

Chapter 7

Some Tests for the Extended Growth Curve Model and Applications in the Analysis of Clustered Longitudinal Data



Jemila S. Hamid and Sayantee Jana

Abstract The Growth Curve Model (GCM) is a Generalized Multivariate Analysis of Variance (GMANOVA) model especially useful in the analysis of longitudinal data, growth curves as well as other response curves. The model is a natural extension of the classical Multivariate Analysis of Variance (MANOVA) model and among other assumptions, relies on the assumption that the mean for each group can be represented as a polynomial of degree q . The assumption that the mean over time for all groups follows a polynomial of the same degree is not always satisfied, since individuals across the different groups may respond differently. An excellent scenario is when we have clustered longitudinal data, where the response over time can be represented by polynomials of different degrees. In such situations, the natural extension is to use the Extended Growth Curve Model (EGCM), where one can assume different shapes to represent different groups or clusters. In this paper, we formulate hypotheses motivated by real life scenarios involving clustered longitudinal data, and propose tests that are motivated by residuals in the EGCM. We then mathematically derive the tests and evaluate performances using simulations. We provide real data examples as illustrations.

7.1 Introduction

The Extended Growth Curve Model (EGCM) is an extension of the Growth Curve Model (GCM), and arises in situations where there are linear restrictions on the mean parameter of the model [22, 26, 27, 30, 31]. Both the GCM and EGCM are also referred to as Generalized Multivariate Analysis of Variance (GMANOVA) models, simply because they are indeed generalizations of the classical Multivariate

J. S. Hamid (✉)
University of Ottawa, Ottawa, ON, Canada
e-mail: jhamid@uottawa.ca

S. Jana
Indian Institute of Management Nagpur, Nagpur, India
e-mail: sayantee@iimnagpur.ac.in

Analysis of Variance (MANOVA) models [1, 9–12, 26]. The term bilinear regression will be used frequently in describing the GCM and its extensions, to indicate the presence of two design matrices: the within and between design matrices, and to describe the bilinear nature of the projections corresponding to these two design matrices. The linear restrictions on MANOVA models lead to a bilinear model, where the model involves projections with respect to two design matrices, as already mentioned [3, 4, 16, 30].

The GCM has been demonstrated to be useful in the analysis of longitudinal data, especially for short to moderate time series, and the analysis is performed assuming that the mean for each of the groups follows a polynomial of degree q [11, 21, 23]. There is extensive literature in the area and various distributions and covariance structures have been considered [6, 13, 14, 18–20, 28]. High-dimensional extensions have also been considered in recent years [5, 8, 24, 25]. In practical applications, however, the mean for different groups might be represented by polynomials of different degrees. This, for instance, happens when we have clustered longitudinal data, where the response over time follows different shapes for different clusters or groups. In such situations, the EGCM is useful since it allows different degrees of polynomials to be fitted within one modeling framework [3, 7, 15, 17, 26, 31].

Consider a GCM with m groups, where measurements are taken from each of the n individuals at p different time points. Suppose also, that the mean (across time) for the i th ($i = 1, 2, \dots, k$) group can be represented by a polynomial function of degree $q - 1$ and can be described as

$$b_{0,i} + b_{1,i}t + b_{2,i}t^2 \cdots + b_{q,i}t^{q-1}, \quad t = t_1, t_2, \dots, t_p.$$

The GCM in matrix format, is given as

$$\mathbf{Y} = \mathbf{ZBX} + \mathbf{E},$$

where $\mathbf{Y} : p \times n$ represents the observation (outcome) matrix, $\mathbf{Z} : p \times q$ and $\mathbf{X} : m \times n$ are the between and within individual design matrices, and $\mathbf{B} : q \times m$ represents the parameter matrix with the coefficients of the polynomials. The columns of the error matrix $\mathbf{E} : p \times n$ are assumed to be distributed as a p -variate normal distribution with mean zero and positive definite covariance matrix Σ . Description of the various matrices in the model above are given by

$$\mathbf{Z}' = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ t_1 & t_2 & t_3 & \cdots & t_p \\ t_1^2 & t_2^2 & t_3^2 & \cdots & t_p^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1^{q-1} & t_2^{q-1} & t_3^{q-1} & \cdots & t_p^{q-1} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_2} & \mathbf{0}_{n_3} & \cdots & \mathbf{0}_{n_m} \\ \mathbf{0}_{n_1} & \mathbf{1}_{n_2} & \mathbf{0}_{n_3} & \cdots & \mathbf{0}_{n_m} \\ \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \mathbf{1}_{n_3} & \cdots & \mathbf{0}_{n_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \mathbf{0}_{n_3} & \cdots & \mathbf{1}_{n_m} \end{pmatrix},$$

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1n} \\ y_{21} & y_{22} & y_{23} & \cdots & y_{2n} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{p1} & y_{p2} & y_{p3} & \cdots & y_{pn} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{01} & b_{02} & b_{03} & \cdots & b_{0m} \\ b_{11} & b_{12} & b_{13} & \cdots & b_{1m} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{(q-1)1} & b_{(q-1)2} & b_{(q-1)3} & \cdots & b_{(q-1)m} \end{pmatrix}$$

and it is assumed that $q \leq p$ and $\text{rank}(\mathbf{X}) + p \leq n$ and $n = n_1 + n_2 + \dots + n_m$.

Inferences for the GCM have been considered by many, and likelihood estimators for both the mean parameter and covariance matrices are already available [4, 6, 11]. The likelihood ratio test for testing the general linear hypotheses was also derived by Khatri [11], under some full rank restrictions. Residuals in the GCM were considered and mathematical decompositions were performed to provide some insights on the characteristics of the various components [29, 30]. This work was further extended to the EGCM, where better understanding of the design space as well as the residual space were achieved [3, 30].

Consider now the decomposed residuals in the GCM [29] and the simple hypothesis $\mathbf{B} = \mathbf{0}$ (mean is zero), an extension of the well-known Lawley-Hotelling trace test was previously derived for testing this hypothesis [4]. The test turned out to be functions of the decomposed residuals, which in turn facilitated appropriate interpretations of the decomposed residuals as well as better understanding of the distribution of the test statistic. In their paper, Hamid and colleagues showed that the distributions of the test statistics for the simple hypothesis of $\mathbf{B} = \mathbf{0}$ as well as the general linear hypothesis $\mathbf{GBF} = \mathbf{0}$ are free of the unknown covariance matrix Σ [4]. This means that the distributions can be generated empirically and the critical value for the tests can be calculated through simulations or parametric bootstrapping. Furthermore, the authors also showed that the distribution of the test can be represented as weighted sums of chi-square random variables, which allowed the authors to provide appropriate approximations.

In this paper, we consider the EGCM and provide some tests for the special case of the model, where we assume the mean growth curves to be clustered in two categories. The mean (over time) for the two clusters follow polynomials of different degrees, e.g., one cluster consisting of groups with linear growth curves and the other cluster consisting of groups with quadratic curves. We first present and discuss the decomposed residuals in the EGCM [3, 30], and formulate two hypotheses based on their practical relevance in analysis of clustered longitudinal data, requiring fitting of the EGCM. We propose potential statistical tests motivated by the residuals, and using a formulation similar to the trace test in the GCM, which as mentioned above is an extension of the well-known Lawley-Hotelling trace test in the MANOVA model. We then provide formal mathematical derivation using restricted maximum likelihood approach and provide some distributional properties of the test statistics. We evaluate performance using extensive simulations and provide real data illustration.

