




The VIF and MSE in Raise Regression

Román Salmerón Gómez ¹, Ainara Rodríguez Sánchez ² and Catalina García García ^{1,*}
and José García Pérez ³

¹ Department of Quantitative Methods for Economics and Business, University of Granada, 18010 Granada, Spain; romansg@ugr.es

² Department of Economic Theory and History, University of Granada, 18010 Granada, Spain; arsanchez@ugr.es

³ Department of Economy and Company, University of Almería, 04120 Almería, Spain; jgarcia@ual.es

* Correspondence: cbgarcia@ugr.es; Tel.: +34-958248790

Received: 1 April 2020; Accepted: 13 April 2020; Published: 16 April 2020



Abstract: The raise regression has been proposed as an alternative to ordinary least squares estimation when a model presents collinearity. In order to analyze whether the problem has been mitigated, it is necessary to develop measures to detect collinearity after the application of the raise regression. This paper extends the concept of the variance inflation factor to be applied in a raise regression. The relevance of this extension is that it can be applied to determine the raising factor which allows an optimal application of this technique. The mean square error is also calculated since the raise regression provides a biased estimator. The results are illustrated by two empirical examples where the application of the raise estimator is compared to the application of the ridge and Lasso estimators that are commonly applied to estimate models with multicollinearity as an alternative to ordinary least squares.

Keywords: detection; mean square error; multicollinearity; raise regression; variance inflation factor

1. Introduction

In the last fifty years, different methods have been developed to avoid the instability of estimates derived from collinearity (see, for example, Kiers and Smilde [1]). Some of these methods can be grouped within a general denomination known as penalized regression.

In general terms, the penalized regression parts from the linear model (with p variables and n observations), $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, and obtains the regularization of the estimated parameters, minimizing the following objective function:

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + P(\boldsymbol{\beta}),$$

where $P(\boldsymbol{\beta})$ is a penalty term that can take different forms. One of the most common penalty terms is the bridge penalty term ([2,3]) is given by

$$P(\boldsymbol{\beta}) = \lambda \sum_{j=1}^p |\beta_j|^\alpha, \alpha > 0,$$

where λ is a tuning parameter. Note that the ridge ([4]) and the Lasso ([5]) regressions are obtained when $\alpha = 2$ and $\alpha = 1$, respectively. Penalties have also been called soft thresholding ([6,7]).

These methods are applied not only for the treatment of multicollinearity but also for the selection of variables (see, for example, Dupuis and Victoria-Feser [8], Li and Yang [9] Liu et al. [10], or Uematsu and Tanaka [11]), which is a crucial issue in many areas of science when the number of variables

exceeds the sample size. Zou and Hastie [12] proposed elastic net regularization by using the penalty terms λ_1 and λ_2 that combine the Lasso and ridge regressions:

$$P(\beta) = \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2.$$

Thus, the Lasso regression usually selects one of the regressors from among all those that are highly correlated, while the elastic net regression selects several of them. In the words of Tutz and Ulbricht [13] “the elastic net catches all the big fish”, meaning that it selects the whole group.

From a different point of view, other authors have also presented different techniques and methods well suited for dealing with the collinearity problems: continuum regression ([14]), least angle regression ([15]), generalized maximum entropy ([16–18]), the principal component analysis (PCA) regression ([19,20]), the principal correlation components estimator ([21]), penalized splines ([22]), partial least squares (PLS) regression ([23,24]), or the surrogate estimator focused on the solution of the normal equations presented by Jensen and Ramirez [25].

Focusing on collinearity, the ridge regression is one of the more commonly applied methodologies and it is estimated by the following expression:

$$\hat{\beta}(K) = (\mathbf{X}^t\mathbf{X} + K \cdot \mathbf{I})^{-1} \mathbf{X}^t\mathbf{Y} \tag{1}$$

where \mathbf{I} is the identity matrix with adequate dimensions and K is the ridge factor (ordinary least squares (OLS) estimators are obtained when $K = 0$). Although ridge regression has been widely applied, it presents some problems with current practice in the presence of multicollinearity and the estimators derived from the penalty come into these same problems whenever $n > p$:

- In relation to the calculation of the variance inflation factors (VIF), measures that quantify the degree of multicollinearity existing in a model from the coefficient of determination of the regression between the independent variables (for more details, see Section 2), García et al. [26] showed that the application of the original data when working with the ridge estimate leads to non-monotone VIF values by considering the VIF as a function of the penalty term. Logically, the Lasso and the elastic net regression inherit this property.
- By following Marquardt [27]: “The least squares objective function is mathematically independent of the scaling of the predictor variables (while the objective function in ridge regression is mathematically dependent on the scaling of the predictor variables)”. That is to say, the penalized objective function will bring problems derived from the standardization of the variables. This fact has to be taken into account both for obtaining the estimators of the regressors and for the application of measures that detect if the collinearity has been mitigated. Other penalized regressions (such as Lasso and elastic net regressions) are not scale invariant and hence yield different results depending on the predictor scaling used.
- Some of the properties of the OLS estimator that are deduced from the normal equations are not verified by the ridge estimator and, among others, the estimated values for the endogenous variable are not orthogonal to the residuals. As a result, the following decomposition is verified

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i(K) - \bar{Y})^2 + \sum_{i=1}^n e_i(K)^2 + 2 \sum_{i=1}^n (\hat{Y}_i(K) - \bar{Y}) \cdot e_i(K).$$

When the OLS estimators are obtained ($K = 0$), the third term is null. However, this term is not null when K is not zero. Consequently, the relationship $TSS(K) = ESS(K) + RSS(K)$ is not satisfied in ridge regression, and the definition of the coefficient of determination may not be suitable. This fact not only limits the analysis of the goodness of fit but also affects the global significance since the critical coefficient of determination is also questioned. Rodríguez et al. [28]

showed that the estimators obtained from the penalties mentioned above inherit the problem of the ridge regression in relation to the goodness of fit.

In order to overcome these problems, this paper is focused on the raise regression (García et al. [29] and Salmerón et al. [30]) based on the treatment of collinearity from a geometrical point of view. It consists in separating the independent variables by using the residuals (weighted by the raising factor) of the auxiliary regression traditionally used to obtain the VIF. Salmerón et al. [30] showed that the raise regression presents better conditions than ridge regression and, more recently, García et al. [31] showed, among other questions, that the ridge regression is a particular case of the raise regression.

This paper presents the extension of the VIF to the raise regression showing that, although García et al. [31] showed that the application of the raise regression guarantees a diminishing of the VIF, it is not guaranteed that its value is lower the threshold traditionally established as troubling. Thus, it will be concluded that an unique application of the raise regression does not guarantee the mitigation of the multicollinearity. Consequently, this extension complements the results presented by García et al. [31] and determines, on the one hand, whether it is necessary to apply a successive raise regression (see García et al. [31] for more details) and, on the other hand, the most adequate variable for raising and the most optimal value for the raising factor in order to guarantee the mitigation of the multicollinearity.

On the other hand, the transformation of variables is common when strong collinearity exists in a linear model. The transformation to unit length (see Belsley et al. [32]) or standardization (see Marquardt [27]) is typical. Although the VIF is invariant to these transformations when it is calculated after estimation by OLS (see García et al. [26]), it is not guaranteed either in the case of the raise regression or in ridge regression as showed by García et al. [26]. The analysis of this fact is one of the goals of this paper.

Finally, since the raise estimator is biased, it is interesting to calculate its mean square error (MSE). It is studied whether the MSE of the raise regression is less than the one obtained by OLS. In this case, this study could be used to select an adequate raising factor similar to what is proposed by Hoerl et al. [33] in the case of the ridge regression. Note that estimators with MSE less than the one from OLS estimators are traditionally preferred (see, for example, Stein [34], James and Stein [35], Hoerl and Kennard [4], Ohtani [36], or Hubert et al. [37]). In addition, this measure allows us to conclude whether the raise regression is preferable, in terms of MSE, to other alternative techniques.

The structure of the paper is as follows: Section 2 briefly describes the VIF and the raise regression, and Section 3 extends the VIF to this methodology. Some desirable properties of the VIF are analyzed, and its asymptotic behavior is studied. It is also concluded that the VIF is invariant to data transformation. Section 4 calculates the MSE of the raise estimator, showing that there is a minimum value that is less than the MSE of the OLS estimator. Section 5 illustrates the contribution of this paper with two numerical examples. Finally, Section 6 summarizes the main conclusions of this paper.

2. Preliminaries

2.1. Variance Inflation Factor

The following model for p independent variables and n observations is considered:

$$\mathbf{Y} = \beta_1 + \beta_2 \mathbf{X}_2 + \cdots + \beta_i \mathbf{X}_i + \cdots + \beta_p \mathbf{X}_p + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (2)$$

where \mathbf{Y} is a vector $n \times 1$ that contains the observations of the dependent variable, $\mathbf{X} = [\mathbf{1} \ \mathbf{X}_2 \ \dots \ \mathbf{X}_i \ \dots \ \mathbf{X}_p]$ (with $\mathbf{1}$ being a vector of ones with dimension $n \times 1$) is a matrix with order $n \times p$ that contains (by columns) the observations of the independent variables, $\boldsymbol{\beta}$ is a vector $p \times 1$ that contains the coefficients of the independent variables, and \mathbf{u} is a vector $n \times 1$ that represents the random disturbance that is supposed to be spherical ($E[\mathbf{u}] = \mathbf{0}$ and $Var(\mathbf{u}) = \sigma^2 \mathbf{I}$, where $\mathbf{0}$ is a vector with zeros with dimension $n \times 1$ and \mathbf{I} the identity matrix with adequate dimensions, in this case $p \times p$).

Given the model in Equation (2), the variance inflation factor (VIF) is obtained as follows:

$$VIF(k) = \frac{1}{1 - R_k^2}, \quad k = 2, \dots, p, \tag{3}$$

where R_k^2 is the coefficient of determination of the regression of the variable X_k as a function of the rest of the independent variables of the model in Equation (2):

$$X_k = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_{k-1} X_{k-1} + \alpha_{k+1} X_{k+1} + \dots + \alpha_p X_p + v = X_{-k} \alpha + v, \tag{4}$$

where X_{-k} corresponds to the matrix X after the elimination of the column k (variable X_k).

If the variable X_k has no linear relationship (i.e., is orthogonal) with the rest of the independent variables, the coefficient of determination will be zero ($R_k^2 = 0$) and the $VIF(k) = 1$. As the linear relationship increases, the coefficient of determination (R_k^2) and consequently $VIF(k)$ will also increase. Thus, the higher the VIF associated with the variable X_k , the greater the linear relationship between this variable and the rest of the independent variables in the model in Equation (2). It is considered that the collinearity is troubling for values of VIF higher than 10. Note that the VIF ignores the role of the constant term (see, for example, Salmerón et al. [38] or Salmerón et al. [39]), and consequently, this extension will be useful when the multicollinearity is essential; that is to say, when there is a linear relationship between at least two independent variables of the model of regression without considering the constant term (see, for example, Marquandt and Snee [40] for the definitions of essential and nonessential multicollinearity).

2.2. Raise Regression

Raise regression, presented by García et al. [29] and more developed further by Salmerón et al. [30], uses the residuals of the model in Equation (4), e_k , to raise the variable k as $\tilde{X}_k = X_k + \lambda e_k$ with $\lambda \geq 0$ (called the raising factor) and to verify that $e_k^T X_{-k} = 0$, where 0 is a vector of zeros with adequate dimensions. In that case, the raise regression consists in the estimation by OLS of the following model:

$$Y = \beta_1(\lambda) + \beta_2(\lambda) X_2 + \dots + \beta_k(\lambda) \tilde{X}_k + \dots + \beta_p(\lambda) X_p + \tilde{u} = \tilde{X} \beta(\lambda) + \tilde{u}, \tag{5}$$

where $\tilde{X} = [1 \ X_2 \dots \tilde{X}_k \dots X_p] = [X_{-k} \ \tilde{X}_k]$. García et al. [29] showed (Theorem 3.3) that this technique does not alter the global characteristics of the initial model. That is to say, the models in Equations (2) and (5) have the same coefficient of determination and experimental statistics for the global significance test.

