

# Estimation of a Linear Model in Terms of Intra-Class Correlations of the Residual Error and the Regressors

Juha Lappi

University of Eastern Finland, Joensuu, Finland  
Email: juha.lappi.sjk@gmail.com

**How to cite this paper:** Lappi, J. (2022) Estimation of a Linear Model in Terms of Intra-Class Correlations of the Residual Error and the Regressors. *Open Journal of Statistics*, 12, 188-199.  
<https://doi.org/10.4236/ojs.2022.122013>

**Received:** January 29, 2022

**Accepted:** April 11, 2022

**Published:** April 14, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

**Objectives:** The objective is to analyze the interaction of the correlation structure and values of the regressor variables in the estimation of a linear model when there is a constant, possibly negative, intra-class correlation of residual errors and the group sizes are equal. Specifically: 1) How does the variance of the generalized least squares (GLS) estimator (GLSE) depend on the regressor values? 2) What is the bias in estimated variances when ordinary least squares (OLS) estimator is used? 3) In what cases are OLS and GLS equivalent. 4) How can the best linear unbiased estimator (BLUE) be constructed when the covariance matrix is singular? The purpose is to make general matrix results understandable. **Results:** The effects of the regressor values can be expressed in terms of the intra-class correlations of the regressors. If the intra-class correlation of residuals is large, then it is beneficial to have small intra-class correlations of the regressors, and vice versa. The algebraic presentation of GLS shows how the GLSE gives different weight to the between-group effects and the within-group effects, in what cases OLSE is equal to GLSE, and how BLUE can be constructed when the residual covariance matrix is singular. Different situations arise when the intra-class correlations of the regressors get their extreme values or intermediate values. The derivations lead to BLUE combining OLS and GLS weighting in an estimator, which can be obtained also using general matrix theory. It is indicated how the analysis can be generalized to non-equal group sizes. The analysis gives insight to models where between-group effects and within-group effects are used as separate regressors.

## Keywords

Best Linear Unbiased Estimator, Ordinary Least-Squares, Generalized Least Squares, Singular Correlation Matrix, Between-Group Effects, Within-Group Effects

## 1. Introduction

Ordinary least squares (OLS) and generalized least squares (GLS) are basic methods to estimate coefficients of regressor variables in regression equations that are linear with respect to coefficients. Let OLSE and GLSE denote the corresponding estimators. OLS minimizes the sum of squared residuals (deviations from the estimated model), while GLS minimizes the estimation variances of the coefficient estimates if the residual errors (deviations from the true model) have non-equal variances and/or are correlated between different observations. OLS estimates are unbiased also for cases where GLS estimates have smaller variances, but the estimated variances of the coefficient estimates can be biased, leading to biased t- and F-tests. GLS estimation utilizes the inverse of the covariance matrix of residual errors. Thus, GLSE does not exist if the inverse does not exist, *i.e.*, the covariance matrix is singular. If the inverse exists, then GLSE is the best linear unbiased estimator, BLUE. It requires deeper matrix algebra to derive BLUE when GLSE does not exist. Further complications are caused if there are linear dependencies between predictor variables, mathematically if the model matrix, to be presented shortly, does not have full column rank. In practice, such difficulties can be avoided by dropping linearly dependent predictors. It is assumed here that there are no such dependencies.

In the history of statistics, attention has been given to analyze what are the necessary and sufficient conditions that guarantee that OLSE is BLUE. Reference [1] describes the influential role of C. R. Rao in this question.

When analyzing grouped data, a common model for correlated residual errors is to assume that there is a constant correlation between all observations of each group, and members of different groups are uncorrelated. Let  $\rho_e$  denote this intra-class correlation, and let  $i$  denote group, and  $j$  denote an observation number within a group. A non-negative intra-class correlation of residual error is implied by assuming a variance component model  $e_{ij} = v_i + \varepsilon_{ij}$ , where  $v_i$  is the random group effect with mean zero and variance  $\sigma_v^2$  and  $\varepsilon_{ij}$  is a random individual effect with variance  $\sigma_\varepsilon^2$  and uncorrelated with  $v_i$ . Thus,  $\rho_e = \sigma_v^2 / (\sigma_v^2 + \sigma_\varepsilon^2)$ . The intra-class correlation of residual errors can be also negative. If  $\rho_e$  is positive, then two observations taken from a given group are, on average, more alike than two observations taken from different groups. When  $\rho_e$  is negative, then two observations taken from a given group are, on average, more different than two observations taken randomly from the whole population. Positive  $\rho_e$  implies that group averages have a larger variance than a random sample of equal size from the whole population, while negative  $\rho_e$  implies that group averages vary less than averages of samples from the whole population. On the extreme, all group means are equal, and all the variation is between individuals. While assuming a positive  $\rho_e$ , it is not necessary to specify the group size (of course, when dealing with data, group size plays an important role). The effect of a negative  $\rho_e$  is always dependent on the group size, and it does not make sense to assume constant negative  $\rho_e$  for different group sizes. The possibility of a neg-

ative  $\rho_e$  is too often ignored. Even if the overall  $\rho_e$  is positive, due to natural variation between groups, the competition between individuals from limited resources is decreasing  $\rho_e$ , and this competition should be described when analyzing grouped data. The lower limit of  $\rho_e$  can be obtained from the condition  $\text{var}\left(\sum_{j=1}^n e_{ij}\right) = n\sigma_e^2 + n(n-1)\rho_e\sigma_e^2 \geq 0$ , which implies that  $\rho_e \geq -1/(n-1)$ . The upper limit of  $\rho_e$  is 1. This paper assumes a constant  $\rho_e$  and a constant group size, which allows a straightforward treatment of negative  $\rho_e$ .

