Heteroscedastic additive models - Estimating the fixed effects and covariance matrix parameters

Adilson Silva*[1] ⬤, Miguel Fonseca[2] ⬤

[1] *Faculty of Science and Technology, University of Cape Verde, Praia, Cape Verde*
[2] *Center for Mathematics and Applications, Faculty of Science and Technology, New University of Lisbon, Caparica, Portugal*

Abstract

This work aims to deduce estimators for the unknown parameters of fixed effects and covariance matrix structure in heteroscedastic additive design. In order to do that, the design will be projected onto the orthogonal complement of the subspace spanned by columns of design matrix for the fixed effects, and the Kronecker product will be used to produced unbiased estimators for the parameters of covariance matrix, and then such estimators used to produce an estimator for the fixed effect vector. Moreover, the coefficient of determination for both fixed effects and covariance structure will be derived. A simulation study will be conducted, and a numerical example will be explored.

1. Introduction

We approach the additive models having the following structure:

$$z = X_o\beta_o + \sum_{i=1}^{s} X_i\beta_i, \tag{1.1}$$

where $\begin{cases} \beta_o \in \mathcal{M}^{c_o \times 1} \text{ is the vector for the fixed effects (the general mean);} \\ X_o \in \mathcal{M}^{m \times c_o} \text{ is the design matrix for the fixed effects;} \\ X_i \in \mathcal{M}^{m \times c_i}, \ i = 1, \ldots, s, \text{ are the design matrix for the random effects;} \\ \beta_i \in \mathcal{M}^{c_i \times 1}, \ i = 1, \ldots, s, \text{ are the random effects due to the design} \\ \qquad \text{different groups of treatments;} \end{cases}$

It is assumed that $\beta_i \sim (0_{c_i}, \ V_i)$, where

$$\begin{aligned} V_i &= \begin{bmatrix} \sigma^2_{i,1,1} & \cdots & \sigma^2_{i,1,c_i} \\ \vdots & \ddots & \vdots \\ \sigma^2_{i,c_i,1} & \cdots & \sigma^2_{i,c_i,c_i} \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2_{i,h,l} \end{bmatrix}, \ h = 1, \ldots, c_i; \ l = 1, \ldots, c_i \text{ and } i = 1, \ldots, s, \end{aligned} \tag{1.2}$$

*Corresponding Author.
Email addresses: adilson.dasilva@docente.unicv.edu.cv (A. Silva), fmig@fct.unl.pt (M. Fonseca)

is a $c_i \times c_i$ matrix whose entry at row h and column l is $\Sigma(\beta_{i,h}, \beta_{i,l})$, with $\beta_{i,h}$ the hth element of β_i, and that the group of treatments are observed independently so that the model follows the following distribution:

$$z \sim \left(X_o \beta_o, \ \sum_{i=1}^{s} X_i V_i X_i^\top \right). \tag{1.3}$$

This is the well known heterocedastic additive model. It is known for its heteroscedastic structure within the different groups of treatments, a problem which occurs more often in data sets that have large range between the largest and smallest observed values (see [2], for example). A blatant example is the household consumption based on income: lower income households are less variable in absolute terms since they need to focus on priority necessities (may not be too many), and higher income households may have need to purchase a wide variety of luxury items, resulting in a broader spread of spending habits.

Dealing with Heteroscedasticity is a major concern in linear mixed models (See [11]), mainly because most of techniques used, such as least squares-based ones, assume that the variability within groups or sub-groups of treatments are constants, i.e. the observed data within groups or sub-groups are uncorrelated. As a matter of fact, according with several recent contributes, as [1, 4, 5, 8], the situation is often different, that is such variabilities may not be constants.

The literature is quite satisfactory with regard to addressing the heteroscedastic in linear mixed models whose the covariance structure has only one unknown matrices (see [6] and [7]). In this sense, in this work we aim to develop tools to estimate the variability in linear mixed models whose covariance structure has an arbitrary number of unknown matrices, that is, in the context of model (1.3), our purpose is to provide the users with estimators for the unknown variation within the different groups of treatments denoted by matrices $V_1, \ldots, V_s$, and for the unknown fixed effects denote by β_o. Firstly, a restricted version of the model will be used to derivate the announced estimators for the matrix components $V_1, \ldots, V_s$ and then, nextly, such estimators together with the general least square estimation will be used to derive an estimator for β_o in (1.1). In what follows, parameters to measure the quality of the adjustment for each one of the estimators will be derived. At the end, a simulation study will be conducted in order to text the estimators performance.

The following *notations* will be used without any additional comments:

- $\Sigma(x)$ denotes the variance-covariance matrix of a random vector x, i.e $\Sigma(x) = E[(x - E(x))(x - E(x))^\top]$;
- $\mathbf{0}_{n,m}$ denotes an $n \times m$ matrix, while $\mathbf{0}_n$ denotes a null vector of dimension n; $\mathbf{1}_n$ denotes a vector of ones with dimension n;
- $\mathbf{J}_n$ denotes a $n \times n$ matrix of ones;
- $z \sim (w, \Sigma)$ denotes a random vector z with mean w, and variance-covariance matrix Σ; z_t denotes its tth element;
- $r(A)$ denotes the rank of a matrix A.
- $\sum_{i \neq j}^{n}$ denotes $\sum_{i=1}^{n} \sum_{j=1}^{n}$ for $i \neq j$;

2. Inference

Let $P_{R(X_o)}$ be the projection matrix onto the subspace spanned by the columns of X_o. Then, there exists a matrix $A \in \mathcal{M}^{m \times n}$, whose columns are the orthogonal eigenvectors associated to the null eigenvalues of $P_{R(X_o)}$, satisfying, therefore,

$$A^\top A = I_{m-r} \text{ and } AA^\top = I_m - P_{R(X_o)} = P_{R(X_o)^\perp},$$

where $r = rank\left(P_{R(X_o)}\right)$, $n = m - r$ and $P_{R(X_o)^\perp}$ is the projection matrix onto the orthogonal complement of the subspace spanned by the columns of X_o.

Lets also consider the restricted mixed linear model

$$y = A^\top z = \sum_{i=1}^{s} A^\top X_i \beta_i = \sum_{i=1}^{s} X_i^o \beta_i, \tag{2.1}$$

with $X_i^o = A^\top X_i$, $i = 1, \ldots, s$, whose distribution is given as

$$y \sim \left(\mathbf{0}_n, \sum_{i=1}^{s} X_i^o V_i X_i^{o\top}\right), \tag{2.2}$$

that is $E(y) = \mathbf{0}_n$ and $\Sigma(y) = E[yy^\top] = \sum_{i=1}^{s} X_i^o V_i X_i^{o\top}$.

Since the parameters of the covariance structure to be estimated, $V_1, \ldots, V_s$, do not depends on the fixed effect structure of the model, and in order to reduce the bias and the complexity of the model for the algebraic manipulation, it is convenient for us to work with the restricted model (2.1) in the search for estimators for variation of the different groups of treatments.

