a better approximation to the small sample covariance matrix of $\widehat{\boldsymbol{\beta}}$ under missing covariates, which accounts for the variability in both $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\gamma}}$, thus reducing the small sample bias and any bias due to the missingness. For complete data, Kackar and Harville (1984) addressed the second source of bias by partitioning the variance of the estimated $\boldsymbol{\beta}$ into two components: $\Phi + \Lambda$. Further Kenward and Roger (1997) addressed the first source of bias and combined both adjustments thus proposing a new estimator of Φ,

$$\widehat{\Phi}_{KR} = \widehat{\Phi} + \widehat{\Lambda},$$

where the estimators of Φ and Λ are such that

$$E(\widehat{\Phi}) = \Phi - \widetilde{\Lambda} + R^* + O(n^{-5/2}), \tag{2}$$

$$E(\widehat{\Lambda}) = \widetilde{\Lambda} + O(n^{-5/2}), \tag{3}$$

with

$$\Phi = \left(X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X^{\star}(\boldsymbol{\gamma})\right)^{-1}, \tag{4}$$

$$\widetilde{\Lambda} = \sum_{l=1}^{r}\sum_{m=1}^{r} \mathrm{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi(Q_{lm} - P_l \Phi P_m)\Phi, \tag{5}$$

and $O(n^r)$ denotes $O(n^r)/n^r$ is a bounded number as $n \to \infty$, for some r. In the above expressions

$$
\begin{aligned}
P_l &= -X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial \Sigma}{\partial \sigma_l}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X^{\star}(\boldsymbol{\gamma}),\\
Q_{lm} &= X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial \Sigma}{\partial \sigma_l}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial \Sigma}{\partial \sigma_m}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X^{\star}(\boldsymbol{\gamma}),\\
R_{lm} &= X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial^2}{\partial \sigma_l \partial \sigma_m}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X^{\star}(\boldsymbol{\gamma}),\\
R^* &= \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r}\mathrm{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi R_{lm}\Phi.
\end{aligned}
$$

We now consider the bias in $\widehat{\Phi}$ as an estimator of $\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))$ thus adjusting for missingness.

Theorem 1 *Under the assumptions given in Appendix A.1, the adjusted variance of* $\widehat{\boldsymbol{\beta}}$ *can be approximated as*

$$\widehat{\Phi}_A = \widehat{\Phi} + \widehat{\Lambda} + \widehat{\Psi}.$$

The estimators of Φ, Λ *and* Ψ *are given by* $\widehat{\Phi}$, $\widehat{\Lambda}$ *and* $\widehat{\Psi}$ *respectively such that (2)-(5) hold and* $E(\widehat{\Psi}) = \Psi + O(n^{-2})$, *where*

$$\Psi = \sum_{i=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i var(\widehat{\gamma}_i)K_i^T\Phi^T(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i cov(\widehat{\gamma}_i, \widehat{\gamma}_j)K_i^T\Phi^T(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})).$$

Proof: The REML based estimator of $\boldsymbol{\beta}$ using the imputed data set is given by

$$\widehat{\boldsymbol{\beta}} = X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star}(\widehat{\boldsymbol{\gamma}})^{-1}X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))\mathbf{Y}. \qquad (6)$$

The covariance of $\widehat{\boldsymbol{\beta}}$ given in (6) can be approximated as $\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \mathrm{var}\{H(\widehat{\boldsymbol{\sigma}})\} + \mathrm{var}(\widehat{\boldsymbol{\beta}}) = \Psi + \Phi + \Lambda$. The estimators of Ψ, Φ and Λ are derived in Appendix A2.

3.2 Approximating the distribution of the test statistic

Suppose we are interested in making inferences about l linear combinations of the elements $\boldsymbol{\beta}$. In other words, we are interested in testing $H_0 : L^T\boldsymbol{\beta} = 0$ where L^T is a fixed matrix of dimension

$(l \times p)$. A common statistic to test H_0 is the Wald statistic given by

$$F = \frac{1}{l}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T \widehat{\Phi}_A L)^{-1} L^T (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{7}$$

where $\widehat{\Phi}_A$ is an adjusted covariance matrix for $\widehat{\boldsymbol{\beta}}$. We will follow a similar procedure as used in Kenward & Roger (1997) and Alnosaier (2007) by approximating the first two moments of the F statistic given in (7) and match the moments to approximate the scaling factor and the denominator degrees of freedom of the test statistic i.e., λ and d such that $\lambda F \sim F(l, d)$ in distribution. Here $F(l, d)$ denotes F-distribution with degrees of freedom (l, d).

Theorem 2 *Under the assumptions given in (A.1), the first two moments of the test statistic (7) under* $H_0 : L^T\boldsymbol{\beta} = 0$ *are*

$$\begin{aligned} E(F) &= 1 + \frac{A_2}{l} - \frac{A_4}{l} + O(n^{-3/2}), & (8)\\ var(F) &= \frac{A_1}{l^2} + \frac{2}{l} + \frac{6A_2}{l^2}, & (9) \end{aligned}$$

where $\widehat{\Theta} = L(L^T\widehat{\Phi}L)^{-1}L^T$, $A_1 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace(\widehat{\Theta}\widehat{\Phi}P_l\widehat{\Phi}) trace(\widehat{\Theta}\widehat{\Phi}P_m\widehat{\Phi})$,
$A_2 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace(\widehat{\Theta}\widehat{\Phi}P_l\widehat{\Phi}\widehat{\Theta}\widehat{\Phi}P_m\widehat{\Phi})$,
$A_3 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace\{\widehat{\Theta}\Phi(Q_{lm} - P_l\widehat{\Phi}P_m - R_{lm}/4)\}$,
$A_4 = \sum_{l=1}^{r}\sum_{m=1}^{r} cov(\widehat{\sigma}_l, \widehat{\sigma}_m) trace(\widehat{\Theta}\widehat{\Psi})$,
and using (8) and (9), we get $\widetilde{d} = 4 + (2 + l)/(l\widetilde{\rho} - 1)$, *where* $\widetilde{\rho} = \widetilde{V}/2\widetilde{E}^2$, *and* $\widetilde{\lambda} = \widetilde{d}/\{\widetilde{E}(\widetilde{d} - 2)\}$, *where* $\widetilde{E}$ *and* $\widetilde{V}$ *are defined in Appendix A3.*

Proof: The expectation and variance are approximated using the following conditional arguments that $E(F) = E\{E(F|\widehat{\boldsymbol{\sigma}})\}$ and $\text{var}(F) = E\{\text{var}(F|\widehat{\boldsymbol{\sigma}})\} + \text{var}\{E(F|\widehat{\boldsymbol{\sigma}})\}$. After using the Taylor series expansion we obtain

$$E(F) = \widetilde{E} + O(n^{-3/2}), \text{ and } \text{var}(F) = \widetilde{V} + O(n^{-3/2}),$$

where the explicit expressions of $\widetilde{E}$ and $\widetilde{V}$ are derived in Appendix A.3. In order to determine λ and d, we match the first two moments of λF with those of $F(l,d)$ i.e.,

$$E\{F(l,d)\} = \frac{d}{d-2}, \text{ and } \mathrm{var}\{F(l,d)\} = 2\left(\frac{d}{d-2}\right)^2 \frac{l+d-2}{l(d-4)},$$

to obtain the expressions for $\widetilde{d}$ and $\widetilde{\lambda}$.

4 Simulation

Simulation design: We considered a linear mixed model $Y = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}$, where $\mathbf{Y}$ is a $n \times 1$ response vector, X is a $n \times p$ fixed covariate matrix, Z is a $n \times q$ random covariate matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\mathbf{b}$ is the $q \times 1$ vector of random effects that follows a $N_q(0, V)$ and $\boldsymbol{\epsilon}$ is the $n \times 1$ error vector that follows a $N(0, \sigma_e^2)$. Also, $\mathbf{b}$ and $\boldsymbol{\epsilon}$ are independent. The model can be rewritten as $Y = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^\star$, where $\boldsymbol{\epsilon}^\star \sim N(0, \Sigma)$ with $\Sigma = \sigma_e^2 I_n + ZVZ^T$.