Even if the matrix formulas of OLSE and GLSE are simple, it may remain intuitively unclear how the correlation structure of residual errors and the values of the regressor variables interact. It is shown that this interaction can be described in terms of the empirical intra-class correlations of the regressor variables. Closed form algebraic formulas are first derived for a special case where a simple linear model is estimated. These derivations are then generalized to a model with several regressors. They give a concrete and easily understood explanation and illustration for the matrix theory results which go beyond the mathematical expertise of most scientists who are just interested to apply statistical methods in research

This paper discusses the following questions: 1) How do the variances of the GLS estimates depend on the intra-class correlation of residuals and on the values of the x-variable in a simple linear model? 2) What are the ratios of the GLS variances and OLS variances? 3) What is the bias in the estimated variances of the coefficient estimates when OLS is used? 4) How does GLS utilize the group means of the regressors and the deviations from the group means? 5) How this formulation shows why OLS is sometimes equal to GLS? 6) How can a BLUE be constructed when the correlation matrix of residual errors is singular, *i.e.*,  $\rho_e = 1$  or  $\rho_e = -1/(n-1)$  and the intra-call correlations of regressors gets any values? Similarly, as with  $\rho_e$ , the extreme values of the regressor correlations need to be considered separately. This last BLUE part may not be of interest to practicing scientists, but it may serve as an introduction to the BLUE problem for researchers, who cannot fluently read matrix theory. For matrix theory experts this part may be trivial or self-evident.

Reference [2] compared the GLS estimation of a simple regression line with OLS estimation using differences in the predictor variable. This work is here put into a general context. The results reported in [3] regarding the bias of F-tests when OLS is used, are exemplified and generalized. Algebraic formulas show how OLSE, GLSE, BLUE and singular correlation matrices are related, and thus illustrate the matrix results presented in [4] [5], and [6]. Reference [5] provides the matrix conditions that are used to show that a suggested estimator is BLUE.

Let us then move to more formal definitions. Let us assume a model  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ , where  $\mathbf{y}$  is a  $N$  dimensional random vector,  $\mathbf{X}$  is  $N \times p$  model matrix,  $\mathbf{b}$  is  $p$  dimensional fixed coefficient vector,  $E(\mathbf{e}) = \mathbf{0}$ , and  $\text{var}(\mathbf{e}) = \sigma^2\mathbf{V}$ . If  $\mathbf{b}$  is estimated with the GLS estimator  $\hat{\mathbf{b}}_{\text{GLS}} = (\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}^{-1}\mathbf{y}$ , then  $\text{var}(\hat{\mathbf{b}}_{\text{GLS}}) = \sigma^2(\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}$ . If  $\mathbf{b}$  is estimated with the OLS estimator

$\hat{\mathbf{b}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$  then  $E(\hat{\mathbf{b}}_{\text{OLS}}) = \mathbf{b}$ , and  $\text{var}(\hat{\mathbf{b}}_{\text{OLS}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$ . In the OLS regression,  $\sigma^2$  is estimated with  $\hat{\sigma}_{\text{OLS}}^2 = (N-p)^{-1} \mathbf{u}'\mathbf{u}$ , where  $\mathbf{u}$  is the residual vector  $\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{OLS}}$ . Then,  $\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{M}\mathbf{e}$  where  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ . The expected value of  $\mathbf{e}'\mathbf{M}\mathbf{e}$  is  $\sigma^2 \text{trace}(\mathbf{M}\mathbf{V})$ . Thus,  $E(\hat{\sigma}_{\text{OLS}}^2) = (N-p)^{-1} \text{trace}(\mathbf{M}\mathbf{V})\sigma^2$ .

It is now assumed that that data consist of  $K \geq 2$  groups, indexed with  $i$ , each group has  $n \geq 2$  observations, indexed with  $j$ . It is denoted that  $N = Kn$ .  $\mathbf{V}$  is block diagonal so that each block  $\mathbf{V}_i$  has a compound symmetry correlation structure where the non-diagonal elements are equal to the intra-class correlation  $\rho_e$ . The term ‘‘intra-class correlation’’ is used for historical reasons, but otherwise the term ‘‘group’’ is used. The intercept is the first parameter in  $\mathbf{b}$ , *i.e.*, the first column in  $\mathbf{X}$  is vector of ones,  $\mathbf{1}_N$ . It is assumed further that other columns are centered, *i.e.*, the sum of elements of the column is zero. This assumption makes the analysis simpler. Total and group averages are denoted as  $\bar{x}$ ,  $\bar{x}_i$ ,  $\bar{y}$  and  $\bar{y}_i$ .

Closed form algebraic equations are initially derived for the simple linear model  $y_{ij} = a + bx_{ij} + e_{ij}$ . A model with several regressors is presented later. Centering means that the working model is  $y_{ij} = a' + b(x - \bar{x}) + e_{ij}$ . Estimates of  $a'$  and  $b$  are uncorrelated both in OLS and GLS. After estimating  $a'$  and  $b$ , an estimate of  $a$  is obtained from  $\hat{a} = \hat{a}' - \hat{b}\bar{x}$ . Thereafter  $\text{var}(\hat{a}) = \text{var}(\hat{a}') + \bar{x}^2 \text{var}(\hat{b})$ , and  $\text{cov}(\hat{a}, \hat{b}) = \text{var}(\hat{a}')$ .