7.2 Residuals in the Extended Growth Curve Model

Suppose we have m groups and a total of n individuals from whom measurements are taken at p time points. Without loss of generality, consider the EGCM, where groups can be categorized into two clusters. Suppose, again without loss of generality, that the mean for the first cluster, consisting of m_1 groups, follows a polynomial of degree $q_1 - 1$ and the mean for the second cluster (consisting of m_2 groups) follows a polynomial of degree $q_1 + q_2 - 1$. For presentation purposes, let $m = m_1 + m_2$ and $q = q_1 + q_2$. The EGCM, in matrix formulations, can be represented as

$$\mathbf{Y} = \mathbf{Z}_1 \mathbf{B}_1 \mathbf{X}_1 + \mathbf{Z}_2 \mathbf{B}_2 \mathbf{X}_2 + \mathbf{E}, \quad (7.1)$$

where $q_1, q_2 \leq p$, $\text{rank}(\mathbf{X}_1) + p \leq n$ and $\mathcal{C}(\mathbf{X}_2') \subseteq \mathcal{C}(\mathbf{X}_1')$, where $\mathcal{C}(\cdot)$ represents the column space of a matrix. The observation matrix \mathbf{Y} has the same dimension and representation as the one described above for the GCM, and the n columns of the error matrix \mathbf{E} are assumed to be distributed as a p -variate normal random variable with mean 0 and covariance $\mathbf{\Sigma}$. The within individual design matrix \mathbf{Z}_1 , now of dimension $p \times q_1$, also has the same representation as \mathbf{Z} in GCM, but with q_1 replacing q ; the parameter matrix \mathbf{B}_1 , of dimension $q_1 \times m$ represents the coefficients of the q_1 polynomials that are common for both clusters consisting of all m groups. The between individual design matrix \mathbf{X}_1 , of dimension $m \times n$, has the same representation as \mathbf{X} in GCM. The descriptions of the remaining matrices involved in the EGCM are given as

$$\begin{aligned} \mathbf{Z}_2' : q_2 \times p &= \begin{pmatrix} t_1^{q_1} & t_2^{q_1} & t_3^{q_1} & \cdots & t_p^{q_1} \\ t_1^{q_1+1} & t_2^{q_1+1} & t_3^{q_1+1} & \cdots & t_p^{q_1+1} \\ t_1^{q_1+2} & t_2^{q_1+2} & t_3^{q_1+2} & \cdots & t_p^{q_1+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1^{q_1+q_2-1} & t_2^{q_1+q_2-1} & t_3^{q_1+q_2-1} & \cdots & t_p^{q_1+q_2-1} \end{pmatrix}, \\ \mathbf{X}_2 : m_2 \times n &= \begin{pmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \cdots & \mathbf{1}_{n_{m_1+1}} & \mathbf{0}_{n_{m_1+2}} & \cdots & \mathbf{0}_{n_m} \\ \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \cdots & \mathbf{0}_{n_{m_1+1}} & \mathbf{1}_{n_{m_1+2}} & \cdots & \mathbf{0}_{n_m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1} & \mathbf{0}_{n_2} & \cdots & \mathbf{0}_{n_{m_1+1}} & \mathbf{0}_{n_{m_1+2}} & \cdots & \mathbf{1}_{n_m} \end{pmatrix}, \\ \mathbf{B}_2 : q_2 \times m_2 &= \begin{pmatrix} b_{q_11} & b_{q_12} & b_{q_13} & \cdots & b_{q_1m_2} \\ b_{(q_1+1)1} & b_{(q_1+1)2} & b_{(q_1+1)3} & \cdots & b_{(q_1+1)m_2} \\ b_{(q_1+2)1} & b_{(q_1+2)2} & b_{(q_1+2)3} & \cdots & b_{(q_1+2)m_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{(q_1+q_2-1)1} & b_{(q_1+q_2-1)2} & b_{(q_1+q_2-1)3} & \cdots & b_{(q_1+q_2-1)m_2} \end{pmatrix}. \end{aligned}$$

Considerable literature is available on GMANOVA models in general, and EGCM in particular [3, 7, 12, 15, 26, 30, 31]. Solutions to the likelihood functions are also provided, and explicit formulae for the estimators are provided [12, 26]. Nevertheless, there is limited work related to hypothesis testing in the context of the mean parameters of EGCM, and hence limited applications of the model exist despite the fact that longitudinal data in practice follow different functions over time across the different groups. The focus of this paper is, therefore, to contribute towards hypothesis testing for the mean parameters of the EGCM. In doing so, we are particularly interested in the decomposed residuals, and we consider a special case of the model, without loss of generality. We use the residuals as motivations to formulate hypotheses relevant to real world applications. We then mathematically derive corresponding test statistics. As such, let us first consider the estimated model, derived using the maximum likelihood (ML) approach, which is always unique [3, 26, 27]. The mathematical expression for the estimated model is given by

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{Z}_1 \hat{\mathbf{B}}_1 \mathbf{X}_1 + \mathbf{Z}_2 \hat{\mathbf{B}}_2 \mathbf{X}_2 \\ &= (\mathbf{I} - \mathbf{T}_1) \mathbf{Y} \mathbf{X}'_1 (\mathbf{X}_1 \mathbf{X}'_1)^{-1} \mathbf{X}_1 + (\mathbf{I} - \mathbf{T}_2) \mathbf{Y} \mathbf{X}'_2 (\mathbf{X}_2 \mathbf{X}'_2)^{-1} \mathbf{X}_2,\end{aligned}\quad (7.2)$$

where \mathbf{A}^- represents a generalized inverse of any matrix \mathbf{A} and

$$\begin{aligned}\mathbf{T}_1 &= \mathbf{I} - \mathbf{Z}_1 (\mathbf{Z}'_1 \mathbf{S}_1^{-1} \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 \mathbf{S}_1^{-1}, \\ \mathbf{T}_2 &= \mathbf{I} - \mathbf{T}_1 \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{T}'_1 \mathbf{S}_2^{-1} \mathbf{T}_1 \mathbf{Z}_2)^{-1} \mathbf{Z}'_2 \mathbf{T}'_1 \mathbf{S}_2^{-1}; \\ \mathbf{S}_1 &= \mathbf{Y} (\mathbf{I} - \mathbf{X}'_1 (\mathbf{X}_1 \mathbf{X}'_1)^{-1} \mathbf{X}_1) \mathbf{Y}'; \\ \mathbf{S}_2 &= \mathbf{S}_1 + \mathbf{T}_1 \mathbf{Y} \mathbf{X}'_1 (\mathbf{X}_1 \mathbf{X}'_1)^{-1} \mathbf{X}_1 (\mathbf{I} - \mathbf{X}'_2 (\mathbf{X}_2 \mathbf{X}'_2)^{-1} \mathbf{X}_2) \\ &\quad \times \mathbf{X}'_1 (\mathbf{X}_1 \mathbf{X}'_1)^{-1} \mathbf{X}_1 \mathbf{Y}' \mathbf{T}'_1.\end{aligned}$$

In vectorized form, this can be re-written as

$$\text{Vec} \hat{\mathbf{Y}} = [\mathbf{P}_{\mathbf{X}_1} \otimes \mathbf{P}_{\mathbf{Z}_1}] \text{Vec} \mathbf{Y} + [\mathbf{P}_{\mathbf{X}_2} \otimes \mathbf{P}_{\mathbf{Z}_2}] \text{Vec} \mathbf{Y} = \mathbf{P} \text{Vec} \mathbf{Y},$$

where Vec represents a vectorized form a matrix, \otimes represents the Kronecker product and

$$\begin{aligned}\mathbf{P}_{\mathbf{X}_1} &= \mathbf{X}'_1 (\mathbf{X}_1 \mathbf{X}'_1)^{-1} \mathbf{X}_1, \\ \mathbf{P}_{\mathbf{X}_2} &= \mathbf{X}'_2 (\mathbf{X}_2 \mathbf{X}'_2)^{-1} \mathbf{X}_2, \\ \mathbf{P}_{\mathbf{Z}_1} &= \mathbf{Z}_1 (\mathbf{Z}'_1 \mathbf{S}_1^{-1} \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 \mathbf{S}_1^{-1}, \\ \mathbf{P}_{\mathbf{Z}_2} &= \mathbf{T}_1 \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{T}'_1 \mathbf{S}_2^{-1} \mathbf{T}_1 \mathbf{Z}_2)^{-1} \mathbf{Z}'_2 \mathbf{T}'_1 \mathbf{S}_2^{-1}.\end{aligned}$$

It is evident that the estimated model is a bilinear projection with respect to the four design matrices, which is equivalent to the column space of the matrix \mathbf{P} [3, 30]. This column space can be represented as sum of two tensor product spaces

$$\mathcal{C}(\mathbf{P}) = \mathcal{C}(\mathbf{X}'_1) \otimes \mathcal{C}_{S_1}(\mathbf{Z}_1) + \mathcal{C}(\mathbf{X}'_2) \otimes \mathcal{C}_{S_2}(\mathbf{T}_1\mathbf{Z}_2),$$

where \otimes represents tensor product between two spaces. The relationship between Kronecker product of two matrices and tensor product space are presented in more details in Kollo and von Rosen [12], and its applications in decomposition of residuals in GMANOVA models is provided in Hamid and von Rosen [3].