Figure 1 illustrates the raise regression for two independent variables being geometrically separated by using the residuals weighted by the raising factor λ . Thus, the selection of an adequate value for λ is essential, analogously to what occurs with the ridge factor K . A preliminary proposal about how to select the raising factor in a model with two independent standardized variables can be found in García et al. [41]. Other recently published papers introduce and highlight the various advantages of raise estimators for statistical analysis: Salmerón et al. [30] presented the raise regression for $p = 3$ standardized variables and showed that it presents better properties than the ridge regression and that the individual inference of the raised variable is not altered, García et al. [31] showed that it is guaranteed that all the VIFs associated with the model in Equation (5) diminish but that it is not possible to quantify the decrease, García and Ramírez [42] presented the successive raise regression, and García et al. [31] showed, among other questions, that ridge regression is a particular case of raise regression.

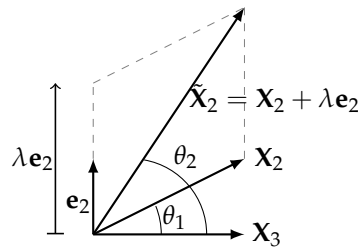


Figure 1. Representation of the raise method.

The following section presents the extension of the VIF to be applied after the estimation by raise regression since it will be interesting whether, after the raising of one independent variable, the VIF falls below 10. It will be also analyzed when a successive raise regression can be recommendable (see García and Ramírez [42]).

3. VIF in Raise Regression

To calculate the VIF in the raise regression, two cases have to be differentiated depending on the dependent variable, X_k , of the auxiliary regression:

1. If it is the raised variable, \tilde{X}_i with $i = 2, \dots, p$, the coefficient of determination, $R_i^2(\lambda)$, of the following auxiliary regression has to be calculated:

$$\begin{aligned} \tilde{X}_i &= \alpha_1(\lambda) + \alpha_2(\lambda)X_2 + \dots + \alpha_{i-1}(\lambda)X_{i-1} + \alpha_{i+1}(\lambda)X_{i+1} + \dots + \alpha_p(\lambda)X_p + \tilde{v} \\ &= \mathbf{X}_{-i}\boldsymbol{\alpha}(\lambda) + \tilde{v}. \end{aligned} \tag{6}$$

2. If it is not the raised variable, X_j with $j = 2, \dots, p$ being $j \neq i$, the coefficient of determination, $R_j^2(\lambda)$, of the following auxiliary regression has to be calculated:

$$\begin{aligned} X_j &= \alpha_1(\lambda) + \alpha_2(\lambda)X_2 + \dots + \alpha_i(\lambda)\tilde{X}_i + \dots + \alpha_{j-1}(\lambda)X_{j-1} + \alpha_{j+1}(\lambda)X_{j+1} \\ &\quad + \dots + \alpha_p(\lambda)X_p + \tilde{v} \\ &= (\mathbf{X}_{-i,-j} \tilde{X}_i) \begin{pmatrix} \alpha_{-i,-j}(\lambda) \\ \alpha_i(\lambda) \end{pmatrix} + \tilde{v}, \end{aligned} \tag{7}$$

where $\mathbf{X}_{-i,-j}$ corresponding to the matrix \mathbf{X} after the elimination of columns i and j (variables X_i and X_j). The same notation is used for $\boldsymbol{\alpha}_{-i,-j}(\lambda)$.

Once these coefficients of determination are obtained (as indicated in the following subsections), the VIF of the raise regression will be given by the following:

$$VIF(k, \lambda) = \frac{1}{1 - R_k^2(\lambda)}, \quad k = 2, \dots, p. \tag{8}$$

3.1. VIF Associated with Raise Variable

In this case, for $i = 2, \dots, p$, the coefficient of determination of the regression in Equation (6) is given by

$$\begin{aligned} R_i^2(\lambda) &= 1 - \frac{(1+2\lambda+\lambda^2)RSS_i^{-i}}{TSS_i^{-i}+(\lambda^2+2\lambda)RSS_i^{-i}} = \frac{ESS_i^{-i}}{TSS_i^{-i}+(\lambda^2+2\lambda)RSS_i^{-i}} \\ &= \frac{R_i^2}{1+(\lambda^2+2\lambda)(1-R_i^2)}, \end{aligned} \tag{9}$$

since:

$$\begin{aligned} TSS_i^{-i}(\lambda) &= \tilde{\mathbf{X}}_i^t \tilde{\mathbf{X}}_i - n \cdot \bar{\tilde{\mathbf{X}}}_i^2 = \mathbf{X}_i^t \mathbf{X}_i + (\lambda^2 + 2\lambda) \mathbf{e}_i^t \mathbf{e}_i - n \cdot \bar{\mathbf{X}}_i^2 \\ &= TSS_i^{-i} + (\lambda^2 + 2\lambda) RSS_i^{-i}, \\ RSS_i^{-i}(\lambda) &= \tilde{\mathbf{X}}_i^t \tilde{\mathbf{X}}_i - \hat{\boldsymbol{\alpha}}(\lambda)^t \mathbf{X}_{-i}^t \tilde{\mathbf{X}}_i = \mathbf{X}_i^t \mathbf{X}_i + (\lambda^2 + 2\lambda) \mathbf{e}_i^t \mathbf{e}_i - \hat{\boldsymbol{\alpha}}^t \mathbf{X}_{-i}^t \mathbf{X}_i \\ &= (\lambda^2 + 2\lambda + 1) RSS_i^{-i}, \end{aligned}$$

where TSS_i^{-i} , ESS_i^{-i} and RSS_i^{-i} are the total sum of squares, explained sum of squares, and residual sum of squares of the model in Equation (4). Note that it has been taken into account that

$$\tilde{\mathbf{X}}_i^t \tilde{\mathbf{X}}_i = (\mathbf{X}_i + \lambda \mathbf{e}_i)^t (\mathbf{X}_i + \lambda \mathbf{e}_i) = \mathbf{X}_i^t \mathbf{X}_i + (\lambda^2 + 2\lambda) \mathbf{e}_i^t \mathbf{e}_i,$$

since $\mathbf{e}_i^t \mathbf{X}_i = \mathbf{e}_i^t \mathbf{e}_i = RSS_i^{-i}$ and

$$\hat{\boldsymbol{\alpha}}(\lambda) = (\mathbf{X}_{-i}^t \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^t \tilde{\mathbf{X}}_i = \hat{\boldsymbol{\alpha}},$$

due to $\mathbf{X}_{-i}^t \tilde{\mathbf{X}}_i = \mathbf{X}_{-i}^t \mathbf{X}_i$.

Indeed, from Equation (9), it is evident that

1. $R_i^2(\lambda)$ decreases as λ increases.
2. $\lim_{\lambda \rightarrow +\infty} R_i^2(\lambda) = 0$.
3. $R_i^2(\lambda)$ is continuous in zero; that is to say, $R_i^2(0) = R_i^2$.

Finally, from properties 1) and 3), it is deduced that $R_i^2(\lambda) \leq R_i^2$ for all λ .

3.2. VIF Associated with Non-Raised Variables

In this case, for $j = 2, \dots, p$, with $j \neq i$, the coefficient of determination of regression in Equation (7) is given by

$$\begin{aligned} R_j^2(\lambda) &= 1 - \frac{RSS_j^{-j}(\lambda)}{TSS_j^{-j}(\lambda)} \\ &= \frac{1}{TSS_j^{-j}} \left(TSS_j^{-j} - RSS_j^{-i,-j} + \frac{RSS_i^{-i,-j} \cdot (RSS_j^{-i,-j} - RSS_j^{-j})}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \right), \end{aligned} \tag{10}$$

Taking into account that $\tilde{\mathbf{X}}_i^t \mathbf{X}_j = (\mathbf{X}_i + \lambda \mathbf{e}_i)^t \mathbf{X}_j = \mathbf{X}_i^t \mathbf{X}_j$ since $\mathbf{e}_i^t \mathbf{X}_j = 0$, it is verified that

$$TSS_j^{-j}(\lambda) = \mathbf{X}_j^t \mathbf{X}_j - n \cdot \bar{\mathbf{X}}_j^2 = TSS_j^{-j},$$

and, from Appendices A and B,

$$\begin{aligned}
 RSS_j^{-j}(\lambda) &= \mathbf{X}_j^t \mathbf{X}_j - \hat{\boldsymbol{\alpha}}(\lambda)^t \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \tilde{\mathbf{X}}_i^t \mathbf{X}_j \end{pmatrix} \\
 &= \mathbf{X}_j^t \mathbf{X}_j - \hat{\boldsymbol{\alpha}}_{-i,-j}(\lambda)^t \mathbf{X}_{-i,-j}^t \mathbf{X}_j - \hat{\boldsymbol{\alpha}}_i(\lambda)^t \mathbf{X}_i^t \mathbf{X}_j \\
 &\stackrel{\text{Appendix A}}{=} \underbrace{\mathbf{X}_j^t \mathbf{X}_j - \mathbf{X}_j^t \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_j}_{\text{Appendix A}} \\
 &\quad - \frac{RSS_i^{-i,-j}}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \cdot \\
 &\quad \cdot \left(RSS_i^{-i,-j} \mathbf{X}_j^t \mathbf{X}_{-i,-j} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j \right. \\
 &\quad \left. + \mathbf{X}_j^t \mathbf{X}_i \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j \right) \\
 &\quad - \frac{RSS_i^{-i,-j}}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \cdot \hat{\boldsymbol{\alpha}}_i^t \mathbf{X}_i^t \mathbf{X}_j \\
 &= \mathbf{X}_j^t \left(\mathbf{I} - \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \right) \mathbf{X}_j \\
 &\quad - \frac{RSS_i^{-i,-j}}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \cdot \\
 &\quad \cdot \left(RSS_i^{-i,-j} \mathbf{X}_j^t \mathbf{X}_{-i,-j} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j \right. \\
 &\quad \left. + \mathbf{X}_j^t \mathbf{X}_i \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + \hat{\boldsymbol{\alpha}}_i^t \mathbf{X}_i^t \mathbf{X}_j \right) \\
 &\stackrel{\text{Appendix B}}{=} \underbrace{RSS_j^{-i,-j}}_{\text{Appendix B}} \\
 &\quad - \frac{RSS_i^{-i,-j}}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \cdot \left(RSS_j^{-i,-j} - RSS_j^{-j} \right),
 \end{aligned}$$

where TSS_j^{-j} and RSS_j^{-j} are the total sum of squares and residual sum of squares of the model in Equation (4) and where $RSS_i^{-i,-j}$ and $RSS_j^{-i,-j}$ are the residual sums of squares of models:

$$\mathbf{X}_i = \mathbf{X}_{-i,-j} \boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{11}$$

$$\mathbf{X}_j = \mathbf{X}_{-i,-j} \boldsymbol{\delta} + \boldsymbol{\nu}. \tag{12}$$

Indeed, from Equation (10), it is evident that

1. $R_j^2(\lambda)$ decreases as λ increases.
2. $\lim_{\lambda \rightarrow +\infty} R_j^2(\lambda) = \frac{TSS_j^{-j} - RSS_j^{-i,-j}}{TSS_j^{-j}}$.
3. $R_j^2(\lambda)$ is continuous in zero. That is to say, $R_j^2(0) = \frac{TSS_j^{-j} - RSS_j^{-j}}{TSS_j^{-j}} = R_j^2$.

Finally, from properties 1) and 3), it is deduced that $R_j^2(\lambda) \leq R_j^2$ for all λ .