In our simulation study, we considered a block design with one random effect ($q = 1$), whose variance is σ_b^2. Here we considered $p = 5$. To simulate the data, we generated the covariate matrix X from a multivariate normal distribution with mean zero and covariance matrix whose $(i,j)^{th}$ element is given by $|0.2|^{i-j}$. We introduced 20% and 40% missingness in one of the columns of X (i.e., missingness in one covariate) using a bernoulli distribution i.e. assuming the missing mechanism to be MCAR. To relax the MCAR assumption, we considered another case to mimic the MAR mechanism where we introduced missingness in one of the covariates such that the probability of an observation being missing depends on the value of another covariate. Then we maximized the observed likelihood of the partially missing covariate to get the estimate of the mean and the variance, and replaced the missing values in the covariate matrix by the estimated mean, thus obtaining the imputed data set. Further, we deployed our method on the imputed data set to test the hypothesis $H_0 : \boldsymbol{\beta} = 0$ vs $H_a : \boldsymbol{\beta} \neq 0$. To test this, we generated $\mathbf{Y}$ from $N(0, \Sigma)$ for different values of σ_b^2/σ_e^2 ranging from 0.25 to 4. For each setting we generated the data 5000 times and computed the simulated size (Type-I error probabilities). Tables 1 and 2 provide the simulated size when the missing mechanism is MCAR for sample sizes 15 and 30 respectively, Table 3 provides the results for the MAR mechanism.

In practice, the distribution of the partially missing covariate is not known. Therefore, to check the robustness of our procedure, we generated the partially missing covariate from heavy tailed distributions like the Laplace distribution with mean parameter as zero and scale parameter as eight, and the Student's t-distribution with one degree of freedom. The corresponding results are given in Table 4.

Implementation: To be able to run the simulation, we need to calculate the following terms: $\partial X^\star(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$ and $\partial\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma}$. The first term is $\partial X^\star(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma} = I(X^\star = \mu(\boldsymbol{\gamma}))$, and using the chain rule of of derivative the second term is $\partial\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma} = (\partial\Sigma/\partial\widehat{\sigma}_e^2)\times(\partial\widehat{\sigma}_e^2/\partial\boldsymbol{\gamma})+(\partial\Sigma/\partial\widehat{\sigma}_b^2)\times(\partial\widehat{\sigma}_b^2/\partial\boldsymbol{\gamma})$. The detailed expression of these terms are given in the Appendix A4.

In our simulation study, we will compare our testing procedure i.e., Kenward Roger adjustment for missing covariates (KRM) to the following methods which are used to deal with missing covariates in a linear mixed model setup on the basis of simulated size: 1) listwise deletion followed by likelihood ratio test (LD-LRT), 2) mean imputation followed by likelihood ratio test (MI-LRT), 3) mean imputation followed by Kenward Roger F-test with no adjustment for missingness (MI-KRT).

Results: From the results in Tables 1- 4, we observe that listwise deletion of observations does not work well since by introducing missingness we are making the sample size even smaller. Mean imputation seems a viable option since it provides a complete data to work with. But the standard likelihood ratio (MI-LRT) still does not work well, again due to small sample sizes. Kenward Roger method after using the imputed data set without adjusting for missingness (MI-KRT) performs reasonably well. We also compared Wald type tests using three different covariance matrices of $\widehat{\boldsymbol{\beta}}$ namely (i) $\widehat{\Phi}$ which does not take the variability of $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\gamma}}$ into account and (ii) $\widehat{\Phi}_{KR}$ which takes the variability in only $\widehat{\boldsymbol{\sigma}}$ into account. In these two cases, no approximation is made to the denominator degrees of freedom of F distribution (we use the residual degrees of freedom). For the sake of brevity and to avoid redundancy, the results from methods (i) and (ii) are not shown in this paper. Finally, our proposed method (KRM) outperforms the other procedures for almost all values of ρ, n and the type and percentage of missingness. Table 4 shows that our testing procedure is robust. We also tested for single fixed effect parameter i.e. $H_0 : \beta_1 = 0$ (results not shown here) in which case the KRM and MI-KRT perform equally well.

5 Real data example

We consider the data from a randomized, double-blind, study of AIDS patients with advanced immune suppression (CD4 counts of less than or equal to 50 $cells/mm^3$). The data description is as in the datasets section in Fitzmaurice (2004): Patients in AIDS Clinical Trial Group (ACTG) Study 193A were randomized to dual or triple combinations of HIV-1 reverse transcriptase inhibitors. Specifically, patients were randomized to one of four daily regimens containing 600mg of zidovudine: zidovudine alternating monthly with 400mg didanosine; zidovudine plus 2.25mg of zalcitabine; zidovudine plus 400mg of didanosine or zidovudine plus 400mg of didanosine plus 400mg of nevirapine (triple therapy). The patients characteristics like age and sex were also recorded. The response variable i.e. measurements of CD4 counts were scheduled to be collected at baseline and at 8-week intervals during follow-up. However, due to skipped visits and dropouts, the measurements could not be taken at regular intervals therefore we have the weeks as another variable. For our use, we use the log transformed CD4 counts log(CD4counts+1) as the response variable. The data is available for 1313 patients, however, since we are testing for small sample studies, we will truncate the data by randomly choosing 40 patients from the study. We introduce 30% missingness in the covariate age and deploy the discussed methods on the obtained data. The model under consideration is

$$Y_{ij} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 t_{ij} + \beta_{20} G_{2i} + \beta_{30} G_{3i} + \beta_{40} G_{4i} + b_{0i} + \epsilon_{ij} \tag{10}$$

with $i = 1, 2, \cdots, 40$ and $j = 1, \cdots, n_i$ n_i varying from 1 to 9, where Y_{ij} is the log transformed CD4 counts, G_{2i} is a dummy variable indicating that the patient was in treatment group 2 and G_{3i}

is the indicator that patient was in treatment group 3 and G_{4i} is defined likewise, X_{1i} represents the gender of the patient (1=male, 0=female), X_{2i} represents the age of the patient in years and t_{ij} is the j^{th} time point (in weeks) at which i^{th} patient's measurement was recorded. Assuming $(b_{0i}) \overset{iid}{\sim} N(0, \sigma_b^2)$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$ is independent of b_{0i}. In the above model (10), we introduce missingness in the variable X_{2i}. The goal of our study is to investigate if there is any significant difference between the impact of the four treatment groups on the CD4 counts. Therefore, we will test $H_0 : (\beta_{20}, \beta_{30}, \beta_{40})^T = 0$ against $H_0 : (\beta_{20}, \beta_{30}, \beta_{40})^T \neq 0$. When we test the hypothesis using all the data available i.e. 1313 patients, we obtain a p-value of approximately 0.008 which suggests that treatment group type is significant in the study. However, in the hope of obtaining the same conclusion from a much smaller size of 40, we run the analysis. The test statistic, denominator degrees of freedom (*ddf*) and p-value for testing the above hypothesis for the different test procedures are (i) KRM: 3.201 (ddf = 62.50, p-value= 0.029), (ii) LD-LRT: 8.068 (p-value= 0.044) (iii) MI-LRT: 8.775 (p-value= 0.033) (iv) MI-KRT: 3.195 (ddf = 62.51, p-value= 0.031). The p-values for (ii) and (iii) are based on a χ^2_3 distribution while the rest are based on the F-distribution with different degrees of freedom. All tests yield the same result that the null hypothesis is rejected at 5% significance level i.e. the treatment group type has a significant impact on the CD4 counts.

6 Conclusion

In this paper, we derived a small sample Kenward-Roger type adjusted test for fixed effects in a linear mixed model with missing covariates. For this purpose, we account for three biases in the estimation of the covariance matrix of $\hat{\beta}$, namely, the variance bias, the small sample bias and the additional bias introduced due to missingness. We further approximated the null distribution of the test statistic. By simulation, we were able to show the superiority of our method to other popular methods. The derivations have been shown for a block design where the number of random effects is one i.e. $q = 1$, but the method can be derived for $q > 1$. Here, we assumed the covariates to be missing at random, but it would be interesting to derive the results in case of missing responses.

References

1. Alnosaier WS, Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models. *Ph.D. thesis, Oregon State University, Department of Statistics* 2007.

2. Chen Q, Ibrahim JG, Chen MH, Senchaudhuri P, Theory and Inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis* 2008; 99(6): 1302-1331.

3. Fitzmaurice G, Laird N, Ware J, Applied Longitudinal Analysis. *John Wiley & Sons: New York* 2004.

4. Gregory KB, A Comparison of Denominator Degrees of Freedom Approximation Methods in the Unbalanced Two-Way Factorial Mixed Model. *Masters paper, Texas A&M University* 2011.