Now consider the residuals in the EGCM, which are defined on the orthogonal complement of the design space, that is a projection onto $(\mathcal{C}(\mathbf{X}'_1) \otimes \mathcal{C}_{S_1}(\mathbf{Z}_1) + \mathcal{C}(\mathbf{X}'_2) \otimes \mathcal{C}_{S_2}(\mathbf{T}_1\mathbf{Z}_2))^\perp$ [3, 30]. This space was mathematically decomposed and four residuals are provided in a previous work. Graphical elucidation of the design and residual spaces as well as the corresponding formulas for the residuals are provided below (Fig. 7.1). More mathematical details about the residuals in the EGCM can be found in Hamid and von Rosen [3].

Note that, some of the spaces in Fig. 7.1 can be decomposed further (e.g., note the broken lines across \mathbf{R}_2 and \mathbf{R}_3). Although further decompositions can also be interpreted in terms of the model characteristics with respect to what the model was not able to explain (i.e., residuals), we focused on these four residuals for now, mainly because of their practical relevance in applications of GMANOVA models:

$$\begin{aligned} \mathbf{R}_1 &= (\mathbf{I} - \mathbf{T}_1)\mathbf{Y}(\mathbf{I} - \mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1), \\ \mathbf{R}_2 &= \mathbf{T}_1\mathbf{Y}(\mathbf{I} - \mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1), \\ \mathbf{R}_3 &= \mathbf{T}_1\mathbf{Y}(\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - \mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2), \\ \mathbf{R}_4 &= (\mathbf{T}_1 + \mathbf{T}_2 - \mathbf{I})\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2. \end{aligned}$$

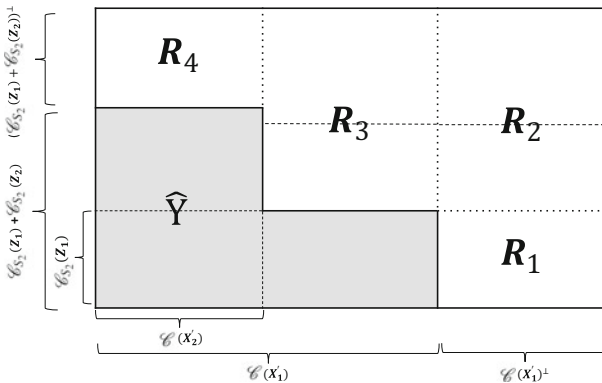


Fig. 7.1 The spaces representing the fitted model (design space) and the residuals in the EGCM

As we can see from the figure and mathematical characteristics of the residuals, $\mathbf{R}_{12} = \mathbf{R}_1 + \mathbf{R}_2$ is equivalent to the residual in MANOVA models and appears in the GCM as a sum of two of the decomposed residuals [29, 30]. This residual represents the difference between the observations and the group mean ($\mathbf{Y} - \mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 = \mathbf{Y}(\mathbf{I} - \mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1)$) and is distributed as a multivariate normal random variable, and hence useful in assessing between individual assumptions as well as distributional assumptions. Note further, that $\mathbf{R}_{12}\mathbf{R}'_{12} = \mathbf{S}_1$, and that the sample covariance matrix is $\frac{1}{n-1}\mathbf{S}_1$. On the other hand, $\mathbf{R}_{34} = \mathbf{R}_3 + \mathbf{R}_4$ provides information about overall model fit, while \mathbf{R}_3 and \mathbf{R}_4 individually provide information on different components of the polynomial fit.

For instance, consider two clusters, where the mean for one of the clusters can be represented by a linear function over time and the mean of the second cluster follows a quadratic curve. For this scenario, \mathbf{R}_3 provides information on the linear component of the model (for both clusters) and \mathbf{R}_4 provides information on the quadratic component of the second cluster. More on this will come while deriving and evaluating the proposed tests.

7.3 Some Tests for the Extended Growth Curve Model

Consider the EGCM defined in (7.1). Without loss of generality, consider only two groups, where each group is considered as a cluster. Suppose measurements from each group are taken at different time points. Suppose the mean for the first group can be represented by a linear function over time and the mean for the second group follows quadratic curve. Note that, this is just to simplify presentation in this paper, nevertheless the tests are derived under general assumptions, and are not restricted to two groups or linear/quadratic polynomials.

Consider now the simple hypothesis that the mean is zero (testing overall significance of the model), which can be formulated as

$$\begin{aligned} H_0 : \mathbf{B}_1 &= \mathbf{0}, \mathbf{B}_2 = \mathbf{0}, \\ H_1 : \mathbf{B}_1 &\neq \mathbf{0}, \mathbf{B}_2 \neq \mathbf{0}. \end{aligned} \quad (7.3)$$

Recall the Lawley-Hotelling trace test for MANOVA and GCM [4, 5]. Both tests are weighted functions of the observed mean and the corresponding residuals (the part of the observed mean that is left unexplained by the fitted model), where the weight is the between individual variation represented by the sample variance-covariance matrix in GCM, which is a function of \mathbf{S} . Using analogous arguments, one can suggest that a test statistic for the simple hypothesis presented above, in the EGCM framework, will be a weighted function of $\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1\mathbf{Y}'$ and $\mathbf{R}_{34}\mathbf{R}'_{34}$, and will have the format

$$\frac{f(\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1\mathbf{Y}')}{f(\mathbf{R}_{34}\mathbf{R}'_{34})}.$$

Below, we will provide mathematical derivation for the test. In doing so, we write the likelihood function as a product of three independent terms. We maximize a certain part of the likelihood to get an estimator for the unknown covariance matrix, which then replaces the covariance matrix in the likelihood function to give the estimated likelihood.

Consider first the EGCM in (7.1), we can rewrite it as

$$\begin{aligned} \mathbf{Y} &= \mathbf{Z}_1(\mathbf{B}_{11} : \mathbf{B}_{12}) \begin{pmatrix} \mathbf{X}_{11} \\ \mathbf{X}_{12} \end{pmatrix} + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2 + \mathbf{E} \\ &= \mathbf{Z}_1\mathbf{B}_{11}\mathbf{X}_{11} + \mathbf{Z}_1\mathbf{B}_{12}\mathbf{X}_{12} + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2 + \mathbf{E}, \end{aligned} \quad (7.4)$$

where $\mathbf{B}_1 = (\mathbf{B}_{11} : \mathbf{B}_{12})$ and $\mathbf{X}'_1 = (\mathbf{X}'_{11} : \mathbf{X}'_{12})$. Considering the two clusters (in our case, two groups as well) separately, the model reduces to

$$\mathbf{Y}_1 = \mathbf{Z}_1\mathbf{B}_{11}\mathbf{X}'_{11} + \mathbf{E}_1,$$

and

$$\mathbf{Y}_2 = \mathbf{Z}_1\mathbf{B}_{12}\mathbf{X}'_{12} + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2 + \mathbf{E}_2,$$

for Group I and Group II, respectively. Here \mathbf{Y}_1 , \mathbf{X}'_{11} and \mathbf{E}_1 are matrices consisting of the first n_1 columns of \mathbf{Y} , \mathbf{X}_{11} and \mathbf{E} , respectively. The matrices \mathbf{Y}_2 , \mathbf{X}'_{12} , \mathbf{X}_2 and \mathbf{E}_2 consist of the last n_2 columns of \mathbf{Y} , \mathbf{X}_{12} , \mathbf{X}_2 and \mathbf{E} , respectively. Observe that $\mathbf{X}_{12} = \mathbf{X}_2$. Moreover, it is possible to show that

$$\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11} = \mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - \mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2.$$

