Communications in Statistic. Theory and Methods manuscript No.
(will be inserted by the editor)

THE RAISE ESTIMATORS. ESTIMATION, INFERENCE AND PROPERTIES

Received: date / Accepted: date

Abstract Several methods using different approaches have been developed to remedy the consequences of collinearity. To the best of our knowledge, only the raise estimator proposed by García et al (2010) deals with this problem from a geometric perspective. This paper fully develops the raise estimator for a model with two standardized explanatory variables. Inference in the raise estimator is examined, showing that it can be obtained from ordinary least squares methodology. In addition, contrary to what happens in ridge regression, the raise estimator maintains the coefficient of determination value constant. The expression of the variance inflation factor for the raise estimator is also presented. Finally, a comparative study of the raise and ridge estimators is carried out using an example.

Keywords Multicollinearity · Ridge estimator · Raise estimator · Variance Inflator Factor · Regression

1 Introduction

The Collinearity problem describes the situation where the non-orthogonality exists among explanatory variables in regression models. The breakdown in orthogonality among explanatory variables results in imprecisions when using a normal equation in ordinary least squares (OLS) estimations. In the presence of collinearity, the 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).
For this reason, a great many techniques have been developed to remedy the consequent symptoms resulting from data collinearity, such as the Stein estimator (Stein et al, 1956), the ridge estimator (RE) (Hoerl and Kennard, 1970a,b; McDonald, 2009, 2010) and the contraction estimator (Liu, 1993; Mayer and Willke, 1973). This last one combines the Stein estimator with the ridge estimator and proposes the following estimator $\hat{\boldsymbol{\beta}}(d) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} + d\hat{\boldsymbol{\beta}})$, where $0 < d < 1$ still depending on OLS estimator which will be unstable. To overcome this situation, Liu (2003) proposed the Liu-type estimator given by the following expression $\hat{\boldsymbol{\beta}}^*(k,d) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} + d\hat{\boldsymbol{\beta}}^*)$, where $k > 0$, $-\infty < d < \infty$ and $\hat{\boldsymbol{\beta}}^*$ can be any estimator of β. Sakallıoğlu and Kaçıranlar (2008) presented the k-d class estimator using the ridge estimator and based on the augmented model provided the following expression $\hat{\boldsymbol{\beta}}(k,d) = (\mathbf{X}'\mathbf{X}+k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{y}+d\hat{\boldsymbol{\beta}}(k))$ where $k > 0$ $-\infty < d < \infty$ and $\hat{\boldsymbol{\beta}}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$. They showed that the k-d class estimator is a general estimator

Address(es) of author(s) should be given

which includes the OLS estimator, the RE and the Liu estimator. By combining the RE and the Liu estimator, Chang and Yang (2012) proposed the two parameter (TP) estimator which includes the OLS, RE and Liu estimators as special cases. All these estimators are founded on the ridge estimator (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. In this same line, Liu et al (2013) recently proposed the improved ridge estimator (IRE).
Another alternative is the principal component regression (PCR) (Batah et al, 2009; Massy, 1965). In the line of the PCR estimator and modifying the TP estimator, Chang and Yang (2012) introduced a new estimator to provide an alternative method for overcoming the multicollinearity problem in linear regression. A similar procedure is to apply the generalized inverse proposed by Marquardt (1970).
Alternative methods for deriving the restricted ridge regression (RRR) estimator have been provided by several authors, Farebrother (1984), Groß (2003), Kaciranlar et al (1998), Özkale (2009), Sarkar (1992), Zhong and Yang (2007) among others. Note that the ridge regression can be obtained by augmenting the equation:

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{u} \\ \mathbf{u}' \end{pmatrix}, \tag{1}$$

to the original equation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ and then using the OLS method. Based on prior information, Swindel (1976) proposed a modified ridge estimator (MRE). Crouse et al (1995) introduced the unbiased ridge regression (URR) estimator as a convex combination of prior information with the RRR estimator.
From the point of view of the conditioned minimum, Hoerl and Kennard (1970a,b) introduced the ridge regression, which minimizes RSS subject to the constraint $\sum |\beta_j| \leq t$. Note that the function $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}'\boldsymbol{\beta}$ is the least square criterion function for the augmented model. Frank and Friedman (1993) introduced bridge regression (BR) subject to the constraint $\sum |\beta_j|^\gamma \leq t$ as a special family of penalized regressions with two very important members: the ridge regression ($\gamma = 2$) and the Lasso regression ($\gamma = 1$) which was treated by Tibshirani (1996). Jensen and Ramirez (2008) and Kapat and Goel (2010) presented some anomalies of these estimators.
As regards the solution of the normal equations, Jensen and Ramirez (2010) introduced the surrogate estimator as a solution to collinearity. In a different line, following works on the least median of squares (LMS) by Rousseeuw (1984, 1985), Pati et al (2015) proposed a new biased estimator of the robust ridge regression called the ridge least median squares (RLMS) estimator to be applied in the presence of multicollinearity and outliers. Other methods, such as continuum regression, (Stone and Brooks, 1990), least angle regression, (Efron et al, 2004) and generalized maximum entropy, (Golan et al, 1996; Golan, 2008), are well suited to cope with collinearity problems.
However, none of the reviewed alternatives focus on the geometric problem expressly stated by Alin (2010): *Multicollinearity refers to the linear relationship among two or more variables, which also means lack of orthogonality among them. This relation is also called collinearity or ill conditioning by some authors like Besley (1991) and Chatterjee and Hadi (2006). 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.* This means that collinearity can be treated from a geometric point of view.
From this geometric point of view, García et al (2010) introduced the raise estimator and now, in this paper, it is analyzed the effect that the raise estimator has on correcting collinearity and the behavior of collinearity measures to diagnose collinearity after applying the raise estimator.
The paper is organized as follows: Section 2 presents the notation. Section 3 reviews the two-variable generalized raise method. Section 4 shows that the raise regression obtained from OLS

keeps constant the value of the sums of squares and presents the expression for the confidence intervals and the global and individual significance tests. A Montecarlo simulation is presented to support the theoretical results. Section 5 presents an expression of the variance inflator factor to be applied after the raise regression. Section 6 illustrate the results with an empirical application. Finally, section 7 summarizes the mean conclusions of the paper. We also present an appendix summarizing the mean characteristics of the ridge estimator.