2.1. Estimating the parameters of covariance matrix - $V_1, \ldots, V_s$

Letting $X_i^o = [x_{i,1}^o \ldots x_{i,c_i}^o]$, $i = 1, \ldots, s$, where $x_{i,t}^o$ denotes the tth column of X_i^o, and recalling V_i in (1.2), we foundthe that

$$X_i^o V_i X_i^{o\top} = \sum_{h=1}^{c_i} \sum_{l=1}^{c_i} \sigma_{i,h,l}^2 x_{i,h}^o x_{i,l}^{o\top}, \tag{2.3}$$

and, therefore,

$$\Sigma(y) = E[yy^\top] = \sum_{i=1}^{s} \sum_{h=1}^{c_i} \sum_{l=1}^{c_i} \sigma_{i,h,l}^2 x_{i,h}^o x_{i,l}^{o\top}. \tag{2.4}$$

According with Proposition A.2 of Appendix, together with vec operator (see Definition A.1) and expectation properties, the following result follows:

Proposition 2.1. $E[vec(yy^\top)] = \sum_{i=1}^{s} \sum_{h=1}^{c_i} \sum_{l=1}^{c_i} \sigma_{i,h,l}^2 x_{i,l}^o \otimes x_{i,h}^o$.

Proof.

$$\begin{aligned}
E\left[vec\left(yy^\top\right)\right] &= vec\left(E\left[yy^\top\right]\right) \\
&= vec\left(\sum_{i=1}^{s} \sum_{h=1}^{c_i} \sum_{l=1}^{c_i} \sigma_{i,h,l}^2 x_{i,h}^o x_{i,l}^{o\top}\right) \\
&= \sum_{i=1}^{s} \sum_{h=1}^{c_i} \sum_{l=1}^{c_i} \sigma_{i,h,l}^2 vec\left(x_{i,h}^o x_{i,l}^{o\top}\right) \\
&= \sum_{i=1}^{s} \sum_{h=1}^{c_i} \sum_{l=1}^{c_i} \sigma_{i,h,l}^2 x_{i,l}^o \otimes x_{i,h}^o. \tag{2.5}
\end{aligned}$$

The third row of (2.5) is due to the linearity of vec operator, while the fourth one due to Proposition A.2 of Appendix. □

Lets define a new vector:

$$\begin{aligned}
y^{oo} &= vec\left[yy^\top\right] \\
&= y \otimes y. \tag{2.6}
\end{aligned}$$

Then we have the following result.

Proposition 2.2. *Let $N_n = \frac{1}{2}K_{nn} + \frac{1}{2}I_{n^2}$, where K_{nn} is a commutation matrix (see Definition A.5 in Appendix) and add normality to the restricted model (2.2), i.e $y^{oo} \sim \mathcal{N}_n\left(\mathbf{0}_n, \ \sum_{i=1}^{s} X_i^o V_i X_i^{o\top}\right)$. Then,*

$$\begin{aligned}
\Sigma(y^{oo}) &= N_n \sum_{i=1}^{s} \sum_{j=1}^{s} \left[\left(X_i^o V_i X_i^{o\top}\right) \otimes \left(X_j^o V_j X_j^{o\top}\right)\right] \\
&= N_n \sum_{i=1}^{s} \sum_{j=1}^{s} \left(X_i^o \otimes X_j^o\right)(V_i \otimes V_j)\left(X_i^o \otimes X_j^o\right)^\top.
\end{aligned} \tag{2.7}$$

Proof. According with Theorem 9.20 of [9], and Proposition A.3 (due to [10]) of Appendix, we found that

$$\begin{aligned}
\Sigma(y^{oo}) &= \Sigma(y \otimes y) \\
&= N_n \left[\left(\sum_{i=1}^{s} X_i^o V_i X_i^{o\top}\right) \otimes \left(\sum_{i=1}^{s} X_i^o V_i X_i^{o\top}\right)\right] \\
&= N_n \sum_{i=1}^{s} \sum_{j=1}^{s} \left[\left(X_i^o V_i X_i^{o\top}\right) \otimes \left(X_j^o V_j X_j^{o\top}\right)\right] \\
&= N_n \sum_{i=1}^{s} \sum_{j=1}^{s} \left(X_i^o \otimes X_j^o\right)(V_i \otimes V_j)\left(X_i^o \otimes X_j^o\right)^\top.
\end{aligned} \tag{2.8}$$

$\square$

Now we define the following Matrix and vector, respectively:

$$\begin{aligned}
X^{oo} &= \left[vec\left(x_{1,1}^o x_{1,1}^{o\top}\right)\ldots\ldots vec\left(x_{s,c_s}^o x_{s,c_s}^{o\top}\right)\right] \\
&= \left[x_{1,1}^o \otimes x_{1,1}^o \ldots\ldots x_{s,c_s}^o \otimes x_{s,c_s}^o\right],
\end{aligned} \tag{2.9}$$

$$\begin{aligned}
\sigma^{oo} &= vec([V_1 \ldots V_s]) \\
&= \left[\sigma_{1,1,1}^2 \ldots\ldots \sigma_{s,c_s,c_s}^2\right]^\top,
\end{aligned} \tag{2.10}$$

recalling that $V_i = \begin{bmatrix} \sigma_{i,1,1}^2 & \cdots & \sigma_{i,1,c_i}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{i,c_i,1}^2 & \cdots & \sigma_{i,c_i,c_i}^2 \end{bmatrix}, \ i = 1,\ldots,s.$

One must note that X^{oo} is a matrix whose columns are the columns of the design matrices for the random effects $X_1,\ldots,X_s$, whereas σ^{oo} is a vector whose entries are the elements of the random effects variance-covariance matrices $V_1,\ldots,V_s$.

According with Proposition 2.1 we have that $E(y^{oo}) = X^{oo}\sigma^{oo}$; because of this, and recalling the covariance structure of y^{oo} in (2.8), we may propose a new model:

$$y^{oo} = X^{oo}\sigma^{oo} + e, \tag{2.11}$$

where

$$e \sim \mathcal{N}_{n^2}(\mathbf{0}_{n^2}, \ \Delta_n), \ \text{with}$$

$$\Delta_n = N_n \sum_{i=1}^{s} \sum_{j=1}^{s} \left[\left(X_i^o V_i X_i^{o\top}\right) \otimes \left(X_j^o V_j X_j^{o\top}\right)\right].$$

In the last square estimation we are interested in choosing the parameter $\widetilde{\sigma^{oo}}$ which minimizes the quadratic error

$$(y^{oo} - X^{oo})^\top (y^{oo} - X^{oo}).$$

Since $\sigma^{oo} \in \mathbb{R}^k$, with $k = \sum_{i=1}^s c_i$, is allowed to be chosen from anywhere in $\mathbb{R}^k$, i.e. in the set of all vectors $\eta = X^{oo}\sigma^{oo}$ ranging over columns space of X^{oo}. Then, we have to choose η closest to the vector y^{oo} so as to minimizes $\|y^{oo} - X^{oo}\sigma^{oo}\|^2$. According with Theorem A.4, the unique closest vector is Py, where P is the projection matrix onto columns space of X^{oo}, which is given by $P = X^{oo}(X^{oo\top}X^{oo})^+X^{oo\top}$; that is, $Py = X^{oo}(X^{oo\top}X^{oo})^+X^{oo\top}y^{oo}$ is the closest vector to the vector η in the columns space of X^{oo}, and so the well known ordinary least square estimator (**OLS**) for σ^{oo} is given by

$$\widetilde{\sigma^{oo}} = (X^{oo\top}X^{oo})^+X^{oo\top}y^{oo} \tag{2.12}$$

2.2. Estimating the fixed effects - β_o

In this section we use the **LSE** for $\sigma^{oo} \in \mathbb{R}^k$ to deduce an estimator for β_o.

