A geometrical interpretation of collinearity: a natural way to justify ridge regression and its anomalies

José García

Department of Economic and Business, Almería University, Almería, Spain.

E-mail: jgarcia@ual.es

María del Mar López-Martín

Department of Didactics of Mathematics, Granada University, Granada, Spain.

Catalina García

Department of Quantitative Methods for Economics and Business, Granada University, Granada, Spain.

Román Salmerón

Department of Quantitative Methods for Economics and Business, Granada University, Granada, Spain.

Summary. Justifying ridge regression from a geometrical perspective is one of the main contributions of this paper. To the best of our knowledge, this question has not been treated previously. This paper shows that ridge regression is a particular case of the raising procedures that provides greater flexibility by transforming the matrix $\mathbf{X}$ associated with the model. Thus, the raising procedures, based on a geometrical idea of the vectorial space associated with the columns of matrix $\mathbf{X}$, lead naturally to ridge regression and justify the presence of the well-known constant k on the main diagonal of matrix $\mathbf{X}'\mathbf{X}$. This paper also analyzes and compares different alternatives to raising with respect to collinearity mitigation. The results are illustrated with an empirical application.

Keywords: Collinearity; Ridge Regression; Raise Regression; Linear Regression

1. Introduction

In the presence of collinearity, the ordinary least square (OLS) estimator is unstable and often causes several problems with the estimator such as inflated variances and covariances, inflated correlations, inflated prediction variance, and the concomitant difficulties in interpreting the significance values and confidence regions for parameters, Willan and Watts (1978). The ridge estimator (RE), Hoerl and Kennard (1970a,b), is an alternative methodology formalized for a multiple linear model as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

for n observations and p independent variables, where it is assumed that $E[\mathbf{u}] = \mathbf{0}$, $E[\mathbf{uu}'] = \sigma^2\mathbf{I}$† according to the following expression:

$$\widehat{\boldsymbol{\beta}}_R(k) = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{y}, \quad k \geq 0, \tag{2}$$

where $\widehat{\boldsymbol{\beta}}_R(k)$ is biased for $k > 0$ and coincides with the OLS estimator when $k = 0$. They showed that ridge estimators satisfy the mean square error (MSE) admissibility condition to improve OLS for some $k \in (0, \infty)$ and characterized by the ridge trace. Thus, ridge regression (RR) solves the problem of the ill-conditioned matrix $\mathbf{X}'\mathbf{X}$ by introducing a constant k in the diagonal.

Based on ridge estimator, other alternative techniques have been developed to remedy the consequent symptoms resulting from data collinearity, such as the Stein estimator (Stein et al., 1956) and the contraction estimator (Liu, 1993; Mayer and Willke, 1973). This last one combines the Stein estimator with the ridge estimator but still depending on OLS estimator which will be unstable. To overcome this situation, Liu (2003) proposed the Liu-type estimator, Sakallıoğlu and Kaçıranlar (2008) presented the k-d class estimator using the ridge estimator. They showed that the k-d class estimator is a general estimator which includes the OLS estimator, the ridge estimator (RE) and the Liu estimator. By combining the RE and the Liu estimator, Chang and Yang (2012) proposed the two parameter estimator which includes the OLS, RE and Liu estimators as special cases and Liu et al. (2013) proposed the improved ridge estimator (IRE). All these estimators are founded on the

†Throughout the article, $\mathbf{0}$ denotes a vector of zeros and $\mathbf{I}$ denotes the identity matrix, both, of adequate dimensions in each case.

RE (Hoerl and Kennard, 1970a,b) and the Stein estimator (Stein et al., 1956) and all try to improve the ill-conditioned matrix $\mathbf{X}'\mathbf{X}$ by adding a constant k as small as possible to reduce the bias.

Apart from the development of alternatives techniques based on ridge regression, another research line is to find a theoretically optimal basis for the ridge procedure. As stated by Piegorsch and Casella (1989) it "has been a lengthy process (Rolph (1976), Strawderman (1978); Casella (1980)) and it is still not fully developed". The ad-hoc solution of Hoerl and Kennard (1970a,b) to overcome the collinearity problem of the design matrix (and consequently the singularity of matrix $\mathbf{X}'\mathbf{X}$) has been motivated posteriorly. A review of all the contributions made in this regard since 1970 is an enormous task and beyond the scope of our work.

Although the geometric interpretation of collinearity is evident, none of the justifications for the ridge estimator are motivated from a geometrical perspective. Some early attempts to geometrically interpret (but not motivate or justify) the ridge estimator can be found in Marquardt and Snee (1975); Swindel (1981). García et al. (2011) introduced the metric number to measure the effect of adding the last column to the other columns in the matrix $\mathbf{X}$. Something similar can be found in Wichers (1975). The basic idea is to measure the angle formed by the vector $\mathbf{X}_{(p)}$ associated with column p and the vectorial space generated by the rest of the columns $\mathbf{X}_{[p]}$ of matrix $\mathbf{X}$. For our purposes, we consider the perspective of Besley (1991): "k variables are collinear or nearly dependent, if one of them lies almost in the space spanned by the remaining $(k-1)$ variables, that is, if the angle between one and its orthogonal projection on the others is small", which is further emphasized by Alin (2010): "In more technical terms, multicollinearity occurs if k vectors lie in a subspace of dimension less than k. This is the definition of exact multicollinearity or exact linear dependence. It is not necessary for multicollinearity to be exact in order to cause a problem. It is enough to have k variables nearly dependent, which occurs if the angle between one variable and its orthogonal projections onto others is small". In our opinion, this means that collinearity can be treated from a geometric perspective.

Raise regression was presented by García et al. (2011) **and developed by García et al. (2016)** as an alternative to ridge regression (RR) to estimate models with collinearity. Posteriorly, **other raising**

procedures have been developed such as the simultaneous raise (SiR), García et al. (2014), and the successive raise regression (SuR), García and Ramirez (2016). None of these procedures lead to RR. In this paper, we present the total successive raise regression (TSuR) as the unique raising procedure that can yield RR. Thus, this raising procedure serves as the geometric and ex ante justification for ridge estimation that differs from the ad hoc justifications presented previously. We also show that successive rather than simultaneous raising is the appropriate means of raising the columns of matrix $\mathbf{X}$.

The paper is structured as follows. In Section 2, we present three alternative raising procedures for a standardized model with **two standardized independent** variables, assuming that collinearity is not sufficiently mitigated after raising the first variable. Section 3 **presents the multivariate case and** how to obtain the ridge regression from the raising procedures. **Section 4 analyzes the connection between TSuR and the generalized ridge regression**. In Section 5, some results on the effect of raising on the matrix that contains the independent variables are discussed, showing that successive raising appears to be preferable. Finally, an empirical application is performed in Section 6, and the main conclusions are summarized in Section 7.

2. Raise regression: types of raising

For the sake of simplicity, we begin by examining the following standardized model with n observations and two standardized exogenous variables:

$$\mathbf{y} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u}, \tag{3}$$

where it is verified that the matrix $\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ is the correlation matrix of the exogenous variables and $\mathbf{X}'\mathbf{y} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$ is the correlation vector of the endogenous variable with every exogenous variable.