2 Notation

Vector and matrices are set in bold type. The transpose and inverse of the matrix $\mathbf{A}$ are $\mathbf{A}'$ and $\mathbf{A}^{-1}$, respectively. The special array are the identity matrix, $\mathbf{I}$, and the null vector, $\mathbf{0}$, of order 2. $\mathbf{x}\perp\mathbf{y}$ indicates that the vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.
The original model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{y}$ is the vector of order $n \times 1$; $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2]$ of dimension $n \times 2$ and x_{ij} is the observation i of the variable j; $\boldsymbol{\beta}$ is the vector of parameters of order 2×1 and $\mathbf{u}$ is the vector of random disturbance for $n \times 1$ where $E(\mathbf{u}) = 0$, $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$. The number of observations is n. Finally, note that we standardize the variables by subtracting the mean for each of the variables and dividing by $(n)^{1/2}$ times the standard deviation. In this case, the matrix $\mathbf{X}'\mathbf{X}$ will be the correlation matrix and the vector $\mathbf{X}'\mathbf{y}$ will be the correlation between the explained variable and every explanatory variable. That is

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & \sum\limits_{i=1}^{n} x_{1i}x_{2i} \\ \sum\limits_{i=1}^{n} x_{2i}x_{1i} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where ρ is the correlation coefficient between the explanatory variables and $\mathbf{X}'\mathbf{y}$ is given by

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} \sum\limits_{i=1}^{n} x_{1i}y_i \\ \sum\limits_{i=1}^{n} x_{2i}y_i \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix},$$

where γ_i, for $i = 1, 2$, is the correlation coefficient between the explained variable and explanatory variables.
The raise model is $\mathbf{y} = \tilde{\mathbf{X}}\boldsymbol{\beta}(\lambda) + \mathbf{w}$ where $\tilde{\mathbf{X}} = [(1+\lambda)\mathbf{x}_1 - \lambda\rho\mathbf{x}_2, \mathbf{x}_2]$ is the raise matrix and $\lambda \geq 0$ and the ridge model is $\mathbf{y}^R = \mathbf{X}^R\boldsymbol{\beta} + \mathbf{v}$ where for $k \geq 0$, $\mathbf{y}^R = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix}$, $\mathbf{X}^R = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I} \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} \mathbf{u} \\ \mathbf{u}' \end{pmatrix}$.
The estimators by OLS for the original model is $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$; for the raise model is $\hat{\boldsymbol{\beta}}(\lambda) = \left\{(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{y}, \quad \lambda \geq 0\right\}$ (García et al, 2010) and for the ridge model is $\hat{\boldsymbol{\beta}}(k) = \left\{(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}, \quad k \geq 0\right\}$ (Marquardt, 1970; Zhang and Ibrahim, 2005).
The explained sum of squares, the residual sum of squares, the total sum of squares, the coefficient of determination and the variance inflator factor are ESS, RSS, TSS, R^2 and VIF for the original model and ESS(λ), RSS(λ) TSS(λ), $R^2(\lambda)$ and VIF(λ) for the raise model. The coefficient of determination and the variance inflator factor for the ridge model are denoted as $R^2(k)$ and VIF(k) respectively.

3 The two-variable generalized raise method. The raise estimators

The common linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ with n observations and two variables will be expressed as:

$$\mathbf{y} = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \mathbf{u}. \tag{2}$$

The collinearity problem arises because the vector $\mathbf{x}_1$ and the vector $\mathbf{x}_2$ are very close geometrically, that is, the angle that determines both vectors, θ_1, is very small (see Figure 1).

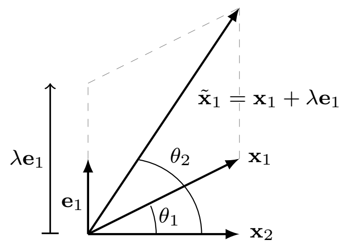


Fig. 1: Representation of raise method

To correct this problem before proceeding to the estimation, we will try to separate them through the following regression

$$\mathbf{x}_1 = \alpha\mathbf{x}_2 + \boldsymbol{\epsilon}, \tag{3}$$

whose estimation by OLS is $\hat{\alpha} = (\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{x}_1$ so that it is verified that $\hat{\alpha} = \rho$. Thus, we can say that $\mathbf{x}_1 = \rho\mathbf{x}_2 + \mathbf{e}_1$ with $\mathbf{e}_1 \perp \mathbf{x}_2$ where $\mathbf{e}_1$ is the residual obtained from regression (3). Then, the raise vector is defined as

$$\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda\mathbf{e}_1. \tag{4}$$

