Linear Algebra and its Applications 543 (2018) 173–204

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

# Principal components regression and r-k class predictions in linear mixed models

M. Revan Özkale *, Özge Kuran

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

ARTICLE INFO

*Article history:*
Received 31 January 2017
Accepted 1 January 2018
Available online xxxx
Submitted by W.B. Wu

*MSC:*
62J07
62J05

*Keywords:*
Multicollinearity
Henderson's predictor
Ridge predictor
Principal components regression predictor
Linear mixed model

ABSTRACT

In this article, we propose the principal components regression and r-k class predictors, which combine the techniques of the ridge and principal components regressions in the linear mixed models. We demonstrate that the Henderson's predictors, the ridge predictors and the principal components regression predictors are special cases of the r-k class predictors. We also research assumption that the variance parameters are not known and get estimators of variance parameters. The necessary and sufficient conditions for the superiorities of the r-k class predictors over each of these three predictors are obtained by the criterion of mean square error matrix. Furthermore, we suggest tests to approve if these conditions are indeed satisfied. Finally, real data analysis and a simulation study are used to illustrate the findings.



## 1. Introduction

Linear mixed models are statistical models containing both fixed and random effects and so, they provide flexibility in fitting models with various combinations of fixed and

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

https://doi.org/10.1016/j.laa.2018.01.001

random effects. These models are often used to analyze data in a broad spectrum of areas 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, $\beta$ 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 $i$th random effects factor, $u = [u_1', \ldots, u_b']'$ is a $q \times 1$ vector of random effects with $u_i$ which is a $q_i \times 1$ vector such that $q = \sum_{i=1}^{b} q_i$, and $\varepsilon$ 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 $\varepsilon$ 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, $\gamma$ and $\rho$ are ŗ $\times 1$ and ş $\times 1$ (with ş $\leq n(n+1)/2$) vectors of variance parameters corresponding to $u$ and $\varepsilon$, respectively. Then, the variance–covariance matrix of $y$ is written as $var(y) = \sigma^2 H$ where $H = ZGZ' + W$.

$\widehat{\beta}$ 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 [3] and Henderson et al. [4]. The Henderson's predictors given by (2) and (3) are called, respectively, as the best linear unbiased estimator (BLUE) and the best linear unbiased predictor (BLUP).

Generally, the variables of design matrix for fixed effects are assumed as linearly independent. However, in practice, we may encounter with strong or near to strong linear dependencies between the variables of design matrix for fixed effects. Then, the problem of multicollinearity is said to exist. In the existence of multicollinearity, at least one main diagonal element of $(X'H^{-1}X)^{-1}$ may be quite large, which in view of $Var(\widehat{\beta}) = \sigma^2(X'H^{-1}X)^{-1}$ means that at least one element of $\widehat{\beta}$ may have a large variance, and $\widehat{\beta}$ may be far from its true value. In order to overcome these multicollinearity problems, predictors alternative to Henderson's predictors can be proposed. Additionally, Liu and Hu [6] introduced ridge method to the problem of predicting for linear mixed models and put forward the ridge predictors in linear mixed models as

$$\widehat{\beta}(k) = (X'H^{-1}X + kI_p)^{-1}X'H^{-1}y \tag{4}$$

$$\widehat{u}(k) = GZ'H^{-1}(y - X\widehat{\beta}(k)) \tag{5}$$