Recalling the linear mixed model (1.1), and letting $\Sigma_o = \sum_{i=1}^s X_i V_i X_i^\top$, the **GLSE** for β_o in such a model is:

$$argmin_{\beta_o \in \mathbb{R}^{c_o}}(z - X_o\beta_o)^\top \Sigma_o^\top (z - X_o\beta_o),$$

which is achieved when

$$\widehat{\beta_o} = \left(X_o^\top \Sigma_o^+ X_o\right)^+ X_o^\top \Sigma_o^+ z.$$

Once $\widehat{\beta_o}$ depends on the unknown parameters $V_1 \dots Vs$, with

$$
\begin{aligned}
V_i &= \begin{bmatrix} \sigma_{i,1,1}^2 & \cdots & \sigma_{i,1,c_i}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{i,c_i,1}^2 & \cdots & \sigma_{i,c_i,c_i}^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{i,h,l}^2 \end{bmatrix}, \ h = 1, \dots, c_i; \ l = 1, \dots, c_i \text{ and } i = 1, \dots, s, \tag{2.13}
\end{aligned}
$$

using $\widehat{\beta_o}$ to estimate β_o is not so practical, since it depends on the unknown parameters $V_1 \dots Vs$.

Now, noting that $\sigma^{oo} = Vec[V_1 \dots V_s]$, and so its **LSE** (deduced in the previous section) given by $\widetilde{\sigma^{oo}} = Vec[\widetilde{V_1} \dots \widetilde{V_s}]$, with

$$
\begin{aligned}
\widetilde{V_i} &= \begin{bmatrix} \widetilde{\sigma_{i,1,1}^2} & \cdots & \widetilde{\sigma_{i,1,c_i}^2} \\ \vdots & \ddots & \vdots \\ \widetilde{\sigma_{i,c_i,1}^2} & \cdots & \widetilde{\sigma_{i,c_i,c_i}^2} \end{bmatrix} \\
&= \begin{bmatrix} \widetilde{\sigma_{i,h,l}^2} \end{bmatrix}, \ h = 1, \dots, c_i; \ l = 1, \dots, c_i \text{ and } i = 1, \dots, s, \tag{2.14}
\end{aligned}
$$

if we replace V_i with its corresponding unbiased estimator $\widetilde{V_i}$, $i = 1, \dots, s$, an unbiased estimator for Σ_o, which we denoted by $\widetilde{\Sigma_o}$, is

$$\widetilde{\Sigma_o} = \sum_{i=1}^s X_i \widetilde{V_i} X_i^\top.$$

We are now in position to propose a new estimator for β_o; it is denoted by $\widetilde{\beta_o}$, and defined as follows:

$$\widetilde{\beta_o} = \left(X_o^\top \widetilde{\Sigma_o}^+ X_o\right)^+ X_o^\top \widetilde{\Sigma_o}^+ z. \tag{2.15}$$

3. The quality of adjustments

In this section we aim to discuss the quality of estimators introduced in the previous section, through their respective coefficient of determination. Several slightly different definitions can be found in the literature, but, usually, they are all equivalent in the essencial (the readers are invited to consult others sources). The coefficient of determination, usually denoted by R^2 (and pronounced "r-squared"), is the proportion of the variance in the dependent (response) variable that is predictable from the independent (explanatory) variables; it is used as a guideline to assess how well a model explain and predict future outcomes, based on the proportion of total variation of outcomes explained by the model.

Precisely, given the fraction of the unexplained variance (FUV), i..e. the fraction of variance of the dependent variable which cannot be correctly predicted (explained) by the explanatory (independent) variables, defined as $FUV = \frac{S_e^2}{S_T^2}$, where S_e^2 is the *residual/regression sums of square (sample variance)* and S_T^2 the *total sum of squares (variance of the observed data)*; it quantifies how much the data points vary around their mean. R^2 is given by

$$R^2 = 1 - FUV = 1 - \frac{S_e^2}{S_T^2}.$$

Once S_T^2 cannot be zero (otherwise R^2 is not defined/applicable), R^2 is a decreasing function of S_e^2, that is the higher the residual sum of square (sample variance) is, the smaller the R^2 is, and so the lower the quality of regression is. R^2 cannot be larger than 1, it is at most equal 1.

- R^2 is equal 1 when the residual sum of square (sample variance) is zero; this indicates that the regression is perfect, that is the fitted model explain perfectly all variability in depend variable;
- R^2 is equal 0 when the sample variance is equal to the variance of the observed data. It indicates that predicting the observed data with the underling fitted model is not better than using the sample mean of observed data as a prediction tool, that is there is no "linear" relationship between the regressors and the response variable, and so the the fitted model may be discarded;
- $R^2 = \alpha$, with $0 < \alpha < 1$, indicates that $100\alpha\%$ of variation in the response variable can be very well explained by the explanatory variable, and the remain $100(1-\alpha)\%$ cannot be explained by such a fitted model

It worths to remark that the coefficient of determination gives an estimate of the relationship between movements of dependent variable based on the independent variable's movement; it does not tell whether the chosen model is good or bad, nor will it tell whether the data and predictors are biased.

3.1. R squared of $\widetilde{\sigma^{oo}}$

Recall that $\widetilde{\sigma^{oo}} = (X^{oo\top} X^{oo})^+ X^{oo\top} y^{oo}$ is the OLS for σ^{oo} in the model

$$y^{oo} = X^{oo}\sigma^{oo} + e.$$

Let y_t^{oo} be the tth element of vector y^{oo} and $\mu = \frac{1}{k}\sum_{t=1}^{k} y_t^{oo} = \frac{1}{k}I_k^\top y^{oo}$ the mean of the observation vector y^{oo}. Then, the *total sum of square, $S_T^2(\sigma)$,* for the model is given by

$$
\begin{aligned}
S_T^2(\sigma) &= \frac{1}{k}[y^{oo} - \mu\mathbf{1}_k]^\top [y^{oo} - \mu\mathbf{1}_k] \\
&= \frac{1}{k}\left(y^{oo\top} y^{oo} - 2\mu\mathbf{1}_k + k\mu^2\right),
\end{aligned}
\tag{3.1}
$$

and the *residual sum of square* , $S_e^2(\sigma)$, given by

$$
\begin{aligned}
S_e^2(\sigma) &= \frac{1}{k}[y^{oo} - X^{oo}\widetilde{\sigma^{oo}}]^\top [[y^{oo} - X^{oo}\widetilde{\sigma^{oo}}] \\
&= \frac{1}{k}y^{oo\top}[I_k - X^{oo}(X^{oo\top}X^{oo})^+ X^{oo\top}]y^{oo} \\
&= \frac{1}{k}y^{oo\top}P_{R(X)^\perp}y^{oo}.
\end{aligned}
\tag{3.2}
$$

Now we are ready to set the coefficient of determination for $\widetilde{\sigma^{oo}}$, denote here by $R^2\left(\widetilde{\sigma^{oo}}\right)$:

$$
\begin{aligned}
R^2\left(\widetilde{\sigma^{oo}}\right) &= 1 - \frac{S_e^2(\sigma)}{S_T^2(\sigma)} = 1 - \frac{\frac{1}{k}y^{oo\top}P_{R(X)^\perp}y^{oo}}{\frac{1}{k}\left(y^{oo\top}y^{oo} - 2\mu\mathbf{1}_k + k\mu^2\right)} \\
&= 1 - \frac{y^{oo\top}P_{R(X)^\perp}y^{oo}}{y^{oo\top}y^{oo} - 2\mu\mathbf{1}_k + k\mu^2}
\end{aligned}
$$