The raise method will be obtained by substituting vector $\mathbf{x}_1$ for the raise vector $\tilde{\mathbf{x}}_1$ in the model (2). That is, the raise method will be the OLS regression with the vectors $\tilde{\mathbf{x}}_1$ and $\mathbf{x}_2$ instead of $\mathbf{x}_1$ and $\mathbf{x}_2$. From Figure 1, it is evident that the angle between $\tilde{\mathbf{x}}_1$ and $\mathbf{x}_2$, θ_2, will be bigger than the angle θ_1. Thus, the correlation between both vectors will be smaller and the correlation problem will diminish. The higher the parameter λ (raising factor) the greater the angle between the vectors and the lower the correlation and, as we will see, the variance inflator factor. García et al (2015b) proposed a criterion to select the raise parameter λ based on select the value of λ that provides the lowest mean square error, analogously to the method proposed to select k in ridge regression. Then, the model to estimate will be given by:

$$\mathbf{y} = \beta_1(\lambda)\tilde{\mathbf{x}}_1 + \beta_2(\lambda)\mathbf{x}_2 + \mathbf{w}, \tag{5}$$

where the estimated parameters depending on λ will be called raise estimators and denoted as $\hat{\beta}_1(\lambda)$ and $\hat{\beta}_2(\lambda)$.
It is possible to obtain the raise estimators taking into account that $\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \lambda\mathbf{e}_1 = \mathbf{x}_1 + \lambda(\mathbf{x}_1 - \rho\mathbf{x}_2) = (1+\lambda)\mathbf{x}_1 - \lambda\rho\mathbf{x}_2$. Thus, the raise matrix $\tilde{\mathbf{X}}$ is defined by $\tilde{\mathbf{X}} = [(1+\lambda)\mathbf{x}_1 - \lambda\rho\mathbf{x}_2,\ \mathbf{x}_2]$. Note that the dimensions of $\mathbf{X}$ are maintained as happens in surrogate model (see Jensen and Ramirez, 2010) and unlike the ridge estimation where it increases.
By using OLS estimation in model (5) we can obtain the raise estimators:

$$\hat{\boldsymbol{\beta}}(\lambda) = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y} = \begin{pmatrix} \frac{\gamma_1 - \rho\gamma_2}{(1+\lambda)(1-\rho^2)} \\ \frac{(1+\lambda)\gamma_2 - \rho\gamma_1 - \rho^2\gamma_2\lambda}{(1+\lambda)(1-\rho^2)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1(\lambda) \\ \hat{\beta}_2(\lambda) \end{pmatrix}. \tag{6}$$

Some important characteristics are:

- If $\gamma_1 - \rho\gamma_2 > 0$, then $\hat{\beta}_1(\lambda)$ will always be positive and decreasing in λ. If $\rho = \frac{\gamma_1}{\gamma_2}$, then $\hat{\beta}_1(\lambda) = 0$ for any value of λ.

- $\hat{\beta}_2(\lambda)$ will be always positive if the numerator is also possitive, which is equivalent to saying that $\gamma_2 > \frac{\rho\gamma_1}{1+\lambda(1-\rho^2)}$. Since the second member of this condition is decreasing in λ, the condition will be verified if $\gamma_2 > \rho\gamma_1$, or equivalently, if $\gamma_2 - \rho\gamma_1 > 0$.
- $\hat{\beta}_2(\lambda)$ is increasing if ρ and $\gamma_1 - \rho\gamma_2$ have the same sign. If they have different signs, it will be decreasing.
- When λ tends to infinity, we obtain that $\lim\limits_{\lambda\to+\infty} \hat{\beta}_1(\lambda) = 0$, similar to ridge estimation. However, $\lim\limits_{\lambda\to+\infty} \hat{\beta}_2(\lambda) = \gamma_2$ that coincides with the OLS estimation when $\rho = 0$.

One of the important characteristics of these estimators is that they are obtained by OLS methodology and therefore, if the classical assumptions are verified, these estimators will maintain all the properties of the OLS estimator and the common statistical tests will be applicable.

3.1 Relation between Raise and OLS estimators

As mentioned above, the raise estimators are obtained by applying the OLS method to the model (5). However, they are not identical to the OLS estimators because the information matrix has changed.
Given that $\tilde{\mathbf{X}} = \mathbf{X} \cdot \mathbf{M}$ where

$$\mathbf{M} = \begin{pmatrix} 1+\lambda & 0 \\ -\lambda\rho & 1 \end{pmatrix},$$

it is clear that

$$\hat{\boldsymbol{\beta}}(\lambda) = \mathbf{M}^{-1} \cdot \hat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{\hat{\beta}_1}{1+\lambda} \\ \hat{\beta}_2 + \frac{\lambda\rho}{1+\lambda}\hat{\beta}_1 \end{pmatrix}, \tag{7}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are OLS estimators of the model (2). Note that if we use a null raising factor ($\lambda = 0$) in expression (7) we obtain that raise and OLS estimators are equal.
The difference between the OLS estimator and the raise estimator can be obtained by:

$$\hat{\beta}_1(\lambda) - \hat{\beta}_1 = \frac{(\rho\gamma_2 - \gamma_1)\lambda}{(1+\lambda)(1-\rho^2)}, \tag{8}$$

$$\hat{\beta}_2(\lambda) - \hat{\beta}_2 = \frac{\lambda\rho(\gamma_1 - \rho\gamma_2)}{(1+\lambda)(1-\rho^2)}. \tag{9}$$

From expressions (8) and (9) it is evident that $\lim\limits_{\lambda\to+\infty} \hat{\beta}_1(\lambda) - \hat{\beta}_1 = -\hat{\beta}_1$ and $\lim\limits_{\lambda\to+\infty} \hat{\beta}_2(\lambda) - \hat{\beta}_2 = \frac{\rho(\gamma_1-\rho\gamma_2)}{1-\rho^2}$.

4 Inference in raise model

The raise estimators (6) have been obtained by applying OLS to the raise model (5) and, due to this fact, it is possible to develop the inference of the raise estimators from the well-known OLS methodology. In this section we present the estimated, residual and total sum of squares (ESS, RSS and TSS respectively), the coefficient of determination, the confidence interval and the global and individual significance test. A Montecarlo simulation is presented to support the theoretical results.