Raise regression was initially presented by García et al. (2011) to estimate a model with near-collinearity, relating the problem of collinearity with the angle between vectors $\mathbf{x}_1$ and $\mathbf{x}_2$; that is, if the angle between the two vectors is small, then the collinearity is higher. To mitigate the collinearity problem, the authors suggest raising variable $\mathbf{x}_1$ taking into account the relation $\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda \mathbf{e}_1$, where $\lambda \geq 0$, $\tilde{\mathbf{x}}_1$ is the raised vector of $\mathbf{x}_1$ and $\mathbf{e}_1$ is the residual vector obtained from the regression of $\mathbf{x}_1$ on variable $\mathbf{x}_2$. That is, $\mathbf{e}_1 = \mathbf{x}_1 - \rho \mathbf{x}_2$ being ρ the correlation coefficient between $\mathbf{x}_1$ and $\mathbf{x}_2$. Then, $\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda (\mathbf{x}_1 - \rho \mathbf{x}_2)$. Note that the correlation between vectors $\tilde{\mathbf{x}}_1$ and $\mathbf{x}_2$ is weaker than the correlation between vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ because the angle is higher. Hence, the collinearity problem has been mitigated (see Figure 1). If the collinearity has not been mitigated after raising the first variable, it is possible to raise the second variable. **This results in different types of raising. With the main purpose of geometrically justifying the ridge regression, the following subsections analyze the different possibilities of raising.**

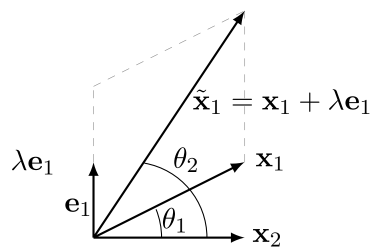

Fig. 1. Representation of raise method

2.1. Simultaneous Raise Regression (SiR)

From model (3), it is possible to simultaneously raise both exogenous variables in the following way: $\widetilde{\mathbf{x}}_j = \mathbf{x}_j + \lambda_j \mathbf{e}_j$ (for $j = 1, 2$), where $\mathbf{e}_j$ is the residual vector obtained from the regression of $\mathbf{x}_j$ on the other exogenous variable and λ_j is the raising factor. In this case, we obtain that $\widetilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda_1 (\mathbf{x}_1 - \rho \mathbf{x}_2)$ and $\widetilde{\mathbf{x}}_2 = \mathbf{x}_2 + \lambda_2 (\mathbf{x}_2 - \rho \mathbf{x}_1)$.
By considering variables $\widetilde{\mathbf{x}}_1$ and $\widetilde{\mathbf{x}}_2$, the transformed model is defined as follows:

$$\mathbf{y} = \beta_1 \widetilde{\mathbf{x}}_1 + \beta_2 \widetilde{\mathbf{x}}_2 + \mathbf{w}. \tag{4}$$

Denoting $\widetilde{\mathbf{X}}_{SiR}$ as the matrix containing the vector obtained from simultaneous raise regression (SiR), the associated matrices are given by the following:

$$\widetilde{\mathbf{X}}'_{SiR}\widetilde{\mathbf{X}}_{SiR} = \begin{pmatrix} 1+\left(\lambda_1^2+2\lambda_1\right)\left(1-\rho^2\right) & \rho-\lambda_1\lambda_2\rho(1-\rho^2) \\ \rho-\lambda_1\lambda_2\rho(1-\rho^2) & 1+\left(\lambda_2^2+2\lambda_2\right)\left(1-\rho^2\right) \end{pmatrix}, \tag{5}$$

$$\widetilde{\mathbf{X}}'_{SiR}\mathbf{y} = \begin{pmatrix} \gamma_1+\left(\gamma_1-\gamma_2\rho\right)\lambda_1 \\ \gamma_2+\left(\gamma_2-\gamma_1\rho\right)\lambda_2 \end{pmatrix}. \tag{6}$$

where γ_i, for $i=1,2$ is the correlation coefficient between the explained variable and explanatory variables.

2.2. Successive Raise Regression (SuR)

The first step of successive raise regression (SuR) is the same as that of SiR. **In the second step,** from $\widetilde{\mathbf{x}}_1 = \mathbf{x}_1+\lambda_1\mathbf{e}_1 = \mathbf{x}_1+\lambda_1\left(\mathbf{x}_1-\rho\mathbf{x}_2\right)$, $\mathbf{x}_2$ is raised taking $\widetilde{\mathbf{x}}_2 = \mathbf{x}_2+\lambda_2\mathbf{e}_2$, where $\mathbf{e}_2$ is the residual obtained from the regression $\mathbf{x}_2 = \alpha\widetilde{\mathbf{x}}_1+\mathbf{v}$. Taking into account that $\widehat{\alpha} = \frac{\rho}{1+\lambda_1(\lambda_1+2)(1-\rho^2)}$ we have that:

$$\begin{aligned} \widetilde{\mathbf{x}}_2 &= \mathbf{x}_2+\lambda_2\left(\mathbf{x}_2-\widehat{\alpha}\widetilde{\mathbf{x}}_1\right) \\ &= \frac{-\left(\lambda_1+1\right)\lambda_2\rho}{1+\lambda_1\left(\lambda_1+2\right)\left(1-\rho^2\right)}\mathbf{x}_1 \\ &\quad + \frac{(1+\lambda_1)^2(1+\lambda_2)-(2+\lambda_1+\lambda_2+\lambda_1\lambda_2)\lambda_1\rho^2}{1+\lambda_1\left(\lambda_1+2\right)\left(1-\rho^2\right)}\mathbf{x}_2. \end{aligned}$$

Denoting $\widetilde{\mathbf{X}}_{SuR}$ as the matrix containing the vector obtained from SuR, the associated matrices are given by:

$$\widetilde{\mathbf{X}}'_{SuR}\widetilde{\mathbf{X}}_{SuR} = \begin{pmatrix} 1+\lambda_1\left(\lambda_1+2\right)\left(1-\rho^2\right) & \rho \\ \rho & 1+\frac{(\lambda_1+1)^2\lambda_2(\lambda_2+2)(1-\rho^2)}{1+\lambda_1(\lambda_1+2)(1-\rho^2)} \end{pmatrix}, \tag{7}$$

$$\widetilde{\mathbf{X}}'_{SuR}\mathbf{y} = \begin{pmatrix} (1+\lambda_1)\gamma_1-\gamma_2\rho\lambda_1 \\ \frac{\gamma_2(\lambda_1+1)^2(\lambda_2+1)-\gamma_1(\lambda_1+1)\lambda_2\rho-\gamma_2\lambda_1(\lambda_2\lambda_1+\lambda_1+\lambda_2+2)\rho^2}{1+\lambda_1(\lambda_1+2)(1-\rho^2)} \end{pmatrix}. \tag{8}$$

2.3. Total Successive Raise Regression (TSuR)

This subsection shows how to obtain the ridge estimator as a particular case of the raise estimator. Hoerl and Kennard (1970a,b) proposed the following transformed matrices:

$$\mathbf{X}'\mathbf{X} + k\mathbf{I} = \begin{pmatrix} 1+k & \sum_{i=1}^{n} \mathbf{x}_{1i}\mathbf{x}_{2i} \\ \sum_{i=1}^{n} \mathbf{x}_{1i}\mathbf{x}_{2i} & 1+k \end{pmatrix} = \begin{pmatrix} 1+k & \rho \\ \rho & 1+k \end{pmatrix}, \quad (9)$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} \sum_{i=1}^{n} \mathbf{x}_{1i} y_i \\ \sum_{i=1}^{n} \mathbf{x}_{2i} y_i \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}. \quad (10)$$

The simultaneous raising (SiR) method leads to expressions (5) and (6), and it is evident that these expressions only coincide with expressions (9) and (10) when $\lambda_1 = \lambda_2 = 0$, that is, only when OLS is applied. Hence, it is possible to conclude that SiR is fundamentally different from RR.
In relation to successive raise regression (SuR), it is possible to obtain the relation between (7) and the information matrix of RR, **selecting** λ_1 **and** λ_2 **to obtain** $k_1 = k_2 = k$ **with** $k_1 = \lambda_1(\lambda_1+2)\left(1-\rho^2\right)$ **and** $k_2 = \frac{(\lambda_1+1)^2\lambda_2(\lambda_2+2)(1-\rho^2)}{1+\lambda_1(\lambda_1+2)(1-\rho^2)}$**. Thus**, it is possible to obtain $\mathbf{X}'\mathbf{X} + k\mathbf{I}$. However, (10) does not coincide with (8), and hence, it is not possible to obtain the ridge estimator with this procedure.
On the other hand, taking into account that:

$$\widetilde{\mathbf{X}}_{SuR} = (\widetilde{\mathbf{x}}_1\ \widetilde{\mathbf{x}}_2) = (\mathbf{x_1} + \lambda_1\mathbf{e_1}\ \mathbf{x_2} + \lambda_2\mathbf{e_2}), \quad (11)$$

it is evident that if $\mathbf{e}_j'\mathbf{y} = 0$, then $\widetilde{\mathbf{X}}_{SuR}'\mathbf{y} = \mathbf{X}'\mathbf{y}$. For this reason, the procedure of Subsection 2.2 has been modified to include the dependent variable as an exogenous variable in the auxiliary regression of the raise methodology. The steps of the procedure are as follows:

- The first variable, $\widetilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda_1\mathbf{e}_1$, is raised, where:

  $$\mathbf{e}_1 = \mathbf{x}_1 - \widehat{\mathbf{x}}_1 = \mathbf{x}_1 - \left(\frac{\rho - \gamma_1\gamma_2}{1-\gamma_2^2}\mathbf{x}_2 + \frac{\gamma_1 - \rho\gamma_2}{1-\gamma_2^2}\mathbf{y}\right),$$

  is the residual vector obtained from regression $\mathbf{x}_1 = \delta_1\mathbf{x}_2 + \delta_2\mathbf{y} + \mathbf{u}_1$. In this case, we obtain the following:

  $$\widetilde{\mathbf{x}}_1 = (1+\lambda_1)\mathbf{x}_1 - \lambda_1\left(\frac{\rho - \gamma_1\gamma_2}{1-\gamma_2^2}\mathbf{x}_2 + \frac{\gamma_1 - \rho\gamma_2}{1-\gamma_2^2}\mathbf{y}\right). \quad (12)$$

- From the residual vector $\mathbf{e}_2$ of regression $\mathbf{x}_2 = \theta_1\widetilde{\mathbf{x}}_1 + \theta_2\mathbf{y} + \mathbf{u}_2$, where $\widetilde{\mathbf{x}}_1$ is (12), $\widetilde{\mathbf{x}}_2$ is obtained as follows:

$$\begin{aligned}\widetilde{\mathbf{x}}_2 &= \mathbf{x}_2 + \lambda_2 \mathbf{e}_2 \quad (13)\\ &= (1+\lambda_2)\mathbf{x}_2 - \lambda_2 \left(\frac{\rho - \gamma_1\gamma_2}{1+k_1-\gamma_2^2}\widetilde{\mathbf{x}}_1 + \frac{\gamma_2(1+k_1) - \rho\gamma_1}{1+k_1-\gamma_1^2}\mathbf{y} \right),\end{aligned}$$

where $k_1 = \frac{(1-\rho^2-\gamma_1^2+2\gamma_1\gamma_2\rho-\gamma_2^2)(2\lambda_1+\lambda_1^2)}{1-\gamma_2^2} = \frac{(1-\rho^2)(1-R^2)(2\lambda_1+\lambda_1^2)}{1-\gamma_2^2}$.

Denoting $\widetilde{\mathbf{X}}_{TSuR}$ as the matrix containing vectors $\widetilde{\mathbf{x}}_1$ and $\widetilde{\mathbf{x}}_2$ given by expressions (12) and (13), respectively, we obtain the following:

$$\widetilde{\mathbf{X}}'_{TSuR}\widetilde{\mathbf{X}}_{TSuR} = \begin{pmatrix} 1+k_1 & \rho \\ \rho & 1+k_2 \end{pmatrix}, \quad (14)$$

$$\widetilde{\mathbf{X}}'_{TSuR}\mathbf{y} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad (15)$$

where $k_2 = \frac{A\mathbf{x}_1 + B\mathbf{x}_2 + C\mathbf{y}}{(1-\gamma_1^2)(1-\gamma_2^2)+(2\lambda_1+\lambda_1^2)(1-\rho^2)(1-R^2)}$ and:

$$\begin{aligned} A &= (\lambda_1+1)\left(1-\gamma_2^2\right)(\gamma_1\gamma_2-\rho), \\ B &= \left(1-\gamma_1^2\right)\left(1-\gamma_2^2\right) + \left(2\lambda_1+\lambda_1^2\right)\left(1-\rho^2\right)\left(1-R^2\right) + \lambda_1(\gamma_1\gamma_2-\rho)^2, \\ C &= \rho\gamma_1 - \gamma_2 - \rho\gamma_1\gamma_2^2 + \gamma_2^3 + \rho\gamma_1\lambda_1 - \gamma_2\lambda_1^2\left(1-R^2\right)\left(1-\rho^2\right) - \\ &\quad -\gamma_2\lambda_1\left[2\left(1-R^2\right)\left(1-\rho^2\right) + \rho^2 - \rho\gamma_1\gamma_2 + \gamma_1^2\right], \end{aligned}$$

being $R^2 = \frac{\gamma_1^2 - 2\gamma_1\gamma_2\rho + \gamma_2^2}{1-\rho^2}$ the coefficient of determination of model (3).
From these results, it is possible to conclude that total successive raise regression (TSuR) leads to the matrices of RR since $\widetilde{\mathbf{X}}'_{TSuR}\widetilde{\mathbf{X}}_{TSuR} = \mathbf{X}'\mathbf{X} + \mathbf{K}$ with $\mathbf{K} = diag(k_1, k_2)$ and $\widetilde{\mathbf{X}}'_{TSuR}\mathbf{y} = \mathbf{X}'\mathbf{y}$. Hence, in the case of two standardized variables, TSuR is the only model that allows one to obtain RR from a raise estimator when adequately selecting λ_1 and λ_2 to obtain $k_1 = k_2 = k$.

3. The multivariate case. A new justification for ridge regression

Given the standardized version of model (1), the following transformed model is considered to obtain the ridge estimator from the raise estimator:

$$\mathbf{y} = \widetilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{w}, \quad (16)$$

where the matrix $\widetilde{\mathbf{X}}$ contains the raised vectors $\tilde{\mathbf{x}}_j = \mathbf{x}_j + \lambda_j \mathbf{e}_j$, with $j = 1, 2, \dots, p$. **Raise estimator is obtained from the ordinary least squares (OLS) estimation of model (16). This section shows how to select the residual vector $\mathbf{e}_j$ and λ_j such that raise estimator coincides with ridge estimator of model** (1)**, that is:**

$$\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}, \tag{17}$$

$$\widetilde{\mathbf{X}}'\mathbf{y} = \mathbf{X}'\mathbf{y}. \tag{18}$$

From these expressions, the following conditions are required:

$$\widetilde{\mathbf{x}}_j'\widetilde{\mathbf{x}}_j = \mathbf{x}_j'\mathbf{x}_j + k, \quad j = 1, \dots, p, \tag{19}$$

$$\widetilde{\mathbf{x}}_i'\widetilde{\mathbf{x}}_j = \mathbf{x}_i'\mathbf{x}_j, \quad i, j = 1, \dots, p, \ i \neq j, \tag{20}$$

$$\widetilde{\mathbf{x}}_j'\mathbf{y} = \mathbf{x}_j'\mathbf{y}, \quad j = 1, \dots, p. \tag{21}$$

In Section 2, we demonstrated **that in the case of two standardized variables**, total successive raise regression (TSuR) is the only raising approach that leads to RR. **In the following subsections we generalize for $p \geq 2$ the different types of raising with the main purpose of this paper for geometrically justifying ridge regression.**
The main idea to develop all the raising procedures is to define the raising matrix, $\widetilde{\mathbf{X}}$, where the set of initial vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ is replaced by the set of raising vectors $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_p\}$ being $\tilde{\mathbf{x}}_j = \mathbf{x}_j + \lambda_j \mathbf{e}_j$.

3.1. Simultaneous Raise Regression (SiR)

In the Simultaneous Raise Regression (SiR), the residual $\mathbf{e}_j$ is obtained from the regression of $\mathbf{x}_j$ over $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \dots, \mathbf{x}_p\}$, $\forall j \in \{1, 2, \dots, p\}$. In this case, it is verified that $\mathbf{e}_j \perp \mathbf{x}_i$ for $i \neq j$, $\mathbf{x}_j'\mathbf{e}_j = \mathbf{e}_j'\mathbf{e}_j$ and

- $\tilde{\mathbf{x}}_j'\tilde{\mathbf{x}}_j = \mathbf{x}_j'\mathbf{x}_j + k_j$ with $k_j = 2\lambda_j \mathbf{e}_j'\mathbf{e}_j + \lambda_j^2 \mathbf{e}_j'\mathbf{e}_j$.
- $\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j = \mathbf{x}_i'\mathbf{x}_j + \lambda_i \lambda_j \mathbf{e}_i'\mathbf{e}_j$.
- $\tilde{\mathbf{x}}_j'\mathbf{y} = \mathbf{x}_j'\mathbf{y} + \lambda_j \mathbf{e}_j'\mathbf{y}$.