3.2. R squared of $\widetilde{\beta_o}$

In order to discuss the R^2 for $\widetilde{\beta_o}$, we recall the initial additive model (1.1):

$$
z = X_o\beta_o + \sum_{i=1}^{s} X_i\beta_i
\tag{3.3}
$$

and the estimator $\widetilde{\beta_o} = \left(X_o^\top\widetilde{\Sigma_o}^+ X_o\right)^+ X_o^\top\widetilde{\Sigma_o}^+ z$ for β_o. Let z_t be the tth elements of vector z and $\zeta = \frac{1}{m}\sum_{t=1}^m z_t = \frac{1}{m}z^\top I_m$ the mean of observations in vector z. Then the *total sum of square*, $S_T^2(\sigma)$, for the model is given by

$$
\begin{aligned}
S_T^2(\sigma) &= \frac{1}{m}[z - \eta\mathbf{1}_m]^\top [z - \zeta\mathbf{1}_m] \\
&= \frac{1}{m}z^\top z - 2\zeta\mathbf{1}_m + m\eta^2
\end{aligned}
\tag{3.4}
$$

and the *residual sum of square* , $S_e^2(\sigma)$, given by

$$
\begin{aligned}
S_e^2(\sigma) &= \frac{1}{k}[z - X_o\widetilde{\beta_o}]^\top [z - X_o\widetilde{\beta_o}] \\
&= \frac{1}{k}z^\top \left[I_m - 2P_o\widetilde{\Sigma_o}^+ + \widetilde{\Sigma_o}^+ P_o^2\widetilde{\Sigma_o}^+\right] z \\
&= \frac{1}{k}z^\top P_o^* z,
\end{aligned}
\tag{3.5}
$$

where $P_o = X_o(X_o^\top\widetilde{\Sigma_o}^+ X_o)^+ X_o^\top$ and $P_o^* = I_m - 2P_o\widetilde{\Sigma_o}^+ + \widetilde{\Sigma_o}^+ P_o^2\widetilde{\Sigma_o}^+$.

Finally, the coefficient of determination for $\widetilde{\beta_o}$, denoted here by $R^2(\widetilde{\beta_o})$, id given by

$$
\begin{aligned}
R^2(\widetilde{\beta_o}) &= 1 - \frac{S_e^2(\beta)}{S_T^2(\beta)} = 1 - \frac{\frac{1}{m}z^\top P_o^* z}{\frac{1}{m}\left(z^\top z - 2\zeta\mathbf{1}_m + m\zeta^2\right)} \\
&= 1 - \frac{z^\top P_o^* z}{z^\top z - 2\zeta\mathbf{1}_m + m\zeta^2}.
\end{aligned}
$$

4. Simulations

4.1. Simulation study

Here we carry simulation study in order to test the unbiased estimator (2.12). It will be conducted in two particular linear mixed model. Specifically, the estimator (2.12) will be applied in a simulated *balanced "one-way" random design* (see subsection 4.1.1) and a simulated *unbalanced "one-way" random design* (subsection 4.1.2).

4.1.1. Balanced "one-way" random design. Lets suppose the data come from the particular random "one-way" balanced design

$$z = X_o\beta_o + X_1\beta_1 + X_2e, \tag{4.1}$$

where $s = 2$, $c_1 = 3$, $c_2 = 9$, $X_o = \mathbf{1}_{c_2}$, $X_o = \mathbf{1}_{c_2}$ a vector of ones, $X_1 = \begin{bmatrix} \mathbf{1}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{1}_3 \end{bmatrix}$ and

$X_2 = \mathbf{I}_9$, with $\beta_1 \sim (\mathbf{0}_{c_1}, V_1)$ and $e \sim (\mathbf{0}_{c_2}, V_2)$. Also suppose that

$$V_1 = \begin{bmatrix} 0.2 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.65 \\ 0.1 & 0.65 & 0.3 \end{bmatrix} \text{ and } V_2 = \begin{bmatrix} 0.1 & 0.5 & 0.3 & 0.3 & 0.2 & 0.3 & 0.2 & 0.3 & 0.7 \\ 0.5 & 0.4 & 0.3 & 0.2 & 0.3 & 0.3 & 0.2 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.2 \\ 0.3 & 0.2 & 0.3 & 0.4 & 0.3 & 0.3 & 0.2 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.3 & 0.3 & 0.5 & 0.3 & 0.3 & 0.2 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.4 & 0.3 & 0.9 & 0.2 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.3 & 0.3 & 0.5 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.2 & 0.9 & 0.3 & 0.5 & 0.3 \\ 0.7 & 0.3 & 0.2 & 0.2 & 0.3 & 0.2 & 0.3 & 0.3 & 0.4 \end{bmatrix}. \tag{4.2}$$

Thus, $P_{R(X_o)} = \mathbf{I}_{c_2} - \frac{J_{c_2}}{c_2}$,

$$A = \begin{bmatrix} 0.9428 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.118 & -0.232 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.906 \\ -0.118 & -0.389 & 0.796 & 0.000 & 0.000 & 0.000 & -0.180 & -0.238 \\ -0.118 & -0.389 & -0.554 & 0.000 & 0.000 & 0.000 & 0.000 & -0.238 \\ -0.118 & -0.389 & -0.243 & 0.000 & 0.000 & 0.000 & 0.779 & -0.238 \\ -0.118 & 0.350 & 0.000 & 0.216 & 0.000 & 0.839 & 0.000 & -0.0481 \\ -0.118 & 0.350 & 0.000 & 0.467 & -0.707 & -0.178 & 0.000 & -0.0481 \\ -0.118 & 0.350 & 0.000 & 0.467 & 0.707 & -0.178 & 0.000 & -0.0481 \\ -0.118 & 0.350 & 0.000 & 0.719 & 0.000 & -0.483 & 0.000 & -0.0481 \end{bmatrix}, \tag{4.3}$$

$$X_1^o = A^\top X_1 = \begin{bmatrix} 0.707 & -0.354 & -0.354 \\ -0.622 & -0.429 & 1.051 \\ 0.796 & -0.796 & 0.000 \\ -0.000 & 0.216 & -0.216 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.839. & -0.839 \\ -0.180 & 0.180 & 0.000 \\ 0.668 & -0.524 & -0.144 \end{bmatrix}, \text{ and}$$

$$X_2^o = A^\top X_2 = \begin{bmatrix} 0.943 & -0.118 & -0.118 & -0.118 & -0.118 & -0.118 & -0.118 & -0.118 & -0.118 \\ 0.000 & -0.232 & -0.389 & -0.389 & -0.389 & -0.389 & -0.389 & -0.389 & -0.389 \\ 0.000 & 0.000 & 0.796 & -0.554 & -0.243 & 0.000 & 0.000. & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.216. & -0.467 & -0.467 & 0.719 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.707 & 0.707 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.839 & -0.180 & -0.180 & -0.483 \\ 0.000 & 0.000 & -0.180 & -0.599 & 0.780 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.906 & -0.238 & -0.238 & -0.238 & -0.481 & -0.048 & -0.048 & -0.048 \end{bmatrix}. \tag{4.4}$$

Then, the restricted design is given as

$$y = X_1^o\beta_1 + X_2^o e. \tag{4.5}$$

We computed

$$X^{o1} = [x_{1,1}^o \otimes x_{1,1}^o \ldots x_{1,c_1}^o \otimes x_{1,c_1}^o] \in \mathcal{M}^{c_2-1 \times c_1^2} \text{ and}$$

$$X^{o2} = [x_{2,1}^o \otimes x_{2,1}^o \ldots x_{2,c_2}^o \otimes x_{2,c_2}^o] \in \mathcal{M}^{(c_2-1)^2 \times c_2^2},$$

(using *kronecker* function in *R Studio*), and then we found $X_{oo} \in \mathcal{M}^{(c_2-1)^2 \times (c_1^2 + c_2^2)}$ (a matrix of dimension 64×90) by concatenating X^{o1} and X^{o2} (using the function *cbindX* in *R Studio*). For the simulation purpose we assume that $\beta_1 \sim \mathcal{N}(\mathbf{0}_{c_1}, V1)$ and $e \sim \mathcal{N}(\mathbf{0}_{c_2}, V_2)$.