## 2. GLS in the Simple Linear Model

First, let us assume that  $-1/(n-1) < \rho_e < 1$ , which means that  $\mathbf{V}$  is non-singular. Then  $\mathbf{V}_i^{-1}$  is a matrix with diagonal elements:

$$\alpha = (1 + (n-2)\rho_e) / ((1 - \rho_e)(1 + (n-1)\rho_e)) \tag{1}$$

The non-diagonal elements are equal to:

$$\beta = -\rho_e / ((1 - \rho_e)(1 + (n-1)\rho_e)) \tag{2}$$

Thereafter  $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$  is

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \begin{pmatrix} N(\alpha + (n-1)\beta) & 0 \\ 0 & \alpha \sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x})^2 + \beta \sum_{i=1}^K \sum_{j=1}^n \sum_{j' \neq j} (x_{ij} - \bar{x})(x_{ij'} - \bar{x}) \end{pmatrix} \tag{3}$$

The second diagonal element in (3) can be expressed as

$\sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x})^2 (\alpha + \beta R_x)$  using the empirical intra-class correlation  $R_x$ , which can be presented in two equivalent forms:

$$R_x = \sum_{i=1}^K \sum_{j=1}^n \sum_{j' \neq j} (x_{ij} - \bar{x})(x_{ij'} - \bar{x}) / (n-1) \sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x})^2 \tag{4}$$

$$R_x = n^2 \sum_i (\bar{x}_i - \bar{x})^2 / (n-1) \sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x})^2 - \frac{1}{n-1} \tag{5}$$

Reference [7] explains how the intra-class correlation, dating back to Fisher, generalizes the correlation idea in order to measure the similarity of group members. Equation (4) appears as covariance divided by variance, and (5) is almost equal to the sample variance of group means divided by the total sample variance, which resembles the definition of  $\rho_e$ . Equation (5) indicates that  $R_x \geq -1/(n-1)$ . The lower limit is obtained when all group means  $\bar{x}_i$  are equal. The upper limit  $R_x = 1$  is obtained when  $x_{ij} = \bar{x}_i$  for all  $i$  and  $j$ . If  $R_x = 0$ , then  $\sum_i^K (\bar{x}_i - \bar{x})^2 / K$ , the population variance of  $\bar{x}_i$ , is equal to the population variance  $\sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x})^2 / N$  divided by  $n$ , resembling the variance of the mean of uncorrelated random variables.

Matrix  $(\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}$  is obtained by taking the reciprocals of the diagonal elements of (3). The variances of the estimates of  $a'$  considered here are proportional to  $C_{a'} = \sigma^2/N$ , and the variances of the estimate of  $b$  are proportional to  $C_b = \sigma^2 / \sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x})^2$ . When  $\rho_e = 0$ ,  $\text{var}(\hat{a}') = C_{a'}$ , and  $\text{var}(\hat{b}) = C_b$ . First, we note that  $\hat{a}'_{\text{GLS}} = \hat{a}'_{\text{OLS}} = \bar{y}$ . Thus, we directly derive that  $\text{var}(\hat{a}') = C_{a'}(1 + (n-1)r_e)$ . If  $r_e = -1/(n-1)$ , then  $\text{var}(\hat{a}') = 0$ . For  $b$ , we obtain:

$$\text{var}(\hat{b}_{\text{GLS}}) = C_b \frac{(1 - \rho_e)(1 + (n-1)\rho_e)}{1 + (n-2)\rho_e - (n-1)R_x\rho_e} \tag{6}$$

If  $\rho_e = 0$  or  $\rho_e = R_x$ , then  $\text{var}(\hat{b}_{\text{GLS}}) = C_b$ . It holds that  $\text{var}(\hat{b}_{\text{GLS}}) > C_b$ , if  $\rho_e$  is between 0 and  $R_x$ . Equation (6) is an increasing function of  $R_x\rho_e$ , an increasing function of  $R_x$  when  $\rho_e > 0$ , and a decreasing function of  $R_x$  when  $\rho_e < 0$ . When  $R_x = 1$ , then  $\text{var}(\hat{b}_{\text{GLS}}) = C_b(1 + (n-1)\rho_e)$ . When  $R_x = -1/(n-1)$ , then  $\text{var}_{\text{GLS}}(\hat{b}) = C_b(1 - \rho_e)$ . The dependency of  $\text{var}(\hat{b}_{\text{GLS}})/C_b$  on  $\rho_e$  for different values of  $R_x$  is shown in **Figure 1**.

### 3. OLS in Simple Linear Model When $\rho_e \neq 0$

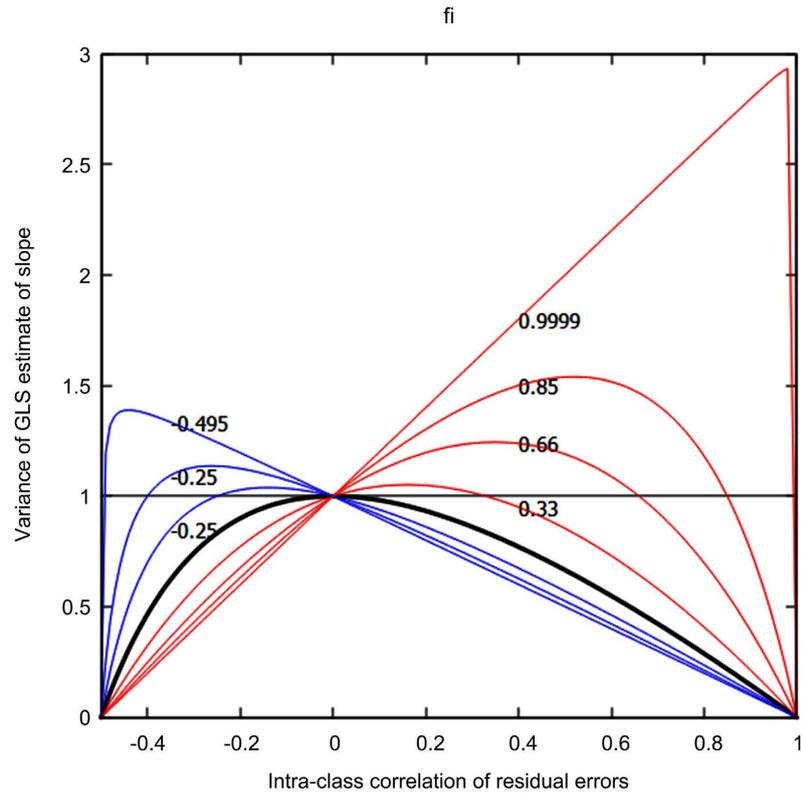
What happens if OLS is used when  $\rho_e \neq 0$ ? Using  $\text{var}(\hat{b}_{\text{OLS}})$ , provided in the introduction section, we obtain:

$$\text{var}(\hat{b}_{\text{OLS}}) = C_b(1 + (n-1)R_x\rho_e) \tag{7}$$

When  $\rho_e \neq 0$  and OLS are used to estimate the variances of the parameter estimates, the bias combines the bias of the estimate of  $\sigma^2$  and the errors of  $\text{var}_{\text{OLS}}(\hat{a}'_{\text{OLS}})/\sigma^2$  and  $\text{var}_{\text{OLS}}(\hat{b}'_{\text{OLS}})/\sigma^2$  where  $\text{var}_{\text{OLS}}$  indicates the variances implied by OLS assumptions (these are not biases, as they are not related to the expected values). In our case:

$$E(\hat{\sigma}_{\text{OLS}}^2) = \left(1 - \frac{(n-1)(1+R_x)\rho_e}{N-2}\right)\sigma^2 \tag{8}$$

It holds that  $E(\hat{\sigma}_{\text{OLS}}^2) < \sigma^2 \Leftrightarrow \rho_e > 0$ . Now,



**Figure 1.** The dependency of  $\text{var}(\hat{b}_{\text{GLS}})/C_b$  in (6) on  $\rho_e$ , the intra-class correlation of residual errors, for different values of  $R_x$ , the intra-class correlation of  $x$  (denoted on the curves) when  $n = 3$ . Blue lines are for negative values of  $R_x$  and red lines for positive values. The thick line is for  $R_x = 0$ . Note that the same variance is obtained for  $\rho_e = 0$  and  $\rho_e = R_x$ .

$$\begin{aligned} \hat{\text{var}}_{\text{OLS}}(\hat{a}'_{\text{OLS}}) &= \hat{\sigma}_{\text{OLS}}^2/N = \hat{\sigma}_{\text{OLS}}^2 \text{var}(\hat{a}'_{\text{OLS}})/(N \text{var}(\hat{a}'_{\text{OLS}})) \\ &= \hat{\sigma}_{\text{OLS}}^2 \text{var}(\hat{a}'_{\text{OLS}})/(1+(n-1)\rho_e) \end{aligned} \tag{9}$$

In (9),  $\hat{\text{var}}_{\text{OLS}}$  denotes the variance estimate we obtain in OLS. If it is combined with (8), we can obtain  $E(\hat{\text{var}}_{\text{OLS}}(\hat{a}'))$  in terms of  $\text{var}(a'_{\text{OLS}})$ . The estimated variance is biased downwards if  $\rho_e > 0$ . If we similarly combine (8) and (7) we obtain:

$$E(\hat{\text{var}}_{\text{OLS}}(\hat{b}_{\text{OLS}})) = \frac{N-2-(n-1)(1+R_x)\rho_e}{(N-2)(1+(n-1)R_x\rho_e)} \text{var}(\hat{b}_{\text{OLS}}) \tag{10}$$

When  $K$  increases, the bias coefficient in (10) approaches  $1/(1+(n-1)\rho_e R_x)$ , which again demonstrates the adverse effect of  $\rho_e R_x$ . The direction of the bias can be seen using the difference between the numerator and denominator of (10). Thus

$$E(\hat{\text{var}}_{\text{OLS}}(\hat{b}_{\text{OLS}})) < \text{var}(\hat{b}_{\text{OLS}}) \Leftrightarrow (N-2)R_x\rho_e + (1+R_x)\rho_e > 0 \tag{11}$$

When  $K$  increases (and thus  $N$  also increases), then the direction of bias is dependent on the sign of  $R_x\rho_e$ . The condition can be stated in a simpler form

as  $(N-1)R_x\rho_e + \rho_e > 0$ , but (11) is needed for comparison with [3] in which two inequalities are derived which taken together, imply that the F-test in OLS leads to  $p$ -values that are too small. In our case, conditions in [3] are  $(n-1)(1+R_x)\rho_e > 0$  and  $(n-1)R_x\rho_e > 0$ , which together imply the validity of the inequality in (11). Moreover, (11) can solve the direction of the bias in the case where sub-conditions indicate opposing directions. When  $\hat{v}\hat{r}_{OLS}(\hat{b}_{OLS})$  is used in a t-test, its bias also produces bias in the computed  $p$ -values. Reversing the inequality in (11), we obtain a condition for obtaining too large  $p$ -values.