Then, the estimation of model (16) **by OLS allows to show that relations** (17) **and** (18) **are not verified since $\mathbf{e}_j'\mathbf{e}_j \neq 0$ and $\mathbf{e}_j'\mathbf{y} \neq 0$. Thus, expressions (20) and (21) are not verified. The relation (19) will be verified selecting λ_j to obtain $k_1 = k_2 = \cdots = k_p = k$.**

However, it is possible to state that the span $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ coincides with the span generated by $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_p\}$. Consequently, the coefficient of determination of both models (expressions (1) and (16)) will coincide.

3.2. Total Simultaneous Raise Regression (TSiR)

Total Simultaneous Raise (TSiR) is based on the raising of every vector considering the residual $\mathbf{e}_j$ obtained from the regression of $\mathbf{x}_j$, $\forall j \in \{1, 2, \dots, p\}$, over vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \dots, \mathbf{x}_p, \mathbf{y}\}$. On contrary to SiR, the endogenous variable is considered joint to the $p-1$ exogenous variables. Thus, it is verified that $\mathbf{e}_j \perp \mathbf{x}_i$ for $i \neq j$, $\mathbf{x}_j'\mathbf{e}_j = \mathbf{e}_j'\mathbf{e}_j$, $\mathbf{e}_j \perp \mathbf{y}$, $\forall i$, and consequently

- $\tilde{\mathbf{x}}_j'\tilde{\mathbf{x}}_j = \mathbf{x}_j'\mathbf{x}_j + k_j$ with $k_j = 2\lambda_j \mathbf{e}_j'\mathbf{e}_j + \lambda_j^2 \mathbf{e}_j'\mathbf{e}_j$.
- $\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j = \mathbf{x}_i'\mathbf{x}_j + \lambda_i\lambda_j \mathbf{e}_i'\mathbf{e}_j$.
- $\tilde{\mathbf{x}}_j'\mathbf{y} = \mathbf{x}_j'\mathbf{y}$.

In this case, from the estimation of model (16) by OLS, the relation (20) is not verified since $\mathbf{e}_j'\mathbf{e}_j \neq 0$. That is, relation (17) is not verified but the relation (18) is verified. Relation (19) will be verified selecting λ_j to obtain $k_1 = k_2 = \cdots = k_p = k$.
Since the vector $\mathbf{y}$ was introduced in the regression to raise every vector, the span $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ does not coincide with the span generated by $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_p\}$. This conclusion is relevant, since the coefficient of determination of model (1) will be different to the coefficient of determination of model (16).

3.3. Successive Raise Regression (SuR)

The steps of this procedure for any value of p are as follows:

- *Step 1.* First, the regression of $\mathbf{x}_1$ is obtained over the set of vectors $\{\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_p\}$ obtaining the residual $\mathbf{e}_1$ that allows to describe the raise vector $\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda_1\mathbf{e}_1$ where $\mathbf{e}_1 \perp \mathbf{x}_i$ with $i \geq 2$ and $\mathbf{x}_1'\mathbf{e}_1 = \mathbf{e}_1'\mathbf{e}_1$. The raise vector will be used instead of $\mathbf{x}_1$.
- *Step 2.* The regression of $\mathbf{x}_2$ is obtained over the set of vectors $\{\tilde{\mathbf{x}}_1, \mathbf{x}_3, \dots, \mathbf{x}_p\}$, obtaining the residual $\mathbf{e}_2$ that allows to raise the second variable $\tilde{\mathbf{x}}_2 = \mathbf{x}_2 + \lambda_2\mathbf{e}_2$ where it is verified that $\mathbf{e}_2 \perp \tilde{\mathbf{x}}_1$, $\mathbf{e}_2 \perp \mathbf{x}_i$ with $i \geq 3$ and $\mathbf{x}_2'\mathbf{e}_2 = \mathbf{e}_2'\mathbf{e}_2$.
- *Step j.* The regression of variable $\mathbf{x}_j$ is obtained over the set of vectors $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_{j-1}, \mathbf{x}_{j+1}, \dots, \mathbf{x}_p\}$ obtaining the residual $\mathbf{e}_j$ that verifies

$\mathbf{e}_j \perp \tilde{\mathbf{x}}_i$, $\forall i \leq j-1$, $\mathbf{e}_j \perp \mathbf{x}_i$, $\forall i \geq j+1$ and $\mathbf{x}_j'\mathbf{e}_j = \mathbf{e}_j'\mathbf{e}_j$. Then, the variable $\mathbf{x}_j$ is replaced by $\tilde{\mathbf{x}}_j = \mathbf{x}_j + \lambda_j \mathbf{e}_j$.

- *Step p.* Finally, the regression of variable $\mathbf{x}_p$ is obtained over the set of vectors $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_{p-1}\}$ obtaining the residual $\mathbf{e}_p$. It is verified that $\mathbf{e}_p \perp \tilde{\mathbf{x}}_i$ $\forall i \leq p-1$. Then, the vector $\mathbf{x}_p$ is replaced by the raise vector $\tilde{\mathbf{x}}_p = \mathbf{x}_p + \lambda_p \mathbf{e}_p$.

The following results are obtained with the estimation of model (16) by OLS after the application of the successive raise regression:

- $\tilde{\mathbf{x}}_j'\tilde{\mathbf{x}}_j = \mathbf{x}_j'\mathbf{x}_j + k_j$ with $k_j = 2\lambda_j \mathbf{e}_j'\mathbf{e}_j + \lambda_j^2 \mathbf{e}_j'\mathbf{e}_j$.
- $\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j = \tilde{\mathbf{x}}_i'(\mathbf{x}_j + \lambda_j \mathbf{e}_j) = \tilde{\mathbf{x}}_i'\mathbf{x}_j + \lambda_j \tilde{\mathbf{x}}_i'\mathbf{e}_j = \tilde{\mathbf{x}}_i'\mathbf{x}_j = (\mathbf{x}_i + \lambda_i \mathbf{e}_i)'\mathbf{x}_j = \mathbf{x}_i'\mathbf{x}_j + \lambda_i \mathbf{e}_i'\mathbf{x}_j = \mathbf{x}_i'\mathbf{x}_j$.
- $\tilde{\mathbf{x}}_j'\mathbf{y} = \mathbf{x}_j'\mathbf{y} + \lambda_j \mathbf{e}_j'\mathbf{y}$.

Thus, when the model (16) is estimated by OLS, the relation (17) is verified but not the relation (18) since $\mathbf{e}_j'\mathbf{y} \neq 0$. To verify the relation (17) will be required to adequately select λ_j to obtain $k_1 = k_2 = \cdots = k_p = k$.
However, it is verified that the span $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ coincides with the span generated by $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_p\}$. Consequently, the coefficient of determination of both models (expressions (1) and (16)) will coincide.

3.4. Total Successive Raise Regression (TSuR)

Total successive raise regression (TSuR) consists in successively raising every independent variable of model (1) but including the dependent variable in the auxiliary regressions to obtain the residual vector $\mathbf{e}_j$.

- *Step 1.* First, the regression of $\mathbf{x}_1$ is obtained over the set of vectors $\{\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_p, \mathbf{y}\}$ obtaining the residual $\mathbf{e}_1$ that allows to describe the raise vector $\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda_1 \mathbf{e}_1$ where it is verified that $\mathbf{e}_1 \perp \mathbf{x}_i$ for $i \geq 2$, $\mathbf{e}_1 \perp \mathbf{y}$ and $\mathbf{x}_1'\mathbf{e}_1 = \mathbf{e}_1'\mathbf{e}_1$. The raise vector will be used instead of $\mathbf{x}_1$.
- *Step 2.* The regression of $\mathbf{x}_2$ is obtained over the set of vectors $\{\tilde{\mathbf{x}}_1, \mathbf{x}_3, \ldots, \mathbf{x}_p, \mathbf{y}\}$, obtaining the residual $\mathbf{e}_2$ that allows to raise the second variable $\tilde{\mathbf{x}}_2 = \mathbf{x}_2 + \lambda_2 \mathbf{e}_2$ where it is verified that $\mathbf{e}_2 \perp \tilde{\mathbf{x}}_1$, $\mathbf{e}_2 \perp \mathbf{x}_i$ with $i \geq 3$, $\mathbf{e}_2 \perp \mathbf{y}$ and $\mathbf{x}_2'\mathbf{e}_2 = \mathbf{e}_2'\mathbf{e}_2$.
- *Step j.* The regression of variable $\mathbf{x}_j$ is obtained over the set of vectors $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_{j-1}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_p, \mathbf{y}\}$ obtaining the residual $\mathbf{e}_j$ that verifies $\mathbf{e}_j \perp \tilde{\mathbf{x}}_i$, $\forall i \leq j-1$, $\mathbf{e}_j \perp \mathbf{x}_i$, $\forall i \geq j+1$, $\mathbf{e}_j \perp \mathbf{y}$ and $\mathbf{x}_j'\mathbf{e}_j = \mathbf{e}_j'\mathbf{e}_j$. Then, the variable $\mathbf{x}_j$ is replaced by $\tilde{\mathbf{x}}_j = \mathbf{x}_j + \lambda_j \mathbf{e}_j$.