3.3. Properties of VIF(k, λ)

From conditions verified by the coefficient of determination in Equations (9) and (10), it is concluded that $VIF(k, \lambda)$ (see expression Equation (8)), verifies that

1. The VIF associated with the raise regression is continuous in zero because the coefficients of determination of the auxiliary regressions in Equations (6) and (7) are also continuous in zero. That is to say, for $\lambda = 0$, it coincides with the VIF obtained for the model in Equation (2) when it is estimated by OLS:

$$VIF(k, 0) = \frac{1}{1 - R_k^2(0)} = \frac{1}{1 - R_k^2} = VIF(k), \quad k = 2, \dots, p.$$

2. The VIF associated with the raise regression decreases as λ increases since this is the behavior of the coefficient of determination of the auxiliary regressions in Equations (6) and (7). Consequently,

$$VIF(k, \lambda) = \frac{1}{1 - R_k^2(\lambda)} \leq \frac{1}{1 - R_k^2} = VIF(k), \quad k = 2, \dots, p, \quad \forall \lambda \geq 0.$$

3. The VIF associated with the raised variable is always higher than one since

$$\lim_{\lambda \rightarrow +\infty} VIF(i, \lambda) = \lim_{\lambda \rightarrow +\infty} \frac{1}{1 - R_i^2(\lambda)} = \frac{1}{1 - 0} = 1, \quad i = 2, \dots, p.$$

4. The VIF associated with the non-raised variables has a horizontal asymptote since

$$\begin{aligned} \lim_{\lambda \rightarrow +\infty} VIF(j, \lambda) &= \lim_{\lambda \rightarrow +\infty} \frac{1}{1 - R_j^2(\lambda)} = \frac{1}{1 - \frac{TSS_j^{-j} - RSS_j^{-i,-j}}{TSS_j^{-j}}} \\ &= \frac{TSS_j^{-j}}{RSS_j^{-i,-j}} = \frac{TSS_j^{-i,-j}}{RSS_j^{-i,-j}} = \frac{1}{1 - R_{ij}^2} = VIF_{-i}(j), \end{aligned}$$

where R_{ij}^2 is the coefficient of determination of the regression in Equation (12) for $j = 2, \dots, p$ and $j \neq i$. Indeed, this asymptote corresponds to the VIF, $VIF_{-i}(j)$, of the regression $\mathbf{Y} = \mathbf{X}_{-i}\boldsymbol{\zeta} + \mathbf{w}$ and, consequently, will also always be equal to or higher than one.

Thus, from properties (1) to (4), $VIF(k, \lambda)$ has the very desirable properties of being continuous, monotone in the raise parameter, and higher than one, as presented in García et al. [26].

In addition, the property (4) can be applied to determine the variable to be raised only considering the one with a lower horizontal asymptote. If the asymptote is lower than 10 (the threshold established traditionally as worrying), the extension could be applied to determine the raising factor by selecting, for example, the first λ that verifies $VIF(k, \lambda) < 10$ for $k = 2, \dots, p$. If none of the $p - 1$ asymptotes is lower than the established threshold, it will not be enough to raise one independent variable and a successive raise regression will be recommended (see García and Ramírez [42] and García et al. [31] for more details). Note that, if it were necessary to raise more than one variable, it is guaranteed that there will be values of the raising parameter that mitigate multicollinearity since, in the extreme case where all the variables of the model are raised, all the VIFs associated with the raised variables tend to one.

3.4. Transformation of Variables

The transformation of data is very common when working with models where strong collinearity exists. For this reason, this section analyzes whether the transformation of the data affects the VIF obtained in the previous section.

Since the expression given by Equation (9) can be expressed with $i = 2, \dots, p$ in the function of R_i^2 :

$$R_i^2(\lambda) = \frac{R_i^2}{1 + (\lambda^2 + 2\lambda) \cdot (1 - R_i^2)},$$

it is concluded that it is invariant to origin and scale changes and, consequently, the VIF calculated from it will also be invariant.

On the other hand, the expression given by Equation (10) can be expressed for $j = 2, \dots, p$, with $j \neq i$ as

$$\begin{aligned} R_j^2(\lambda) &= 1 - \frac{RSS_j^{-i,-j}}{TSS_j^{-j}} + \frac{1}{TSS_j^{-j}} \cdot \frac{RSS_i^{-i,-j} \cdot (RSS_j^{-i,-j} - RSS_j^{-j})}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \\ &= R_{ij}^2 + \frac{RSS_i^{-i,-j}}{RSS_i^{-i,-j} + (\lambda^2 + 2\lambda) \cdot RSS_i^{-i}} \cdot \left(\frac{RSS_j^{-i,-j}}{TSS_j^{-i,-j}} - \frac{RSS_j^{-j}}{TSS_j^{-j}} \right) \\ &= R_{ij}^2 + \frac{R_j^2 - R_{ij}^2}{1 + (\lambda^2 + 2\lambda) \cdot \frac{RSS_i^{-i}}{RSS_i^{-i,-j}}}, \end{aligned} \tag{13}$$

where it was applied that $TSS_j^{-j} = TSS_j^{-i,-j}$.

In this case, by following García et al. [26], transforming the variable X_i as

$$x_i = \frac{X_i - a_i}{b_i}, \quad a_i \in \mathbb{R}, \quad b_i \in \mathbb{R} - \{0\}, \quad i = 2, \dots, p,$$

it is obtained that $RSS_i^{-i}(T) = \frac{1}{b_i^2} RSS_i^{-i}$ and $RSS_i^{-i,-j}(T) = \frac{1}{b_i^2} RSS_i^{-i,-j}$ where $RSS_i^{-i}(T)$ and $RSS_i^{-i,-j}(T)$ are the residual sum of squares of the transformed variables.

Taking into account that X_i is the dependent variables in the regressions of RSS_i^{-i} and $RSS_i^{-i,-j}$, the following is obtained:

$$\frac{RSS_i^{-i}}{RSS_i^{-i,-j}} = \frac{RSS_i^{-i}(T)}{RSS_i^{-i,-j}(T)}.$$

Then, the expression given by Equation (13) is invariant to data transformations (As long as the dependent variables are transformed from the regressions of RSS_i^{-i} and $RSS_i^{-i,-j}$ in the same form. For example, (a) for considering that a_i is its mean and b_i is its standard deviation (typification), (b) for considering that a_i is its mean and b_i is its standard deviation multiplied by the square root of the number of observations (standardization), or (c) for considering that a_i is zero and b_i is the square root of the squares sum of observations (unit length).) and, consequently, the VIF calculated from it will also be invariant.

4. MSE for Raise Regression

Since the estimator β obtained from Equation (5) is biased, it is interesting to study its Mean Square Error (MSE).

Taking into account that, for $k = 2, \dots, p$,

$$\begin{aligned} \tilde{\mathbf{X}}_k &= \mathbf{X}_k + \lambda \mathbf{e}_k \\ &= (1 + \lambda)\mathbf{X}_k - \lambda (\hat{\alpha}_0 + \hat{\alpha}_1\mathbf{X}_1 + \dots + \hat{\alpha}_{k-1}\mathbf{X}_{k-1} + \hat{\alpha}_{k+1}\mathbf{X}_{k+1} + \dots + \hat{\alpha}_p\mathbf{X}_p), \end{aligned}$$

it is obtained that matrix $\tilde{\mathbf{X}}$ of the expression in Equation (5) can be rewritten as $\tilde{\mathbf{X}} = \mathbf{X} \cdot \mathbf{M}_\lambda$, where

$$\mathbf{M}_\lambda = \begin{pmatrix} 1 & 0 & \dots & 0 & -\lambda\hat{\alpha}_0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & -\lambda\hat{\alpha}_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & -\lambda\hat{\alpha}_{k-1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 + \lambda & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & -\lambda\hat{\alpha}_{k+1} & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & -\lambda\hat{\alpha}_p & 0 & \dots & 1 \end{pmatrix}. \tag{14}$$

Thus, we have $\hat{\beta}(\lambda) = (\tilde{\mathbf{X}}^t \cdot \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^t \cdot \mathbf{Y} = \mathbf{M}_\lambda^{-1} \cdot \hat{\beta}$, and then, the estimator of β obtained from Equation (5) is biased unless $\mathbf{M}_\lambda = \mathbf{I}$, which only occurs when $\lambda = 0$, that is to say, when the raise regression coincides with OLS. Moreover,

$$\begin{aligned} tr \left(Var \left(\hat{\beta}(\lambda) \right) \right) &= tr \left(\mathbf{M}_\lambda^{-1} \cdot Var(\hat{\beta}) \cdot (\mathbf{M}_\lambda^{-1})^t \right) = \sigma^2 tr \left((\tilde{\mathbf{X}}^t \tilde{\mathbf{X}})^{-1} \right), \\ (E[\hat{\beta}(\lambda)] - \beta)^t (E[\hat{\beta}(\lambda)] - \beta) &= \beta^t (\mathbf{M}_\lambda^{-1} - \mathbf{I})^t (\mathbf{M}_\lambda^{-1} - \mathbf{I}) \beta, \end{aligned}$$

where tr denotes the trace of a matrix.

In that case, the MSE for raise regression is

$$\begin{aligned} \text{MSE} \left(\hat{\beta}(\lambda) \right) &= tr \left(Var \left(\hat{\beta}(\lambda) \right) \right) + (E[\hat{\beta}(\lambda)] - \beta)^t (E[\hat{\beta}(\lambda)] - \beta) \\ &= \sigma^2 tr \left((\tilde{\mathbf{X}}^t \tilde{\mathbf{X}})^{-1} \right) + \beta^t (\mathbf{M}_\lambda^{-1} - \mathbf{I})^t (\mathbf{M}_\lambda^{-1} - \mathbf{I}) \beta \\ &\stackrel{\text{Appendix C}}{=} \underbrace{\sigma^2 tr \left((\mathbf{X}_{-k}^t \mathbf{X}_{-k})^{-1} \right)}_{\text{Appendix C}} + \left(1 + \sum_{j=0, j \neq k}^p \hat{\alpha}_j^2 \right) \cdot \beta_k^2 \cdot \frac{\lambda^2 + h}{(1 + \lambda)^2}, \end{aligned}$$

where $h = \frac{\sigma^2}{\hat{\beta}_k^2 \cdot RSS_k^{-k}}$.

We can obtain the MSE from the estimated values of σ^2 and β_k from the model in Equation (2).

On the other hand, once the estimations are obtained and taking into account the Appendix C, $\lambda_{min} = \frac{\hat{\sigma}^2}{\hat{\beta}_k^2 \cdot RSS_k^{-k}}$ minimizes $\text{MSE} \left(\hat{\beta}(\lambda) \right)$. Indeed, it is verified that $\text{MSE} \left(\hat{\beta}(\lambda_{min}) \right) < \text{MSE} \left(\hat{\beta}(0) \right)$; that is to say, if the goal is exclusively to minimize the MSE (as in the work presented by Hoerl et al. [33]), λ_{min} should be selected as the raising factor.

Finally, note that, if $\lambda_{min} > 1$, then $\text{MSE} \left(\hat{\beta}(\lambda) \right) < \text{MSE} \left(\hat{\beta}(0) \right)$ for all $\lambda > 0$.

5. Numerical Examples

To illustrate the results of previous sections, two different set of data will be used that collect the two situations shown in the graphs of Figures A1 and A2. The second example also compares results obtained by the raise regression to results obtained by the application of ridge and Lasso regression.