#### 4. BLUE for Several Regressors, Nonsingular $V$

Let us now assume a general linear model  $y = Xb + e$ , the first column of  $X$  being  $\mathbf{1}_N$  and the other columns centered. It is here assumed that  $V$  is nonsingular, i.e.,  $-1/(n-1) < \rho_e < 1$ . Let  $x$  and  $z$  denote two regressors. Both  $\alpha$  and  $\beta$  are inversely proportional to  $(1-\rho_e)(1+(n-1)\rho_e)$ . Thus,  $\alpha$  and  $\beta$  increase when  $\rho_e$  approaches  $-1/(n-1)$  or 1, leading to singular  $V_i$  when either of these limits are obtained. Let us define

$W_i = (1-\rho_e)(1+(n-1)\rho_e)V_i^{-1}$ . The diagonal elements of  $W_i$  are equal to  $(1+(n-2)\rho_e)$ , the nondiagonal elements are equal to  $-\rho_e$ . Let  $W$  combine all blocks. Using  $W$ , we obtain a linear estimator:

$$\hat{b}_W = (X'WX)^{-1} X'Wy \tag{12}$$

Element  $ij$  of  $X'W$  is for  $x$ :

$$\begin{aligned} u_{ij} &= (1+(n-2)\rho_e)(x_{ij} - \bar{x}) - \rho_e \sum_{j' \neq j}^n (x_{ij'} - \bar{x}) \\ &= (1+(n-1)\rho_e)(x_{ij} - \bar{x}_i) + (1-\rho_e)(\bar{x}_i - \bar{x}) \end{aligned} \tag{13}$$

The first diagonal element of  $X'WX$ , corresponding to the intercept, is  $N(1-\rho_e)$ , and an element for  $x$  is:

$$X'WX_{xx} = \sum_{i=1}^K \sum_{j=1}^n \left[ (1+(n-1)\rho_e)(x_{ij} - \bar{x}_i)^2 + (1-\rho_e)(\bar{x}_i - \bar{x})^2 \right] \tag{14}$$

Equation (14) explains why (6) is an increasing function of  $R_x\rho_e$ : with a large  $R_x\rho_e$  GLS places large weighting on the component with small variation. The non-diagonal elements are zero for the first row and column, and generally:

$$\begin{aligned} X'WX_{xz} &= \sum_{i=1}^K \sum_{j=1}^n (1+(n-1)\rho_e)(x_{ij} - \bar{x}_i)(z_{ij} - \bar{z}_i) \\ &\quad + \sum_{i=1}^K \sum_{j=1}^n (1-\rho_e)(\bar{x}_i - \bar{x})(\bar{z}_i - \bar{z}) \end{aligned} \tag{15}$$

Note that on the second-row summation over  $n$  could be dropped by multiplying with  $n$ .

The first element of  $X'Wy$  is  $N(1-\rho_e)\bar{y}$ , and the others are:

$$\begin{aligned} X'Wy_x &= \sum_{i=1}^K \sum_{j=1}^n (1+(n-1)\rho_e)(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) \\ &\quad + \sum_{i=1}^K \sum_{j=1}^n (1-\rho_e)(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \end{aligned} \tag{16}$$

The averages  $\bar{y}_i$  and  $\bar{y}$  do not contribute, but they may increase (or decrease) understanding. They do not indicate centering of  $y$  because they are not involved in the first element of  $\mathbf{X}'\mathbf{W}\mathbf{y}$ . In GLS, the same estimates are obtained with any scaling of  $\mathbf{V}^{-1}$ , thus  $\hat{\mathbf{b}}_w = \hat{\mathbf{b}}_{\text{GLS}}$  for a non-singular  $\mathbf{V}$ , but may be computationally more stable. Variances are then computed using

$$\text{var}(\hat{\mathbf{b}}_w) = ((1 - \rho_e)(1 + 1/(n-1)))(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{17}.$$

## 5. BLUE When $\mathbf{V}$ is Singular

### 5.1. BLUE in the Simple Linear Regression with Singular $\mathbf{V}$

Let us introduce the estimation with a singular  $\mathbf{V}$  using the simple linear model.  $y_{ij} = a + bx_{ij} + e_{ij}$ . If  $\rho_e = 1$  and  $R_x \neq 1$  we can estimate  $b$  in the model with zero variance with  $\hat{b} = (y_{ij} - y_{i'j'}) / (x_{ij} - x_{i'j'})$  using two observations such that  $x_{ij} \neq x_{i'j'}$ . As  $\rho_e = 1$  implies that  $e_{ij} = \bar{e}_i$  for all observations,  $a$  can be estimated by computing the arithmetic mean (which is OLS estimator) of  $y_{ij} - \hat{b}x_{ij}$  or  $y_i - \hat{b}\bar{x}_i$ , the latter computation implying correct variance estimator  $\sigma^2/K$ .

If  $R_x \neq 1$ , we can find at least two observations in the same group having different  $x$  values. If there are more observations deviating from the group means, there is an infinite number of estimating equations which all produce this same estimate with zero variance.

If  $\rho_e = 1$  and  $R_x = 1$ , then all  $y$  and  $x$  values are equal to the group means in all groups (for  $y$  with probability one), and  $b$  needs to be estimated with OLS using group means. This means that  $\text{var}(\hat{b})$  is  $n$  times larger than we would get for  $\rho_e = 0$ , as could be anticipated from **Figure 1** for  $n = 3$ . The same estimates are produced using observation level OLS regression, which produces biased estimates for the variances  $\text{var}(\hat{b})$  and  $\text{var}(\hat{a})$ .