The design (4.5) is reproduced 1000 times, and for each observed y the estimator (2.12) is applied and the matrices V_1 and V_2 are estimated. After that, the average of the estimated values for the parameters are computed, as well as the correspondents standard deviations of the estimated values. The achieved (unbiased) estimates for V_1 and V_2 are respectively

$$\tilde{V}_1 = \begin{bmatrix} 0.196 & 0.410 & 0.089 \\ 0.410 & 0.519 & 0.671 \\ 0.089 & 0.671 & 0.296 \end{bmatrix} \text{ and } \tilde{V}_2 = \begin{bmatrix} 0.131 & 0.516 & 0.270 & 0.325 & 0.201 & 0.311 & 0.182 & 0.320 & 0.64 \\ 0.516 & 0.42 & 0.314 & 0.200 & 0.303 & 0.293 & 0.221 & 0.293 & 0.292 \\ 0.270 & 0.314 & 0.421 & 0.309 & 0.283 & 0.313 & 0.332 & 0.294 & 0.202 \\ 0.325 & 0.200 & 0.309 & 0.389 & 0.317 & 0.289 & 0.204 & 0.311 & 0.215 \\ 0.201 & 0.303 & 0.283 & 0.317 & 0.521 & 0.302 & 0.332 & 0.183 & 0.310 \\ 0.311 & 0.293 & 0.313 & 0.289 & 0.302 & 0.400 & 0.306 & 1.058 & 0.222 \\ 0.182 & 0.221 & 0.332 & 0.204 & 0.332 & 0.306 & 0.485 & 0.319 & 0.273 \\ 0.320 & 0.293 & 0.294 & 0.311 & 0.183 & 1.058 & 0.319 & 0.511 & 0.313 \\ 0.640 & 0.292 & 0.202 & 0.215 & 0.310 & 0.222 & 0.273 & 0.313 & 0.455 \end{bmatrix}.$$

Taking a look at the absolute value of the "bias quantities" of V_1 and V_2, respectively given by

$$b_{V_1} = |V_1 - \tilde{V}_1| = \begin{bmatrix} 0.004 & 0.01 & 0.011 \\ 0.010 & 0.019 & 0.021 \\ 0.011 & 0.021 & 0.004 \end{bmatrix} \text{ and }$$

$$b_{V_2} = |V_2 - \tilde{V}_2| = \begin{bmatrix} 0.031 & 0.016 & 0.030 & 0.025 & 0.001 & 0.011 & 0.018 & 0.020 & 0.060 \\ 0.016 & 0.020 & 0.014 & 0.000 & 0.003 & 0.007 & 0.021 & 0.007 & 0.008 \\ 0.030 & 0.014 & 0.021 & 0.009 & 0.017 & 0.013 & 0.032 & 0.006 & 0.002 \\ 0.025 & 0.000 & 0.009 & 0.011 & 0.017 & 0.011 & 0.004 & 0.011 & 0.015 \\ 0.001 & 0.003 & 0.017 & 0.017 & 0.021 & 0.002 & 0.032 & 0.017 & 0.010 \\ 0.011 & 0.007 & 0.013 & 0.011 & 0.002 & 0.000 & 0.006 & 0.158 & 0.022 \\ 0.018 & 0.021 & 0.032 & 0.004 & 0.032 & 0.006 & 0.015 & 0.019 & 0.027 \\ 0.020 & 0.007 & 0.006 & 0.011 & 0.017 & 0.158 & 0.019 & 0.011 & 0.013 \\ 0.060 & 0.008 & 0.002 & 0.015 & 0.010 & 0.022 & 0.027 & 0.013 & 0.055 \end{bmatrix}$$

it becomes clear that our estimator (see (2.12)) produces estimates significantly accurate.

4.1.2. Unbalanced "one-way" random design. Now we suppose that the data come from the particular *random "one-way" unbalanced design*

$$z^* = X_o^* \beta_o + X_1 * \beta_1 + X_2^* e, \tag{4.6}$$

where $s = 2$, $c_1 = 4$, $c_2 = 9$, $X_o^* = \mathbf{1}_{c_2}$, $X_2^* = \mathbf{1}_{c_2}$ a vector of ones, $X_1^* = \begin{bmatrix} \mathbf{1}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_4 & \mathbf{1}_4 & \mathbf{0}_4 & \mathbf{0}_4 \\ \mathbf{0}_2 & \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 \\ \mathbf{0}_1 & \mathbf{0}_1 & \mathbf{0}_1 & \mathbf{1}_1 \end{bmatrix}$

and $X_2 = \mathbf{I}_9$, with $\beta_1 \sim (\mathbf{0}_{c_1}, V_1^*)$ and $e \sim (\mathbf{0}_{c_2}, V_2^*)$. Also suppose that

$$V_1^* = \begin{bmatrix} 0.2 & 0.4 & 0.1 & 0.5 \\ 0.4 & 0.5 & 0.65 & 0.45 \\ 0.1 & 0.65 & 0.3 & 0.5 \\ 0.5 & 0.45 & 0.5 & 0.2 \end{bmatrix} \text{ and } V_2^* = V_2 = \begin{bmatrix} 0.1 & 0.5 & 0.3 & 0.3 & 0.2 & 0.3 & 0.2 & 0.3 & 0.7 \\ 0.5 & 0.4 & 0.3 & 0.2 & 0.3 & 0.3 & 0.2 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.2 \\ 0.3 & 0.2 & 0.3 & 0.4 & 0.3 & 0.3 & 0.2 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.3 & 0.3 & 0.5 & 0.3 & 0.3 & 0.2 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.4 & 0.3 & 0.9 & 0.2 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.3 & 0.3 & 0.5 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.2 & 0.9 & 0.3 & 0.5 & 0.3 \\ 0.7 & 0.3 & 0.2 & 0.2 & 0.3 & 0.2 & 0.3 & 0.3 & 0.4 \end{bmatrix}.$$

It worts to remind that V_1^* was obtained by introducing the fourth raw $[0.0\ 0.45\ 0.5\ 0.2]$ and the fourth column $[0.0\ 0.45\ 0.5\ 0.2]^\top$ into V_1.