4.1 ESS, RSS, TSS and coefficient of determination

Since all variables of the raise model (5) are centered $\text{ESS}(\lambda) = \hat{\boldsymbol{\beta}}(\lambda)'\tilde{\mathbf{X}}'\mathbf{y}$. By operating it is possible to prove that $\text{ESS}(\lambda)$ will be equal to ESS:

$$\text{ESS}(\lambda) = \hat{\boldsymbol{\beta}}(\lambda)'\tilde{\mathbf{X}}'\mathbf{y} = \frac{\gamma_1^2 + \gamma_2^2 - 2\rho\gamma_1\gamma_2}{1-\rho^2} = \hat{\beta}_1\gamma_1 + \hat{\beta}_2\gamma_2 = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \text{ESS}.$$

Evidently, the explained variable in the raise model (5) will also coincide with the one in model (2) estimated by OLS, it is to say $\text{TSS} = \text{TSS}(\lambda) = 1$. Therefore, the RSS of both models will also be equal:

$$\text{RSS}(\lambda) = \text{RSS} = \text{TSS} - \text{ESS} = \frac{1 - \rho^2 - \gamma_1^2 - \gamma_2^2 + 2\rho\gamma_1\gamma_2}{1-\rho^2}.$$

In consequence, the coefficient of determination and the estimated variances will also be similar:

$$R^2(\lambda) = \frac{\text{ESS}(\lambda)}{\text{TSS}(\lambda)} = \frac{\gamma_1^2 + \gamma_2^2 - 2\rho\gamma_1\gamma_2}{1-\rho^2} = \frac{\text{ESS}}{\text{TSS}} = R^2, \tag{10}$$

$$\hat{\sigma}^2(\lambda) = \frac{\text{RSS}(\lambda)}{n-2} = \frac{\text{RSS}}{n-2} = \hat{\sigma}^2. \tag{11}$$

Note that the raise estimator maintains the value of the determination coefficient, R^2, regardless of the value of the raising parameter, λ. Note that this is contrary to what happens in ridge regression (where it decreases, see McDonald, 2010) and it will be an important characteristic of the raise estimator.

4.2 Confidence intervals

Since the raise estimators have been obtained by OLS, we can obtain the confidence intervals for the raise estimators from this well known methodology. Then, we can state that $\hat{D}(\hat{\beta}_i(\lambda)) = \sqrt{\hat{\sigma}^2 a_{ii}}$ where a_{ii} is the element (i, i) in the matrix $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$. Using the confidence interval concept, we can say that the parameter $\beta_i(\lambda)$ will be, with a $(1-\varepsilon)100\%$ of confidence, in the confidence interval

$$\hat{\beta}_i(\lambda) \pm t_{1-\frac{\varepsilon}{2}} \hat{D}(\hat{\beta}_i(\lambda)).$$

For $i = 1$

$$\beta_1(\lambda) \in \frac{\gamma_1 - \rho\gamma_2}{(1+\lambda)(1-\rho^2)} \pm t_{1-\frac{\varepsilon}{2}} \frac{1}{1+\lambda}\sqrt{\frac{\hat{\sigma}^2}{1-\rho^2}} = \frac{1}{1+\lambda}[\hat{\beta}_1 \pm t_{1-\frac{\varepsilon}{2}} \hat{D}(\hat{\beta}_1)]. \tag{12}$$

Note that confidence intervals have been associated for the first coefficient models (2) and (5) so that the first has been reduced by dividing it by $(1+\lambda)$. The larger the parameter λ, the smaller the confidence interval, which becomes zero when λ tends to infinity. Analogously, for $i = 2$

$$\beta_2(\lambda) \in \left(\frac{\gamma_2 - \rho\gamma_1}{1-\rho^2} + \frac{\lambda\rho(\gamma_1 - \rho\gamma_2)}{(1+\lambda)(1-\rho^2)}\right) \pm t_{1-\frac{\varepsilon}{2}} \sqrt{\hat{\sigma}^2 \frac{(1+\lambda)^2 - \rho^2(2\lambda+\lambda^2)}{(1+\lambda)^2(1-\rho^2)}}. \tag{13}$$

In this second parameter the interval has also been transformed by adding $\hat{\beta}_2$, the bias of the raise estimator, to the center of the interval. The radio of the interval is smaller than the radio corresponding to the OLS estimator, since the interval for the OLS estimator will be:

$$\hat{\beta}_2 \pm t_{1-\frac{\varepsilon}{2}} \sqrt{\hat{\sigma}^2 \frac{1}{1-\rho^2}}, \tag{14}$$

and by comparing the radio of both intervals, expressions (13) and (14), we can state that:

$$\frac{(1+\lambda^2) - \rho^2(2\lambda+\lambda^2)}{(1+\lambda)^2(1-\rho^2)} \leq \frac{1}{1-\rho^2}.$$

4.3 Global and individual significance test

As shown in subsection 4.1 the sums of squares are constant in raise regression. For this reason, the experimental F will be also similar in OLS and raise models:

$$F_{exp}(\lambda) = \frac{\text{ESS}}{\frac{\text{RSS}}{n-2}} = (n-2)\frac{\gamma_1^2 + \gamma_2^2 - 2\rho\gamma_1\gamma_2}{1-\rho^2-\gamma_1^2-\gamma_2^2+2\rho\gamma_1\gamma_2} = F_{exp}. \tag{15}$$