Now consider the likelihood function for the EGCM

$$L = \gamma |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - (\mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1 + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2))(\mathbf{Y} - (\mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1 + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2))'\}}, \quad (7.5)$$

where $\gamma = (2\pi)^{-\frac{1}{2}np}$. It can be rewritten as a product of three terms,

$$L = L_1 \times L_2 \times L_3,$$

where

$$\begin{aligned} L_1 &= \gamma \exp\left\{-\frac{1}{2}tr\{\boldsymbol{\Sigma}^{-1}(\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2 - (\mathbf{Z}_1\mathbf{B}_{12}\mathbf{X}_{12} + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2))\right. \\ &\quad \left. \times (\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2 - (\mathbf{Z}_1\mathbf{B}_{12}\mathbf{X}_{12} + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2))'\}\right\}, \\ L_2 &= \exp\left\{-\frac{1}{2}tr\{\boldsymbol{\Sigma}^{-1}(\mathbf{Y}(\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - \mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2) - (\mathbf{Z}_1\mathbf{B}_{11}\mathbf{X}_{11}))\right. \\ &\quad \left. \times (\mathbf{Y}(\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - \mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2) - (\mathbf{Z}_1\mathbf{B}_{11}\mathbf{X}_{11}))'\}\right\}, \\ L_3 &= |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}tr\{\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I} - \mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1)\mathbf{Y}'\}\right\}. \end{aligned}$$

Let us now consider L_3 , which is free of the parameters specified in the hypothesis and maximize the expression to get an estimator for the unknown covariance matrix Σ . It is possible to show that the estimator that maximizes L_3 is $(n-1)\hat{\Sigma} = \mathbf{S}_1$. We now replace Σ in (7.5) by its estimator, to get the estimated likelihood, denoted by EL, and then maximize EL under H_o and $H_o \cup H_1$, to get the desired test. The maximum of the EL under H_o and $H_o \cup H_1$ are respectively given by

$$\gamma_1 |\mathbf{S}_1|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\{n\mathbf{S}_1^{-1}\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1\mathbf{Y}'\}} \quad (7.6)$$

and

$$\begin{aligned} & \gamma_1 |\mathbf{S}_1|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}ntr\left\{\mathbf{S}_1^{-1}\left(\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - (\mathbf{Z}_1\hat{\mathbf{B}}_1\mathbf{X}_1 + \mathbf{Z}_2\hat{\mathbf{B}}_2\mathbf{X}_2)\right.\right.\right. \\ & \quad \left.\left.\left.\times \left(\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - (\mathbf{Z}_1\hat{\mathbf{B}}_1\mathbf{X}_1 + \mathbf{Z}_2\hat{\mathbf{B}}_2\mathbf{X}_2)\right)'\right\}\right\}, \end{aligned} \quad (7.7)$$

where $\gamma_1 = n^{\frac{n}{2}}(2\pi)^{-\frac{1}{2}np}e^{-\frac{1}{2}np}$. Note that we can rewrite \mathbf{R}_3 and \mathbf{R}_4 as follows:

$$\mathbf{R}_3 = \mathbf{S}_1\mathbf{Z}'_1(\mathbf{Z}'_1\mathbf{S}_1\mathbf{Z}'_1)^{-1}\mathbf{Z}'_1\mathbf{Y}(\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - \mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2), \quad (7.8)$$

$$\mathbf{R}_4 = \mathbf{S}_2\mathbf{Z}'_2(\mathbf{Z}'_2\mathbf{S}_2\mathbf{Z}'_2)^{-1}\mathbf{Z}'_2\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2,$$

where \mathbf{Z}'_1 and \mathbf{Z}'_2 are matrices of full rank spanning the orthogonal complements of the column spaces of the matrices \mathbf{Z}_1 and $\mathbf{Z} = (\mathbf{Z}_1 : \mathbf{T}_1\mathbf{Z}_2)$, respectively. It is possible to show that \mathbf{R}_{34} , which denotes the sum of the residuals \mathbf{R}_3 and \mathbf{R}_4 , can be written as a difference between the observed and estimated means, i.e.,

$$\mathbf{R}_{34} = \mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1 - (\mathbf{Z}_1\hat{\mathbf{B}}_1\mathbf{X}_1 + \mathbf{Z}_2\hat{\mathbf{B}}_2\mathbf{X}_2).$$

A test statistic is defined by taking the ratio between (7.6) and (7.7), which can be simplified as

$$\frac{e^{-\frac{1}{2}ntr\{\mathbf{S}_1^{-1}\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1\mathbf{Y}'\}}}{e^{-\frac{1}{2}ntr\{\mathbf{S}_1^{-1}\mathbf{R}_{34}\mathbf{R}'_{34}\}}}, \quad (7.9)$$

where the hypothesis is rejected when the value of the ratio is small, i.e., close to zero. Note that the ratio has values between zero and one. One can also define an equivalent test by taking the logarithm of the test statistic, which can be re-written as

$$tr\{\mathbf{S}_1^{-1}\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1\mathbf{Y}'\} - tr\{\mathbf{S}_1^{-1}\mathbf{R}_{34}\mathbf{R}'_{34}\}, \quad (7.10)$$

and the hypothesis will be rejected for large values of (7.10). This formulation allows a relatively easier understanding of some distributional characteristics of the

proposed test. As such, consider the first term in (7.10) and write it as a sum of two terms as follows:

$$\begin{aligned} tr\{\mathbf{S}_1^{-1}\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1\mathbf{Y}'\} &= tr\{\mathbf{S}_1^{-1}\mathbf{Y}\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11}\mathbf{Y}'\} \\ &\quad + tr\{\mathbf{S}_1^{-1}\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{X}'_2\}, \end{aligned} \quad (7.11)$$

where \mathbf{X}_{11} is as in (7.4). Similarly, use the expressions for \mathbf{R}_3 and \mathbf{R}_4 and write the second term in (7.10) as

$$\begin{aligned} tr\{\mathbf{S}_1^{-1}\mathbf{R}_{34}\mathbf{R}'_{34}\} &= tr\{\mathbf{Y}\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11}\mathbf{Y}'\mathbf{S}_1^{-1}\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{S}_1^{-1}\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{S}_1^{-1}\} \\ &\quad + tr\{\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{S}_2^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{S}_1^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}_1^{-1}\}. \end{aligned} \quad (7.12)$$

By subtracting (7.12) from (7.11), the test statistic reduces to

$$\begin{aligned} \phi_1(\mathbf{Y}) &= tr\{\mathbf{Y}\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11}\mathbf{Y}'\mathbf{S}_1^{-1}\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{S}_1^{-1}\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{S}_1^{-1}\} \\ &\quad + tr\{\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{S}_2^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{S}_1^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}_1^{-1}\}, \end{aligned} \quad (7.13)$$

where $\mathbf{Z} = (\mathbf{Z}_1 : \mathbf{T}_1\mathbf{Z}_2)$. The hypothesis is rejected when the value of $\phi_1(\mathbf{Y})$ is large. Here it is important to note that the column spaces of $(\mathbf{Z}_1 : \mathbf{Z}_2)$ and $(\mathbf{Z}_1 : \mathbf{T}_1\mathbf{Z}_2)$ are identical [2, 3].