- *Step p.* Finally, the regression of variable $\mathbf{x}_p$ is obtained over the set of vectors $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_{p-1}, \mathbf{y}\}$ obtaining the residual $\mathbf{e}_p$. It is verified that $\mathbf{e}_p \perp \tilde{\mathbf{x}}_i$, $\forall i \leq p-1$, $\mathbf{e}_p \perp \mathbf{y}$ and $\mathbf{x}'_p \mathbf{e}_p = \mathbf{e}'_p \mathbf{e}_p$. Then, the vector $\mathbf{x}_p$ is replaced by the raise vector $\tilde{\mathbf{x}}_p = \mathbf{x}_p + \lambda_p \mathbf{e}_p$.

In this case, it is verified that:

- $\tilde{\mathbf{x}}'_j \tilde{\mathbf{x}}_j = \mathbf{x}'_j \mathbf{x}_j + k_j$ with $k_j = 2\lambda_j \mathbf{e}'_j \mathbf{e}_j + \lambda_j^2 \mathbf{e}'_j \mathbf{e}_j$.
- $\tilde{\mathbf{x}}'_i \tilde{\mathbf{x}}_j = \mathbf{x}'_i \mathbf{x}_j$.
- $\tilde{\mathbf{x}}'_j \mathbf{y} = \mathbf{x}'_j \mathbf{y}$.

Thus, the relations (17) and (18) are verified if $k_1 = k_2 = \cdots = k_p = k$.
In summary, after the application of the total successive raise regression the matrix $\widetilde{\mathbf{X}}_{TSuR}$ is obtained and the model (16) is estimated by OLS. In this situation, the estimator $\widehat{\boldsymbol{\beta}}_{TSuR}$ of model (1) is given by:

$$\widehat{\boldsymbol{\beta}}_{TSuR} = \left(\widetilde{\mathbf{X}}'_{TSuR} \widetilde{\mathbf{X}}_{TSuR}\right)^{-1} \widetilde{\mathbf{X}}_{TSuR} \mathbf{y} = \left(\mathbf{X}'\mathbf{X} + \mathbf{K}\right)^{-1} \mathbf{X}'\mathbf{y}, \quad (22)$$

where $\mathbf{K} = diag(k_1, k_2, \dots, k_p)$. From expression $k_j = 2\lambda_j \mathbf{e}'_j \mathbf{e}_j + \lambda_j^2 \mathbf{e}'_j \mathbf{e}_j$, we obtain that $k_1 = k_2 = \cdots = k_p = k$ considering

$$\lambda_j = -1 + \sqrt{1 + \frac{k}{e'_j e_j}}, \quad (23)$$

and then

$$\widehat{\boldsymbol{\beta}}_{TSuR} = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1} \mathbf{X}'\mathbf{y}, \quad (24)$$

that is, $\widehat{\boldsymbol{\beta}}_{TSuR}$ coincides with the expression of the classical ridge estimator.
In conclusion, the expression associated with the ridge estimator $\widehat{\beta}_R(k) = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1} \mathbf{X}'\mathbf{y}$ has been obtained from TSuR with the appropriate choice of raising factors λ_j. **The RR is a particular case of the various possibilities that the raising procedure provides.**

3.5. Anomalies of the ridge regression

From the geometrical interpretation given when the ridge regression is obtained from TSuR, it is possible to show that the span generated by TSuR, $\{\widetilde{\mathrm{x}}_1, \widetilde{\mathrm{x}}_2, \dots, \widetilde{\mathrm{x}}_p\}$, does not coincide with the original span, $\{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_p\}$. Consequently, the application of TSuR, which is the same that the ordinary ridge regression, neither maintains the initial coefficient of determination or the experimental F statistic; see García et al. (2017). This generates doubts about the capacity and integrity of the ridge estimation. In our opinion, the ridge

estimator is questioned as an estimator in the strict sense or as a regularization process to obtain a system solution of normal equations when collinearity exists.
These results are in line with results found by García et al. (2017), where it is shown that

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{\mathbf{Y}})^2 &= \sum_{i=1}^{n}(\hat{Y}_i(k) - \overline{\mathbf{Y}} + e_i(k))^2 \\ &= \sum_{i=1}^{n}(\hat{Y}_i(k) - \overline{\mathbf{Y}})^2 + \sum_{i=1}^{n} e_i(k)^2 + 2 \cdot \sum_{i=1}^{n}(\hat{Y}_i(k) - \overline{\mathbf{Y}}) \cdot e_i(k),\end{aligned}$$

with $e_i(k) = Y_i - \hat{Y}_i(k) = Y_i - \mathbf{X}\hat{\beta}_i(k)$. **Contrary to OLS, in this case, it is not verified that** $\sum_{i=1}^{n}(\hat{Y}_i(k) - \overline{\mathbf{Y}}) \cdot e_i(k) = 0$ **for** $k > 0$. **Then, the sum of squares decomposition is not verified. Since the coefficient of determination is based on this decomposition, we consider that the coefficient of determination has no sense in ridge regression. For this same reason, the application of the experimental statistic F is also questioned.**

4. Relation with the generalized ridge regression (GRR)

Although the ridge regression obtained from TSuR satisfies the main purpose of this paper for geometrically justifying ridge regression, it is interesting to extend the analysis to the possible relation between TSuR and generalized ridge regression (GRR).

Given the model (1), it is possible to establish the decomposition $\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T} = \mathbf{\Gamma}$ **where** $\mathbf{\Gamma} = diag(\gamma_1, \gamma_2, ..., \gamma_p)$ **and** $\mathbf{T}$, $\mathbf{T}'\mathbf{T} = \mathbf{I} = \mathbf{T}\mathbf{T}'$ **are the eigenvalues and eigenvectors matrices of** $\mathbf{X}'\mathbf{X}$**, respectively.**
The canonical version of model (1) is:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}, \quad (25)$$

where $\mathbf{Z} = \mathbf{X}\mathbf{T}$ **and** $\boldsymbol{\alpha} = \mathbf{T}'\boldsymbol{\beta}$**. The OLS estimator of** $\boldsymbol{\alpha}$ **is given by:**

$$\widehat{\boldsymbol{\alpha}}_{OLS} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{\Gamma}^{-1}\mathbf{Z}'\mathbf{y}. \quad (26)$$

Then, the OLS estimator of $\boldsymbol{\beta}$ **is given by** $\widehat{\boldsymbol{\beta}}_{OLS} = \mathbf{T}\widehat{\alpha}_{OLS}$**.**
In this case, by following Hoerl and Kennard (1970b), the generalized ridge estimators are given by:

$$\widehat{\boldsymbol{\alpha}}_R(\mathbf{K}) = (\mathbf{Z}'\mathbf{Z} + \mathbf{K})^{-1}\mathbf{Z}'\mathbf{y}, \quad (27)$$

where $\mathbf{K} = diag(k_1, k_2, ..., k_p)$.
From expression (27), it is evident that the application of TSuR to the orthogonal model (25) leads to GRR, which is the ordinary ridge regression obtained as a particular case when adequately selecting λ_j to obtain $k_1 = k_2 = \cdots = k_p = k$.
On the other hand, since $\boldsymbol{\alpha} = \mathbf{T}'\boldsymbol{\beta}$, it is verified that the GRR estimator in the non-orthogonal model (1) is given by:

$$\widehat{\boldsymbol{\beta}}_R(\mathbf{K}) = \mathbf{T}\widehat{\boldsymbol{\alpha}}_R(\mathbf{K}) = (\mathbf{X}'\mathbf{X} + \mathbf{TKT}')^{-1}\mathbf{X}'\mathbf{y}. \tag{28}$$

Comparing this expression with the expression (22), it is evident that the application of TSuR to the non-orthogonal model does not lead to GRR. However, when $\mathbf{K} = k\mathbf{I}$, the estimator given by expression (28) can be expressed as $\widehat{\boldsymbol{\beta}}_R(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$; as shown in Section 3, the application of TSuR to the non-orthogonal model (1) also leads to RR.
In summary, GRR is obtained by applying TSuR to the orthogonal model, and consequently, RR is also obtained as a particular case by only selecting adequate raising factors. However, the application of TSuR to the non-orthogonal model does not lead to GRR although it leads to RR only adequately selecting raising factors.
Expression (22) introduces k_i in the diagonal of matrix $\mathbf{X}'\mathbf{X}$ that is supposed to be ill-conditioned. This can be considered more coherent with the work of Marshall et al. (1979) and Piegorsch and Casella (1989) that established the origins of ridge regression in the results provided by Riley (1955) about the conditioning of matrices $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{X} + k\mathbf{I}$.