Therefore, we can say that the global significance tests in the original model, expression (2), and the raise model, expression (5), are identical.
We should remember that one of the symptoms of multicollinearity is a globally significant model but with individually insignificant parameters. This situation could be corrected with the raise method by choosing a right raising parameter, λ. Then the raise estimator allows us to match the global significance test with the individual significance test by choosing a determined raising parameter. Hence, it could be more interesting to analysis the behavior of the individual significance test for raise estimators, which can be obtained by:

$$t_{exp}(\beta_i(\lambda)) = \frac{\hat{\beta}_i(\lambda)}{\sqrt{\hat{\sigma}^2 a_{ii}}}. \tag{16}$$

In this situation we obtain:

$$t_{exp}(\beta_1(\lambda)) = \frac{\gamma_1 - \rho\gamma_2}{\sqrt{\hat{\sigma}^2(1-\rho^2)}}, \tag{17}$$

$$t_{exp}(\beta_2(\lambda)) = \frac{(1+\lambda)\gamma_2 - \rho\gamma_1 - \rho^2\gamma_2\lambda}{\sqrt{\hat{\sigma}^2(1-\rho^2)[(1+\lambda)^2 - \rho^2(2\lambda+\lambda^2)]}}. \tag{18}$$

We can see that t_{exp} corresponding to the raise variable, $t_{exp}(\beta_1(\lambda))$, does not depend on λ. Thus, it remains constant after the successive raising, while $t_{exp}(\beta_2(\lambda))$, depends on λ and it will vary when λ increases. Also:

$$\lim_{\lambda\to+\infty} t_{exp}(\beta_2(\lambda)) = \frac{\gamma_2}{\hat{\sigma}} = \frac{\gamma_2\sqrt{(1-\rho^2)(n-2)}}{\sqrt{1-\rho^2-\gamma_1^2-\gamma_2^2+2\rho\gamma_1\gamma_2}}.$$

4.4 Montecarlo simulation

As indicated above, the model (5) will be globally significant if it was the initial model (2), and the same for the estimated parameter of the raised variable. However, the individual significance of the not raised variable will depend on the parameter λ taking into account the expression (18). This section analyzes, trough a Montecarlo simulation, the conditions required to obtain an individually significant estimated parameter of the not raised variable, it is to say $|t_{exp}(\beta_2(\lambda))| > t_{n-k}\left(1-\frac{\alpha}{2}\right)$. The simulation is performed under the following considerations:

- $k = 2$ and $\alpha = 0.05$.
- $\rho \in \{\pm 0.95, \pm 0.96, \pm 0.97, \pm 0.98, \pm 0.99\}$, then it is verified that VIF > 10.
- $\lambda \in [0, 10]$ with an interval of 0.1.
- For γ_1 and γ_2 it is possible to distinguish three cases:
  - Case A: $\gamma_1, \gamma_2 \in \{0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4, \pm 0.5, \pm 0.6, \pm 0.7, \pm 0.8, \pm 0.9, \pm 1\}$.
  - Case B: $\gamma_1, \gamma_2 \in \{\pm 0.6, \pm 0.7, \pm 0.8, \pm 0.9, \pm 1\}$.
  - Case C: $\gamma_1, \gamma_2 \in \{0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4\}$.
- $n = 20, 35, 40, 60, 160$[1].

[1] Taking into account that, keeping the rest of values constant, when n increases $\hat{\sigma}^2$ decreases, then $t_{exp}(\beta_2(\lambda))$ will increase. Thus, the tendency to reject the null hypothesis will be greater as the number of observations increases.

From these considerations, it is analyzed how many times it is verified that $|t_{exp}(\beta_2(\lambda))| > t_{n-k}\left(1-\frac{\alpha}{2}\right)$. Table 1 resumes the results of the simulation for $\lambda = 10$.

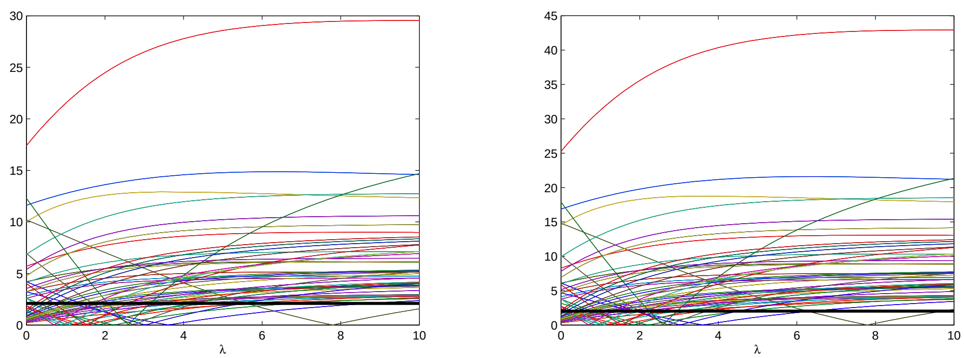
n	$t_{n-k}\left(1-\frac{\alpha}{2}\right)$	Case A	Case B	Case C
20	2.101	54.04%	97.91%	28.70%
35	2.035	64.07%	100%	43.06%
40	2.024	68.52%	100%	48.61%
60	2	74.65%	100%	56.35%
160	1.975	84.67%	100%	74.03%

Table 1: Results of the Montecarlo simulation ($\lambda = 10$)

Note that:

- As expected, the percentage of cases in which the null hypothesis is rejected increases as the number of observations increases.
- When the correlation between the independent variables and the dependent variable is higher (Case B), the percentage of cases in which the null hypothesis is rejected is higher, becoming 100% for $n \geq 35$.

Figure 2 is presented to analyze the effect on t-student experimental of using different values of λ. Note that the number of cases where the variable remains no significant (under solid line that represents the theoretical value $t_{n-k}\left(1-\frac{\alpha}{2}\right)$) decrease as λ increases and for $\lambda = 10$ all cases conclude that the variable is individually significant.