The test statistic given in (7.13) is always greater or equal to zero. Moreover, it is possible to see from the expression in (7.9) that the numerator is a function of $\mathbf{Y}\mathbf{X}'_1(\mathbf{X}_1\mathbf{X}'_1)^{-1}\mathbf{X}_1$, which is the observed mean. On the other hand, in the denominator we have a function of \mathbf{R}_{34} , which is the residual obtained by subtracting the estimated mean from the observed mean. This shows that the test compares the observed and estimated means, in some weighted fashion, and rejects the hypothesis when the difference between them is “small”, which is quite intuitive and in agreement with the formulation of the Lawley-Hotelling trace tests in the MANOVA and GMANOVA models.

Below, we show that the distribution of $\phi_1(\mathbf{Y})$ under the null hypothesis is independent of the unknown covariance matrix, $\mathbf{\Sigma}$. This is extremely important in applications, since the empirical distribution can be generated without the knowledge of the covariance matrix, and hence significance testing can be performed and p -values can be provided. Performance evaluation through simulations can also be performed by assuming, without loss of generality, that $\mathbf{\Sigma} = \mathbf{I}$. On the other hand, it is possible to show that the distribution under the alternative depends on $\mathbf{\Sigma}$, and hence the power of the test depends on the variance-covariance matrix. We will empirically show this through the simulations.

Now, to show that the distribution of the test statistic in (7.13) under the null is independent of $\mathbf{\Sigma}$, consider the first part of the expression and let \mathbf{Z}_1^o be a matrix of

full rank spanning the orthogonal complement to the space generated by the columns of \mathbf{Z}_1 . We can write first expression in (7.13) as

$$\begin{aligned} & tr\{\mathbf{Y}\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11}\mathbf{Y}'\mathbf{S}_1^{-1}\} \\ & - tr\{\mathbf{Y}\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11}\mathbf{Y}'\mathbf{Z}'_1(\mathbf{Z}'_1\mathbf{S}_1\mathbf{Z}'_1)^{-1}\mathbf{Z}'_1\mathbf{Y}\}. \end{aligned} \quad (7.14)$$

The first term in (7.14) is invariant under the transformation $\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}$. It is, therefore, possible to replace \mathbf{Y} by $\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}$, which shows that the distribution of this term is independent of $\boldsymbol{\Sigma}$. For the second term, by using the property of the trace function, we can rewrite it as

$$tr\{\mathbf{X}'_{11}(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11}\mathbf{Y}'\mathbf{Z}'_1(\mathbf{Z}'_1\mathbf{S}_1\mathbf{Z}'_1)^{-1}\mathbf{Z}'_1\mathbf{Y}\}.$$

Now, write $\mathbf{Z}'_1\mathbf{Y}$ as

$$(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{\frac{1}{2}}(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}\mathbf{Z}'_1\mathbf{Y}$$

and observe that we can rewrite $(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{\frac{1}{2}}(\mathbf{Z}'_1\mathbf{S}_1\mathbf{Z}'_1)^{-1}(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}$ as

$$((\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}\mathbf{Z}'_1\mathbf{Y}(\mathbf{I} - \mathbf{X}_{11}'(\mathbf{X}_{11}\mathbf{X}'_{11})^{-1}\mathbf{X}_{11})\mathbf{Y}'\mathbf{Z}'_1(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}})^{-1}.$$

Consequently, it remains to show that the distribution of $(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}\mathbf{Z}'_1\mathbf{Y}$ is independent of $\boldsymbol{\Sigma}$. Note that the expression is a linear function of a multivariate normal random variable. As a result, it is enough to show that the mean and dispersion matrices are independent of $\boldsymbol{\Sigma}$, which are shown below.

Under the null hypothesis $E[\mathbf{Y}] = \mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1 + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2 = \mathbf{0}$ which implies

$$E[(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}\mathbf{Z}'_1\mathbf{Y}] = \mathbf{0}.$$

The dispersion matrix, \mathbf{D} , is given by

$$\mathbf{D}[(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}\mathbf{Z}'_1\mathbf{Y}] = (\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}}\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1(\mathbf{Z}'_1\boldsymbol{\Sigma}\mathbf{Z}'_1)^{-\frac{1}{2}} = \mathbf{I}.$$

Suppose now, that we want to check if the quadratic term in the growth curves of the individuals in the second group (cluster) is significantly different from zero. We can formulate the hypotheses (in terms of the mean parameters) as

$$\begin{aligned} H_0 & : \mathbf{B}_2 = \mathbf{0} \\ H_1 & : \mathbf{B}_2 \neq \mathbf{0}. \end{aligned} \quad (7.15)$$

Note that this hypothesis is associated with the residual, \mathbf{R}_4 . Using a similar motivation as in the Lawley-Hotelling trace test in MANOVA, the trace test in the

GCM and the test provided in (7.13), one can argue that the statistic for testing the hypothesis in (7.15) should have the format

$$\frac{f(\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}')}{f(\mathbf{R}_4\mathbf{R}'_4)}.$$

In order to provide a formal derivation, recall the likelihood function of the EGCM given in (7.5) and maximize the product, $L_2 \times L_3$, which is independent of \mathbf{R}_4 . This gives us the estimator \mathbf{S}_2 for $n\boldsymbol{\Sigma}$. Update the likelihood by using the estimator and proceed by taking the ratio of the maximum of the estimated likelihood under H_0 and $H_0 \cup H_1$. The test statistic provided in (7.17) can be obtained by taking the logarithm of the ratio and doing some algebraic manipulations similar to $\phi_1(\mathbf{Y})$, including using the property of the trace function (e.g., for any two matrices $tr(AB) = tr(BA)$) as well as the fact that

$$\begin{aligned} \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}'_1\mathbf{S}_2^{-1}\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{S}_2^{-1} &= \mathbf{T}_1\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{T}'_1\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{T}'_1\mathbf{S}_2^{-1} \\ &+ \mathbf{S}_2\mathbf{Z}'_2(\mathbf{Z}'_2\mathbf{S}_2\mathbf{Z}'_2)^{-1}\mathbf{Z}_2, \end{aligned} \quad (7.16)$$

where \mathbf{T}_1 and \mathbf{Z}_1^o are as presented in (7.2) and (7.8), respectively. Note that the two terms on the right hand side of (7.16) are orthogonal to each other. The test statistic for testing the hypothesis in (7.15) can be formally written as

$$\phi_2(\mathbf{Y}) = tr\{\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{T}'_1\mathbf{S}_2^{-1}\mathbf{T}_1\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{T}'_1\mathbf{S}_2^{-1}\}, \quad (7.17)$$

and the hypothesis is rejected when the value of $\phi_2(\mathbf{Y})$ is large. The above test statistic is always greater or equal to zero. Similar to $\phi_1(\mathbf{Y})$, the distribution of $\phi_2(\mathbf{Y})$ is under the null hypothesis independent of the unknown covariance matrix, $\boldsymbol{\Sigma}$. To show this, note that the test statistic in (7.17) can be written as the difference between two terms as

$$\begin{aligned} &tr\{\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{G}_1(\mathbf{G}'_1\mathbf{W}_2\mathbf{G}_1)^{-1}\mathbf{G}'_1\} \\ &- tr\{\mathbf{Y}\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{G}_2(\mathbf{G}'_2\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}'_2\}, \end{aligned} \quad (7.18)$$

where

$$\begin{aligned} \mathbf{G}_{r+1} &= \mathbf{G}_r(\mathbf{G}'_r\mathbf{Z}_{r+1})^o, \quad \mathbf{G}_0 = \mathbf{I}, \\ \mathbf{W}_{r+1} &= \mathbf{Y}(\mathbf{I} - \mathbf{X}'_r(\mathbf{X}_r\mathbf{X}'_r)^{-1}\mathbf{X}_r)\mathbf{Y}', \quad r = 0, 1, 2, \dots, m-1. \end{aligned}$$