5.1. Example 1: $h < 1$

The data set includes different financial variables for 15 Spanish companies for the year 2016 (consolidated account and results between €800,000 and €9,000,000) obtained from the dabase Sistema de Análisis de Balances Ibéricos (SABI) database. The relationship is studied between the number of employees, E , and the fixed assets (€), FA ; operating income (€), OI ; and sales (€), S . The model is expressed as

$$E = \beta_1 + \beta_2 FA + \beta_3 OI + \beta_4 S + u. \quad (15)$$

Table 1 displays the results of the estimation by OLS of the model in Equation (15). The presence of essential collinearity in the model in Equation (15) is indicated by the determinant close to zero (0.0000919) of the correlation matrix of independent variables

$$R = \begin{pmatrix} 1 & 0.7264656 & 0.7225473 \\ 0.7264656 & 1 & 0.9998871 \\ 0.7225473 & 0.9998871 & 1 \end{pmatrix},$$

and the VIFs (2.45664, 5200.315, and 5138.535) higher than 10. Note that the collinearity is provoked fundamentally by the relationship between OI and S .

In contrast, due to the fact that the coefficients of variation of the independent variables (1.015027, 0.7469496, and 0.7452014) are higher than 0.1002506, the threshold established as troubling by Salmerón et al. [39], it is possible to conclude that the nonessential multicollinearity is not troubling. Thus, the extension of the VIF seems appropriate to check if the application of the raise regression has mitigated the multicollinearity.

Remark 1. $\lambda^{(1)}$ and $\lambda^{(2)}$ will be the raising factor of the first and second raising, respectively.

Table 1. Estimations of the models in Equations (15)–(18): Standard deviation is inside the parenthesis, R^2 is the coefficient of determination, $F_{3,11}$ is the experimental value of the joint significance contrast, and $\hat{\sigma}^2$ is the variance estimate of the random perturbation.

	Model (15)	<i>p</i> -Value	Model (16) for $\lambda_{vif}^{(1)} = 24.5$	<i>p</i> -Value	Model (17) for $\lambda_{min}^{(1)} = 0.42$ and $\lambda_{vif}^{(2)} = 17.5$	<i>p</i> -Value	Model (18) for $\lambda_{mse}^{(1)} = 1.43$ and $\lambda_{vif}^{(2)} = 10$	<i>p</i> -Value
Intercept	994.21 (17940)	0.957	4588.68 (17,773.22)	0.801	5257.84 (1744.26)	0.772	5582.29 (17740.18)	0.759
FA	−1.28 (0.55)	0.039	−1.59 (0.50)	0.009	−1.59 (0.51)	0.009	−1.58 (0.51)	0.009
OI	−81.79 (52.86)	0.150						
$\widetilde{OI}_{\lambda_{vif}^{(1)}}$			−3.21 (2.07)	0.150				
$\widetilde{OI}_{\lambda_{min}^{(1)}}$					1.67 (2.28)	0.478		
$\widetilde{OI}_{\lambda_{mse}^{(1)}}$							1.51 (2.24)	0.517
S	87.58 (53.29)	0.129	8.38 (2.35)	0.004				
$\widetilde{S}_{\lambda_{vif}^{(2)}}$					3.42 (2.03)	0.120	3.55 (1.99)	0.103
R^2	0.70		0.70		0.70		0.70	
$F_{3,11}$	8.50		8.50		8.50		8.50	
$\hat{\sigma}^2$	1,617,171,931		1,617,171,931		1,617,171,931		1,617,171,931	
MSE	321,730,738		321,790,581		336,915,567		325,478,516	

5.1.1. First Raising

A possible solution could be to apply the raise regression to try to mitigate the collinearity. To decide which variable is raised, the thresholds for the VIFs associated with the raise regression are calculated with the goal of raising the variable that the smaller horizontal asymptotes present. In addition to raising the variable that presents the lowest VIF, it would be interesting to obtain a lower mean squared error (MSE) after raising. For this, the $\lambda_{min}^{(1)}$ is calculated for each case. Results are shown in Table 2. Note that the variable to be raised should be the second or third since their asymptotes are lower than 10, although in both cases $\lambda_{min}^{(1)}$ is lower than 1 and it is not guaranteed that the MSE of the raise regression will be less than the one obtained from the estimation by the OLS of the model in Equation (15). For this reason, this table also shows the values of $\lambda^{(1)}$ that make the MSE of the raise regression coincide with the MSE of the OLS regression, $\lambda_{mse}^{(1)}$, and the minimum value of $\lambda^{(1)}$ that leads to values of VIF less than 10, $\lambda_{vif}^{(1)}$.

Table 2. Horizontal asymptotes for variance inflation factors (VIF) after raising each variable and $\lambda_{min}^{(1)}$, $\lambda_{mse}^{(1)}$, and $\lambda_{vif}^{(1)}$.

Raised	$\lim_{\lambda^{(1)} \rightarrow +\infty} VIF(FA, \lambda^{(1)})$	$\lim_{\lambda^{(1)} \rightarrow +\infty} VIF(OI, \lambda^{(1)})$	$\lim_{\lambda^{(1)} \rightarrow +\infty} VIF(S, \lambda^{(1)})$
Variable 1	1	4429.22	4429.22
Variable 2	2.09	1	2.09
Variable 3	2.12	2.12	1
Raised	$\lambda_{min}^{(1)}$	$\lambda_{mse}^{(1)}$	$\lambda_{vif}^{(1)}$
Variable 1	0.18	0.45	#
Variable 2	0.42	1.43	24.5
Variable 3	0.37	1.18	24.7

Figure 2 displays the VIF associated with the raise regression for $0 \leq \lambda^{(1)} \leq 900$ after raising the second variable. It is observed that VIFs are always higher than its corresponding horizontal asymptotes.

The model after raising the second variable will be given by

$$E = \beta_1(\lambda) + \beta_2(\lambda)FA + \beta_3(\lambda)\widetilde{OI} + \beta_4(\lambda)S + \tilde{u}, \tag{16}$$

where $\widetilde{OI} = OI + \lambda^{(1)} \cdot e_{OI}$ with e_{OI} the residual of regression:

$$OI = \alpha_1 + \alpha_2FA + \alpha_3S + v.$$

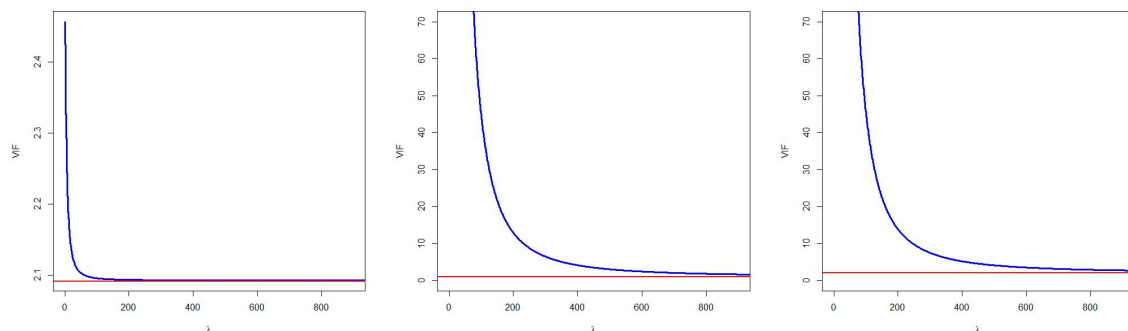


Figure 2. VIF of the variables after raising OI.

Remark 2. The coefficient of variation of $\widetilde{\mathbf{OI}}$ for $\lambda^{(1)} = 24.5$ is equal to 0.7922063; that is to say, it was lightly increased.

As can be observed from Table 3, in Equation (16), the collinearity is not mitigated by considering $\lambda^{(1)}$ equal to $\lambda_{min}^{(1)}$ and $\lambda_{mse}^{(1)}$. For this reason, Table 1 only shows the values of the model in Equation (16) for the value of $\lambda^{(1)}$ that leads to VIF lower than 10.

Table 3. VIF of regression Equation (16) for $\lambda^{(1)}$ equal to $\lambda_{min}^{(1)}$, $\lambda_{mse}^{(1)}$, and $\lambda_{vif}^{(1)}$.

	$VIF(\mathbf{FA}, \lambda^{(1)})$	$VIF(\widetilde{\mathbf{OI}}, \lambda^{(1)})$	$VIF(\mathbf{S}, \lambda^{(1)})$
$\lambda_{min}^{(1)}$	2.27	2587.84	2557.66
$\lambda_{mse}^{(1)}$	2.15	878.10	868.58
$\lambda_{vif}^{(1)}$	2.09	9.00	9.99

5.1.2. Transformation of Variables

After the first raising, it is interesting to verify that the VIF associated with the raise regression is invariant to data transformation. With this goal, the second variable has been raised, obtaining the $VIF(\mathbf{FA}, \lambda^{(1)})$, $VIF(\widetilde{\mathbf{OI}}, \lambda^{(1)})$, and $VIF(\mathbf{S}, \lambda^{(1)})$ for $\lambda^{(1)} \in \{0, 0.5, 1, 1.5, 2, \dots, 9.5, 10\}$, supposing original, unit length, and standardized data. Next, the three possible differences and the average of the VIF associated with each variable are obtained. Table 4 displays the results from which it is possible to conclude that differences are almost null and that, consequently, the VIF associated with the raise regression is invariant to the most common data transformation.

Table 4. Effect of data transformations on VIF associated with raise regression.

	$VIF(\mathbf{FA}, \lambda^{(1)})$	$VIF(\widetilde{\mathbf{OE}}, \lambda^{(1)})$	$VIF(\mathbf{S}, \lambda^{(1)})$
Original–Unit length	$9.83 \cdot 10^{-16}$	$1.55 \cdot 10^{-11}$	$1.83 \cdot 10^{-10}$
Original–Standardized	$-1.80 \cdot 10^{-16}$	$-3.10 \cdot 10^{-10}$	$2.98 \cdot 10^{-10}$
Unit length–Standardized	$-1.16 \cdot 10^{-15}$	$-3.26 \cdot 10^{-10}$	$1.15 \cdot 10^{-10}$

5.1.3. Second Raising

After the first raising, we can use the results obtained from the value of λ that obtains all VIFs less than 10 or consider the results obtained for λ_{min} or λ_{mse} and continue the procedure with a second raising. By following the second option, we part from the value of $\lambda^{(1)} = \lambda_{min}^{(1)} = 0.42$ obtained after the first raising. From Table 5, the third variable is selected to be raised. Table 6 shows the VIF associated with the following model for $\lambda_{min}^{(2)}$, $\lambda_{mse}^{(2)}$, and $\lambda_{vif}^{(2)}$:

$$E = \beta_1(\lambda) + \beta_2(\lambda)\mathbf{FA} + \beta_3(\lambda)\widetilde{\mathbf{OI}} + \beta_4(\lambda)\widetilde{\mathbf{S}} + \widetilde{\mathbf{u}}, \tag{17}$$

where $\widetilde{\mathbf{S}} = \mathbf{S} + \lambda^{(2)} \cdot \mathbf{e}_S$ with \mathbf{e}_S the residuals or regression:

$$\mathbf{S} = \alpha_1(\lambda) + \alpha_2(\lambda)\mathbf{FA} + \alpha_3(\lambda)\widetilde{\mathbf{OI}} + \widetilde{\mathbf{v}}.$$

Remark 3. The coefficient of variation of $\widetilde{\mathbf{OI}}$ for $\lambda^{(1)} = 0.42$ is equal to 0.7470222, and the coefficient of variation of $\widetilde{\mathbf{S}}$ for $\lambda^{(2)} = 17.5$ is equal to 0.7473472. In both cases, they were slightly increased.

Note that it is only possible to state that collinearity has been mitigated when $\lambda^{(2)} = \lambda_{vif}^{(2)} = 17.5$. Results of this estimation are displayed in Table 1.

Table 5. Horizontal asymptote for VIFs after raising each variable in the second raising for $\lambda_{min}^{(2)}$, $\lambda_{mse}^{(2)}$ and $\lambda_{vif}^{(2)}$.