The matrix A is given as in the previous case (see (4.3)).

$$X_1^{0*} = A^\top X_1^* = \begin{bmatrix} 0.825 & -0.471 & -0.236 & -0.118 \\ -0.232 & -0.818 & 0.701 & 0.350 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.216 & -0.935 & 0.718 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.839 & -0.356 & -0.483 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.906 & -0.762 & -0.096 & -0.481 \end{bmatrix} \quad \text{and } X_2^{o*} = X_2^o \text{ (see (4.4))}.$$

Then, the restricted design is given as

$$y^* = X_1^{o*}\beta_1 + X_2^{o*}e. \tag{4.7}$$

To compute $X^{oo*} \in \mathcal{M}^{(c_2-1)^2 \times c_2^2}$ we proceed as in the case of X^{oo} (previous case). For the simulation purpose we assume that $\beta_1 \sim \mathcal{N}(\mathbf{0}_{c_1}, V1)$ and $e \sim \mathcal{N}(\mathbf{0}_{c_2}, V2)$.

The design (4.5) is reproduced 1000 times, and for each observed y^* the estimator (2.12) is applied and the matrices V_1^* and V_2* are estimated. After that, as in the previous case, the average of the estimated values for the parameters are computed, as well as the correspondents standard deviations of the estimated values. The achieved (unbiased) estimates for V_1^* and V_2^* are respectively

$$V_1^* = \begin{bmatrix} 0.209 & 0.421 & 0.093 & 0.517 \\ 0.421 & 0.481 & 0.622 & 0.434 \\ 0.093 & 0.622 & 0.288 & 0.511 \\ 0.517 & 0.434 & 0.511 & 0.214 \end{bmatrix} \text{ and}$$

$$V_2^* = \begin{bmatrix} 0.142 & 0.492 & 0.285 & 0.312 & 0.240 & 0.317 & 0.222 & 0.295 & 0.811 \\ 0.492 & 0.408 & 0.315 & 0.214 & 0.331 & 0.286 & 0.213 & 0.295 & 0.319 \\ 0.285 & 0.315 & 0.428 & 0.333 & 0.301 & 0.310 & 0.298 & 0.296 & 0.211 \\ 0.312 & 0.214 & 0.333 & 0.398 & 0.273 & 0.292 & 0.198 & 0.310 & 0.192 \\ 0.240 & 0.331 & 0.301 & 0.273 & 0.497 & 0.324 & 0.286 & 0.222 & 0.319 \\ 0.317 & 0.286 & 0.310 & 0.292 & 0.324 & 0.412 & 0.321 & 1.101 & 0.214 \\ 0.222 & 0.213 & 0.298 & 0.198 & 0.286 & 0.321 & 0.531 & 0.301 & 0.294 \\ 0.295 & 0.295 & 0.296 & 0.310 & 0.222 & 1.101 & 0.301 & 0.511 & 0.302 \\ 0.811 & 0.319 & 0.211 & 0.192 & 0.319 & 0.214 & 0.294 & 0.302 & 0.445 \end{bmatrix}.$$

Taking a look at the absolute value of the bias quantities of V_1^* and V_2^*, respectively given by

$$b_{V_1^*} = |V_1^* - \tilde{V}_1^*| = \begin{bmatrix} 0.009 & 0.021 & 0.007 & 0.017 \\ 0.021 & 0.019 & 0.028 & 0.016 \\ 0.007 & 0.028 & 0.012 & 0.011 \\ 0.017 & 0.016 & 0.011 & 0.014 \end{bmatrix} \text{ and}$$

$$b_{V_2^*} = |V_2^* - \tilde{V}_2^*| = \begin{bmatrix} 0.042 & 0.008 & 0.015 & 0.012 & 0.040 & 0.017 & 0.022 & 0.005 & 0.111 \\ 0.008 & 0.008 & 0.015 & 0.014 & 0.031 & 0.014 & 0.013 & 0.005 & 0.019 \\ 0.015 & 0.015 & 0.028 & 0.033 & 0.001 & 0.010 & 0.002 & 0.004 & 0.011 \\ 0.012 & 0.014 & 0.033 & 0.002 & 0.027 & 0.008 & 0.002 & 0.010 & 0.008 \\ 0.040 & 0.031 & 0.001 & 0.027 & 0.003 & 0.024 & 0.014 & 0.022 & 0.019 \\ 0.017 & 0.014 & 0.010 & 0.008 & 0.024 & 0.012 & 0.021 & 0.201 & 0.014 \\ 0.022 & 0.013 & 0.002 & 0.002 & 0.014 & 0.021 & 0.031 & 0.001 & 0.006 \\ 0.005 & 0.005 & 0.004 & 0.010 & 0.022 & 0.201 & 0.001 & 0.011 & 0.002 \\ 0.111 & 0.019 & 0.011 & 0.008 & 0.019 & 0.014 & 0.006 & 0.002 & 0.045 \end{bmatrix}.$$

Clearly, as we may see here, our estimator (see (2.12)) also produces estimates significantly accurate in unbalanced designs.

4.2. A numerical example

In this section we give an example of a situation in which the tools discussed in the previous sections are useful.

Lets consider that, for some country an engineer had the duty to study the strain (stress) of glass cathode supports on the production machines. Since the available source is restricted it was allowed to (randomly) choose at most four machines, say M_i, $i = 1, 2, 3, 4$, all over the country. For each machine M_i, a_i heads was randomly selected, and for each head i, a_{ij}, $i = 1, 2, 3, 4$, $j = 1, \ldots, a_i$, glass was observed and the results organized in the following tables. This is an example of an unbalanced "two-way" neste design (the heads are nested within the heads).

Table 1. Selected heads for each Machine: $a_1 = 2$ heads within M_1; $a_2 = 3$ heads within M_2; $a_3 = 4$ heads within M_3; $a_4 = 2$ heads within M_4.

Machines	M_1		M_2			M_3				M_4	
Heads	H_{11}	H_{12}	H_{21}	H_{22}	H_{23}	H_{31}	H_{32}	H_{33}	H_{34}	H_{41}	H_{42}

Table 2. Observed glasses for each head: $a_{11} = 5$ glasses for head H_{11}; $a_{12} = 3$ glasses for head H_{12}; $a_{21} = 4$ glasses for head H_{21}; $a_{22} = 3$ glasses for head H_{22}; $a_{23} = 3$ glasses for head H_{23}; $a_{31} = 6$ glasses for head H_{31}; $a_{32} = 3$ glasses for head H_{32}; $a_{33} = 3$ glasses for head H_{33}; $a_{34} = 2$ glasses for head H_{34}; $a_{41} = 3$ glasses for head H_{41}; $a_{42} = 4$ glasses for head H_{42}.