Such approaches have been discussed in a more general form in von Rosen [27]. For special cases, we refer to Hamid [2] and Hamid and von Rosen [3]. The two terms in (7.18) can be, respectively, rewritten as

$$tr\{\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{G}_1(\mathbf{G}'_1\mathbf{W}_2\mathbf{G}_1)^{-1}\mathbf{G}'_1\mathbf{Y}\}$$

and

$$\text{tr}\{\mathbf{X}'_2(\mathbf{X}_2\mathbf{X}'_2)^{-1}\mathbf{X}_2\mathbf{Y}'\mathbf{G}_2(\mathbf{G}'_2\mathbf{W}_2\mathbf{G}_2)^{-1}\mathbf{G}'_2\mathbf{Y}\}.$$

In order to show that the distribution of $\phi_2(\mathbf{Y})$ under the null hypothesis is independent of $\boldsymbol{\Sigma}$, we want to show that the distributions of the above two expressions under the null hypothesis are independent of $\boldsymbol{\Sigma}$, which is equivalent to showing that the distributions of $\mathbf{G}'_1\mathbf{Y}$ and $\mathbf{G}'_2\mathbf{Y}$ under the null hypothesis are independent of $\boldsymbol{\Sigma}$. As such, we write $\mathbf{G}'_1\mathbf{Y}$ as

$$(\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{\frac{1}{2}}(\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}}\mathbf{G}'_1\mathbf{Y}',$$

and it remains to show that the distribution of $(\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}}\mathbf{G}'_1\mathbf{Y}'$, which is a linear function of a multivariate normal random variable, is independent of $\boldsymbol{\Sigma}$. Once again, because of normality, it is enough to show that the mean and dispersion matrices are independent of $\boldsymbol{\Sigma}$.

Under the null hypothesis, $\mathbf{B}_2 = \mathbf{0}$, we have that $\mathbf{G}'_1\mathbf{Z}_1 = \mathbf{Z}'_1\mathbf{Z}_1 = \mathbf{0}$. Consequently,

$$E[(\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}}\mathbf{G}'_1\mathbf{Y}] = (\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}}\mathbf{G}'_1(\mathbf{Z}_1\mathbf{B}_1\mathbf{X}_1 + \mathbf{Z}_2\mathbf{B}_2\mathbf{X}_2) = \mathbf{0}.$$

Furthermore,

$$D[(\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}}\mathbf{G}'_1\mathbf{Y}] = (\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}}\mathbf{G}'_1\boldsymbol{\Sigma}\mathbf{G}_1(\mathbf{G}_1\boldsymbol{\Sigma}\mathbf{G}'_1)^{-\frac{1}{2}} = \mathbf{I}.$$

Similar calculations can show that the distribution of $\mathbf{G}'_2\mathbf{Y}$ is independent of $\boldsymbol{\Sigma}$.

7.4 Simulations

We performed simulations to evaluate the performance of the tests. We first generated the empirical distributions of both tests under the corresponding null hypotheses, and calculated the critical values for the tests based on 50,000 simulations. The empirical level and power of the tests were then calculated from second sets of 10,000 simulations. Several scenarios were considered in terms of sample size (n), departure from the null hypotheses and number of time points, p . We have also considered several scenarios in terms of the covariance matrix $\boldsymbol{\Sigma}$.

For the null hypothesis in (7.3), the distribution of the test statistic, $\phi_1(\mathbf{Y})$ under the null hypothesis is skewed to the left (Fig. 7.2), which is consistent across different sample sizes, dimensions of p , degrees of polynomial q and magnitudes of $\boldsymbol{\Sigma}$.

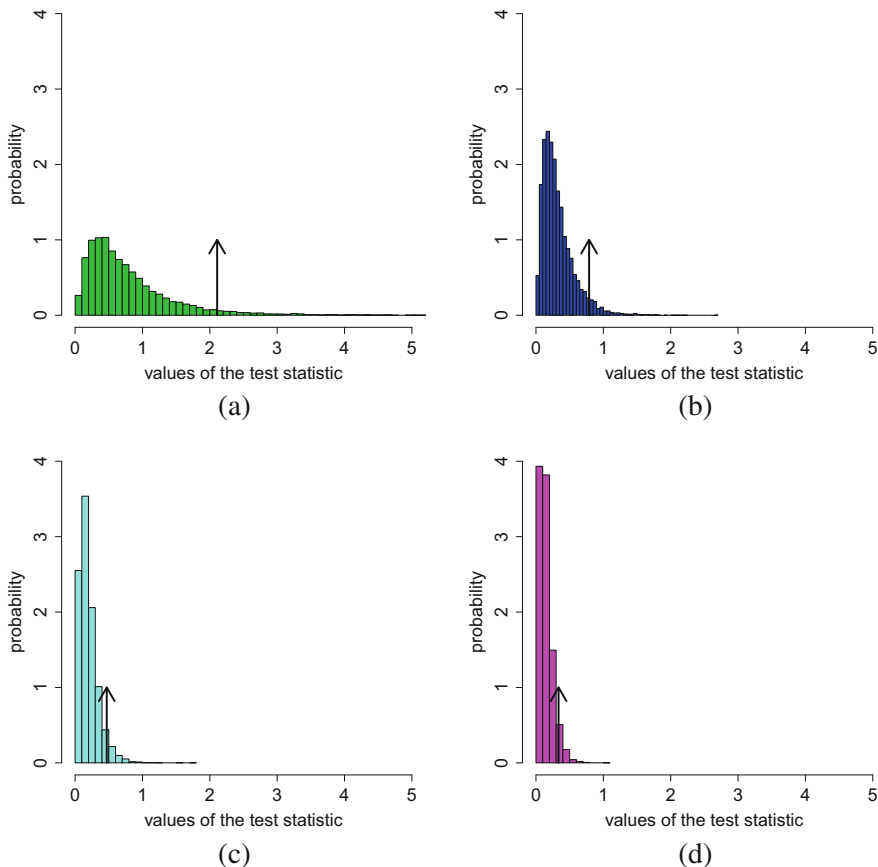


Fig. 7.2 The null distribution of $\phi_1(Y)$, where $p = 8$, and (a) $n=20$, (b) $n=30$, (c) $n=40$ and (d) $n=50$. The arrows show the critical values corresponding to the tests

The results of the simulation show that $\phi_1(Y)$ performs well and possesses all the desirable properties of a test statistic. Similar to the trace test in the GCM, our simulation results show that the test maintains the nominal level $\alpha = 0.05$ (Fig. 7.3), for all the scenarios considered.

The results also show that the test is unbiased, symmetric, monotone with respect to both n and departures from the null hypothesis (Fig. 7.4). Similar to our previous studies on the GCM and EGCM, departure from the null hypotheses is measured using the Euclidean norm of the parameter matrix [5–8]. The results also show that

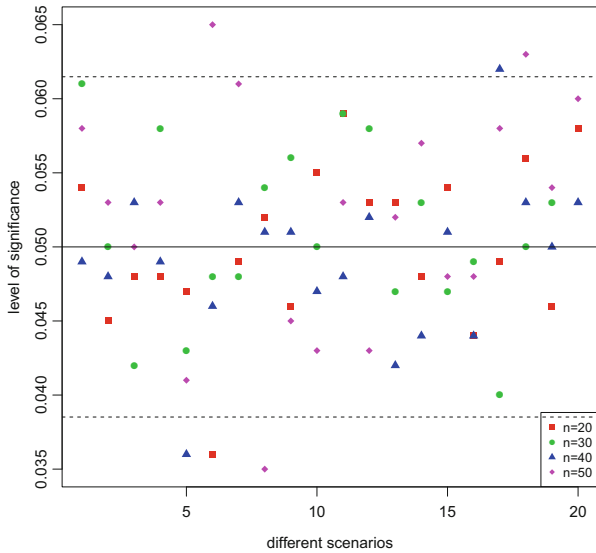


Fig. 7.3 Empirical level of $\phi_1(Y)$ and the corresponding 95% confidence bands