Raised	$\lim_{\lambda^{(2)} \rightarrow +\infty} VIF(\mathbf{FA}, \lambda^{(2)})$	$\lim_{\lambda^{(2)} \rightarrow +\infty} VIF(\widetilde{\mathbf{OI}}, \lambda^{(2)})$	$\lim_{\lambda^{(2)} \rightarrow +\infty} VIF(\mathbf{S}, \lambda^{(2)})$
Variable 1	1	2381.56	2381.56
Variable 3	2.12	2.12	1
Raised	$\lambda_{min}^{(2)}$	$\lambda_{mse}^{(2)}$	$\lambda_{vif}^{(2)}$
Variable 1	0.15	0.34	\nexists
Variable 3	0.35	1.09	17.5

Table 6. VIFs of regression Equation (16) for $\lambda^{(2)}$ equal to $\lambda_{min}^{(2)}$, $\lambda_{mse}^{(2)}$, and $\lambda_{vif}^{(2)}$.

	$VIF(\mathbf{FA}, \lambda^{(2)})$	$VIF(\widetilde{\mathbf{OI}}, \lambda^{(2)})$	$VIF(\widetilde{\mathbf{S}}, \lambda^{(2)})$
$\lambda_{min}^{(2)}$	2.20	1415.06	1398.05
$\lambda_{mse}^{(2)}$	2.15	593.98	586.20
$\lambda_{vif}^{(2)}$	2.12	9.67	8.47

Considering that, after the first raising, it is obtained that $\lambda^{(1)} = \lambda_{mse}^{(1)} = 1.43$, from Table 7, the third variable is selected to be raised. Table 8 shows the VIF associated with the following model for $\lambda_{min}^{(2)}$, $\lambda_{mse}^{(2)}$, and $\lambda_{vif}^{(2)}$:

$$E = \beta_1(\lambda) + \beta_2(\lambda)\mathbf{FA} + \beta_3(\lambda)\widetilde{\mathbf{OI}} + \beta_4(\lambda)\widetilde{\mathbf{S}} + \tilde{\mathbf{u}}, \tag{18}$$

where $\widetilde{\mathbf{S}} = \mathbf{S} + \lambda \cdot \mathbf{e}_3$.

Remark 4. The coefficient of variation of $\widetilde{\mathbf{OI}}$ for $\lambda^{(1)} = 1.43$ is equal to 0.7473033, and the coefficient of variation of $\widetilde{\mathbf{S}}$ for $\lambda^{(2)} = 10$ is equal to 0.7651473. In both cases, they were lightly increased.

Remark 5. Observing the coefficients of variation of $\widetilde{\mathbf{OI}}$ for different raising factor. it is concluded that the coefficient of variation increases as the raising factor increases: 0.7470222 ($\lambda = 0.42$), 0.7473033 ($\lambda = 1.43$), and 0.7922063 ($\lambda = 24.5$).

Note that it is only possible to state that collinearity has been mitigated when $\lambda^{(2)} = \lambda_{vif}^{(2)} = 10$. Results of the estimations of this model are shown in Table 1.

Table 7. Horizontal asymptote for VIFs after raising each variables in the second raising for $\lambda_{min}^{(2)}$, $\lambda_{mse}^{(2)}$ and $\lambda_{vif}^{(2)}$.

Raised	$\lim_{\lambda^{(2)} \rightarrow +\infty} VIF(\mathbf{FA}, \lambda^{(2)})$	$\lim_{\lambda^{(2)} \rightarrow +\infty} VIF(\widetilde{\mathbf{OI}}, \lambda^{(2)})$	$\lim_{\lambda^{(2)} \rightarrow +\infty} VIF(\mathbf{S}, \lambda^{(2)})$
Variable 1	1	853.40	853.40
Variable 3	2.12	2.12	1
Raised	$\lambda_{min}^{(2)}$	$\lambda_{mse}^{(2)}$	$\lambda_{vif}^{(2)}$
Variable 1	0.12	0.27	\nexists
Variable 3	0.32	0.92	10

Table 8. VIFs of regression Equation (16) for $\lambda^{(2)}$ equal to $\lambda_{min}^{(2)}$, $\lambda_{mse}^{(2)}$, and $\lambda_{vif}^{(2)}$.

	$VIF(\mathbf{FA}, \lambda^{(2)})$	$VIF(\widetilde{\mathbf{OI}}, \lambda^{(2)})$	$VIF(\widetilde{\mathbf{S}}, \lambda^{(2)})$
$\lambda_{min}^{(2)}$	2.14	508.54	502.58
$\lambda_{mse}^{(2)}$	2.13	239.42	236.03
$\lambda_{vif}^{(2)}$	2.12	9.36	8.17

5.1.4. Interpretation of Results

Analyzing the results of Table 1, it is possible to conclude that

1. In the model in Equation (16) (in which the second variable is raised considering the smallest λ that makes all the VIFs less than 10, $\lambda^{(1)} = 24.5$), the variable sales have a coefficient significantly different from zero, where in the original model this was not the case. In this case, the MSE is superior to the one obtained by OLS.
2. In the model in Equation (17) (in which the second variable is raised considering the value of λ that minimizes the MSE, $\lambda^{(1)} = 0.42$, and after that, the third variable is raised considering the smallest λ that makes all the VIFs less than 10, $\lambda^{(2)} = 17.5$), there is no difference in the individual significance of the coefficient.
3. In the model in Equation (18) (in which the second variable is raised considering the value of λ that makes the MSE of the raise regression coincide with that of OLS, $\lambda^{(1)} = 1.43$, and next, the third variable is raised considering the smallest λ that makes all the VIFs less than 10, $\lambda^{(2)} = 10$), there is no difference in the individual significance of the coefficient.
4. Although the coefficient of variable **OI** is not significantly different from zero in any case, the not expected negative sign obtained in model in Equation (15) is corrected in models Equations (17) and (18).
5. In the models with one or two raisings, all the global characteristics coincide with that of the model in Equation (15). Furthermore, there is a relevant decrease in the estimation of the standard deviation for the second and third variable.
6. In models with one or two raisings, the MSE increases, with the model in Equation (16) being the one that presents the smallest MSE among the biased models.

Thus, in conclusion, the model in Equation (16) is selected as it presents the smallest MSE and there is an improvement in the individual significance of the variables.

5.2. Example 2: $h > 1$

This example uses the following model previously applied by Klein and Goldberger [43] about consumption and salaries in the United States from 1936 to 1952 (1942 to 1944 were war years, and data are not available):

$$\mathbf{C} = \beta_1 + \beta_2\mathbf{WI} + \beta_3\mathbf{NWI} + \beta_4\mathbf{FI} + \mathbf{u}, \tag{19}$$

where **C** is consumption, **WI** is wage income, **NWI** is non-wage, non-farm income, and **FI** is the farm income. Its estimation by OLS is shown in Table 9.

However, this estimation is questionable since no estimated coefficient is significantly different to zero while the model is globally significant (with 5% significance level), and the VIFs associated with each variable (12.296, 9.23, and 2.97) indicate the presence of severe essential collinearity. In addition, the determinant of the matrix of correlation

$$\mathbf{R} = \begin{pmatrix} 1 & 0.9431118 & 0.8106989 \\ 0.9431118 & 1 & 0.7371272 \\ 0.8106989 & 0.7371272 & 1 \end{pmatrix},$$

is equal to 0.03713592 and, consequently, lower than the threshold recommended by García et al. [44] ($1.013 \cdot 0.1 + 0.00008626 \cdot n - 0.01384 \cdot p = 0.04714764$ being $n = 14$ and $p = 4$); it is maintained the conclusion that the near multicollinearity existing in this model is troubling.

Once again, the values of the coefficients of variation (0.2761369, 0.2597991, and 0.2976122) indicate that the nonessential multicollinearity is not troubling (see Salmerón et al. [39]). Thus, the extension of the VIF seems appropriate to check if the application of the raise regression has mitigated the near multicollinearity.

Next, it is presented the estimation of the model by raise regression and the results are compared to the estimation by ridge and Lasso regression.

5.2.1. Raise Regression

When calculating the thresholds that would be obtained for VIFs by raising each variable (see Table 10), it is observed that, in all cases, they are less than 10. However, when calculating λ_{min} in each case, a value higher than one is only obtained when raising the third variable. Figure 3 displays the MSE for $\lambda \in [0, 37]$. Note that $MSE(\hat{\beta}(\lambda))$ is always less than the one obtained by OLS, 49.434, and presents an asymptote in $\lim_{\lambda \rightarrow +\infty} MSE(\hat{\beta}(\lambda)) = 45.69422$.

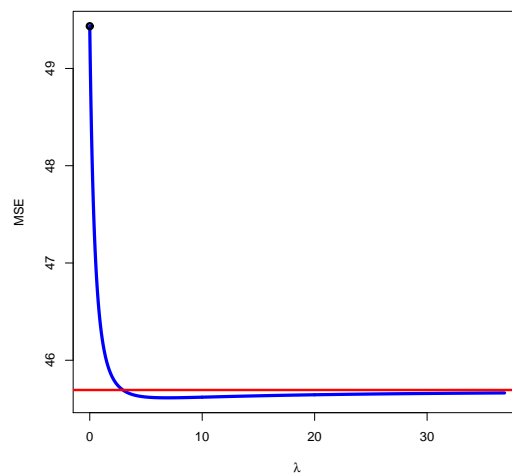


Figure 3. Mean square error (MSE) for the model in Equation (19) after raising third variable.

Table 9. Estimation of the original and raised models: Standard deviation is inside the parentheses, R^2 is the coefficient of determination, $F_{3,10}$ is the experimental value of the joint significance contrast, and $\hat{\sigma}^2$ is the variance estimate of the random perturbation.

	Model (19)	p-Value	Model (20) for $\lambda_{min} = 6.895$	p-Value	Model (21) for $\lambda_{min} = 0.673$	p-Value
Intercept	18.7021 (6.8454)	0.021	19.21507 (6.67216)	0.016	18.2948 (6.8129)	0.023
WI	0.3803 (0.3121)	0.251	0.43365 (0.26849)	0.137		
\tilde{WI}					0.2273 (0.1866)	0.251
NWI	1.4186 (0.7204)	0.077	1.38479 (0.71329)	0.081	1.7269 (0.5143)	0.007
FI	0.5331 (1.3998)	0.711			0.8858 (1.2754)	0.503
\tilde{FI}			0.06752 (0.17730)	0.711		
R^2	0.9187		0.9187		0.9187	
$\hat{\sigma}$	6.06		6.06		6.06	
$F_{3,10}$	37.68		37.68		37.68	
MSE	49.43469		45.61387		48.7497	

Table 10. Horizontal asymptote for VIFs after raising each variable and λ_{min} .

Raised	$\lim_{\lambda \rightarrow +\infty} VIF(WI, \lambda)$	$\lim_{\lambda \rightarrow +\infty} VIF(NWI, \lambda)$	$\lim_{\lambda \rightarrow +\infty} VIF(FI, \lambda)$	λ_{min}
Variable 1	1	2.19	2.19	0.673
Variable 2	2.92	1	2.92	0.257
Variable 3	9.05	9.05	1	6.895

The following model is obtained by raising the third variable:

$$C = \beta_1(\lambda) + \beta_2(\lambda)WI + \beta_3(\lambda)NWI + \beta_4(\lambda)\tilde{FI} + \tilde{u}, \tag{20}$$

where $\tilde{FI} = FI + \lambda \cdot e_{FI}$ being e_{FI} the residuals of regression:

$$FI = \alpha_1 + \alpha_2WI + \alpha_3NWI + v.$$

Remark 6. The coefficient of variation \tilde{FI} for $\lambda^{(1)} = 6.895$ is 1.383309. Thus, the application of the raise regression has mitigated the nonessential multicollinearity in this variable.