5. Harville DA, Maximum likelihood approaches to variance component estimation and related problems. *Journal of American Statistical Association* 1977; 72: 320-340.

6. Harville DA, Decomposition of prediction error. *Journal of the American Statistical Association* 1985; 80: 132-138.

7. Harville DA, Jeske DR, Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* 1992; 87:724-731.

8. Henry K, Erice A et al., A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. AIDS Clinical Trial Group 193A Study

Team. *Journal of acquired immune deficiency syndromes and human retrovirology.* 1998; 19(4):339-349.

9. Ibrahim JG, Incomplete data in generalized linear models. *Journal of the American Statistical Association* 1990; 85(411):765-769.

10. Ibrahim JG, Chen MH, Lipsitz SR, Missing responses in generalized linear mixed models when the missing data mechanism is non ignorable. *Biometrika* 2001; 88(2):551-564.

11. Ibrahim JG, Chen MH, Lipsitz SR, Herring AH, Missing-Data methods for generalized linear models. *Journal of the American Statistical Association* 2005; 100(469):332-346.

12. Kackar AN, Harville DA, Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* 1984; 79: 853-862.

13. Kenward MG, Roger JH, Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; 53: 983-997.

14. Laird NM, Ware JH, Random-effects models for longitudinal data. *Biometrics* 1982; 38: 963-974.

15. Leeper JD, Chang SW, Comparison of general multivariate and random-effects models for growth curve analysis: incomplete data small sample situations. *Journal of Statistical Computation and Simulation* 1992; 44(1-2):93-104.

16. Little RJA, Rubin DB, Statistical Analysis with Missing Data. *John Viley & Sons: New York* 1987.

17. McKeon JJ, F approximations to the distribution of Hotelling's T_0^2. *Biometrika* 1974; 61(2): 381-383.

18. Patterson HD, Thompson R, Recovery of inter-block information when block sizes are unequal. *Biometrika* 1971; 58: 545-554.

19. Satterthwaite FF. Synthesis of variance. *Psychometrika* 1941; 6: 309-316.

20. Stubbendick AL, Ibrahim JG, Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* 2003; 59(4):1140-1150.

21. Thompson R. Maximum likelihood estimation of variance components. *Mathematische Operationsforschung und Statistik, Series Statistics* 1980; 11: 545-561.

22. Verbeke G, Molenberghs G, Linear mixed models for longitudinal data. *Springer: New York* 2000.

23. Zucker DM, Lieberman O, Manor O, Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society, Series B* 2000; 62(4): 827-838.

Appendix

A1 Assumptions

We impose the following assumptions about the model as done in Alnosaier (2007)

C.1 Σ is a block diagonal and non singular matrix. Also $\Sigma^{-1}, \partial\Sigma/\partial\sigma_l, \partial\Sigma/\partial\gamma_i, \partial^2\Sigma/\partial\sigma_l\partial\sigma_m$ and $\partial^2\Sigma/\partial\gamma_i\partial\gamma_j$ are bounded.

C.2 $E(\widehat{\boldsymbol{\sigma}}) = \boldsymbol{\sigma} + O(n^{-3/2})$.

C.3 The possible dependence between $\widehat{\boldsymbol{\sigma}}(\widehat{\gamma})$ and $\widehat{\boldsymbol{\beta}}$ is ignored.

C.4 $\Phi = (X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}))^{-1} = O(n^{-1}),\ (L^T\Phi L) = O(n)$.

C.5 $\partial\widetilde{\boldsymbol{\beta}}/\partial\sigma_l = O(n^{-1/2}), \partial^2\widetilde{\boldsymbol{\beta}}/\partial\sigma_l\partial\sigma_m = O(n^{-1/2})$ where $\widetilde{\boldsymbol{\beta}} = \{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\mathbf{Y}$.

C.6 The expectation of $\widehat{\boldsymbol{\beta}}$ exists.

A2 Estimates of Ψ, Φ and Λ in Theorem 1

Let $\boldsymbol{\sigma}=(\sigma_1,\cdots,\sigma_r)^T$ and $\boldsymbol{\gamma}=(\gamma_1,\cdots,\gamma_k)^T$. We begin by expanding the REML based estimate of $\boldsymbol{\beta}$ for the imputed data set which is given by $\widehat{\boldsymbol{\beta}} = \left(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^\star(\widehat{\boldsymbol{\gamma}})\right)^{-1}X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))\mathbf{Y} = C_1+C_2+C_3+C_4+C_5+C_6$, where

$$
\begin{aligned}
C_1 =&\left(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^\star(\widehat{\boldsymbol{\gamma}})\right)^{-1}X^{\star T}(\widehat{\boldsymbol{\gamma}})(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))\mathbf{Y},\\
C_2 =&\left(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^\star(\widehat{\boldsymbol{\gamma}})\right)^{-1}(X^{\star T}(\widehat{\boldsymbol{\gamma}})-X^{\star T}(\boldsymbol{\gamma}))\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y},\\
C_3 =&\left(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^\star(\widehat{\boldsymbol{\gamma}})\right)^{-1}-(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^{\star T}(\boldsymbol{\gamma})\big)^{-1})X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y},\\
C_4 =&\left(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^\star(\boldsymbol{\gamma})\right)^{-1}-(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^{\star T}(\boldsymbol{\gamma})\big)^{-1})X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y},\\
C_5 =&\left(X^{\star T}(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^\star(\boldsymbol{\gamma})\right)^{-1}-(X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^{\star T}(\boldsymbol{\gamma})\big)^{-1})X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y},\\
C_6 =&\left(X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^\star(\boldsymbol{\gamma})\right)^{-1}X^{\star T}(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}.
\end{aligned}
$$

For notational convenience, we will denote $X^\star$ as X. By the Taylor series expansion around $\boldsymbol{\gamma}$,

$$
\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) = \sum_{i=1}^{k}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}}{\partial\gamma_i}(\widehat{\gamma}_i-\gamma_i)+\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}}{\partial\gamma_i\partial\gamma_j}(\widehat{\gamma}_i-\gamma_i)(\widehat{\gamma}_j-\gamma_j)+\cdots.
$$

Since $E(\widehat{\boldsymbol{\gamma}})=\boldsymbol{\gamma}+O_p(n^{-1/2})$ and $E(\mathbf{Y})=X\boldsymbol{\beta}$, we get

$$
\begin{aligned}
\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) &= \sum_{i=1}^{k}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}}{\partial\gamma_i}(\widehat{\gamma}_i-\gamma_i)+O_p(n^{-1})\\
&=-\sum_{i=1}^{k}\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}(\widehat{\gamma}_i-\gamma_i)+O_p(n^{-1}). \qquad \text{(A1)}
\end{aligned}
$$

Substituting (A1) in the expression for C_1 we get $C_1=\widetilde{C}_1+O(n^{-1/2})$, where
$\widetilde{C_1}=\{X^T(\boldsymbol{\gamma})\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\sum_{i=1}^{k}\{-\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}\{\partial\Sigma(\widehat{\boldsymbol{\sigma}})/\partial\gamma_i\}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}})(\widehat{\gamma}_i-\gamma_i)\}X\boldsymbol{\beta}$.
Let $\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))=\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}$, then

$$
C_1 = -\ \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\sum_{i=1}^{k}\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma(\widehat{\boldsymbol{\sigma}})^{-1}(\widehat{\gamma}_i-\gamma_i)X\boldsymbol{\beta}+O_p(n^{-1/2}). \qquad \text{(A2)}
$$

Consider $C_2 = (X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^T(\widehat{\boldsymbol{\gamma}}))^{-1}(X^T(\widehat{\boldsymbol{\gamma}}) - X^T(\boldsymbol{\gamma}))\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))Y$.

By the Taylor series expansion,

$$X^T(\widehat{\boldsymbol{\gamma}}) - X^T(\boldsymbol{\gamma}) = \sum_{i=1}^{k} \frac{\partial X^T}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i) + O_p(n^{-1}).$$

Therefore,

$$\begin{aligned} C_2 &= \widetilde{C_2} + O_p(n^{-1/2}), \text{ where} \\ \widetilde{C_2} &= \Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma}) \sum_{i=1}^{k} \frac{\partial X^T}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i)\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X\boldsymbol{\beta}. \end{aligned} \tag{A3}$$

Consider $C_3 = \{(X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^T(\widehat{\boldsymbol{\gamma}}))^{-1} - (X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X^T(\boldsymbol{\gamma}))^{-1}\}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))Y$.