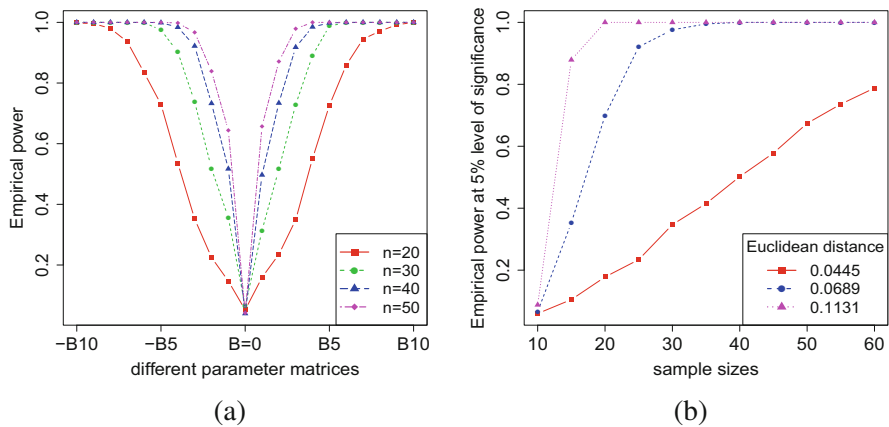


Fig. 7.4 Empirical power of $\phi_1(Y)$ with x -axis depicting (a) positive and negative scenarios to show symmetry, (b) Euclidean norm of the parameter matrix, used to measure departures from the null

the test has a reasonably good statistical power, where the test detects very small departures from the null hypothesis, even with relatively small sample sizes (Fig. 7.4 and Table 7.1). Please note, the results are consistent across all the scenarios we considered.

Table 7.1 Empirical power of $\phi_1(\mathbf{Y})$ for multiple n , p and \mathbf{B}

Euclidean distance	$p = 8$				$p = 4$			
	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
0	0.052	0.064	0.04	0.042	0.045	0.06	0.051	0.042
0.054	0.349	0.729	0.918	0.979	0.112	0.167	0.221	0.32
0.061	0.551	0.891	0.985	1.000	0.157	0.234	0.333	0.415
0.069	0.726	0.988	0.998	1.000	0.188	0.304	0.417	0.556
0.077	0.859	0.999	1.000	1.000	0.239	0.395	0.521	0.684
0.086	0.945	1.000	1.000	1.000	0.289	0.501	0.669	0.795
0.095	0.971	1.000	1.000	1.000	0.371	0.593	0.757	0.884
0.104	0.992	1.000	1.000	1.000	0.42	0.723	0.859	0.944
0.113	1.000	1.000	1.000	1.000	0.521	0.800	0.935	0.981

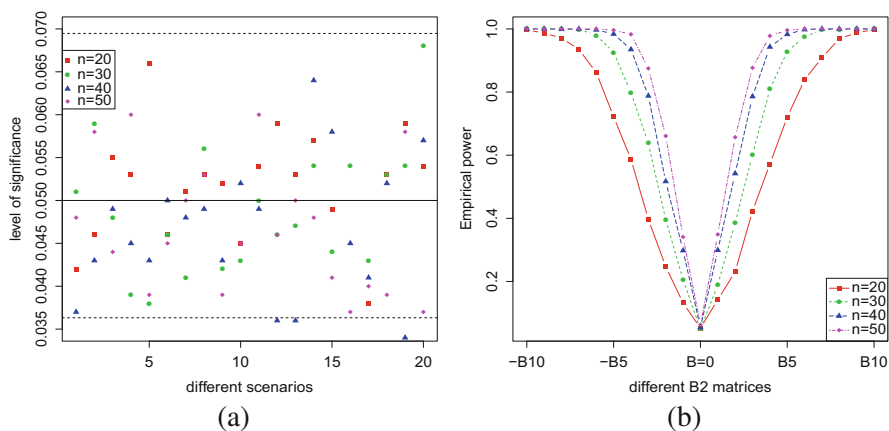


Fig. 7.5 (a) Empirical level and (b) empirical power, of $\phi_2(Y)$

Similar results were also obtained for $\phi_2(\mathbf{Y})$, where the simulation results show that the distribution of the test statistic is skewed to the left. The test statistic maintains the nominal level and has all the other desirable properties such as unbiasedness, symmetry and monotonicity (Figs. 7.5 and 7.6). Furthermore, our simulations demonstrate that the test has a very good performance, as it detects extremely small departures from the null hypothesis, with a reasonably small sample size (Fig. 7.6).

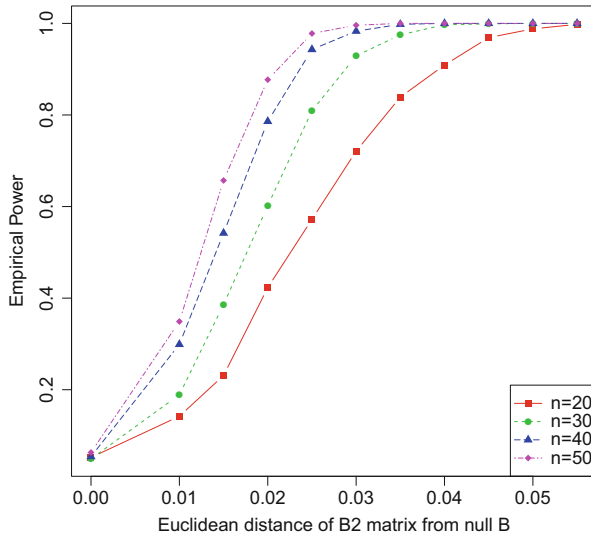


Fig. 7.6 Empirical power of $\phi_2(Y)$ with respect to Euclidean norm of the parameter matrix across four sample sizes

7.5 Empirical Example

As an empirical illustration, we consider the glucose data [32]. Data consists of measurements taken at 8 time points from 13 controls and 20 obese patients. The mean profile plots for the two groups are provided in Fig. 7.7 below. For illustration purposes, we assume two clusters, where the mean of one of the clusters (consisting of individuals in the obese group) follows a quadratic growth curve and the mean for the second cluster (consisting of individuals in the control group) follows a cubic growth curve.

Results show that the null hypothesis of zero mean for both groups (i.e., $\mathbf{B}_1 = \mathbf{0}$, $\mathbf{B}_2 = \mathbf{0}$) is rejected, with observed test statistic = 62.76, critical value = 0.80, and p -value $r < 0.0001$. For the hypothesis that the coefficient of the cubic term is zero, the observed test statistic value is 0.142 and the critical value is 0.173 (Fig. 7.8), indicating that there is no evidence to reject the null hypothesis (p -value = 0.07328) at 5% level of significance.