Table 9 shows the results for the model in Equation (20), being $\lambda = 6.895$. In this case, the MSE is the lowest possible for every possible value of λ and lower than the one obtained by OLS for the model in Equation (19). Furthermore, in this case, the collinearity is not strong once all the VIF are lower than 10 (9.098, 9.049, and 1.031, respectively). However, the individual significance in the variable was not improved.

With the purpose of improving this situation, another variable is raised. If the first variable is selected to be raised, the following model is obtained:

$$C = \beta_1(\lambda) + \beta_2(\lambda)\tilde{WI} + \beta_3(\lambda)NWI + \beta_4(\lambda)FI + \tilde{u}, \tag{21}$$

where $\tilde{WI} = WI + \lambda \cdot e_{WI}$ being e_{WI} the residuals of regression:

$$WI = \alpha_1 + \alpha_2NWI + \alpha_3FI + v.$$

Remark 7. The coefficient of variation of \tilde{WI} for $\lambda^{(1)} = 0.673$ is 0.2956465. Thus, it is noted that the raise regression has lightly mitigated the nonessential mutlicollinearity of this variable.

Table 9 shows the results for the model in Equation (21), being $\lambda = 0.673$. In this case, the MSE is lower than the one obtained by OLS for the model in Equation (19). Furthermore, in this case, the collinearity is not strong once all the VIF are lower than 10 (5.036024, 4.705204, and 2.470980, respectively). Note that raising this variable, the values of VIFs are lower than raising the first variable but the MSE is higher. However, this model is selected as preferable due to the individual significance being better in this model and the MSE being lower than the one obtained by OLS.

5.2.2. Ridge Regression

This subsection presents the estimation of the model in Equation (19) by ridge regression (see Hoerl and Kennard [4] or Marquardt [45]). The first step is the selection of the appropriate value of K .

The following suggestions are addressed:

- Hoerl et al. [33] proposed the value of $K_{HKB} = p \cdot \frac{\hat{\sigma}^2}{\hat{\beta}^T \hat{\beta}}$ since probability higher than 50% leads to a MSE lower than the one from OLS.
- García et al. [26] proposed the value of K , denoted as K_{VIF} , that leads to values of VIF lower than 10 (threshold traditionally established as troubling).

- García et al. [44] proposed the following values:

$$\begin{aligned}
 K_{exp} &= 0.006639 \cdot e^{1-\det(\mathbf{R})} - 0.00001241 \cdot n + 0.005745 \cdot p, \\
 K_{linear} &= 0.01837 \cdot (1 - \det(\mathbf{R})) - 0.00001262 \cdot n + 0.005678 \cdot p, \\
 K_{sq} &= 0.7922 \cdot (1 - \det(\mathbf{R}))^2 - 0.6901 \cdot (1 - \det(\mathbf{R})) - 0.000007567 \cdot n \\
 &\quad - 0.01081 \cdot p,
 \end{aligned}$$

where $\det(\mathbf{R})$ denotes the determinant of the matrix of correlation, \mathbf{R} .

The following values are obtained $K_{HKB} = 0.417083$, $K_{VIF} = 0.013$, $K_{exp} = 0.04020704$, $K_{linear} = 0.04022313$, and $K_{sq} = 0.02663591$.

Tables 11 and 12 show (The results for K_{linear} are not considered as they are very similar to results obtained by K_{exp} .) the estimations obtained from ridge estimators (expression (1)) and the individual significance intervals obtained by bootstrap considering percentiles 5 and 95 for 5000 repeats. It is also calculated the goodness of the fit by following the results shown by Rodríguez et al. [28] and the MSE.

Note that only the constant term can be considered significantly different to zero and that, curiously, the value of K proposed by Hoerl et al. [33] leads to a value of MSE higher than the one from OLS while the values proposed by García et al. [26] and García et al. [44] lead to a value of MSE lower than the one obtained by OLS. All cases lead to values of VIF lower than 10; see García et al. [26] for its calculation:

$$\begin{aligned}
 &2.0529, 1.8933 \text{ and } 1.5678 \quad \text{for } K_{HKB}, \\
 &9.8856, 7.5541 \text{ and } 2.7991 \quad \text{for } K_{VIF}, \\
 &7.1255, 5.6191 \text{ and } 2.5473 \quad \text{for } K_{exp}, \\
 &8.2528, 6.4123 \text{ and } 2.65903 \quad \text{for } K_{sq}.
 \end{aligned}$$

In any case, the lack of individual significance justifies the selection of the raise regression as preferable in comparison to the models obtained by ridge regression.

Table 11. Estimation of the ridge models for $K_{HKB} = 0.417083$ and $K_{VIF} = 0.013$. Confidence interval, at 10% confidence, is obtained from bootstrap inside the parentheses, and R^2 is the coefficient of determination obtained from Rodríguez et al. [28].

	Model (19) for $K_{HKB} = 0.417083$	Model (19) for $K_{VIF} = 0.013$
Intercept	12.2395 (6.5394, 15.9444)	18.3981 (12.1725, 24.1816)
WI	0.3495 (−0.4376, 1.2481)	0.3787 (−0.4593, 1.216)
NWI	1.6474 (−0.1453, 3.4272)	1.4295 (−0.2405, 3.2544)
FI	0.8133 (−1.5584, 3.028)	0.5467 (−1.827, 2.9238)
R^2	0.8957	0.9353
MSE	64.20028	47.99713

Table 12. Estimation of the ridge models for $K_{exp} = 0.04020704$ and $K_{sq} = 0.02663591$. Confidence interval, at 10% confidence, is obtained from bootstrap inside the parentheses, and R^2 is the coefficient of determination obtained from Rodríguez et al. [28].

	Model (19) for $K_{exp} = 0.04020704$	Model (19) for $K_{sq} = 0.02663591$
Intercept	17.7932 (11.4986, 22.9815)	18.0898 (11.8745, 23.8594)
WI	0.3756 (−0.4752, 1.2254)	0.3771 (−0.4653, 1.2401)
NWI	1.4512 (−0.2249, 3.288)	1.4406 (−0.2551, 3.2519)
FI	0.5737 (−1.798, 2.9337)	0.5605 (−1.6999, 2.9505)
R^2	0.918034	0.9183955
MSE	45.76226	46.75402

5.2.3. Lasso Regression

The Lasso regression (see Tibshirani [5]) is a method initially designed to select variables constraining the coefficient to zero, being specially useful in models with a high number of independent variables. However, this estimation methodology has been widely applied in situation where the model presents worrying near multicollinearity.

Table 13 shows results obtained by the application of the Lasso regression to the model in Equation (19) by using the package *glmnet* of the programming environment R Core Team [46]. Note that these estimations are obtained for the optimal value of $\lambda = 0.1258925$ obtained after a k-fold cross-validation.

Table 13. Estimation of the Lasso model for $\lambda = 0.1258925$: Confidence interval at 10% confidence (obtained from bootstrap inside the parentheses).

Model (19) for $\lambda = 0.1258925$	
Intercept	19.1444 (13.5814489, 24.586207)
WI	0.4198 (−0.2013491, 1.052905)
NWI	1.3253 (0.0000000, 2.752345)
FI	0.4675 (−1.1574169, 2.151648)

The inference obtained by bootstrap methodology (with 5000 repeats) allows us to conclude that in, at least, the 5% of the cases, the coefficient of **NWI** is constrained to zero. Thus, this variable should be eliminated from the model.

However, we consider that this situation should be avoided, and as an alternative to the elimination of variable, that is, as an alternative from the following model, the estimation by raise or ridge regression is proposed.

$$C = \pi_1 + \pi_2 WI + \pi_3 FI + \epsilon, \tag{22}$$

It could be also appropriate to apply the residualization method (see, for example, York [47], Salmerón et al. [48], and García et al. [44]), which consists in the estimation of the following model:

$$C = \tau_1 + \tau_2 WI + \tau_3 FI + \tau_4 \text{res}_{NWI} + \epsilon, \tag{23}$$

where, for example, res_{NWI} represents the residuals of the regression of **NWI** as a function of **WI** that will be interpreted as the part of **NWI** not related to **WI**. In this case (see García et al. [44]), it is verified that $\hat{\pi}_i = \hat{\tau}_i$ for $i = 1, 2, 3$. That is to say, the model in Equation (23) estimates the same relationship between **WI** and **FI** with **C** as in the model in Equation (22) with the benefit that the variable **NWI** is not eliminated due to a part of it being considered..

6. Conclusions

The Variance Inflation Factor (VIF) is one of the most applied measures to diagnose collinearity together with the Condition Number (CN). Once the collinearity is detected, different methodologies can be applied as, for example, the raise regression, but it will be required to check if the methodology has mitigated the collinearity effectively. This paper extends the concept of VIF to be applied after the raise regression and presents an expression of the VIF that verifies the following desirable properties (see García et al. [26]):

1. continuous in zero. That is to say, when the raising factor (λ) is zero, the VIF obtained in the raise regression coincides with the one obtained by OLS;
2. decreasing as a function of the raising factor (λ). That is to say, the degree of collinearity diminishes as λ increases, and
3. always equal or higher than 1.

The paper also shows that the VIF in the raise regression is scale invariant, which is a very common transformation when working with models with collinearity. Thus, it yields identical results regardless of whether predictions are based on unstandardized or standardized predictors. Contrarily, the VIFs obtained from other penalized regressions (ridge regression, Lasso, and Elastic Net) are not scale invariant and hence yield different results depending on the predictor scaling used.

Another contribution of this paper is the analysis of the asymptotic behavior of the VIF associated with the raised variable (verifying that its limit is equal to 1) and associated with the rest of the variables (presenting an horizontal asymptote). This analysis allows to conclude that

- It is possible to know a priori how far each of the VIFs can decrease simply by calculating their horizontal asymptote. This could be used as a criterion to select the variable to be raised, the one with the lowest horizontal asymptote being chosen.
- If there is asymptote under the threshold established as worrying, the extension of the VIF can be applied to select the raising factor considering the value of λ that verifies $VIF(k, \lambda) < 10$ for $k = 2, \dots, p$.
- It is possible that the collinearity is not mitigated with any value of λ . This can happen when at least one horizontal asymptote is greater than the threshold. In that case, a second variable has to be raised. García and Ramírez [42] and García et al. [31] show the successive raising procedure.

On the other hand, since the raise estimator is biased, the paper analyzes its Mean Square Error (MSE), showing that there is a value of λ that minimizes the possibility of the MSE being lower than the one obtained by OLS. However, it is not guaranteed that the VIF for this value of λ presents a value less than the established thresholds. The results are illustrated with two numerical examples, and in the second one, the results obtained by OLS are compared to the results obtained with the raise, ridge, and Lasso regressions that are widely applied to estimated models with worrying multicollinearity. It is showed that the raise regression can compete and even overcome these methodologies.

Finally, we propose as future lines of research the following questions:

- The examples showed that the coefficients of variation increase after raising the variables. This fact is associated with an increase in the variability of the variable and, consequently, with a decrease of the near nonessential multicollinearity. Although a deeper analysis is required, it seems that raise regression mitigates this kind of near multicollinearity.
- The value of the ridge factor traditionally applied, K_{HKB} , leads to estimators with smaller MSEs than the OLS estimators with probability greater than 0.5. In contrast, the value of the raising factor λ_{min} always leads to estimators with smaller MSEs than OLS estimators. Thus, it is deduced that the ridge regression provides estimators with MSEs higher than the MSEs of OLS estimators with probability lower than 0.5. These questions seem to indicate that, in terms of MSE, the raise regression can present better behaviour than the ridge regression. However, the confirmation of this judgment will require a more complete analysis, including other aspects such as interpretability and inference.