5. Raise Regression and VIF

In this section, we will study the behaviour of the VIF in the raising methods. García et al. (2011) showed that the VIF associated with the raised variable diminishes when raise regression is applied. The following theorem shows that the VIF of the remaining variables also diminishes.

Theorem 1 *When A is the set of vectors determined by vectors $\{\mathbf{x}_1\ \mathbf{x}_2 \dots \mathbf{x}_p\}$ and B is the set of vectors $\{\tilde{\mathbf{x}}_1\ \mathbf{x}_2 \dots \mathbf{x}_p\}$ where $\tilde{\mathbf{x}}_1$ is the raised vector of $\mathbf{x}_1$, it is verified that the VIF of every vector in the set of vectors B is less that or equal to the VIF associated with the vectors of the set of vectors A.*

Proof. García et al. (2011, Theorem 4.2) proved that "upon raising a variable its VIF diminishes"; that is, the VIF of vector $\mathbf{x}_1$ in the set of vectors

$\boldsymbol{\mathcal{A}} = \{\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_p\}$ is higher than the VIF of vector $\tilde{\mathbf{x}}_1$ in the set of vectors $\boldsymbol{\mathcal{B}} = \{\tilde{\mathbf{x}}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_p\}$. The main goal of this section is to show that the VIFs of vectors $\mathbf{x}_j$ with $j \neq 1$ in the set of vectors $\boldsymbol{\mathcal{B}}$ are lower than the VIFs of a vector in the set of vectors $\boldsymbol{\mathcal{A}}$. Without loss of generality, the VIF associated with vector $\mathbf{x}_2$ is obtained. Thus, we will first estimate vector $\mathbf{x}_2$ on the rest of the set of vectors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ and then compare the associated VIFs in each set. The estimations of $\mathbf{x}_2$ in each set of vectors are as follows:

$$\hat{\mathbf{x}}_2(\boldsymbol{\mathcal{A}}) = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{x}_2, \tag{29}$$

$$\hat{\mathbf{x}}_2(\boldsymbol{\mathcal{B}}) = \left(\mathbf{T}'\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{x}_2, \tag{30}$$

where

$$\mathbf{Z} = \begin{pmatrix} \mathbf{x}_{11} & \mathbf{x}_{31} & \cdots & \mathbf{x}_{p1} \\ \mathbf{x}_{12} & \mathbf{x}_{32} & \cdots & \mathbf{x}_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{1n} & \mathbf{x}_{3n} & \cdots & \mathbf{x}_{pn} \end{pmatrix}_{n\times(p-1)},$$

and

$$\mathbf{T} = \begin{pmatrix} \mathbf{x}_{11} + \lambda_1\mathbf{e}_{11} & \mathbf{x}_{31} & \cdots & \mathbf{x}_{p1} \\ \mathbf{x}_{12} + \lambda_1\mathbf{e}_{12} & \mathbf{x}_{32} & \cdots & \mathbf{x}_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{1n} + \lambda_1\mathbf{e}_{1n} & \mathbf{x}_{3n} & \cdots & \mathbf{x}_{pn} \end{pmatrix}_{n\times(p-1)}.$$

Given the relation $\mathbf{T}'\mathbf{T} = \mathbf{Z}'\mathbf{Z} + \mathbf{K}$ (with $\mathbf{K}$ a square matrix of order $p-1$ for which the unique element different from zero is in the position (1,1) where its value is equal to $k = \lambda_1^2\mathbf{e}_1'\mathbf{e}_1 + 2\lambda_1\mathbf{e}_1'\mathbf{x}_1$ or alternatively $k = \lambda_1^2\mathbf{e}_1'\mathbf{e}_1 + 2\lambda_1\mathbf{e}_1'\mathbf{e}_1$ and $\mathbf{T}'\mathbf{x}_2 = \mathbf{Z}'\mathbf{x}_2$ (since $\mathbf{x}_2 \perp \mathbf{e}_1$), it is then verified that $\widehat{\mathbf{x}}_2(\boldsymbol{\mathcal{B}}) = \left(\mathbf{T}'\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{x}_2 = \left(\mathbf{Z}'\mathbf{Z} + \mathbf{K}\right)^{-1}\mathbf{Z}'\mathbf{x}_2 = \hat{\mathbf{x}}_2^R(\boldsymbol{\mathcal{A}})$. That is, the estimation of vector $\mathbf{x}_2$ in the set of vectors $\boldsymbol{\mathcal{B}}$ is the same as applying the generalized ridge estimation (with $k_1 = k$ and $k_i = 0$ for $i \neq 1$) of $\mathbf{x}_2$ in the set of vectors $\boldsymbol{\mathcal{A}}$.

However, the error sum of squares of the set of vectors are, respectively:

$$\begin{aligned} \text{RSS}(\boldsymbol{\mathcal{A}}) &= \mathbf{x}_2'\mathbf{x}_2 - \hat{\mathbf{x}}_2(\boldsymbol{\mathcal{A}})'\mathbf{Z}'\mathbf{x}_2 = \mathbf{x}_2'\mathbf{x}_2 - \mathbf{x}_2'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{x}_2, \\ \text{RSS}(\boldsymbol{\mathcal{B}}) &= \mathbf{x}_2'\mathbf{x}_2 - \hat{\mathbf{x}}_2(\boldsymbol{\mathcal{B}})'\mathbf{T}'\mathbf{x}_2 = \mathbf{x}_2'\mathbf{x}_2 - \mathbf{x}_2'\mathbf{T}\left(\mathbf{T}'\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{x}_2, \end{aligned}$$

the difference of which is as follows:

$$\begin{aligned} \text{RSS}(\boldsymbol{\mathcal{B}}) - \text{RSS}(\boldsymbol{\mathcal{A}}) &= \mathbf{x}_2'\mathbf{Z}\left[\left(\mathbf{Z}'\mathbf{Z}\right)^{-1} - \left(\mathbf{T}'\mathbf{T}\right)^{-1}\right]\mathbf{Z}'\mathbf{x}_2 \\ &= \mathbf{a}'\left[\left(\mathbf{Z}'\mathbf{Z}\right)^{-1} - \left(\mathbf{T}'\mathbf{T}\right)^{-1}\right]\mathbf{a}, \end{aligned}$$

where $\mathbf{Z}'\mathbf{x}_2 = \mathbf{a}$.