(a) $n = 20, 0 \leq \lambda \leq 10$ (b) $n = 40, 0 \leq \lambda \leq 10$

Fig. 2: t-student experimental of raise for $\hat{\beta}_2(\lambda)$ for different values of λ and the theoretical value (solid line)

On the other hand, the conclusion of this simulation can serve as a criterion to select the variable to raise. From model (2), the following scenarios are analyzed:

- If the estimated parameter of variable $\mathbf{x}_1$ is individually significant and the estimated parameter of variable $\mathbf{x}_2$ is not, it is recommendable to raise the variable $\mathbf{x}_1$ to maintain the characteristics associated to its individual inference in the raise model.

– If both estimated parameters are not individually significant, it is recommendable to raise the variable considered less important since the not raised variable (considered more relevant) can become individually significant after the application of the raise estimation.
– Finally, it is not considered the case where both estimated parameters are individually significant since this case is not common in models with collinearity. Anyway, in this case a criterion could be to raise the variable with the higher VIF (in a general model with more than two variables).

5 The VIF in the raise regression

The variance inflator factor (VIF) is usually applied to diagnose the presence of collinearity. Thus, if the VIF exceeds the threshold generally accepted in the literature, it could be advisable to apply some procedure, such as the ridge or the raise regression, to mitigate the collinearity. Once we have applied this procedure, it is necessary to check if it was effective, and this justifies the need to recalculate the value of the VIF after applying the ridge or raise regression. Salmerón et al (2015) presented the expression of the VIF extended to ridge regression and we will now present the VIF extended to raise regression. Thus, the VIF associated with the raise regression is given by:

$$\text{VIF}(\lambda) = \frac{1}{1 - R^2_{aux}}, \tag{19}$$

where R^2_{aux} is the coefficient of determination of the regression of $\mathbf{x}_2$ from $\tilde{\mathbf{x}}_1 = (1+\lambda)\mathbf{x}_1 - \lambda\rho\mathbf{x}_2$ (or vice versa). That is:

$$R^2_{aux} = \frac{\rho^2}{(1+\lambda)^2 - (\lambda^2 + 2\lambda)\rho^2}. \tag{20}$$

Therefore:

$$\text{VIF}(\lambda) = \frac{1}{1 - \frac{\rho^2}{(1+\lambda)^2 - (\lambda^2+2\lambda)\rho^2}} = \frac{(1+\lambda)^2(1-\rho^2) + \rho^2}{(1+\lambda)^2(1-\rho^2)}. \tag{21}$$

From (21) we have:
– For $\lambda = 0$, $\text{VIF}(0) = \frac{1}{1-\rho^2} = \text{VIF}$ matches with the VIF of (2) calculated from OLS.
– $\text{VIF}(\lambda)$ is decreasing in λ: if λ increases, the R^2_{aux} decreases. In this case $1 - R^2_{aux}$ increases and therefore $\text{VIF}(\lambda)$ decreases.
– $\text{VIF}(\lambda)$ is always greater or equal to 1 since $\text{VIF}(\lambda) = 1 \Leftrightarrow \rho = 0$ and also if $\text{VIF}(\lambda) < 1$ then $(1+\lambda)^2(1-\rho^2) + \rho^2 < (1+\lambda)^2(1-\rho^2)$ then $\rho^2 < 0$, which is impossible.
– When λ tends to infinity then the limit of $\text{VIF}(\lambda)$ is equal to one.
– An equivalent expression is $\text{VIF}(\lambda) = 1 + \frac{\rho^2}{(1+\lambda)^2(1-\rho^2)}$.

Note that $\text{VIF}(\lambda)$ is always greater or equal to 1, is decreasing in λ and continuous at $\lambda = 0$. That is, check the desired conditions for all VIF (see García et al, 2015a).

6 Empirical application

To illustrate the contribution of this paper, we will use the empirical application previously applied by McDonald and Schwing (1973) and McDonald (2010). In this example the total mortality rate, $\mathbf{y}$, is related to the nitrogen oxide pollution potential, $\mathbf{x}_1$, and the hydrocarbon pollution potential, $\mathbf{x}_2$, for 60 cities where $\rho = 0.984$, $\gamma_1 = -0.077$ and $\gamma_2 = -0.177$.
Figure[2] 3 shows the values of the ridge, raise and OLS estimators. Note that when k and λ are equal to zero, the values obtained by the raise method coincide with those obtained by the ridge estimation[3] and with those obtained from OLS in model (2). However, when the value is different

[2] All figures are represented from a discretization of 1001 points equally distributed in the range [0, 10].
[3] Every characteristic shown in this section on ridge estimation has been calculated using the expressions in the appendix.

from zero the results are very distant. Since $\gamma_1 - \rho\gamma_2 = 0.0972 > 0$, $\hat{\beta}_1(\lambda)$ is always positive. In contrast, as shown in Figure 3, $\hat{\beta}_2(\lambda)$ is always negative and increasing since ρ and $\gamma_1 - \rho\gamma_2$ have the same sign and the $\lim_{\lambda \to +\infty} \hat{\beta}_2(\lambda) = \gamma_2 < 0$. Indeed, the estimation obtained from the ridge regression is always less (in absolute value) than the estimation obtained from the raise regression. Note that the ridge estimators converge rapidly, while the raise estimators converge more slowly.