Heads	H_{11}	H_{12}	H_{21}	H_{22}	H_{23}	H_{31}	H_{32}	H_{33}	H_{34}	H_{41}	H_{42}
	2	10	10	2	7	5	4	7	7	1	6
	8	5	12	1	1	0	1	5	8	4	0
	0	3	9	1	12	0	7	2		3	15
	1		7			2					9
	5					3					
						1					

The "two-way" design for this data is

$$z = \beta X_o + X_1\alpha + X_2\beta + X_3\epsilon, \tag{4.8}$$

where:

- β is the model general mean;
- $z = [z_{111} \ldots z_{123}\ z_{211} \ldots z_{233}\ z_{311} \ldots z_{342}\ z_{411} \ldots z_{424}]^\top$ is the vector for the observed glass cathode supports (data);
- $\alpha = [\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4]^\top$ is the vector for the effects due to the four machines;
- $\beta = [\beta_{11}\ \beta_{12}\ \beta_{21}\ \beta_{22}\ \beta_{23}\ \beta_{31} \ldots \beta_{34} \ldots \beta_{41}\ \beta_{42}]^\top$ is the vector for the effects due to the $\sum_{i=1}^4 a_i = 11$ heads;
- $\epsilon = [\epsilon_{111} \ldots \epsilon_{123}\ \epsilon_{211} \ldots \epsilon_{233}\ \epsilon_{311} \ldots \epsilon_{342}\ \epsilon_{411} \ldots \epsilon_{424}]^\top$ is the vector for the errors;
- $X_o = \mathbf{1}_{39}$ is the design matrix (vector) for general mean;
- $X_1 = \begin{bmatrix} \mathbf{1}_8 & \mathbf{0}_8 & \mathbf{0}_8 & \mathbf{0}_8 \\ \mathbf{0}_{10} & \mathbf{1}_{10} & \mathbf{0}_{10} & \mathbf{0}_{10} \\ \mathbf{0}_{14} & \mathbf{0}_{14} & \mathbf{1}_{14} & \mathbf{0}_{14} \\ \mathbf{0}_7 & \mathbf{0}_7 & \mathbf{0}_7 & \mathbf{1}_7 \end{bmatrix}$ is the design matrix for the effects due to the different machines effects;
- $X_2 = \begin{bmatrix} \mathbf{1}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\ \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_3 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{1}_4 & \mathbf{0}_4 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_4 \\ \mathbf{0}_3 & \ldots & \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_3 \\ \mathbf{0}_3 & \ldots & \ldots & \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_3 \\ \mathbf{0}_6 & \ldots & \ldots & \ldots & \mathbf{0}_6 & \mathbf{1}_6 & \mathbf{0}_6 & \ldots & \ldots & \ldots & \mathbf{0}_6 \\ \mathbf{0}_3 & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 & \ldots & \ldots & \mathbf{0}_3 \\ \mathbf{0}_3 & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 & \ldots & \mathbf{0}_3 \\ \mathbf{0}_2 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_2 & \mathbf{1}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_3 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 \\ \mathbf{0}_4 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0}_4 & \mathbf{1}_4 \end{bmatrix}$ is the design matrix for the effects due to the different heads nested within different machines;
- $X_3 = I_{39}$ the design matrix for the errors effects.

It is assumed that the α, β and ϵ are independently observed. More over,

$$\alpha \sim (\mathbf{0}_4,\ V_1), \text{ where } V_1 = \begin{bmatrix} \sigma_{1,1,1}^2 & \cdots & \sigma_{1,1,4}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{1,4,1}^2 & \cdots & \sigma_{1,4,4}^2 \end{bmatrix};$$

$$\beta \sim (\mathbf{0}_{11},\ V_2), \text{ where } V_2 = \begin{bmatrix} \sigma_{2,1,1}^2 & \cdots & \sigma_{2,1,4}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{2,11,1}^2 & \cdots & \sigma_{1,11,11}^2 \end{bmatrix};$$

$$\epsilon \sim (\mathbf{0}_{39},\ V_3), \text{ where } V_3 = \begin{bmatrix} \sigma_{3,1,1}^2 & \cdots & \sigma_{3,1,39}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{3,39,1}^2 & \cdots & \sigma_{3,39,39}^2 \end{bmatrix},$$

with $\sigma_{3,i,j}^2 = 0$, for $i \neq j$.

V_i, $i = 1, 2, 3$, and μ are the model unknown parameters. In the next steps we will use the estimators proposed in the previous section to estimated them.

Let B be the matrix whose the columns are the eigenvectors associated to the null eigenvalues of $P_{R(X_o)} = \frac{1}{39}\mathbf{J}_{39}$. Then,

$$B^\top B = I_{38} \text{ and } BB^\top = P_{R(X_o)^\perp} = I_{39} - \frac{1}{39}\mathbf{J}_{39}.$$

The new model is:

$$y = B^\top z \sim (\mathbf{0}_{38}, \Sigma^o), \tag{4.9}$$

where $\Sigma^o = \sum_{i=1}^3 X_i^o V_i X_i^{o\top}$, with

$$X_i^o = B^\top X_i = [x_{i,1}^o \ldots x_{i,c_i}^o], i = 1, 2, 3, \text{ and } c_1 = 4,\ c_2 = 11, c_3 = 39.$$

We found that $\widetilde{V_1} = \begin{bmatrix} 0.0004 & -0.0528 & 0.0018 & 0.0505 \\ -0.0528 & 1.3374 & -0.1997 & -1.0849 \\ 0.0018 & -0.1997 & 0.0091 & 0.1888 \\ 0.0505 & -1.0849 & 0.1888 & 0.8456 \end{bmatrix}$,

$$\widetilde{V_2} = \begin{bmatrix} 2.15 & -1.84 & -6.73 & 4.84 & -2.64 & 4.15 & 0.95 & 0.02 & -3.81 & 2.73 & 0.17 \\ -1.84 & 1.52 & 5.89 & -3.84 & 2.43 & -3.54 & -0.84 & -0.06 & 3.13 & -2.50 & -0.35 \\ -6.73 & 5.89 & 20.59 & -16.02 & 7.64 & -12.95 & -2.86 & 0.09 & 12.22 & -8.00 & 0.13 \\ 4.84 & -3.84 & -16.02 & 9.11 & -7.02 & 9.38 & 2.33 & 0.32 & -7.84 & 7.12 & 1.62 \\ -2.64 & 2.43 & 7.64 & -7.02 & 2.44 & -5.07 & -0.99 & 0.18 & 4.99 & -2.61 & 0.66 \\ 4.15 & -3.54 & -12.95 & 9.38 & -5.07 & 7.97 & 1.86 & 0.05 & -7.41 & 5.26 & 0.29 \\ 0.95 & -0.84 & -2.86 & 2.33 & -0.99 & 1.86 & 0.39 & -0.03 & -1.69 & 1.02 & -0.13 \\ 0.02 & -0.06 & 0.09 & 0.32 & 0.18 & 0.05 & -0.03 & -0.04 & -0.09 & -0.19 & -0.25 \\ -3.81 & 3.13 & 12.22 & -7.84 & 4.99 & -7.41 & -1.69 & -0.09 & 6.29 & -5.09 & -0.71 \\ 2.73 & -2.50 & -8.00 & 7.12 & -2.61 & 5.26 & 1.02 & -0.19 & -5.09 & 2.82 & -0.54 \\ 0.17 & -0.35 & 0.13 & 1.62 & 0.66 & 0.29 & -0.13 & -0.25 & -0.71 & -0.54 & -0.89 \end{bmatrix},$$

and

$$\widetilde{V_3} = \begin{bmatrix} V_{31} \\ V_{32} \\ V_{33} \\ V_{34} \end{bmatrix}, \text{ where}$$

$$V_{31} = [4.55 \quad 11.02 \quad 18.40 \quad 10.48 \quad -1.22 \quad 28.48 \quad -1.91 \quad -0.06 \quad 6.58]^\top,$$

$$V_{32} = [32.73 \quad -3.50 \quad -17.66 \quad -2.71 \quad 3.21 \quad 3.21 \quad 1.85 \quad 7.39 \quad 52.24 \quad -6.40]^\top,$$

$$V_{32} = [13.22 \quad 13.22 \quad -0.63 \quad -4.55 \quad 5.30 \quad 0.04 \quad 11.81 \quad 6.27 \quad 6.40 \quad 0.24]^\top,$$

and

$$V_{34} = [6.01 \quad -1.21 \quad 4.86 \quad 9.07 \quad -2.70 \quad -0.77 \quad 2.31 \quad 19.85 \quad 6.39 \quad 0.23]^\top.$$

The negatives values in the estimated matrices variance components is due to the non-uniqueness of the generalized inverse (instead of Moore-Penrose inverse) used to compute the inverse of $\widetilde{\Sigma}_o$.