To show that the VIFs associated with the vector of set of vectors $\boldsymbol{\mathcal{B}}$ are lower than the VIFs of the vectors of the set of vectors $\boldsymbol{\mathcal{A}}$, it is sufficient to check that $\mathbf{a}' \left[(\mathbf{Z}'\mathbf{Z})^{-1} - (\mathbf{T}'\mathbf{T})^{-1} \right] \mathbf{a}$ is a positive quadratic form since the total sum of squares is equal in both regressions and a greater RSS value leads to a lower VIF value.
Using partitioned matrices, for example, write $\mathbf{Z} = (\mathbf{x}_1|\mathbf{Z}_{[1]})$ and $\mathbf{T} = \mathbf{Z} + (\lambda_1\mathbf{e}_1|\mathbf{0})$ with $\mathbf{e}_1 \perp \mathbf{X}_{[1]}$ and taking into consideration that it is verified that $\mathbf{T}'\mathbf{T} = \mathbf{Z}'\mathbf{Z} + \mathbf{K}$, it is possible to apply the Sherman-Morrison formula for matrix inverses $(\mathbf{A} + \mathbf{w}\mathbf{w}')^{-1} = \mathbf{A}^{-1} - \frac{1}{1+\mathbf{w}'\mathbf{A}^{-1}\mathbf{w}}\mathbf{A}^{-1}\mathbf{w}\mathbf{w}'\mathbf{A}^{-1}$ with $\mathbf{Z}'\mathbf{Z}$ as $\mathbf{A}$ and $\mathbf{w}' = (\sqrt{k}|\mathbf{0})$ to conclude the following:

$$(\mathbf{Z}'\mathbf{Z})^{-1} - (\mathbf{T}'\mathbf{T})^{-1} = \frac{1}{1 + k(\mathbf{Z}'\mathbf{Z})^{-1}_{1,1}} (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{K}(\mathbf{Z}'\mathbf{Z})^{-1} \tag{31}$$

which has the form $\mathbf{V}'\mathbf{V}$ and thus is nonnegative definite for $k \geq 0$.
In addition, the expression (31) will be equal to zero only when $\mathbf{a}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{K}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{a} = 0$ and taking into account the form of matrix $\mathbf{K}$, this will only occur when the first coordinate of $\mathbf{a}'$ is zero, that is, when $\mathbf{x}_2 \perp \mathbf{x}_1$. That is, raising vector $\mathbf{x}_1$ by the raise regression diminishes the VIFs of the rest of the vectors that are not perpendicular to $\mathbf{x}_1$. □

In conclusion, the successive raise regression (SuR) and the total successive raise regression (TSuR) guarantee the mitigation of collinearity. However, the simultaneous raise regression (SiR) can not guarantee this fact as will be shown in the empirical application **(see Tables 2 to 4). This information is summarized in the following corollaries:**.

Corollary 1 *The VIFs associated with all variables (both raised and unraised) diminish in every step of successive raise regression (SuR). It is to say, the successive raise regression guarantees that collinearity is mitigated. This fact is also verified in the Total Successive raise regression (TSuR).*

Corollary 2 *The case of simultaneous raise regression (SiR) and total simultaneous raise regression (TSiR), the simultaneous raising of all vector does not guarantees the mitigation of collinearity.*

6. Application

To illustrate the contributions of this paper, an empirical application is presented by using the Economic Report of the President cited by Wissel (2009)

but enlarging the sample with data from 1995 to 2011. We study the linear relationship between the variable of outstanding mortgage debt (in trillions of dollars), $\mathbf{y}$, and the independent variables of personal consumption (in trillions of dollars), $\mathbf{x}_1$, personal income (in trillions of dollars), $\mathbf{x}_2$, and outstanding consumer credit (in trillions of dollars), $\mathbf{x}_3$.
The standardized estimated model is:

$$\begin{array}{rcccc} \widehat{\mathbf{y}} & = & \text{-}0.4099\ \mathbf{x}_1 & +\ 0.5071\ \mathbf{x}_2 & +\ 0.8827\ \mathbf{x}_3, \\ & & (0.6791) & (0.3331) & (0.8144) \end{array}$$

where the coefficient of determination is $R^2 = 0.92325$ and $F_{exp} = 107.3174$, and this is therefore a globally significant model. Indeed, the standard deviation indicates that all the independent variables have coefficients not individually significant at the 95% confidence level. Finally, the VIF confirms the presence of high collinearity since VIF($\mathbf{x}_1$) = 154.9488, VIF($\mathbf{x}_2$) = 37.2682 and VIF($\mathbf{x}_3$) = 222.8143.
Considering that $k = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 10, 11, 12, 13, 14, 15, 20, 50, 100, Table 1 shows the values of λ_i, $i = 1, 2, 3$ that make TSuR coincide with RR, applying only the expression (23). From these values, Tables 2 to 4 show the values calculated for the VIFs of each independent variable in RR (following the methodology presented by García et al. (2016)), SiR, SuR and TSuR. Thus, note that the VIFs in RR and TSuR have been calculated for the same values of k and the VIFs in SiR, SuR and TSuR have been calculated for the same values of the raising factor, λ_i. In addition, k increases as the raising factor increases.
From this information, it is possible to obtain the following conclusions:

- The values of VIF calculated from SuR decrease as the raising factors increase. However, in the case of SiR, the values of the VIF decrease until **a certain value of** k ($k = 10$ for $\mathbf{x}_1$ and $\mathbf{x}_3$ and $k = 11$ for $\mathbf{x}_2$), and from these values, the value of the VIF increases. Therefore, SuR seems to exhibit better behavior than SiR, as mentioned in Section 5. Indeed, since SuR maintains the residual, explained and total sums of squares of the original model, the coefficient of determination and the global significance are not altered by the successive raising of the vectors.
- Given the requirement that RR and TSuR have the same values of λ_i, $i = 1, 2, 3$, using the expression (23) verifies that $\widetilde{\mathbf{X}}'_{TSuR}\widetilde{\mathbf{X}}_{TSuR} = \mathbf{X}'\mathbf{X} + k\mathbf{I}$. Because the calculation of the VIF is based on these matrices, the obtained values have to coincide. García et al. (2016) established that the extension of the concept of VIF to RR is continuous, monotone decreasing and equal to or higher than one. Consequently, the VIFs obtained from TSuR will also automatically verify these desirable properties.

Table 1. Values of λ_i, $i = 1, 2, 3$ that make total successive taise (TSuR) coincide with ridge regression (RR)

k	λ_1	λ_2	λ_3
0	0	0	0
0.1	3.11073554	1.29575071	1.1093676
0.2	4.72680482	2.08811668	1.54302882
0.3	5.97814016	2.71515356	1.90555742
0.4	7.03695134	3.25067562	2.22582099
0.5	7.97166281	3.72589788	2.51635375
0.6	8.81778387	4.15751619	2.78432199
0.7	9.59655731	4.55570274	3.03435342
0.8	10.3218891	4.92719924	3.26966227
0.9	11.0034712	5.27674657	3.49259561
1	11.648378	5.60782895	3.70493207
10	38.8850181	19.6782531	13.0627471
11	40.8305646	20.685218	13.7401913
12	42.6895594	21.6474547	14.3878392
13	44.4726189	22.5704418	15.0093077
14	46.1883517	23.4586233	15.6075361
15	47.8438533	24.3156627	16.1849518
20	55.3970685	28.2262793	18.8213019
50	88.1631838	45.1944926	30.2770127
100	125.091819	64.3211863	43.2025271

Table 2. VIFs of $\mathbf{x}_1$ obtained with ridge regression (RR), the simultaneous raise method (SiR), successive raise method (SuR) and total successive raise method (TSuR) from the values of k and λ of Table 1

k	RR	SiR	SuR	TSuR
0	154.9488	154.9488	154.9488	154.9488
0.1	6.7663	7.6507	5.7482	6.7663
0.2	3.9217	4.3713	3.2431	3.9217
0.3	2.9195	3.1894	2.4079	2.9195
0.4	2.4088	2.2337	1.9986	2.4088
0.5	2.1003	2.3317	1.7587	2.1003
0.6	1.8943	2.0907	1.6025	1.8943
0.7	1.7476	2.0487	1.4936	1.7476
0.8	1.6380	1.7886	1.4138	1.6380
0.9	1.5533	1.6882	1.3531	1.5533
1	1.4860	1.6082	1.3056	1.4860
10	1.0150	1.1854	1.0071	1.0150
11	1.0127	1.2064	1.0059	1.0127
12	1.0109	1.2284	1.0050	1.0109
13	1.0094	1.2512	1.0043	1.0094
14	1.0082	1.2745	1.0038	1.0082
15	1.0072	1.2983	1.0033	1.0072
20	1.0042	1.4209	1.0019	1.0042
50	1.0007	2.1621	1.0003	1.0007
100	1.0002	3.3042	1.0001	1.0002