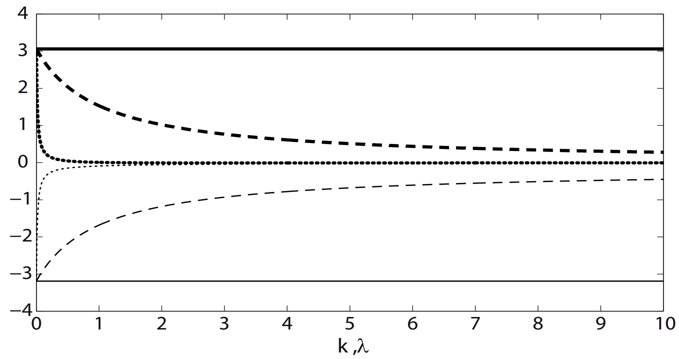


Fig. 3: OLS estimations ($\hat{\beta}_1$ thick solid; $\hat{\beta}_2$ thin solid), raise estimation ($\hat{\beta}_1(\lambda)$ thick dashed; $\hat{\beta}_2(\lambda)$ thin dashed) and ridge estimations ($\hat{\beta}_1(k)$ thick dotted; $\hat{\beta}_2(k)$ thin dotted)

Multicollinearity is often manifested in regressions that lead to different results in the global and individual significance test, i.e. the model is globally significant while one, several or all variables are not individually significant. The behavior of the ridge estimation in the global significance test is not adequate since the coefficient of determination is decreasing (remember that it is decreasing in k when k is increasing, see Figure 4). Therefore, the F_{exp} of the global significance test is also decreasing. Thus, if we choose a sufficiently high value of k, in this case $k = 0.33$, the $F_{exp}(k)$, in this case 3.9571, will be less than its critical value, 4.0069 (solid line), and the model will not be globally significant (see Figure 5). This does not occur with the raise estimation because the coefficient of determination (equal to 0.3288) remains constant for all values of λ. As a result, if the model was globally significant in the beginning, it will remain as occurs in this example where $F_{exp} = 28.4071$.

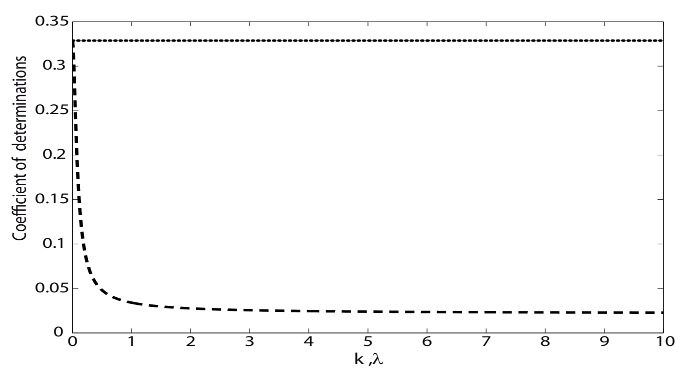


Fig. 4: Representing the coefficient of determination of raise estimation (dashed) and ridge estimation (dotted)

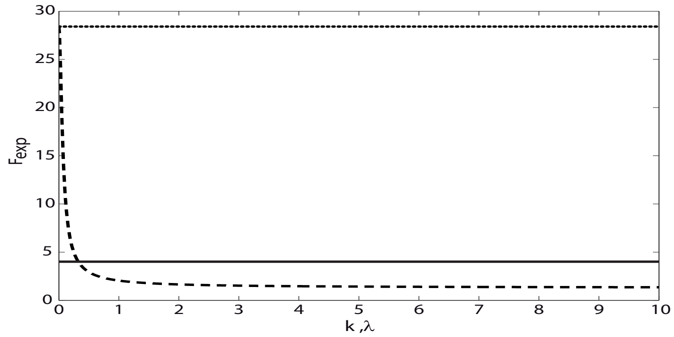


Fig. 5: Representing the F experimental of raise estimation (dashed) and ridge estimation (dotted) and theoretical value (solid)

Figure 6a shows the absolute value for the experimental t-student test for raise and ridge estimations. Note that the experimental t-student of the β_1 raise estimator (5.0695, constant and equal to the OLS estimator) is always over the theoretical value, 2.0017 (displayed with a dot-dot.dashed line), and thus the estimator will always be individually significant. This does not occur with the ridge estimator whose experimental value for $k = 0.28$ is 1.965 and less than the theoretical value. Thus, the coefficient of the first variable is not significantly different from zero. On the other hand, the ridge experimental t-student value for the second variable is also higher than the theoretical value as λ increases. However, in the ridge regression the second coefficient is not significantly different from zero from $k = 2.43$ onwards since the experimental value is equal to 2.0013.

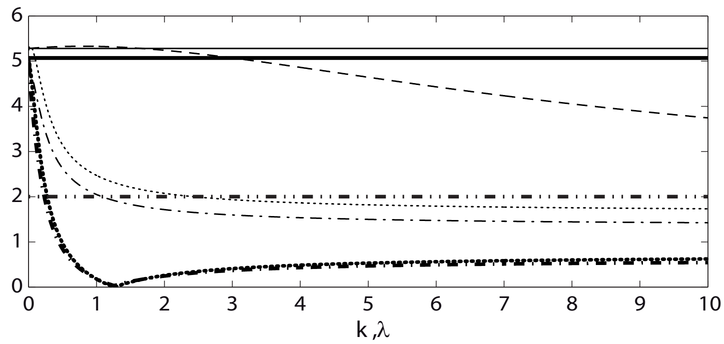
Thus, in the ridge regression for k equal or higher than 2.43 none of the coefficients are significantly different from zero and the model is not globally significant. However, in the raise regression the model is globally significant, the first coefficient is individually significant and the second coefficient will also be individually significant except for high values of λ (see Figure 6b) since:

$$\lim_{\lambda \to +\infty} |t_{exp}\left(\beta_2(\lambda)\right)| = \frac{0.177}{0.1076} = 1.6453 < 2.0017.$$