Again, using the Taylor series expansion,

$$\begin{aligned} (X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\widehat{\boldsymbol{\gamma}}))^{-1} - (X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}))^{-1} &= \sum_{i=1}^{k} \frac{\partial}{\partial \gamma_i}(X^T(\widehat{\boldsymbol{\gamma}})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\gamma}}))X(\boldsymbol{\gamma}))^{-1}(\widehat{\gamma}_i - \gamma_i) \\ &\quad + O_p(n^{-1}). \end{aligned}$$

Hence,

$$\begin{aligned} C_3 &= \widetilde{C}_3 + O_p(n^{-1/2}), \text{ where} \\ \widetilde{C}_3 &= \sum_{i=1}^{k} -(X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}))^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta} \\ &= -\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\sum_{i=1}^{k} X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i}(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta}. \end{aligned} \tag{A4}$$

Similarly,

$$\begin{aligned} \widetilde{C}_4 &= \sum_{i=1}^{k} -\Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})X^T(\boldsymbol{\gamma})\frac{\partial \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \gamma_i}X(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta} \\ &= \sum_{i=1}^{k} \Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial \Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta}, \qquad \text{(A5)} \\ \widetilde{C}_5 &= \sum_{i=1}^{k} -\Phi(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\gamma})\frac{\partial X^T}{\partial \gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\widehat{\gamma}_i - \gamma_i)\boldsymbol{\beta}, \qquad \text{(A6)} \\ C_6 &= (X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma}))^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y} = \breve{\beta}. \end{aligned}$$

Let $H(\widehat{\boldsymbol{\sigma}}) = \widetilde{C}_1 + \widetilde{C}_2 + \widetilde{C}_3 + \widetilde{C}_4 + \widetilde{C}_5$. Further, we need to compute the variance of $\widehat{\boldsymbol{\beta}}$.

$$V(\widehat{\boldsymbol{\beta}}) = V(H(\widehat{\boldsymbol{\sigma}})) + V(\breve{\beta}).$$

To take into account the variability in $\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})$ when $\boldsymbol{\gamma}$ is fixed, we follow the same proof as in Kenward & Roger (1997) and Alnosaier (2007),

$$\begin{aligned} V(\breve{\beta}) &= V(\widetilde{\boldsymbol{\beta}}) + V(\breve{\beta} - \widetilde{\boldsymbol{\beta}}), \text{where } \widetilde{\boldsymbol{\beta}} = (X^T(\boldsymbol{\gamma})\Sigma(\boldsymbol{\sigma}(\boldsymbol{\gamma}))^{-1}X(\boldsymbol{\gamma}))^{-1}X^T(\boldsymbol{\gamma})\Sigma(\boldsymbol{\sigma}(\boldsymbol{\gamma}))^{-1}\mathbf{Y}. \\ V(\breve{\beta}) &= \Phi + \Lambda. \end{aligned}$$

The estimates of Φ and Λ are given by $\widehat{\Phi}$ and $\widehat{\Lambda}$ such that

$$\begin{aligned} E(\widehat{\Phi}) &= \Phi - \widetilde{\Lambda} + R^* + O(n^{-5/2}), \\ E(\widehat{\Lambda}) &= \widetilde{\Lambda} + O(n^{-5/2}), \text{ where} \\ \Phi &= (X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}))^{-1}, \\ \widetilde{\Lambda} &= \sum_{l=1}^{r}\sum_{m=1}^{r} \operatorname{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi(Q_{lm} - P_l\Phi P_m)\Phi. \end{aligned}$$

In the above expressions

$$\begin{aligned} P_l &= -X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial\Sigma}{\partial\sigma_l}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}), \\ Q_{lm} &= X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial\Sigma}{\partial\sigma_l}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial\Sigma}{\partial\sigma_m}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}), \\ R_{lm} &= X^T(\boldsymbol{\gamma})\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))\frac{\partial^2}{\partial\sigma_l\partial\sigma_m}\Sigma^{-1}(\boldsymbol{\sigma}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma}), \\ R^* &= \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r} \operatorname{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\Phi R_{lm}\Phi. \end{aligned}$$

Now, from (A2)-(A6)

$$\begin{aligned} H(\widehat{\boldsymbol{\sigma}}) &= \sum_{i=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\big[\{-X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\} + \{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\} \\ &\quad -\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial\gamma_i}\} + \{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\} \\ &\quad -\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}\big]\boldsymbol{\beta}(\widehat{\gamma}_i-\gamma_i) \\ &= \sum_{i=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i(\widehat{\gamma}_i-\gamma_i), \end{aligned}$$

where

$$\begin{aligned} K_i &= \big[\{-X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\} + \{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\} \\ &\quad -\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial\gamma_i}\} + \{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial\Sigma(\widehat{\boldsymbol{\sigma}})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\} \\ &\quad -\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial\gamma_i}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}\big]\boldsymbol{\beta}, \end{aligned}$$

and

$$\Psi = V(H(\widehat{\boldsymbol{\sigma}})) = \sum_{i=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_iV(\gamma_i)K_i^T\Phi^T(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))K_i\mathrm{cov}(\gamma_i,\gamma_j)K_j^T\Phi^T(\widehat{\boldsymbol{\sigma}},\boldsymbol{\gamma}).$$

Now using the Taylor series expansion about $\boldsymbol{\sigma}$ we get,

$$\Psi(\widehat{\boldsymbol{\sigma}}) = \Psi(\boldsymbol{\sigma}) + \sum_{l=1}^{r}\frac{\partial\Psi}{\partial\sigma_l}(\widehat{\sigma}_l-\sigma_l) + \sum_{l=1}^{r}\sum_{m=1}^{r}\frac{\partial^2\Psi}{\partial\sigma_l\partial\sigma_m}(\widehat{\sigma}_l-\sigma_l)(\widehat{\sigma}_m-\sigma_m) + \cdots . \tag{A7}$$

Consider

$$\begin{aligned} \frac{\partial\Psi(\boldsymbol{\sigma})}{\partial\sigma_l} &= \frac{\partial}{\partial\sigma_l}\left(\sum_{i=1}^{k}\Phi K_i\mathrm{var}(\widehat{\gamma}_i)K_i^T\Phi^T + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi K_i\mathrm{cov}(\widehat{\gamma}_i,\widehat{\gamma}_j)K_j^T\Phi^T\right) \\ &= \sum_{i=1}^{k}\frac{\partial\Phi K_i}{\partial\sigma_l}\mathrm{var}(\widehat{\gamma}_i)K_i^T\Phi^T + \sum_{i=1}^{k}\Phi K_i\mathrm{var}(\widehat{\gamma}_i)\frac{\partial K_i^T\Phi^T}{\partial\sigma_l} \\ &\quad + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\frac{\partial\Phi K_i}{\partial\sigma_l}\mathrm{cov}(\widehat{\gamma}_i,\widehat{\gamma}_j)\Phi K_j + \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\Phi K_i\mathrm{cov}(\widehat{\gamma}_i,\widehat{\gamma}_j)\frac{\partial K_j^T\Phi^T}{\partial\sigma_l}. \end{aligned} \tag{A8}$$

Now,

$$\frac{\partial \Phi K_i}{\partial \sigma_l} = \sum_{i=1}^{k} \frac{\partial \Phi}{\partial \sigma_l} K_i + \sum_{i=1}^{k} \Phi \frac{\partial K_i}{\partial \sigma_l} = \sum_{i=1}^{k} (-\Phi P_l \Phi) K_i + \sum_{i=1}^{k} \Phi \frac{\partial K_i}{\partial \sigma_l}. \tag{A9}$$

and

$$\begin{aligned} \frac{\partial K_i}{\partial \sigma_l} &= \{X^T(\boldsymbol{\gamma}) \frac{\partial}{\partial \sigma_l} (\frac{\partial \Sigma^{-1}}{\partial \gamma_i}) X(\boldsymbol{\gamma}) + \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \gamma_i} \frac{\partial \Sigma^{-1}}{\partial \sigma_l} X(\boldsymbol{\gamma}) - X^T(\boldsymbol{\gamma}) \frac{\partial \Sigma^{-1}}{\partial \sigma_l} \frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i} \\ &\quad - X^T(\boldsymbol{\gamma}) \frac{\partial}{\partial \sigma_l} (\frac{\partial \Sigma^{-1}}{\partial \gamma_i}) X(\boldsymbol{\gamma}) - \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \gamma_i} \frac{\partial \Sigma^{-1}}{\partial \sigma_l} X(\boldsymbol{\gamma})\} \boldsymbol{\beta} \\ &= -X^T(\boldsymbol{\gamma}) \frac{\partial \Sigma^{-1}}{\partial \sigma_l} \frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i} \boldsymbol{\beta}. \end{aligned}$$

Let us denote $M_{li} = \Phi X^T(\boldsymbol{\gamma}) \partial \Sigma^{-1} / \partial \sigma_l \times \partial X(\boldsymbol{\gamma}) / \partial \gamma_i \times \boldsymbol{\beta}$.