Author Contributions: conceptualization, J.G.P., C.G.G. and R.S.G. and A.R.S.; methodology, R.S.G. and A.R.S.; software, A.R.S.; validation, J.G.P., R.S.G. and C.G.G.; formal analysis, R.S.G. and C.G.G.; investigation, R.S.G. and A.R.S.; writing—original draft preparation, A.R.S. and C.G.G.; writing—review and editing, C.G.G.; supervision, J.G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank the anonymous referees for their useful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Given the linear model in Equation (7), it is obtained that

$$\begin{aligned} \hat{\alpha}(\lambda) &= \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} & \mathbf{X}_{-i,-j}^t \tilde{\mathbf{X}}_i \\ \tilde{\mathbf{X}}_i^t \mathbf{X}_{-i,-j} & \tilde{\mathbf{X}}_i^t \tilde{\mathbf{X}}_i \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \tilde{\mathbf{X}}_i^t \mathbf{X}_j \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} & \mathbf{X}_{-i,-j}^t \mathbf{X}_i \\ \mathbf{X}_i^t \mathbf{X}_{-i,-j} & \mathbf{X}_i^t \mathbf{X}_i + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} \\ &= \begin{pmatrix} A(\lambda) & B(\lambda) \\ B(\lambda)^t & C(\lambda) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} \\ &= \begin{pmatrix} A(\lambda) \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + B(\lambda) \cdot \mathbf{X}_i^t \mathbf{X}_j \\ B(\lambda)^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + C(\lambda) \cdot \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_{-i,-j}(\lambda) \\ \hat{\alpha}_i(\lambda) \end{pmatrix}, \end{aligned}$$

Since it is verified that $\mathbf{e}_i^t \mathbf{X}_{-i,-j} = \mathbf{0}$, then $\tilde{\mathbf{X}}_i^t \mathbf{X}_{-i,-j} = (\mathbf{X}_i + \lambda \mathbf{e}_i)^t \mathbf{X}_{-i,-j} = \mathbf{X}_i^t \mathbf{X}_{-i,-j}$, where

$$\begin{aligned} C(\lambda) &= \left(\mathbf{X}_i^t \mathbf{X}_i + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i} - \mathbf{X}_i^t \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_i \right)^{-1} \\ &= \left(\mathbf{X}_i^t \left(\mathbf{I} - \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \right) \mathbf{X}_i + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i} \right)^{-1} \\ &= \left(\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i} \right)^{-1}, \\ B(\lambda) &= - \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_i \cdot C(\lambda) = \frac{\text{RSS}_i^{-i,-j}}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot B, \\ A(\lambda) &= \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} + \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_i \cdot C(\lambda) \cdot \mathbf{X}_i^t \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \\ &= \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} + \frac{(\text{RSS}_i^{-i,-j})^2}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot B \cdot B^t. \end{aligned}$$

Then,

$$\begin{aligned} \hat{\alpha}_{-i,-j}(\lambda) &= \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_j + \frac{(\text{RSS}_i^{-i,-j})^2}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ &\quad + \frac{\text{RSS}_i^{-i,-j}}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot B \cdot \mathbf{X}_i^t \mathbf{X}_j \\ &= \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ &\quad + \frac{\text{RSS}_i^{-i,-j} \left(\text{RSS}_i^{-i,-j} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + B \cdot \mathbf{X}_i^t \mathbf{X}_j \right)}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}}, \\ \hat{\alpha}_i(\lambda) &= \frac{\text{RSS}_i^{-i,-j}}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ &\quad + \frac{1}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot \mathbf{X}_i^t \mathbf{X}_j \\ &= \frac{\text{RSS}_i^{-i,-j}}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot \left(B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + (\text{RSS}_i^{-i,-j})^{-1} \mathbf{X}_i^t \mathbf{X}_j \right) \\ &= \frac{\text{RSS}_i^{-i,-j}}{\text{RSS}_i^{-i,-j} + (\lambda^2 + 2\lambda) \text{RSS}_i^{-i}} \cdot \hat{\alpha}_i. \end{aligned}$$

Appendix B

Given the linear model

$$\mathbf{X}_j = \mathbf{X}_{-j} \boldsymbol{\alpha} + \mathbf{v} = (\mathbf{X}_{-i,-j} \ \mathbf{X}_i) \begin{pmatrix} \boldsymbol{\alpha}_{-i,-j} \\ \alpha_i \end{pmatrix} + \mathbf{v},$$

it is obtained that

$$\begin{aligned} \hat{\alpha} &= \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} & \mathbf{X}_{-i,-j}^t \mathbf{X}_i \\ \mathbf{X}_i^t \mathbf{X}_{-i,-j} & \mathbf{X}_i^t \mathbf{X}_i \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} = \begin{pmatrix} A & B \\ B^t & C \end{pmatrix} \cdot \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} \\ &= \begin{pmatrix} A \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + B \cdot \mathbf{X}_i^t \mathbf{X}_j \\ B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + C \cdot \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_{-i,-j} \\ \hat{\alpha}_i \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} C &= \left(\mathbf{X}_i^t \mathbf{X}_i - \mathbf{X}_i^t \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_i \right)^{-1} \\ &= \left(\mathbf{X}_i^t \left(\mathbf{I} - \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \right) \mathbf{X}_i \right)^{-1} = \left(\text{RSS}_i^{-i,-j} \right)^{-1}, \\ B &= - \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_i \cdot C, \\ A &= \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \cdot \left(\mathbf{I} + \mathbf{X}_{-i,-j}^t \mathbf{X}_i \cdot C \cdot \mathbf{X}_i^t \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \right) \\ &= \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} + \frac{1}{C} \cdot B \cdot B^t. \end{aligned}$$

In that case, the residual sum of squares is given by

$$\begin{aligned} \text{RSS}_j^{-j} &= \mathbf{X}_j^t \mathbf{X}_j - \begin{pmatrix} A \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + B \cdot \mathbf{X}_i^t \mathbf{X}_j \\ B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + C \cdot \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix}^t \begin{pmatrix} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \mathbf{X}_i^t \mathbf{X}_j \end{pmatrix} \\ &= \mathbf{X}_j^t \mathbf{X}_j - \mathbf{X}_j^t \mathbf{X}_{-i,-j} \cdot A^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j - \mathbf{X}_j^t \mathbf{X}_i \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j - \hat{\alpha}_i^t \mathbf{X}_i^t \mathbf{X}_j \\ &= \left(\mathbf{X}_j^t \mathbf{X}_j - \mathbf{X}_j^t \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j} \right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \right) \\ &\quad - \text{RSS}_i^{-i,-j} \mathbf{X}_j^t \mathbf{X}_{-i,-j} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j - \mathbf{X}_j^t \mathbf{X}_i \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j - \hat{\alpha}_i^t \mathbf{X}_i^t \mathbf{X}_j \\ &= \text{RSS}_j^{-i,-j} - \left(\text{RSS}_i^{-i,-j} \mathbf{X}_j^t \mathbf{X}_{-i,-j} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + \mathbf{X}_j^t \mathbf{X}_i \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j \right. \\ &\quad \left. + \hat{\alpha}_i^t \mathbf{X}_i^t \mathbf{X}_j \right), \end{aligned}$$

and consequently

$$\text{RSS}_j^{-i,-j} - \text{RSS}_j^{-j} = \text{RSS}_i^{-i,-j} \mathbf{X}_j^t \mathbf{X}_{-i,-j} \cdot B \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + \mathbf{X}_j^t \mathbf{X}_i \cdot B^t \cdot \mathbf{X}_{-i,-j}^t \mathbf{X}_j + \hat{\alpha}_i^t \mathbf{X}_i^t \mathbf{X}_j.$$

Appendix C

First, parting from the expression Equation (14), it is obtained that

$$\mathbf{M}_\lambda^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -\frac{\lambda}{1+\lambda} \hat{\alpha}_0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \frac{\lambda}{1+\lambda} \hat{\alpha}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & (-1)^{k-1} \frac{\lambda}{1+\lambda} \hat{\alpha}_{k-1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{1+\lambda} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & (-1)^{k+1} \frac{\lambda}{1+\lambda} \hat{\alpha}_{k+1} & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & (-1)^p \frac{\lambda}{1+\lambda} \hat{\alpha}_p & 0 & \cdots & 1 \end{pmatrix},$$

and then,

$$(\mathbf{M}_\lambda^{-1} - \mathbf{I})^t(\mathbf{M}_\lambda^{-1} - \mathbf{I}) = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & a(\lambda) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

where $a(\lambda) = \frac{\lambda^2}{(1+\lambda)^2} \cdot (\hat{\alpha}_0 + \hat{\alpha}_1 + \dots + \hat{\alpha}_{k-1}^2 + 1 + \hat{\alpha}_{k+1}^2 + \dots + \hat{\alpha}_p^2)$. In that case,

$$\boldsymbol{\beta}^t(\mathbf{M}_\lambda^{-1} - \mathbf{I})^t(\mathbf{M}_\lambda^{-1} - \mathbf{I})\boldsymbol{\beta} = a(\lambda) \cdot \beta_k^2.$$

Second, partitioning $\tilde{\mathbf{X}}$ in the form $\tilde{\mathbf{X}} = [\mathbf{X}_{-k} \tilde{\mathbf{X}}_k]$, it is obtained that

$$(\tilde{\mathbf{X}}^t\tilde{\mathbf{X}})^{-1} = \begin{pmatrix} (\mathbf{X}_{-k}^t\mathbf{X}_{-k})^{-1} + \frac{\hat{\alpha}\hat{\alpha}^t}{(1+\lambda)^2 \cdot \mathbf{e}_k^t\mathbf{e}_k} & -\frac{\hat{\alpha}}{(1+\lambda)^2 \cdot \mathbf{e}_k^t\mathbf{e}_k} \\ -\frac{\hat{\alpha}^t}{(1+\lambda)^2 \cdot \mathbf{e}_k^t\mathbf{e}_k} & \frac{1}{(1+\lambda)^2 \cdot \mathbf{e}_k^t\mathbf{e}_k} \end{pmatrix},$$

and then,

$$tr((\tilde{\mathbf{X}}^t\tilde{\mathbf{X}})^{-1}) = tr((\mathbf{X}_{-k}^t\mathbf{X}_{-k})^{-1}) + \frac{1}{(1+\lambda)^2 \cdot \mathbf{e}_k^t\mathbf{e}_k} \cdot (tr(\hat{\alpha}\hat{\alpha}^t) + 1).$$

Consequently, it is obtained that

$$MSE(\hat{\boldsymbol{\beta}}(\lambda)) = \sigma^2 tr((\mathbf{X}_{-k}^t\mathbf{X}_{-k})^{-1}) + \left(1 + \sum_{j=0, j \neq k}^p \hat{\alpha}_j^2\right) \cdot \beta_k^2 \cdot \frac{\lambda^2 + h}{(1+\lambda)^2}, \tag{A1}$$

where $h = \frac{\sigma^2}{\beta_k^2 \cdot RSS_k^{-k}}$.

Third, taking into account that the first and second derivatives of expression Equation (A1) are, respectively,

$$\begin{aligned} \frac{\partial}{\partial \lambda} MSE(\hat{\boldsymbol{\beta}}(\lambda)) &= \left(1 + \sum_{j=0, j \neq k}^p \hat{\alpha}_j^2\right) \cdot \beta_k^2 \cdot \frac{2(\lambda - h)}{(1+\lambda)^3}, \\ \frac{\partial^2}{\partial \lambda^2} MSE(\hat{\boldsymbol{\beta}}(\lambda)) &= -2 \left(1 + \sum_{j=0, j \neq k}^p \hat{\alpha}_j^2\right) \cdot \beta_k^2 \cdot \frac{2\lambda - (1+3h)}{(1+\lambda)^4}. \end{aligned}$$

Since $\lambda \geq 0$, it is obtained that $MSE(\hat{\boldsymbol{\beta}}(\lambda))$ is decreasing if $\lambda < h$ and increasing if $\lambda > h$, and it is concave if $\lambda > \frac{1+3h}{2}$ and convex if $\lambda < \frac{1+3h}{2}$.