Finally, using the estimator in (2.15), we found that $\widetilde{\beta} = 5.3360$ is an estimate for β. As it can be quickly seen, the mean of z ($\sum_{i=1}^{39} z_i$, with z_i the ith observation of vector z) is 4.7692, which is 0.5 less then $\widetilde{\beta}$. We also found that:

- $R^2\left(\widetilde{\sigma^{oo}}\right) = 1$, that is the linear model (2.11) perfectly explain all variability in the vector $y^{oo} = y \otimes y$;
- $R^2\left(\widetilde{\beta}\right) = 0.9596$, that is 95.96% of variation in the response variable z is very well explained by the the linear mixed model (4.8), while the remaining 4.04% cannot be explained by such a model.

5. Final comments

This work reveals an efficient method to estimate the variability in Heteroscedastic linear mixed models whose covariance structure has an arbitrary number of unknown matrices (Section 2.2); it is well known that the hetoroscidasticity is a major concern in this kind of models, mainly because of the limitation of the available techniques. The method proposed here has successfully passed the quality of adjustments test as we may see through the theoretical results (Section 3), which is corroborated by the simulation study carried out at Section 4. Thus, it seems that the discussed method constitutes an useful tool for variances components in such a kind of models.

References

[1] A.C. Atkinson and R.D. Cook, *D-optimum designs for heteroscedastic linear models*, Amer.Statist. Assoc. **90** (429), 204-212, 1995.

[2] G.E. Battese and B.P. Bonyhady, *Estimation of household expenditure functions: An application of a class of heteroscedastic regression models*, Economic Record **57** (1), 1-22, 1981.

[3] C. Brenton, *Linear Models. The Theory and Application of Analysis of Variance*, John Wiley & Sons, Inc., 2008.

[4] M. Carapeto and W. Holt, *Testing for heteroscedasticity in regression models*, J. Appl. Stat. **30** (1), 13-20, 2003.

[5] K.S. Gordon, *An efficient algorithm for REML in heteroscedastic regression*, J. Comput. Graph. Statist. **11** (4), 836-847, 2002.

[6] A.C. Harvey, *Estimating regression models with multiplicative heteroscedasticity*, Econometrica **44** (3), 461-465, 1976.

[7] S.D. Horn, R.A. Horn and D.B. Duncan, *Estimating heteroscedastic variances in linear models*, J. Amer. Statist. Assoc. **70** (350), 380-385, 1975.

[8] D.J. Nott, M. Tran and C. Leng, *Variational approximation for heteroscedastic linear models and matching pursuit algorithms*, Stat. Comput. **22** (2), 497-512, 2012.

[9] J. Schott, *Matrix Analysis for Statistics*, John Wiley & Sons, Inc., 1997.

[10] A. Silva, *Variance Components Estimation in Mixed Linear Models*, Ph.D Thesis, New University of Lisbon, 2017.

[11] A.H. Welsh, R.J. Carroll and D. Ruppert, *Fitting heteroscedastic regression models*, J. Amer. Statist. Assoc. **89** (425), 100-116, 1994.

Appendix

Here we recall a few needed notions on specific topics of mathematics, and discuss a few results that was used in this work main body without additional comments.

A.1. Vectorization of a matrix - vec operator

Definition A.1. The vectorization of a $n \times m$ matrix (the vec operator), denoted $vec()$, is a linear transformation which converts such a matrix into a $nm \times 1$ (column) vector, by stacking its columns on top of one another. More precise, given a $n \times m$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \text{ its vectorization is given as}$$

$$vec(A) = [a_{11}\ a_{12}\dots a_{1m}a_{21}\ a_{22}\dots a_{2m}\dots a_{n1}\ a_{n2}\dots a_{nm}]^{\top}. \tag{A.1}$$

A.2. Kronecker product

Proposition A.2. *Let a and b be any vectors of reals. Then,*

(1) $vec[ba^{\top}] = a \otimes b$;

(2) $vec[aa^{\top}] = a \otimes a$, *when, in particular, $a = b$.*

See [9].

Proposition A.3. *Let $A_j \in \mathcal{M}^{m_j \times h_j}$, $B_j \in \mathcal{M}^{h_j \times q_j}$, $C_a \in \mathcal{M}^{o \times p}$, and $D_b \in \mathcal{M}^{u \times v}$, $j = 1, \dots, r > 2$, $a = 1, \dots, s$, $b = 1, \dots, t$, with s and t denoting natural numbers. Then,*

(a)

$$\left(\bigotimes_{j=1}^{r} A_j \right) \left(\bigotimes_{i=1}^{r} B_i \right) = \left(\bigotimes_{j=1}^{r} A_j B_j \right) \in \mathcal{M}^{m \times q},$$

where $m = \prod_{j=1}^{r} m_j$ and $q = \prod_{j=1}^{r} q_j$;

(b)

$$\left(\sum_{a=1}^{s} C_a \right) \otimes \left(\sum_{b=1}^{t} D_b \right) = \sum_{a=1}^{s} \sum_{b=1}^{t} C_a \otimes D_b$$

Proof: See [9] and [10].

A.3. Orthogonal projection

Theorem A.4. *Let $y \in \mathbb{R}^n$ and suppose that the subspace $\mathcal{S} \subset \mathbb{R}$ has an orthogonal projection matrix P. Then the closest vector in $\mathcal{S}$ to y, in euclidean distance, is Py uniquely.*

Proof: Lev $x \in \mathcal{S}$. Put

$$y - x = (Py - x) + (y - Py)$$

and note that

$$(Py - x) \in \mathcal{S} \text{ and } y - Py = (I_n - P)y \in \mathcal{S}^\perp.$$

With $\|.\|$ denoting the euclidean norm, and according with Theorem 2.9 of [3], we have that

$$\begin{aligned} \|y - x\|^2 &= \|P(y - x)\|^2 + \|(I_n - P)(y - x)\|^2 \\ &= \|Py - x\|^2 + \|y - Py\|^2 \\ &\geq \|y - Py\|^2. \end{aligned} \tag{A.2}$$

The equality in (A.2) holds if and only if $x = Py$, and so the proof is completed.

A.4. Commutation matrix - K_{nn}

Let H_{ij} be the $n \times n$ matrix that has its only nonzero elements, a one, in the $(i, j)th$ position. Such a matrix can be conveniently expressed in terms of columns from the identity matrix I_n, in the following manner: if e_{ij} denotes the ith column of I_n, then $H_{ij} = e_{in} e_{jn}^\top$.

Definition A.5. The $n^2 \times n^2$ commutation matrix, denoted by K_{nn}, is given by

$$K_{nn} = \sum_{i=1}^{n} \sum_{j=1}^{n} H_{ij} \otimes H_{ij}^\top \tag{A.3}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(e_{in} e_{jn}^\top \right) \otimes \left(e_{jn} e_{in}^\top \right). \tag{A.4}$$

The commutation matrix is a special case of permutation matrices. Additional consideration can be found in [9].