Therefore,

$$\frac{\partial \Phi K_i}{\partial \sigma_l} = \sum_{i=1}^{k} (-\Phi P_l \Phi) K_i + \sum_{i=1}^{k} \Phi - X^T(\boldsymbol{\gamma}) \frac{\partial \Sigma^{-1}}{\partial \sigma_l} \frac{\partial X(\boldsymbol{\gamma})}{\partial \gamma_i} \boldsymbol{\beta} = \sum_{i=1}^{k} (-\Phi P_l \Phi) K_i - M_{li}. \tag{A10}$$

Using (A9) and (A10), (A8) becomes,

$$\begin{aligned} \frac{\partial \Psi(\boldsymbol{\sigma})}{\partial \sigma_l} &= \sum_{i=1}^{k} \left\{ (-\Phi P_l \Phi K_i - M_{li}) \mathrm{var}(\widehat{\gamma}_i) K_i^T \Phi^T + \Phi K_i \mathrm{var}(\widehat{\gamma}_i) (-\Phi P_l \Phi K_i - M_{li})^T \right\} \\ &\quad + \sum_{\substack{i=1 \\ i \neq j}}^{k} \sum_{i=1}^{k} \left\{ (-\Phi P_l \Phi K_i - M_{li}) \mathrm{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j) K_j^T \Phi^T + \Phi K_i \mathrm{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j) (-\Phi P_l \Phi K_j - M_{lj})^T \right\}. \end{aligned}$$

Substituting the above in (A7), we get

$$\begin{aligned} \Psi(\widehat{\boldsymbol{\sigma}}) &= \Psi(\boldsymbol{\sigma}) + \sum_{l=1}^{r} \left[\sum_{i=1}^{k} \left\{ (-\Phi P_l \Phi K_i - M_{li}) \mathrm{var}(\widehat{\gamma}_i) K_i^T \Phi^T + \Phi K_i \mathrm{var}(\widehat{\gamma}_i) (-\Phi P_l \Phi K_i - M_{li})^T \right\} \right. \\ &\quad \left. + \sum_{\substack{i=1 \\ i \neq j}}^{k} \sum_{i=1}^{k} \left\{ (-\Phi P_l \Phi K_i - M_{li}) \mathrm{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j) K_j^T \Phi^T + \Phi K_i \mathrm{cov}(\widehat{\gamma}_i, \widehat{\gamma}_j) (-\Phi P_l \Phi K_j - M_{lj})^T \right\} \right] \\ &\quad + \sum_{l=1}^{r} \sum_{m=1}^{r} \frac{\partial^2 \Psi}{\partial \sigma_l \partial \sigma_m} (\widehat{\sigma}_l - \sigma_l)(\widehat{\sigma}_m - \sigma_m) \\ &= \Psi(\sigma) + O_p(n^{-2}). \end{aligned} \tag{A11}$$

Therefore, the adjusted covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by $\widehat{\Phi}_A = \widehat{\Phi} + 2\widehat{\Lambda} - \widehat{R}^\star + \widehat{\Psi}$ and $E(\widehat{\Phi}_A) = \mathrm{var}(\widehat{\boldsymbol{\beta}}) + O(n^{-2})$. This completes the proof.

A3 Derivation of $\widetilde{E}$ and $\widetilde{V}$ used in Theorem 2

We begin by computing the expected value and variance of the F statistic

$$F = \frac{1}{l}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T \widehat{\Phi}_A L)^{-1} L^T (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

The expected value is approximated using the conditional expectation arguments.

$$\begin{aligned} E(F) &= E\{E(F|\widehat{\boldsymbol{\sigma}})\}, \\ E(F|\widehat{\boldsymbol{\sigma}}) &= E\{\frac{1}{l}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T\widehat{\Phi}_A L)^{-1} L^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})|\widehat{\boldsymbol{\sigma}}\} \\ &= \frac{1}{l}\left[E\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L\}(L^T\widehat{\Phi}_A L)^{-1} E\{L^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\} + \text{trace}\{(L^T\widehat{\Phi}_A L)^{-1}\text{var}(L^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}))\}\right]. \end{aligned}$$

Since $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$, therefore

$$lE(F|\widehat{\boldsymbol{\sigma}}) = \text{trace}\{(L^T\widehat{\Phi}_A L)^{-1}(L^T V L)\}, \text{where } V = \text{var}(\widehat{\boldsymbol{\beta}}) = \Phi + \Lambda + \Psi. \tag{A12}$$

Since $\widehat{\Phi}_A = \widehat{\Phi} + \widehat{A}^* + \widehat{\Psi}$, where $\widehat{A}^* = 2\widehat{\Phi}\big\{\sum_{l=1}^r \sum_{m=1}^r \text{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)(\widehat{Q}_{lm} - \widehat{P}_l\widehat{\Phi}\widehat{P}_m - \frac{1}{4}\widehat{R}_{lm})\big\}\widehat{\Phi}$ (Alnosaier 2007), we get

$$\begin{aligned} (L^T\widehat{\Phi}_A L)^{-1}(L^T V L) &= \{(L^T\widehat{\Phi}L) + (L^T\widehat{A}^*L) + (L^T\widehat{\Psi}L)\}^{-1}(L^T V L) \\ &= \big[(L^T\widehat{\Phi}L)\big\{I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big\}\big]^{-1}(L^T V L) \\ &= \big\{I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big\}^{-1}(L^T\widehat{\Phi}L)^{-1}(L^T V L) \\ &= \big\{I - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L) - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L) + O_p(n^{-2})\big\}(L^T\widehat{\Phi}L)^{-1}(L^T V L) \\ &= (L^T\widehat{\Phi}L)^{-1}(L^T V L) - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^T V L) - \\ &\quad (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T V L). \end{aligned}$$

From Alnosaier (2007), we notice that

$$E[\text{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^T V L)\}] = l + A_2 + 2A_3 + O_p(n^{-3/2}), \tag{A13}$$

$$E[\text{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^T V L)\}] = E\{\text{trace}(\Theta\widehat{A}^*)\} = 2A_3 + O(n^{-3/2}), \tag{A14}$$

where

$$\begin{aligned}
\Theta &= L(L^T\Phi L)^{-1}L^T, \\
A_2 &= \sum_{l=1}^{r}\sum_{m=1}^{r}\text{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\text{trace}(\Theta\Phi P_l\Phi\Theta\Phi P_m\Phi), \\
A_3 &= \sum_{l=1}^{r}\sum_{m=1}^{r}\text{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\text{trace}\big(\Theta\Phi(Q_{lm}-P_l\Phi P_m-\frac{1}{4}R_{lm})\big).
\end{aligned}$$

Next, we need to consider

$$\begin{aligned}
(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^TVL) &= (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)+ \\
&\quad (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Lambda L)+ \\
&\quad (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Psi L)+ \\
&= (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)+O(n^{-2}).
\end{aligned}$$