Indeed, given that

$$\begin{aligned} \lim_{\lambda \rightarrow +\infty} MSE(\hat{\boldsymbol{\beta}}(\lambda)) &= \sigma^2 tr((\mathbf{X}_{-k}^t\mathbf{X}_{-k})^{-1}) + \left(1 + \sum_{j=0, j \neq k}^p \hat{\alpha}_j^2\right) \cdot \beta_k^2, \\ MSE(\hat{\boldsymbol{\beta}}(0)) &= \sigma^2 tr((\mathbf{X}_{-k}^t\mathbf{X}_{-k})^{-1}) + \left(1 + \sum_{j=0, j \neq k}^p \hat{\alpha}_j^2\right) \cdot \beta_k^2 \cdot h, \end{aligned} \tag{A2}$$

if $h > 1$, then $MSE(\hat{\beta}(0)) > \lim_{\lambda \rightarrow +\infty} MSE(\hat{\beta}(\lambda))$, and if $h < 1$, then $MSE(\hat{\beta}(0)) < \lim_{\lambda \rightarrow +\infty} MSE(\hat{\beta}(\lambda))$. That is to say, if $h > 1$, then the raise estimator presents always a lower MSE than the one obtained by OLS for all λ , and comparing expressions Equations (A1) and (A2) when $h < 1$, $MSE(\hat{\beta}(\lambda)) \leq MSE(\hat{\beta}(0))$ if $\lambda \leq \frac{2 \cdot h}{1-h}$.

From this information, the behavior of the MSE is represented in Figures A1 and A2. Note that the MSE presents a minimum value for $\lambda = h$.

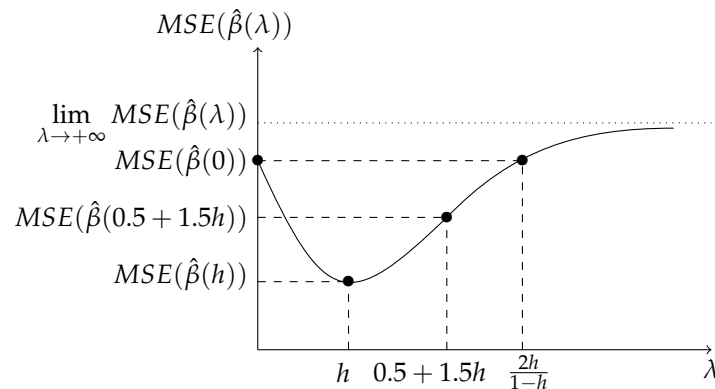


Figure A1. $MSE(\hat{\beta}(\lambda))$ representation for $h = \frac{\sigma^2}{(e_k^+ e_k) \cdot \beta_k^2} < 1$.

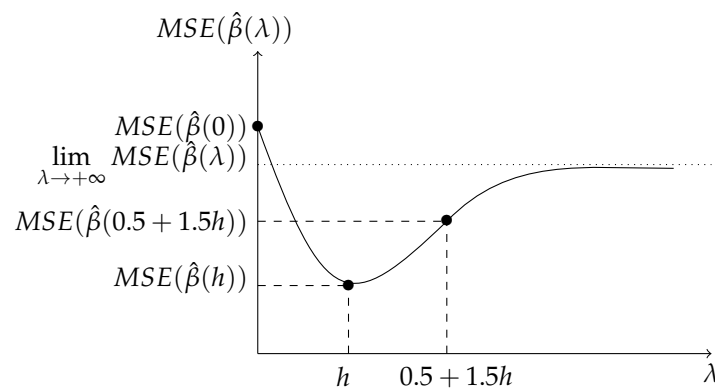


Figure A2. $MSE(\hat{\beta}(\lambda))$ representation for $h = \frac{\sigma^2}{(e_k^+ e_k) \cdot \beta_k^2} > 1$.

References

1. Kiers, H.; Smilde, A. A comparison of various methods for multivariate regression with highly collinear variables. *Stat. Methods Appl.* **2007**, *16*, 193–228. [CrossRef]
2. Frank, L.E.; Friedman, J.H. A statistical view of some chemometrics regression tools. *Technometrics* **1993**, *35*, 109–135. [CrossRef]
3. Fu, W.J. Penalized regressions: the bridge versus the lasso. *J. Comput. Graph. Stat.* **1998**, *7*, 397–416.
4. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]
5. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [CrossRef]
6. Donoho, D.L.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **1995**, *90*, 1200–1224. [CrossRef]
7. Klinger, A. Inference in high dimensional generalized linear models based on soft thresholding. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2001**, *63*, 377–392. [CrossRef]
8. Dupuis, D.; Victoria-Feser, M. Robust VIF regression with application to variable selection in large data sets. *Ann. Appl. Stat.* **2013**, *7*, 319–341. [CrossRef]

9. Li, Y.; Yang, H. A new Liu-type estimator in linear regression model. *Stat. Pap.* **2012**, *53*, 427–437. [[CrossRef](#)]
10. Liu, Y.; Wang, Y.; Feng, Y.; Wall, M. Variable selection and prediction with incomplete high-dimensional data. *Ann. Appl. Stat.* **2016**, *10*, 418–450. [[CrossRef](#)]
11. Uematsu, Y.; Tanaka, S. High-dimensional macroeconomic forecasting and variable selection via penalized regression. *Econom. J.* **2019**, *22*, 34–56. [[CrossRef](#)]
12. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2005**, *67*, 301–320. [[CrossRef](#)]
13. Tutz, G.; Ulbricht, J. Penalized regression with correlation-based penalty. *Stat. Comput.* **2009**, *19*, 239–253. [[CrossRef](#)]
14. Stone, M.; Brooks, R.J. Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *J. R. Stat. Soc. Ser. B (Methodol.)* **1990**, *52*, 237–269. [[CrossRef](#)]
15. Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least angle regression. *Ann. Stat.* **2004**, *32*, 407–499.
16. Golan, A.; Judge, G.; Miller, D. *Maximum Entropy Econometrics: Robust Estimation With Limited Data*; John Wiley and Sons: Chichester, UK, 1997.
17. Golan, A. Information and entropy econometrics review and synthesis. *Found. Trends Econom.* **2008**, *2*, 1–145. [[CrossRef](#)]
18. Macedo, P. Ridge Regression and Generalized Maximum Entropy: An improved version of the Ridge–GME parameter estimator. *Commun. Stat.-Simul. Comput.* **2017**, *46*, 3527–3539. [[CrossRef](#)]
19. Batah, F.S.M.; Özkale, M.R.; Gore, S. Combining unbiased ridge and principal component regression estimators. *Commun. Stat. Theory Methods* **2009**, *38*, 2201–2209. [[CrossRef](#)]
20. Massy, W.F. Principal components regression in exploratory statistical research. *J. Am. Stat. Assoc.* **1965**, *60*, 234–256. [[CrossRef](#)]
21. Guo, W.; Liu, X.; Zhang, S. The principal correlation components estimator and its optimality. *Stat. Pap.* **2016**, *57*, 755–779. [[CrossRef](#)]
22. Aguilera-Morillo, M.; Aguilera, A.; Escabias, M.; Valderrama, M. Penalized spline approaches for functional logit regression. *Test* **2013**, *22*, 251–277. [[CrossRef](#)]
23. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
24. De Jong, S. SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 251–263. [[CrossRef](#)]
25. Jensen, D.; Ramirez, D. Surrogate models in ill-conditioned systems. *J. Stat. Plan. Inference* **2010**, *140*, 2069–2077. [[CrossRef](#)]
26. García, J.; Salmerón, R.; García, C.; López Martín, M.D.M. Standardization of variables and collinearity diagnostic in ridge regression. *Int. Stat. Rev.* **2016**, *84*, 245–266. [[CrossRef](#)]
27. Marquardt, D. You should standardize the predictor variables in your regression models. Discussion of: A critique of some ridge regression methods. *J. Am. Stat. Assoc.* **1980**, *75*, 87–91.
28. Rodríguez, A.; Salmerón, R.; García, C. The coefficient of determination in the ridge regression. *Commun. Stat. Simul. Comput.* **2019**. [[CrossRef](#)]
29. García, C.G.; Pérez, J.G.; Liria, J.S. The raise method. An alternative procedure to estimate the parameters in presence of collinearity. *Qual. Quant.* **2011**, *45*, 403–423. [[CrossRef](#)]
30. Salmerón, R.; García, C.; García, J.; López, M.D.M. The raise estimator estimation, inference, and properties. *Commun. Stat. Theory Methods* **2017**, *46*, 6446–6462. [[CrossRef](#)]
31. García, J.; López-Martín, M.; García, C.; Salmerón, R. A geometrical interpretation of collinearity: A natural way to justify ridge regression and its anomalies. *Int. Stat. Rev.* **2020**. [[CrossRef](#)]
32. Belsley, D.A.; Kuh, E.; Welsch, R.E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*; John Wiley & Sons: Hoboken, NJ, USA, 2005; Volume 571.
33. Hoerl, A.; Kannard, R.; Baldwin, K. Ridge regression: some simulations. *Commun. Stat. Theory Methods* **1975**, *4*, 105–123. [[CrossRef](#)]
34. Stein, C. 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: Contributions to the Theory of Statistics*; University of California Press: Berkeley, CA, USA, 1956; pp. 197–206.

35. James, W.; Stein, C. Estimation with Quadratic Loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*; University of California Press: Berkeley, CA, USA, 1961; pp. 361–379.
36. Ohtani, K. An MSE comparison of the restricted Stein-rule and minimum mean squared error estimators in regression. *Test* **1998**, *7*, 361–376. [[CrossRef](#)]
37. Hubert, M.; Gijbels, I.; Vanpaemel, D. Reducing the mean squared error of quantile-based estimators by smoothing. *Test* **2013**, *22*, 448–465. [[CrossRef](#)]
38. Salmerón, R.; García, C.; García, J. Variance Inflation Factor and Condition Number in multiple linear regression. *J. Stat. Comput. Simul.* **2018**, *88*, 2365–2384. [[CrossRef](#)]
39. Salmerón, R.; Rodríguez, A.; García, C. Diagnosis and quantification of the non-essential collinearity. *Comput. Stat.* **2019**. [[CrossRef](#)]
40. Marquardt, D.; Snee, R. Ridge regression in practice. *Am. Stat.* **1975**, *29*, 3–20.
41. García, C.B.; Garcí, J.; Salmerón, R.; López, M.M. Raise regression: Selection of the raise parameter. In *Proceedings of the International Conference on Data Mining, Vancouver, BC, Canada, 30 April–2 May 2015*.
42. García, J.; Ramírez, D. The successive raising estimator and its relation with the ridge estimator. *Commun. Stat. Simul. Comput.* **2016**, *46*, 11123–11142. [[CrossRef](#)]
43. Klein, L.; Goldberger, A. *An Economic Model of the United States, 1929–1952*; North Holland Publishing Company: Amsterdam, The Netherlands, 1964.
44. García, C.; Salmerón, R.; García, C.; García, J. Residualization: Justification, properties and application. *J. Appl. Stat.* **2019**. [[CrossRef](#)]
45. Marquardt, D. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **1970**, *12*, 591–612. [[CrossRef](#)]
46. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
47. York, R. Residualization is not the answer: Rethinking how to address multicollinearity. *Soc. Sci. Res.* **2012**, *41*, 1379–1386. [[CrossRef](#)]
48. Salmerón, R.; García, J.; García, C.; García, C. Treatment of collinearity through orthogonal regression: An economic application. *Boletín Estadística Investig. Oper.* **2016**, *32*, 184–202.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).