If  $\rho_e = -1/(n-1)$  and  $R_x \neq -1/(n-1)$ , then we can find at least two groups  $i$  and  $i'$  so that  $\bar{x}_i \neq \bar{x}_{i'}$ , and  $b$  can be estimated with  $(\bar{y}_i - \bar{y}_{i'}) / (\bar{x}_i - \bar{x}_{i'})$  with zero variance. If there are more than two differing group means, there is an infinite number of estimating equations producing same estimates. After estimating  $b$ ,  $a$  can be estimated with the arithmetic mean of  $\bar{y}_i - b\bar{x}_i$  with variance  $\sigma^2/K$ . The same estimate is obtained with the arithmetic mean of  $y_{ij} - bx_{ij}$ , but standard OLS computations would produce a biased estimate for the estimation variance.

If  $\rho_e = -1/(n-1)$  and  $R_x = -1/(n-1)$ , *i.e.* all group means of  $\bar{e}_i$  and  $\bar{x}_i$  are equal (for  $\bar{e}_i$  with probability one), then BLUE of  $b$  and  $a$  can be obtained by dropping one observation from each group and doing OLS regression in the remaining data. This means that  $\text{var}(\hat{b})$  is equal to  $n/(n-1)$  times the OLS variance obtained for  $\rho_e = 0$ . For  $n = 3$ ,  $n/(n-1) = 1.5$  as could be anticipated from **Figure 1**.

### 5.2. BLUE in the Multiple Regression When $\rho_e = 1$

When  $\rho_e = 1$ , then  $\mathbf{V}$  is singular. Also  $\mathbf{W}$  is singular, all of its diagonal ele-

ments are equal to  $n - 1$ , and nondiagonal elements are equal to  $-1$ . The estimator (12) cannot be applied directly because the first diagonal element of  $\mathbf{X}'\mathbf{W}\mathbf{X}$  is  $N(1 - \rho_e) = 0$  thus producing singular  $\mathbf{X}'\mathbf{W}\mathbf{X}$ . According to (14) the diagonal elements are zero also for such predictors  $x$  having  $R_x = 1$ , *i.e.* no variation in  $x_{ij} - \bar{x}_i$ . If all predictors have  $R_x = 1$  then OLSE is BLUE, because  $C(\mathbf{V}\mathbf{X}) \subset C(\mathbf{X})$ , see [4].

A general BLUE can be obtained using the decomposition  $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$  where  $\mathbf{X}_1$  contains  $\mathbf{1}_N$  and all columns having  $R_x = 1$  and  $\mathbf{X}_2$  contains other columns. It is now required that  $K \geq 3$ . Let  $\mathbf{b} = (\mathbf{b}'_1 \ \mathbf{b}'_2)'$  decompose  $\mathbf{b}$  similarly. If we weight columns of  $\mathbf{X}_1$  as done in OLS, *i.e.* replace columns of  $\mathbf{X}'_1\mathbf{W}$  in Equation (12) with  $\mathbf{X}_1$  we get an estimator

$$\hat{\mathbf{b}}_{\text{GOLS}} = \mathbf{B}\mathbf{y} = \begin{pmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{X}'_1\mathbf{X}_2 \\ \mathbf{X}'_2\mathbf{W}\mathbf{X}_1 & \mathbf{X}'_2\mathbf{W}\mathbf{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2\mathbf{W} \end{pmatrix} \mathbf{y} \tag{18}$$

I suggest that the estimator is called a GOLS estimator as it combines OLS and GLS estimation principles in the same estimator. Using the standard formula for computing the inverse of a partitioned matrix, the estimator can be presented as

$$\hat{\mathbf{b}}_{\text{GOLS}} = \mathbf{B}\mathbf{y} = \begin{pmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1 - (\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1\mathbf{X}_2 (\mathbf{X}'_2\mathbf{W}\mathbf{X}_2)^{-1} \mathbf{X}'_2\mathbf{W} \\ (\mathbf{X}'_2\mathbf{W}\mathbf{X}_2)^{-1} \mathbf{X}'_2\mathbf{W} \end{pmatrix} \mathbf{y} \tag{19}$$

Thus

$$\hat{\mathbf{b}}_2 = (\mathbf{X}'_2\mathbf{W}\mathbf{X}_2)^{-1} \mathbf{X}'_2\mathbf{W}\mathbf{y} \tag{20}$$

Thus,  $\hat{\mathbf{b}}_2$  is the GLSE of  $\mathbf{b}_2$  as estimated solely using  $\mathbf{X}_2$ . Now  $\text{var}(\hat{\mathbf{b}}_2) = (\mathbf{X}'_2\mathbf{W}\mathbf{X}_2)^{-1} \mathbf{X}'_2\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{X}_2 (\mathbf{X}'_2\mathbf{W}\mathbf{X}_2)^{-1} = \mathbf{0}$ . Noting that the second row of  $\mathbf{B}$  is part of the first row, we obtain:

$$\hat{\mathbf{b}}_1 = (\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1\mathbf{y} - (\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1\hat{\mathbf{b}}_2 \tag{21}$$

$$\hat{\mathbf{b}}_1 = \hat{\mathbf{b}}_{1,\text{OLS}} - (\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1\mathbf{X}_2\hat{\mathbf{b}}_2 \tag{22}$$

where  $\hat{\mathbf{b}}_{1,\text{OLS}}$  is the OLS estimator of  $\mathbf{b}_1$  obtained ignoring  $\mathbf{X}_2$ .

Using (22), we get  $\text{var}(\hat{\mathbf{b}}_1) = \sigma^2 n (\mathbf{X}'_1\mathbf{X}_1)^{-1}$ , *i.e.* the same variance what we get when  $\mathbf{b}_2 = 0$  and we regress group means of  $y$  on group means of all variables in  $\mathbf{X}_1$ . This is natural because all  $y$  values and  $x$  values are equal in all observations in each group.