Using the Taylor series expansion for $(L^T\widehat{\Phi}L)^{-1}$ about $\boldsymbol{\sigma}$, we get

$$\begin{aligned}
&(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)+O(n^{-2}) \\
= &\Big\{(L^T\Phi L)^{-1}+\sum_{l=1}^{r}(\widehat{\sigma}_l-\sigma_l)\frac{\partial(L^T\Phi L)^{-1}}{\partial\sigma_l}+\frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r}(\widehat{\sigma}_l-\sigma_l)(\widehat{\sigma}_m-\sigma_m)\frac{\partial^2(L^T\Phi L)^{-1}}{\partial\sigma_l\partial\sigma_m}\Big\}(L^T\widehat{\Psi}L)^{-1} \\
\times &\Big\{(L^T\Phi L)^{-1}+\sum_{l=1}^{r}(\widehat{\sigma}_l-\sigma_l)\frac{\partial(L^T\Phi L)^{-1}}{\partial\sigma_l}+\frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{r}(\widehat{\sigma}_l-\sigma_l)(\widehat{\sigma}_m-\sigma_m)\frac{\partial^2(L^T\Phi L)^{-1}}{\partial\sigma_l\partial\sigma_m}\Big\}(L^T\Phi L)+O_p(n^{-2}) \\
= &(L^T\Phi L)^{-1}(L^T\widehat{\Psi}L)+O_p(n^{-3/2}).
\end{aligned}$$

As shown in Theorem 1, $E(\widehat{\Psi})=\Psi+O_p(n^{-2})$, therefore

$$E[\text{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^TVL)\}]=E\{\text{trace}(\Theta\widehat{\Psi})\}=A_4+O(n^{-3/2}), \tag{A15}$$

where

$$A_4=\sum_{l=1}^{r}\sum_{m=1}^{r}\text{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\text{trace}(\Theta\widehat{\Psi}).$$

Therefore, substituting (A13)-(A15) in (A12), we obtain the approximation of the expected value of the F statistics (7).

$$E(F)=1+\frac{A_2}{l}-\frac{A_4}{l}+O(n^{-3/2}). \tag{A16}$$

Now, we need to compute the variance of the F statistic.

$$\text{var}(F) = E\{\text{var}(F|\widehat{\boldsymbol{\sigma}})\} + \text{var}\{E(F|\widehat{\boldsymbol{\sigma}})\}. \quad \text{(A17)}$$

Consider

$$\begin{aligned}\text{var}(F|\widehat{\boldsymbol{\sigma}}) &= \frac{1}{l^2}\text{var}\big\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T L(L^T\widehat{\Phi}_A L)^{-1}L^T(\widehat{\boldsymbol{\beta}})|\widehat{\boldsymbol{\sigma}}\big\} \\ &= \frac{2}{l^2}\text{trace}\big\{(L^T\widehat{\Phi}_A L)^{-1}(L^T V L)\big\}^2, \text{ since it is a quadratic form.} \quad \text{(A18)}\end{aligned}$$

Using similar arguments as used in Alnosaier (2007), we approximate the following

$$\begin{aligned}(L^T\widehat{\Phi}_A L)^{-1}(L^T V L) &= \{(L^T\widehat{\Phi}L) + (L^T\widehat{A}^* L) + (L^T\widehat{\Psi}L)\}^{-1}(L^T V L) \\ &= \big\{(L^T\widehat{\Phi}L)\big(I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^* L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big)\big\}^{-1}(L^T V L) \\ &= \big\{I + (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^* L) + (L^T\widehat{\Phi}L)(L^T\widehat{\Psi}L)\big\}^{-1}(L^T\widehat{\Phi}L)^{-1}(L^T V L) \\ &= \big\{I - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^* L) - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L) + O_p(n^{-2})\big\}(L^T\widehat{\Phi}L)^{-1}(L^T V L) \\ &= (L^T\widehat{\Phi}L)^{-1}(L^T V L) - (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^* L)(L^T\widehat{\Phi}L)^{-1}(L^T V L) - \\ &\quad (L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T V L).\end{aligned}$$

Therefore

$$
\begin{aligned}
& \operatorname{trace}\big\{(L^T\widehat{\Phi}_A L)^{-1}(L^TVL)\big\}^2 \\
= \;& \operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^TVL)\big\}^2 - 2\operatorname{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^TVL)(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^TVL)\} - \\
& 2\operatorname{trace}\{(L^T\widehat{\Phi}L)^{-1}(L^TVL)(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^TVL)\} + O_p(n^{-2}) \\
= \;& \operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L + L^T\Lambda L + L^T\Psi L)\big\} + \\
& O_p(n^{-2}) \\
= \;& \operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)\big\} + \operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi}L)^{-1}(L^T\Lambda L)\big\} + \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi}L)^{-1}(L^T\Psi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)(L^T\widehat{\Phi}L)^{1}(L^T\Phi L)\big\} + O_p(n^{-2})
\end{aligned}
$$

Using the Taylor series expansion about $\boldsymbol{\sigma}$, we get

$$
\begin{aligned}
= \;& \operatorname{trace}\Big\{I + 2\sum_{l=1}^{r}(\widehat{\sigma}_l - \sigma_l)\frac{\partial}{\partial\sigma_l}(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L) + \sum_{l=1}^{r}\sum_{m=1}^{r}\operatorname{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\frac{\partial^2}{\partial\sigma_l\partial\sigma_m}(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)\Big\} + \\
& \operatorname{trace}\Big\{\sum_{l=1}^{r}\sum_{m=1}^{r}\operatorname{cov}(\widehat{\sigma}_l,\widehat{\sigma}_m)\frac{\partial}{\partial\sigma_l}(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)\frac{\partial}{\partial\sigma_m}(L^T\widehat{\Phi}L)^{-1}(L^T\Phi L)\Big\} + 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Lambda L)\big\} + \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\Psi L)\big\} - \\
& 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{A}^*L)\big\} - 2\operatorname{trace}\big\{(L^T\widehat{\Phi}L)^{-1}(L^T\widehat{\Psi}L)\big\} + O_p(n^{-3/2}).
\end{aligned}
$$

Using $E(\widehat{\Psi}) = \Psi + O(n^{-2})$ and (A18), we approximate the expected value of the conditional variance as follows,

$$E\{\operatorname{var}(F|\widehat{\boldsymbol{\sigma}})\} = \frac{2}{l^2}(l + 3A_2) + O(n^{-3/2}). \tag{A19}$$

Also, using the expression of the mean i.e. (A16), we obtain

$$\operatorname{var}\{E(F|\widehat{\boldsymbol{\sigma}})\} = \frac{A_1}{l^2} + O(n^{-3/2}), \text{ where} \tag{A20}$$

$$A_1 = \sum_{l=1}^{r} \sum_{m=1}^{r} \text{cov}(\widehat{\sigma}_l, \widehat{\sigma}_m)\text{trace}(\Theta\Phi P_l \Phi)\text{trace}(\Theta\Phi P_m \Phi).$$

Using (A19) and (A20) in (A17) we obtain

$$\text{var}(F) = \frac{A_1}{l^2} + \frac{2}{l} + \frac{6A_2}{l^2} + O(n^{-3/2}).$$

Therefore the approximate expected value and variance of the F statistic that will be used for matching moments of λF with those of $F(l, d)$ distribution are given as

$$\widetilde{E} = 1 + \frac{A_2}{l} - \frac{A_4}{l},$$
$$\widetilde{V} = \frac{A_1}{l^2} + \frac{2}{l} + \frac{6A_2}{l^2}.$$

This completes the proof.

A4 Implementation of Simulation Study

We derive the the required quantities for a block design experiment, where the fixed effect covariates are generated from a Gaussian distribution. Therefore,

$$\frac{\partial X^\star(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1} = \begin{pmatrix} 0 & : X^\star = X \\ 1 & : X^\star = \boldsymbol{\gamma}_1 \end{pmatrix}$$

and

$$\frac{\partial X^\star(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_2} = 0.$$

Also, now we need to compute $\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}$ and $\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}$. In the following, $\Sigma = \sigma_e^2 D_2 + \sigma_b^2 D_1$, and for our case we take $D_2 = I$; therefore

$$\frac{\partial \Sigma}{\partial \widehat{\sigma}_e^2} = D_2 = I$$
$$\frac{\partial \Sigma}{\partial \widehat{\sigma}_b^2} = D_1.$$

The REML equations are given by:

$$2\frac{\partial l(\boldsymbol{\sigma})}{\partial \sigma_e^2} = -\text{trace}(G) + \mathbf{Y}^T GG\mathbf{Y} \text{ since} D_2 = I, \quad \text{(A21)}$$

$$2\frac{\partial l(\boldsymbol{\sigma})}{\partial \sigma_b^2} = -\text{trace}(GD_1) + \mathbf{Y}^T GD_1G\mathbf{Y}, \quad \text{(A22)}$$

where $G = \Sigma^{-1} - \Sigma^{-1}X^\star(X^\star\Sigma^{-1}X^\star)^{-1}X^{\star T}\Sigma^{-1}$.