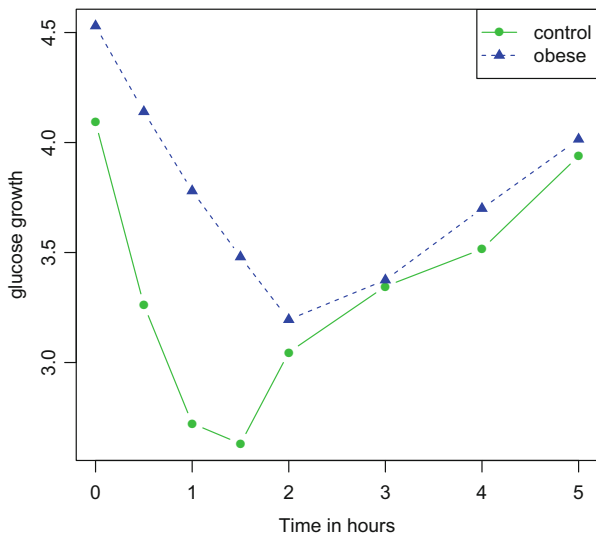


Fig. 7.7 Mean profile plots for the glucose data

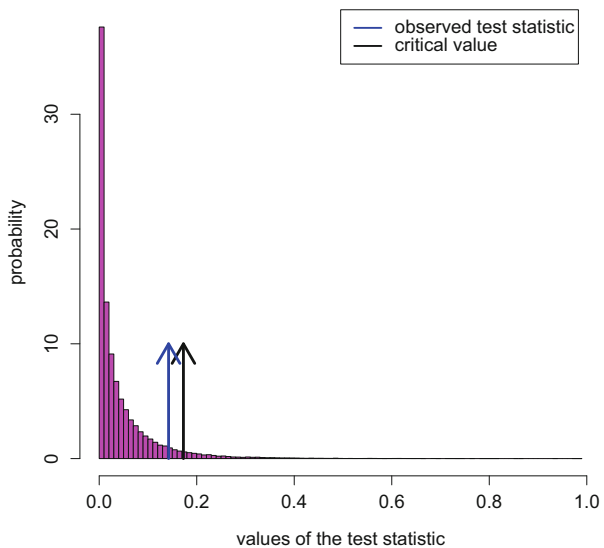


Fig. 7.8 The null distribution of $\phi_2(\mathbf{Y})$ for the glucose data

7.6 Summary

We considered the Extended Growth Curve Model, which can be applied in situations where longitudinal measurements from groups of individuals can be clustered into several groups. Using decomposed residuals as a motivation, we derived test statistics for two hypotheses related to the mean parameters of the model. The two corresponding tests can be shown to be extensions of the trace test in the Growth Curve Model, which in turn is the Lawley-Hotelling trace test.

Simulation results showed that the two tests possess all the desirable properties such as unbiasedness, symmetry and monotonicity with respect to both sample size and departures from the null hypotheses, where departure is measured using Euclidean norm. Results also demonstrate that the tests have good performances, where the tests were able to detect very small departures from the null hypotheses. The results were consistent under all scenarios considered (e.g., different covariance matrices, different values of p and q).

Although the tests are derived under special scenarios, where we assumed two clusters, the approach used in our manuscript allows extensions to several clusters (the general EGCM) to be made. Moreover, the formulation we provided in this study can also be extended to allow a more general linear hypothesis to be tested.

Acknowledgments Parts of this work was done under the supervision of Professor Dietrich von Rosen and we would like to acknowledge his contributions. This work is supported by funds from the Natural Sciences and Engineering Council of Canada (NSERC).

References

1. Fujikoshi, Y: Asymptotic expansions of the non-null distributions of three statistics in GMANOVA. *Ann. Inst. Stat. Math.* **26**, 289–297 (1974)
2. Hamid, J.S.: New residuals in multivariate bilinear models. PhD dissertation, Swedish University of Agricultural Sciences (2005). ISBN 91-576-6982-1
3. Hamid, J.S., von Rosen, D.: Residuals in the extended growth curve model. *Scand. J. Stat.* **33**(1), 121–138 (2006)
4. Hamid, J.S., Beyene, J., von Rosen, D.: A novel trace test for the mean parameters in a multivariate growth curve model. *J. Multivar. Anal.* **102**(2), 238–251 (2011)
5. Jana, S., Balakrishnan, N., Hamid, J.S.: High dimensional extension of the growth curve model and its application in genetics. *Stat. Methods Appl.* **6**(2), 273–292 (2017)
6. Jana, S., Balakrishnan, N., Hamid, J.S.: Inference in the growth curve model under multivariate skew normal distribution. *Sankhya B.* (2018). <https://doi.org/10.1007/s13571-018-0174-1>
7. Jana, S., Balakrishnan, N., Hamid, J.S.: Estimation of the parameters of the extended growth curve model under multivariate skew normal distribution. *J. Multivar. Anal.* **166**(C), 111–128 (2018)
8. Jana, S., Balakrishnan, N., Hamid, J.S.: Bayesian growth curve model useful for high-dimensional longitudinal data. *J. Appl. Stat.* **46**(5), 814–834 (2019)
9. Kariya, T.: The general MANOVA problem. *Ann. Stat.* **6**(1), 200–214 (1978)
10. Kariya, T.: *Testing in the Multivariate General Linear Model*. Kinokuniya Co., Tokyo (1985)

11. Khatri, C.G.: A note on a MANOVA model applied to the problems in growth curve. *Ann. Inst. Stat. Math.* **18**, 75–86 (1966)
12. Kollo, T., von Rosen, D.: *Advanced Multivariate Statistics with Matrices*. Springer, New York (2005)
13. Kshirsagar, A.M., Smith, W.B.: *Growth Curves*. Marcel Dekker, New York (1995)
14. Liski, E.P.: Detecting influential measurements in a growth curve model. *Biometrics* **47**, 659–668 (1991)
15. Nzabanita, J., Singull, M., von Rosen, D.: Estimation of parameters in the extended growth curve model with a linearly structured covariance matrix. *Acta Comment. Univ. Tartu. Math.* **16**, 13–32 (2012)
16. Nzabanita, J., von Rosen, D., Singull, M.: Bilinear regression model with Kronecker and linear structures for the covariance matrix. *Afr. Stat.* **10**, 827–837 (2015)
17. Nzabanita, J., von Rosen, D., Singull, M.: Extended GMANOVA model with a linearly structured covariance matrix. *Math. Methods Stat.* **24**, 280–291 (2015)
18. Ohlson, M., von Rosen, D.: Explicit estimators of parameters in the growth curve model with linearly structured covariance matrices. *J. Multivar. Anal.* **101**(5), 1284–1295 (2010)
19. Pan, J.X., Fang, K.T.: Multiple outlier detection in growth curve model with unstructured covariance matrix. *Ann. Inst. Stat. Math.* **47**, 137–153 (1995)
20. Pan, J.X., Fang, K.T.: Influential observations in the growth curve model with unstructured covariance matrix. *Comput. Stat. Data Anal.* **22**, 71–87 (1996)
21. Pan, J.X., Fang, K.T.: *Growth Curve Model and Statistical Diagnostics*. Springer, New York (2002)
22. Potthoff, R.F., Roy, S.N.: A generalized multivariate analysis of variance model useful especially for growth curves. *Biometrika* **51**, 313–326 (1964)
23. Rao, C.R.: The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–458 (1965)
24. Srivastava, M.S., Singull, M.: Test for the mean matrix in a growth curve model for high dimensions. *Commun. Stat. Theory Methods.* **46**(13), 6668–6683 (2017)
25. Srivastava, M.S., Singull, M.: Testing sphericity and intraclass covariance structures under a growth curve model in high dimension. *Commun. Stat. Simul. Comput.* **46**(7), 5740–5751 (2017)
26. von Rosen, D.: Maximum likelihood estimators in multivariate linear normal models. *J. Multivar. Anal.* **31**, 187–200 (1989)
27. von Rosen, D.: Moments of a multivariate linear model with an application to the growth curve model. *J. Multivar. Anal.* **35**, 243–259 (1990)
28. von Rosen, D.: The growth curve model: a review. *Commun. Stat. Theory Methods.* **20**, 2791–1822 (1991)
29. von Rosen, D.: Residuals in the growth curve model. *Ann. Inst. Stat. Math.* **47**, 129–136 (1995)
30. von Rosen, D.: *Bilinear Regression Analysis: An Introduction*. Springer International Publishing, Part of Springer Nature, Cham (2018)
31. Verbyla, A.P., Venables, W.N.: An extension of the growth curve model. *Biometrika* **75**(1), 129–138 (1988)
32. Zerbe, G.O.: Randomization analysis of the completely randomized design extended to growth and response curves. *J. Am. Stat. Assoc.* **74**, 215–221 (1979)