Denote that  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . An estimator  $\mathbf{A}\mathbf{y}$  is BLUE for  $\mathbf{X}\mathbf{b}$ , if and only if  $\mathbf{A}(\mathbf{X} \ \mathbf{V}\mathbf{M}) = (\mathbf{X} \ \mathbf{0})$ , see [5]. If  $\mathbf{X}$  has full column rank (as we have), then  $\mathbf{b}$  is estimable and  $\mathbf{B}\mathbf{y}$  is BLUE for  $\mathbf{b}$  if and only if  $\mathbf{B}(\mathbf{X} \ \mathbf{V}\mathbf{M}) = (\mathbf{I} \ \mathbf{0})$ . Then, that  $\mathbf{B}\mathbf{X} = \mathbf{I}$  can be seen directly from (19) taking into account that  $\mathbf{X}'_2\mathbf{W}\mathbf{X}_1 = \mathbf{0}$ . Noting that  $\mathbf{W}\mathbf{V} = \mathbf{0}$  and  $\mathbf{X}'_1\mathbf{V} = n\mathbf{X}'_1$  we obtain that:

$$\mathbf{B}\mathbf{V}\mathbf{M} = \left( n(\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1 - n(\mathbf{X}'_1\mathbf{X}_1)^{-1} \mathbf{X}'_1\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}', \mathbf{0} \right)' \tag{23}$$

Matrix  $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  is a projector to the space spanned by  $\mathbf{X}$ . Now

$C(n(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1) \subset C(\mathbf{X})$ , thus  $n(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}' = n(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$ , which completes the demonstration that  $\mathbf{BVM} = \mathbf{0}$ .

The GOLS estimator can be put into general matrix theory context using [5]. Now  $C(\mathbf{X}_1)$  and  $C(\mathbf{X}_2)$  are disjoint and  $C(\mathbf{X}_1) \subset C(\mathbf{V})$  and  $C(\mathbf{X}_2) \not\subset C(\mathbf{V})$ . In such a case (20) provides BLUE for  $\mathbf{b}_2$ . The estimator  $\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}$  is unbiased for  $\mathbf{X}_1\mathbf{b}_1$ . This unbiased estimator is updated into BLUE in (22), being an application of proposition 10.5 on p. 228 of [5].

### 5.3. BLUE When $\rho_e = -1/(n-1)$

If  $\rho_e = -1/(n-1)$  then all elements of  $\mathbf{W}_i$  are equal to  $1/(n-1)$ . Let us organize  $\mathbf{X}$  into  $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$ , where  $\mathbf{X}_1$  contains all such predictors for which  $R_i = -1/(n-1)$  and  $\mathbf{X}_2$  contains all other columns. Equation (20) provides again a BLUE with  $\text{var}(\hat{\mathbf{b}}_2) = 0$  and the same arguments can be used to prove it also for this case. Now  $\text{var}(\hat{\mathbf{b}}_1) = \sigma^2 \frac{n}{n-1}(\mathbf{X}'_1\mathbf{X}_1)^{-1}$ , which is the same variance we obtain if  $\mathbf{b}_2 = 0$  and  $\mathbf{b}_1$  is estimated from data where one redundant observation is dropped from each group.

It may be a useful exercise to get confidence in multiple regressor derivations to show how the estimators derived in Section 5.1 for the simple linear model are produced also with the matrix formulas. Matrix formulas use in computations of all observations while formulas in section 5.1 utilize only nonredundant information.

## 6. Discussion

When  $y_{ij}$  is regressed on  $x_{ij}$ , it is implicitly assumed that the regression of  $y_{ij}$  on  $x_{ij} - \bar{x}_i$ , and the regression of  $\bar{y}_i$  on  $\bar{x}_i - \bar{x}$ , have the same slope. GLS estimation provides different weights to  $x_{ij} - \bar{x}_i$  and  $\bar{x}_i - \bar{x}$  when attempting to utilize the correlation structure, although  $(1+(n-1)\rho_e)(x_{ij} - \bar{x}_i)$  and  $(1-\rho_e)(\bar{x}_i - \bar{x})$  are put into the same column of  $\mathbf{X}$  as shown in (14). When developing mixed models for grouped data, it is often necessary to consider both  $x_{ij} - \bar{x}_i$  and  $\bar{x}_i$  as separate predictors, see [8]. It is natural to assume that  $y_{ij}$  is related to  $x_{ij} - \bar{X}_i$  and  $x_{ij} - \bar{X}_i$ , where  $\bar{X}_i$  is the mean of  $x$  in the whole group. Reference [8] suggests a solution, based on a multivariate mixed model, which solves the measurement error bias problem that occurs when  $\bar{x}_i$  computed from a sample from group  $i$  is used to “measure”  $\bar{X}_i$ .

The mean of variable  $x_{ij} - \bar{x}_i$  is zero in each group, thereby indicating a negative within-group correlation. If  $y_{ij}$  is linearly related to  $x_{ij} - \bar{x}_i$ , the negative within-group correlation of  $x_{ij} - \bar{x}_i$  is also transmitted to  $y_{ij}$ . Hence, the addition of predictor  $x_{ij} - \bar{x}_i$  to the models makes the estimated variance of the random intercept larger, as the “negative variance” is subtracted from the residual variance, see [8] and [9].

The mixed model formulation always leads to non-negative intra-class correlations. However, negative intra-class correlations are needed in situations where

the group means have a smaller variance than that implied by the assumption of uncorrelated individuals. In the marginal interpretation of mixed model equations, a negative definite variance matrix of random effects that maximize the likelihood is allowed, leading to negative correlations and negative variances, provided the marginal variance of the  $y$ -vector is positive semi-definite, see [10]. The significance of a negative intra-class coefficient is dependent on the group size.