Again, for notational convenience, we will replace $X^\star$ by X. To obtain MLE of σ_e^2 and σ_b^2, we need to set the equations (A21) and (A22) to zero.

$$\begin{aligned}\left|\frac{\partial l}{\partial \sigma_e^2}\right|_{\sigma_e^2=\widehat{\sigma}_e^2, \sigma_b^2=\widehat{\sigma}_b^2} &= 0,\\ \implies & \text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) = \mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\\ \implies & \frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) = \frac{\partial}{\partial \boldsymbol{\gamma}}\{\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\}. \quad \text{(A23)}\end{aligned}$$

and

$$\begin{aligned}\left|\frac{\partial l}{\partial \sigma_b^2}\right|_{\sigma_e^2=\widehat{\sigma}_e^2, \sigma_b^2=\widehat{\sigma}_b^2} &= 0,\\ \implies & \text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1) = \mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\\ \implies & \frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1) = \frac{\partial}{\partial \boldsymbol{\gamma}}\{\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\}. \quad \text{(A24)}\end{aligned}$$

Let us consider the LHS (left hand side) of equation (A21)

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{\gamma}}\text{trace}\{G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\}\\ =& [\text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) - \text{trace}\{\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})(X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X)^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\}]\\ =& \{\frac{\partial}{\partial \widehat{\sigma}_e^2}\text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \frac{\partial}{\partial \widehat{\sigma}_b^2}\text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\} - \frac{\partial}{\partial \boldsymbol{\gamma}}(\text{trace}(M(\boldsymbol{\gamma}))), \text{where}\\ & M = \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})(X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X)^{-1})X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})),\\ =& \text{trace}(\frac{\partial \text{trace}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \Sigma}\frac{\partial \Sigma}{\partial \widehat{\sigma}_e^2})\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \text{trace}(\frac{\partial \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))}{\partial \Sigma}\frac{\partial \Sigma}{\partial \widehat{\sigma}_b^2}) - \frac{\partial}{\partial \boldsymbol{\gamma}}(\text{trace}(M(\boldsymbol{\gamma})))\\ =& \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T}D_2)\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T}D_1)\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} - \text{trace}(\frac{\partial M(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}). \quad \text{(A25)}\end{aligned}$$

We need to explicitly solve one of the terms in (A25),

$$\frac{\partial M(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} = \frac{\partial \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \tag{A26}$$

$$\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))(\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \tag{A27}$$

$$\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\frac{\partial \{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}}{\partial \boldsymbol{\gamma}}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \tag{A28}$$

$$\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) + \tag{A29}$$

$$\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\frac{\partial \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}}. \tag{A30}$$

Further, we need to solve all the terms A26-A30 separately using standard matrix calculus. Therefore consider (A26),

$$\begin{aligned}
&\left(\frac{\partial \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \widehat{\sigma}_e^2} \cdot \frac{\partial \sigma_e^2}{\partial \boldsymbol{\gamma}} + \frac{\partial \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \widehat{\sigma}_b^2} \cdot \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\right) X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
= &\left\{(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}) - (\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}})\right\} \\
&\cdot X\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})).
\end{aligned} \tag{A31}$$

Solving (A28) we get,

$$\begin{aligned}
&-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\left\{\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\right. \\
&-X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} - X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \\
&\left.+X^T\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right\}\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})).
\end{aligned} \tag{A32}$$

Solving (A30) we get,

$$\begin{aligned}
&\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\left\{ - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}\right. \\
&\left.-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\right\}.
\end{aligned} \tag{A33}$$

Substituting the expressions from (A27), (A29), (A31), (A32) and (A33) we get,

$$
\begin{aligned}
&\frac{\partial M(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} \\
= \; & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& + \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \Big\} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& + \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \Big\} \\
& + \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& + \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})). \qquad \text{(A34)}
\end{aligned}
$$

Now substituting (A34) in (A25), we obtain the following,

$$
\begin{aligned}
& \frac{\partial \text{trace}(G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))}{\partial \boldsymbol{\gamma}} \\
= \quad & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T} D_2) \\
& + \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& - \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& + \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \Big\} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T} D_1) \\
& + \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& - \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& + \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \Big\} \\
& - \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& + \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& + \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& - \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} \frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))). \qquad \text{(A35)}
\end{aligned}
$$

Moving on, we compute the RHS (right hand side) of equation (A23),

$$
\frac{\partial}{\partial \boldsymbol{\gamma}} (\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y}) \quad = \quad \mathbf{Y}^T \frac{\partial G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y} + \mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} \mathbf{Y}. \quad \text{(A36)}
$$

Consider,

$$
\begin{aligned}
\frac{\partial G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} &= \frac{\partial}{\partial \boldsymbol{\gamma}} \{\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma})\}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\} \\
&= -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} - \frac{\partial M(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}}. \qquad \text{(A37)}
\end{aligned}
$$

Now using (A34) in (A37), we get,

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}}(\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\gamma))G(\widehat{\boldsymbol{\sigma}}(\gamma))\mathbf{Y}) \\
= \; & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}\Big\{ -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\Big\} \\
& +\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\Big\{ -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\Big\} \\
& -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})). \qquad \text{(A38)}
\end{aligned}
$$

For notational convenience let us define three new quantities U, U_1 and V as follows,

$$
\begin{aligned}
U = \; & \Big\{\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
& +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\Big\}, \qquad \text{(A39)}
\end{aligned}
$$

$$
\begin{aligned}
U_1 &= \Big\{\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
&\quad -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
&\quad +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\Big\}, \qquad \text{(A40)}
\end{aligned}
$$

and

$$
\begin{aligned}
V &= -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
&\quad +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
&\quad +\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\frac{\partial X(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Phi(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \\
&\quad -\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\{X^T(\boldsymbol{\gamma})\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))X(\boldsymbol{\gamma})\}^{-1}\frac{\partial X^T(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})). \qquad \text{(A41)}
\end{aligned}
$$

Now, substituting the equations (A38), (A39), (A40), (A41) in (A36) we get RHS of (A23) as

$$
\begin{aligned}
& \mathbf{Y}^T\{\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))+U)+\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))+U_1)+V\}G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y} \\
& +\mathbf{Y}^T G(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\{\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))+U)+\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))+U_1)+V\}\mathbf{Y} \\
= & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}}\{\mathbf{Y}^T U G\mathbf{Y}+\mathbf{Y}^T G U\mathbf{Y}-\mathbf{Y}^T\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))-\mathbf{Y}^T G\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_2\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\} \\
& +\frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}}\{\mathbf{Y}^T U_1 G\mathbf{Y}+\mathbf{Y}^T G U_1\mathbf{Y}-\mathbf{Y}^T\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))-\mathbf{Y}^T G\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))D_1\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))\mathbf{Y}\} \\
& +\mathbf{Y}^T V G\mathbf{Y}+\mathbf{Y}^T G V\mathbf{Y}. \qquad \text{(A42)}
\end{aligned}
$$

Equating the RHS (A42) and LHS (A35) of (A21) we get,

$$
\begin{aligned}
& \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T} D_2) \\
& + \text{trace}(U) - \mathbf{Y}^T U G \mathbf{Y} - \mathbf{Y}^T G U \mathbf{Y} + \mathbf{Y}^T \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) G \mathbf{Y} + \mathbf{Y}^T G \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \mathbf{Y} \Big\} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ \text{trace}(-\Sigma(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))^{-2T} D_1) \\
& + \text{trace}(U_1) - \mathbf{Y}^T U_1 G \mathbf{Y} - \mathbf{Y}^T G U_1 \mathbf{Y} + \mathbf{Y}^T \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) G \mathbf{Y} + \mathbf{Y}^T G \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \mathbf{Y} \Big\} \\
= \; & \mathbf{Y}^T V G \mathbf{Y} + \mathbf{Y}^T G V \mathbf{Y} - \text{trace}(V). \qquad \text{(A43)}
\end{aligned}
$$