Table 3. VIFs of $\mathbf{x}_2$ obtained with ridge regression (RR), the simultaneous raise method (SiR), successive raise method (SuR) and total successive raise method (TSuR) from the values of k and λ of Table 1

k	RR	SiR	SuR	TSuR
0	37.2682	37.2682	37.2682	37.2682
0.1	6.1599	6.9596	5.1850	6.1599
0.2	3.7429	2.8261	3.0800	3.7429
0.3	2.8327	3.2352	2.3310	2.8327
0.4	2.3568	2.5667	1.9539	2.3568
0.5	2.0652	2.3917	1.7295	2.0652
0.6	1.8689	2.1603	1.5819	1.8689
0.7	1.7281	1.9988	1.4783	1.7281
0.8	1.6226	1.8657	1.4019	1.6226
0.9	1.5407	1.7662	1.3437	1.5407
1	1.4755	1.6863	1.2980	1.4755
10	1.0149	1.0774	1.0070	1.0149
11	1.0125	1.0773	1.0058	1.0125
12	1.0107	1.0782	1.0050	1.0107
13	1.0093	1.0800	1.0043	1.0093
14	1.0081	1.0825	1.0037	1.0081
15	1.0072	1.0854	1.0033	1.0072
20	1.0042	1.1053	1.0019	1.0042
50	1.0007	1.2652	1.0003	1.0007
100	1.0002	1.5310	1.0001	1.0002

Table 4. VIFs of $\mathbf{x}_3$ obtained with ridge regression (RR), the simultaneous raise method (SiR), successive raise method (SuR) and total successive raise method (TSuR) from the values of k and λ of Table 1

k	RR	SiR	SuR	TSuR
0	222.8143	222.8143	222.8143	154.9488
0.1	7.1161	9.7907	4.6014	7.1161
0.2	4.0248	4.5303	2.3020	4.0248
0.3	2.9695	3.5731	1.6768	2.9695
0.4	2.4388	1.5525	1.4155	2.4388
0.5	2.1205	2.5478	1.2813	2.1205
0.6	1.9090	2.2312	1.2031	1.9090
0.7	1.7588	2.0815	1.1536	1.7588
0.8	1.6469	1.8388	1.1202	1.6469
0.9	1.5606	1.7096	1.0967	1.5606
1	1.4921	1.6073	1.0794	1.4921
10	1.0151	1.1234	1.0009	1.0151
11	1.0128	1.1550	1.0007	1.0128
12	1.0109	1.1878	1.0006	1.0109
13	1.0095	1.2215	1.0005	1.0095
14	1.0083	1.2559	1.0005	1.0083
15	1.0073	1.2908	1.0004	1.0073
20	1.0043	1.4695	1.0002	1.0043
50	1.0007	2.5405	1.0000	1.0007
100	1.0002	4.1864	1.0000	1.0002

Table 5. Main characteristics of different raising methods

	SuR	TSuR	SiR
It leads to ridge regression	No	Yes	No
The VIF diminishes in every step	Yes	Yes	No
The R^2 and the F_{exp} are maintained	Yes	No	Yes

- Analogously, it is possible to impose the condition that the values of k in SuR and TSuR coincide. In this case, we obtain that $\widetilde{\mathbf{X}}'_{SuR}\widetilde{\mathbf{X}}_{SuR} = \mathbf{X}'\mathbf{X} + k\mathbf{I} = \widetilde{\mathbf{X}}'_{TSuR}\widetilde{\mathbf{X}}_{TSuR}$, and for the same reason as above, the VIFs obtained from the two methodologies coincide.

- Note that the VIFs associated with SuR are smaller than and decrease faster than the VIFs associated with TSuR.

To conclude, Table 5 shows the main characteristics of the different raising methods.

7. Conclusions

This paper shows that the RR is obtained from the total successive raise (TSuR), which is a particular case of the various possibilities and flexibilities that the raising procedure provides. From this geometrical interpretation, the span generated by the TSuR, $\{\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \dots, \widetilde{\mathbf{x}}_p\}$, is shown not to coincide with the original span, $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$. This fact could explain the anomalies of RR in the coefficient of determination or the experimental F statistic. Additionally, it is shown that in TSuR, the collinearity is mitigated since the VIFs associated with all variables, both raised and not raised, diminish.

Secondly, the paper discusses the simultaneous raise regression (SiR) presented by García et al. (2014) and the successive raise regression (SuR) presented by García and Ramirez (2016). The study concludes that SuR is the preferred method, as it guarantees that the VIFs associated with all variables, both raised and not raised, diminish; that is, SuR ensures that collinearity is mitigated. At the same time, the SuR guarantees that the span generated coincides with the original, and consequently, it is possible to obtain the coefficient of determination.

Acknowledgements

We would like to thank referees for their detailed, encouraging and constructive reviews of our paper which clearly contributed to improving both the structure and the content of this manuscript.

References

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(3), 370–374.

Besley, D. A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression.* New York: John Wiley.

Casella, G. (1980). Minimax ridge regression estimation. *The Annals of Statistics*, 1036–1056.

Chang, X. and H. Yang (2012). Combining two-parameter and principal component regression estimators. *Statistical Papers 53*(3), 549–562.

García, C., J. García, M. D. M. López, and R. Salmerón (2014). El método de alzado sucesivo y su relación con el método cresta. In *Anales de economía aplicada 2014*, pp. 1323–1338. Asociación Española de Economía Aplicada, ASEPELT.

García, C., R. Salmerón, A. Rodríguez, and J. García (2017). Regresión cresta: algunos inconvenientes. In *Anales de economía aplicada 2017*, pp. 15–25. Asociación Española de Economía Aplicada, ASEPELT.

García, C. B., J. García, and J. Soto (2011). The Raise Method: An alternative procedure to estimate the parameters in presence of collinearity. *Quality and Quantity 45*(2), 403–423.

García, J., C. García, R. Salmerón, and M. M. López (2016). The raise estimators. Estimation, inference and properties. *Communications in Statistics–Theory and Methods In press*, 1–19.

García, J. and D. E. Ramirez (2016). The successive raising estimator and its relation with the ridge estimator. *In press.*

García, J., R. Salmerón, C. García, and M. M. López (2016). Standardization of Variables and Collinearity Diagnostic in Ridge Regression. *International Statistical Review 84, 2*, 245–266.

Hoerl, A. E. and R. W. Kennard (1970a). Ridge regression: Applications to nonorthogonal problems. *Technometrics 12*, 69–82.

Hoerl, A. E. and R. W. Kennard (1970b). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*, 55–67.

Liu, K. (1993). A new class of blased estimate in linear regression. *Communications in Statistics-Theory and Methods 22*(2), 393–402.

Liu, K. (2003). Using liu-type estimator to combat collinearity. *Communications in Statistics-Theory and Methods 32*(5), 1009–1020.

Liu, X.-Q., F. Gao, and Z.-F. Yu (2013). Improved ridge estimators in a linear regression model. *Journal of Applied Statistics 40*(1), 209–220.

Marquardt, D. W. and S. R. Snee (1975). Ridge regression in practice. *The American Statistician 29*(1), 3–20.

Marshall, A. W., I. Olkin, and B. C. Arnold (1979). *Inequalities: theory of majorization and its applications*, Volume 143. Springer.

Mayer, L. S. and T. A. Willke (1973). On biased estimation in linear models. *Technometrics 15*(3), 497–508.

Piegorsch, W. W. and G. Casella (1989). The early use of matrix diagonal increments in statistical problems. *SIAM review 31*(3), 428–434.

Riley, J. (1955). Solving systems of linear equations with a positive definite, symmetric but possibly ill-conditiones matrix. *Math. Tables Aids Comput. 9*, 96–101.

Rolph, J. E. (1976). Choosing shrinkage estimators for regression problems. *Communications in Statistics-Theory and Methods 5*(9), 789–802.

Sakallıoğlu, S. and S. Kaçıranlar (2008). A new biased estimator based on ridge estimation. *Statistical Papers 49*(4), 669–689.

Spanos, A. and A. McGuirk (2002). The problem of near-multicollinearity revisited: erratic vs systematic volatility. *Journal of Econometrics 108*(2), 365–393.

Stein, C. et al. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 197–206.

Strawderman, W. E. (1978). Minimax adaptive generalized ridge regression estimators. *Journal of the American Statistical Association 73*(363), 623–627.

Swindel, B. F. (1981). Geometry of ridge regression illustrated. *The American Statistician 35*(1), 12–15.

Wichers, C. R. (1975). The detection of multicollinearity: a comment. *The Review of Economics and Statistics*, 366–368.

Willan, A. R. and D. G. Watts (1978). Meaningful multicollinearity measures. *Technometrics 20*(4), 407–412.

Wissel, J. (2009). A new biased estimator for multivariate regression models with highly collinear variables.