When the  $x$ -variable is a random variable with a theoretical intra-class correlation  $\rho_x$ , the empirical intra-class correlation  $R_x$  is a random variable. The expected value of  $R_x$  approaches  $\rho_x$  when  $K$  increases. Simulations with normally distributed variables show that, for given values of  $n$  and  $K$ ,  $\text{var}(R_x)$  can be well described with function  $c(\rho_x + 1/(n-1))^s (1-\rho_x)^q$ . Negatively correlated variables can be simulated using principal components.

Reference [2] discussed the estimation of the slope in a simple linear model assuming that  $n=2$ ,  $x_{i1}=0$  and  $x_{i2}=t_i$  for all  $i$ . To compare the formulas presented above with the formulas in [2], denote that  $p = K\bar{t}^2 / \sum_{i=1}^K t_i^2$  where  $\bar{t}$  is the average of  $t_i$ , then  $0 < p \leq 1$ , and  $p=1$  when all  $t_i$ 's values are equal. It holds that  $R_x = -p/(2-p)$ . They compare the variance of GLSE to the variance of the OLS estimate obtained by regressing  $y_{i2} - y_{i1}$  on  $t_i$ . The standard OLSE of the slope has, however, a smaller variance than their OLSE based on differences when  $\rho_e > 0.5$ . For  $\rho_e = 0$ , the ratio of the variance of the standard OLSE to the variance of their OLSE is  $2-p$ . When all  $t_i$ 's are equal, then  $p=1$ ,  $R_x = -1$ , and all three methods provide the same estimate.

If the group sizes  $n_i$  vary in a data set, it is not reasonable to assume the same negative intra-class coefficient  $\rho_e$  for all groups. Thus  $\rho_e$  needs to be made group specific  $\rho_{ei}$ . Reference [7] generalizes the intra-class correlation to non-equal group sizes, but this generalization does not provide meaningful analysis for the interaction of the intra-class correlation of the residual and the values of regressors. However, it is possible to define a measure for the between-group variation of  $x$  which gets value of zero if there is no variation between groups, and use this measure to analyze how the intra-class correlation of residual errors and the between-group and within-group variation of predictors interact. This analysis can be extended to auto-correlation structures, and is under development.

In experiments, (6) informs us that  $R_x$  should be  $-1/(n-1)$  (all  $\bar{x}_i$  should be equal) if it is anticipated that  $\rho_e > 0$ . If it is anticipated that  $\rho_e < 0$ , then the same treatment level should be applied to the whole group.

The derivations of this paper are fully covered with the general matrix theory. The purpose of the paper is to make general matrix formula understandable using algebraic derivations which show how between-group variation and within-group variations of residual errors are connected to the between-group and within-group variations of regressors in the estimation of a general linear model.

## Acknowledgements

I thank Ronald Christensen for taking seriously the first version dealing with the

simple linear model, Simo Puntanen for patient guidance to the conditions which prove that an estimator is BLUE, and Jarkko Isotalo for indicating how the GOLS estimator is related to the general BLUE theory. I take full responsibility if I have not understood their explanations correctly. They fly smoothly in linear spaces and subspaces where I need to walk watching each step.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Markiewicz, A., Puntanen, S. and Styan, P.H. (2022) The Legend of the Equality of OLSE and BLUE: Highlighted by C. R. Rao in 1967. In: Arnold, B.C., Balakrishnan, N. and Coelho, C.A., Eds., *Methodology and Applications of Statistics*, Springer, Cham.
- [2] Jeske, D.R. and Myhre, J.M. (2018) Regression Using Pairs vs. Regression on Differences: A Real-life Case Study for a Master's Level Methods Class, *The American Statistician*, **72**, 163-168. <https://doi.org/10.1080/00031305.2017.1292956>
- [3] Christensen, R. (2003) Significantly insignificant  $F$  tests. *The American Statistician*. **57**, 27-32. <https://doi.org/10.1198/0003130031108>
- [4] Puntanen, S. and Styan, G.P.H. (1989) The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator. *The American Statistician*, **43**, 153-161. <https://doi.org/10.1080/00031305.1989.10475644>
- [5] Puntanen, S., Styan, G.P.H. and Isotalo, J. (2011) Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty. Springer, Heidelberg. <https://doi.org/10.1007/978-3-642-10473-2>
- [6] Christensen, R. (2011) Plane Answers to Complex Questions. The Theory of Linear Models. 4th Edition, Springer, New York. <https://doi.org/10.1007/978-1-4419-9816-3>
- [7] Kendall, M. and Stuart, A. (1979) The Advanced Theory of Statistics: Volume 2: Inference and Relationship. 4th Edition, Charles Griffin, London.
- [8] Mehtätalo, L. and Lappi, J. (2020) Biometry for Forestry and Environmental Data with Examples in R. Chapman and Hall/CRC, New York. <https://doi.org/10.1201/9780429173462>
- [9] Snijders, T. and Bosker, R.J. (1994) Modeled Variance in Two-Level Models. *Sociological Methods Research*, **22**, 342-363. <https://doi.org/10.1177/0049124194022003004>
- [10] Pryseley, A., Tchonlafi, C., Verbeke, G. and Molenberghs, G. (2011) Estimating Negative Variance Data Components from Gaussian and Non-Gaussian Data: A Mixed Models Approach. *Computational Statistics and Data Analysis*, **55**, 1071-1085. <https://doi.org/10.1016/j.csda.2010.09.002>