Similarly, we need to solve both sides of the equation (A22). Consider the LHS

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}} \{ \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \{ X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) X(\boldsymbol{\gamma}) \}^{-1} X^T(\boldsymbol{\gamma}) \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1) \} \\
= \; & \frac{\partial}{\partial \widehat{\sigma}_e^2} \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1) \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} + \frac{\partial}{\partial \widehat{\sigma}_b^2} \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1) \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} - \frac{\partial}{\partial \boldsymbol{\gamma}} (\text{trace}(M(\boldsymbol{\gamma}) D_1)) \\
= \; & \text{trace}\{ \frac{\partial}{\partial \Sigma} \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1) \frac{\partial \Sigma}{\partial \widehat{\sigma}_e^2} \} + \text{trace}\{ \frac{\partial}{\partial \Sigma} \text{trace}(\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1) \frac{\partial \Sigma}{\partial \widehat{\sigma}_b^2} \} \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \\
& - \frac{\partial}{\partial \boldsymbol{\gamma}} \text{trace}(M(\boldsymbol{\gamma}) D_1). \qquad \text{(A44)}
\end{aligned}
$$

Now,

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\gamma}} \text{trace}(M(\boldsymbol{\gamma}) D_1) &= \text{trace}(\frac{\partial M D_1}{\boldsymbol{\gamma}}) \\
&= \text{trace}(\frac{\partial M}{\partial \boldsymbol{\gamma}} \cdot D_1) \\
&= \text{trace}\{ (-\frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} U - \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} U_1 - V) \cdot D_1 \} \\
&= \text{trace}(-U D_1) \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} - \text{trace}(U_1 D_1) \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} - \text{trace}(V D_1). \qquad \text{(A45)}
\end{aligned}
$$

Using (A45) in (A44), we get

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\gamma}} \text{trace}(G D_1) = \; & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \{ -\text{trace}(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))^T D_2 + \text{trace}(U D_1) \} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \{ -\text{trace}(-\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))^T D_1 + \text{trace}(U_1 D_1) \} \\
& + \text{trace}(V D_1). \qquad \text{(A46)}
\end{aligned}
$$

Solving RHS of (A24) we get,

$$\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\gamma}}(\mathbf{Y}^T G D_1 G \mathbf{Y}) \\
= \quad & \mathbf{Y}^T (\frac{\partial G}{\partial \boldsymbol{\gamma}}) D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 \frac{\partial G}{\partial \boldsymbol{\gamma}} \mathbf{Y} \\
= \quad & \mathbf{Y}^T \Big\{ \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} (U - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} (U_1 - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) + V \Big\} D_1 G \mathbf{Y} \\
& + \mathbf{Y}^T G D_1 \Big\{ \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} (U - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} (U_1 - \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma}))) + V \Big\} \mathbf{Y} \\
= \quad & \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ \mathbf{Y}^T U D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 U \mathbf{Y} - \mathbf{Y}^T \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 G \mathbf{Y} \\
& - \mathbf{Y}^T G D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y} \Big\} + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ \mathbf{Y}^T U_1 D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 U_1 \mathbf{Y} \\
& - \mathbf{Y}^T \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 G \mathbf{Y} - \mathbf{Y}^T G D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y} \Big\} \\
& + \mathbf{Y}^T V D_1 G \mathbf{Y} - \mathbf{Y}^T G D_1 V \mathbf{Y}. \qquad \text{(A47)}
\end{aligned}$$

Therefore, equating LHS (A46) = RHS (A47) we get,

$$\begin{aligned}
& \frac{\partial \widehat{\sigma}_e^2}{\partial \boldsymbol{\gamma}} \Big\{ - \mathrm{trace}((\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))^T D_2) + \mathrm{trace}(U D_1) - \mathbf{Y}^T U D_1 G \mathbf{Y} \\
& - \mathbf{Y}^T G D_1 U \mathbf{Y} + \mathbf{Y}^T \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_2 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y} \Big\} \\
& + \frac{\partial \widehat{\sigma}_b^2}{\partial \boldsymbol{\gamma}} \Big\{ - \mathrm{trace}((\Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})))^T D_1) + \mathrm{trace}(U_1 D_1) - \mathbf{Y}^T U_1 D_1 G \mathbf{Y} \\
& - \mathbf{Y}^T G D_1 U_1 \mathbf{Y} + \mathbf{Y}^T \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) D_1 \Sigma^{-1}(\widehat{\boldsymbol{\sigma}}(\boldsymbol{\gamma})) \mathbf{Y} \Big\} \\
& = \mathbf{Y}^T V D_1 G \mathbf{Y} + \mathbf{Y}^T G D_1 V \mathbf{Y} - \mathrm{trace}(V D_1). \qquad \text{(A48)}
\end{aligned}$$

Finally, we can solve the equations (A43) and (A48) simultaneously to obtain $\partial \widehat{\sigma}_e^2 / \partial \boldsymbol{\gamma}$ and $\partial \widehat{\sigma}_b^2 / \partial \boldsymbol{\gamma}$.

Remark: Here we have derived the quantities to implement the simulation for a study with one random effect. In the studies, which have more than one random effect, similar method can be used to derive the required quantities.

Simulation Study Results

Table 1: Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 15$ and $n_i = 2$ for all $i = 1, \ldots, 15$. Here KRM, LD-LRT, MI-LRT and MI-KRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test, mean imputation followed by likelihood ratio test, mean imputation followed by Kenward-Roger test with no adjustment for missingness.

Percentage of missing data	ρ	KRM	LD-LRT	MI-LRT	MI-KRT
20	0.25	0.049	0.148	0.113	0.057
	0.5	0.055	0.246	0.122	0.065
	1	0.059	0.152	0.120	0.068
	2	0.057	0.188	0.127	0.067
	4	0.063	0.173	0.129	0.075
40	0.25	0.042	0.298	0.112	0.054
	0.5	0.057	0.205	0.126	0.070
	1	0.052	0.375	0.123	0.066
	2	0.058	0.375	0.130	0.070
	4	0.063	0.415	0.124	0.073

Table 2: Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 30$ and $n_i = 2$ for all $i = 1, \ldots, 30$. Here KRM, LD-LRT, MI-LRT, MI-KRT and MI-WT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test, mean imputation followed by likelihood ratio test, mean imputation followed by Kenward-Roger test with no adjustment for missingness.

Percentage of missing data	ρ	KRM	LD-LRT	MI-LRT	MI-KRT
20	0.25	0.044	0.085	0.078	0.049
	0.5	0.055	0.083	0.087	0.060
	1	0.050	0.094	0.082	0.056
	2	0.055	0.106	0.083	0.059
	4	0.047	0.088	0.079	0.052
40	0.25	0.047	0.100	0.073	0.045
	0.5	0.048	0.143	0.081	0.053
	1	0.048	0.104	0.072	0.051
	2	0.051	0.217	0.079	0.055
	4	0.048	0.169	0.076	0.053

Table 3: Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 15$ and $n_i = 2$ for all $i = 1, \ldots, 30$ when the missing mechanism is MAR. Here KRM, LD-LRT, MI-LRT and MI-KRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test, mean imputation followed by likelihood ratio test, mean imputation followed by Kenward-Roger test with no adjustment for missingness.

ρ	KRM	LD-LRT	MI-LRT	MI-KRT
0.25	0.058	0.103	0.102	0.057
0.5	0.053	0.114	0.091	0.051
1	0.050	0.164	0.103	0.063
2	0.049	0.117	0.089	0.047
4	0.063	0.141	0.099	0.065

Table 4: Here we present Type-I error probabilities for the test $H_0 : \boldsymbol{\beta} = 0$ at the 5% level of significance when $\rho = \sigma_b^2/\sigma_e^2$, $m = 15$ and $n_i = 2$ for all $i = 1, \ldots, 30$ when the partially missing (20% missing) covariate is generated from heavy tailed distributions. Here KRM, LD-LRT, MI-LRT and MI-KRT refer to Kenward-Roger adjustment for missing covariates, listwise deletion followed by likelihood ratio test, mean imputation followed by likelihood ratio test, mean imputation followed by Kenward-Roger test with no adjustment for missingness.

Distribution	ρ	KRM	LD-LRT	MI-LRT	MI-KRT
Laplace(0,2)	0.25	0.052	0.106	0.095	0.053
	0.5	0.056	0.115	0.102	0.056
	1	0.049	0.185	0.094	0.047
	2	0.049	0.170	0.099	0.047
	4	0.051	0.199	0.100	0.055
Student's t(1)	0.25	0.050	0.153	0.099	0.046
	0.5	0.051	0.105	0.096	0.057
	1	0.047	0.137	0.104	0.047
	2	0.055	0.142	0.105	0.061
	4	0.059	0.139	0.107	0.061