Linear Algebra and its Applications 508 (2016) 22–47

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Gilmour's approach to mixed and stochastic restricted ridge predictions in linear mixed models

Özge Kuran *, M. Revan Özkale

Çukurova University, Faculty of Science and Letters, Department of Statistics, Adana, 01330, Turkey

ARTICLE INFO

Article history:
Received 11 March 2015
Accepted 24 June 2016
Available online 29 June 2016
Submitted by W.B. Wu

MSC:
62J07
62J05

Keywords:
Multicollinearity
Mixed predictor
Stochastic restricted ridge predictor
Mean square error

ABSTRACT

This article is concerned with the predictions in linear mixed models under stochastic linear restrictions. Mixed and stochastic restricted ridge predictors are introduced by using Gilmour's approach. We also investigate assumptions that the variance parameters are not known under stochastic linear restrictions and attain estimators of variance parameters. Superiorities the linear combinations of the predictors are done in the sense of mean square error matrix criterion. Finally, a hypothetical data set is considered to illustrate the findings.

1. Introduction

Statistical models are used for the identification of the data, the determination of the mathematical form that best represents the data structure and approximately the definition of the relationship between the variables. Although the most widely-used statistical model is the linear regression model, it is sometimes possible to encounter with the data structures that accord with the models that include both fixed and random effects. In

* Corresponding author.
E-mail addresses: ozge-kuran@hotmail.com (Ö. Kuran), mrevan@cu.edu.tr (M.R. Özkale).

http://dx.doi.org/10.1016/j.laa.2016.06.040

this case, the linear mixed models are appeared and they provide flexibility in fitting models with various combinations of fixed and random effects.

Linear mixed models are also an important tool for the analysis of a broad variety of data including clustered data such as longitudinal data, repeated measures and multilevel data.

Let us consider the linear mixed model

$$y = X\beta + Zu + \varepsilon \tag{1}$$

where y is an $n \times 1$ vector of responses, X is an $n \times p$ known design matrix for the fixed effects, β is a $p \times 1$ parameter vector of fixed effects, $Z = [Z_1, \ldots, Z_b]$ with Z_i is an $n \times q_i$ design matrix for the ith random effects factor, $u = [u_1', \ldots, u_b']'$ is a $q \times 1$ vector of random effects with u_i is a $q_i \times 1$ vector such that $q = \sum\limits_{i=1}^{b} q_i$, and ε is an $n \times 1$ vector of random errors, with $E(u) = 0$ and $E(\varepsilon) = 0$. In addition it is assumed that u and ε follow independent and multivariate Gaussian distributions such that

$$\begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} G(\gamma) & 0 \\ 0 & W(\rho) \end{bmatrix} \right)$$

where $G = G(\gamma)$ and $W = W(\rho)$ are positive definite (pd) matrices, γ and ρ are $r \times 1$ and $s \times 1$ (with $s \leq n(n+1)/2$) vectors of variance parameters corresponding to u and ε, respectively. Following Patterson and Thompson [18], we write the variance–covariance matrix of y as $var(y) = \sigma^2 H$ where $H = ZGZ' + W$.

Under the assumption that the variance parameters γ, ρ and σ^2 are known, $\widehat{\beta}$ and $\widehat{u}$ are obtained as

$$\widehat{\beta} = (X'H^{-1}X)^{-1}X'H^{-1}y \tag{2}$$

$$\widehat{u} = GZ'H^{-1}(y - X\widehat{\beta}) \tag{3}$$

by Henderson [8] and Henderson et al. [9]. These predictors are called as, respectively, the best linear unbiased estimator (BLUE) and the best linear unbiased predictor (BLUP).

In linear regression model, we usually assume that the variables of fixed effects design matrix are independent. However, in practice, there may be strong or near to strong linear relationships among the variables of fixed effects design matrix. In that case the independence assumptions are no longer valid, which causes the problem of multicollinearity.

Multicollinearity is one of the serious problem in the linear regression analysis. In literature, there are various multicollinearity diagnostics. Multicollinearity diagnostics used in linear regression model were extended to linear mixed models by Sandra [23]. Gilmour et al. [5] gave an efficient computational strategy for calculating predictors and their standard errors in linear mixed models and they also considered estimation and estimability for the case when the fixed effects design matrix is not of full rank.