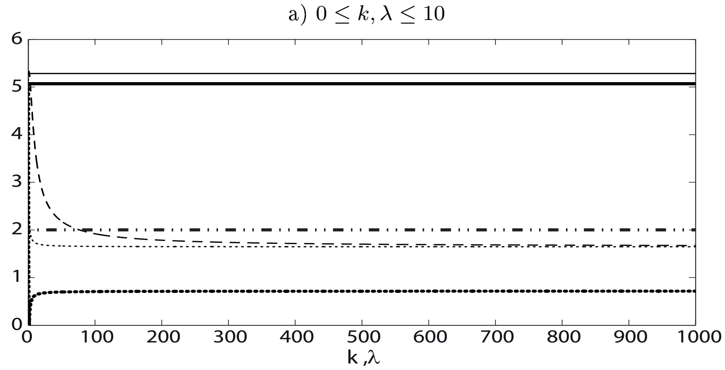
Figure 6b also shows that if we use the variance of the random perturbation associated with the ridge estimator, $\hat{\sigma}^2(k)$ (see expression (28) in Appendix A), to obtain the experimental values of individual significance test in the ridge regression, we will always obtain smaller values than if we use $\hat{\sigma}^2$. Thus, the probability of rejecting the null hypotheses will be lower in the first case.

Figure 7 displays the VIF values for the raise and ridge estimations. Note that for values of k to 0.02, the ridge VIF is equal to 6.4199, less than 10. However, the raise VIF will be less than 10 only from values of λ equal or higher than 0.85. Remember that the value 10 is usually applied as the limit to consider the collinearity problem mitigated. Thus, by using the ridge regression the collinearity will be solved for values of k equal or higher than 0.02 while values of λ equal to or higher than 0.85 are necessary in the raise regression.

To sum up, when k is equal to 0.09, the value of the VIF obtained by the ridge estimation, 0.9426, is less than one which is not possible (García et al, 2015a). The VIF obtained from the raise regression will have a much more gradual decline and will never be less than 1.



a) $0 \leq k, \lambda \leq 10$



b) $0 \leq k, \lambda \leq 1000$

Fig. 6: t-student experimental of: OLS for $\hat{\beta}_1$ (thick solid) and $\hat{\beta}_2$ (thin solid); Raise for $\hat{\beta}_1(\lambda)$ (thick dashed) and $\hat{\beta}_2(\lambda)$ (thin dashed); Ridge with σ^2 for $\hat{\beta}_1(k)$ (thick dotted) and $\hat{\beta}_2(k)$ (thin dotted); Ridge with $\sigma^2(k)$ for $\hat{\beta}_1(k)$ (thick dot-dashed) and $\hat{\beta}_2(k)$ (thin dot-dashed); and theoretical value (dot-dot-dashed)

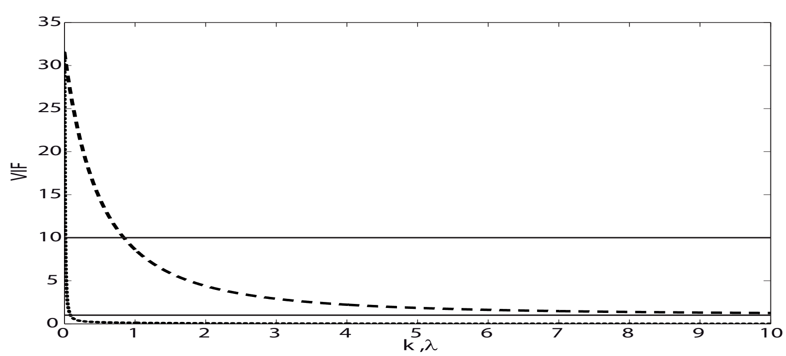


Fig. 7: Representing the VIF of raise estimation (dashed) and ridge estimation (dotted)

Although the goal of this paper is not the selection of the optimal value of λ, we now present the results for $\lambda = 0.85$ that was the selected as the optimal value of λ for this

same example in García et al (2015b). Thus, the following model is obtained from $\lambda = 0.85$**:**

$$\begin{array}{cccccc} \hat{\mathbf{y}} & = & \mathbf{1.6546}\ \tilde{\mathbf{x}}_1 & - & \mathbf{1.8051}\ \mathbf{x}_2, \\ & & \mathbf{(0.3264)} & & \mathbf{(0.3387)} \\ t_{exp} & & \mathbf{5.0695} & & \mathbf{-5.3269} \end{array}$$

with $\hat{\sigma} = 0.1076$**. Note that both coefficients are significantly different to zero and, consequently, we can state that the nitroxen oxide pollution has a positive influence on the total mortality rate while the hydrocarbon pollution potential has a negative influence. On the other hand, the model is globally significant due to the experimental value,** $F_{exp} = 28.4071$**, is higher than the theoretical one, 4.0069.**

7 Conclusions

In econometric practice it is common to find a globally significant model but where one or more variables are not individually significant. This seems contradictory and a usual explanation is the presence of collinearity. Ridge regression has been widely applied to estimate models with collinearity. However, what happens is that the experimental F diminishes and the model becomes globally insignificant. Thus, although the collinearity is mitigated, the model is neither globally or individually significant. In this paper we show that the raise estimator has a great advantage in this regard. The experimental F remains constant, so if the initial model is globally significant it will also remain significant in the raise regression. Moreover, the experimental t of one of the explanatory variables remains constant and so its corresponding parameter will be individually significant if it was already in the initial model. Since the rest of t will be increasing or decreasing, we can obtain a globally significant model in which the variables will also be individually significant, with only the value of λ varying. To conclude, the raise estimator mitigates collinearity while maintains the global and individual significance of the initial model.

Appendix

In this appendix the main expressions about the ridge estimator are collected.

A The ridge estimator

The ridge estimator $\hat{\boldsymbol{\beta}}(k)$, is given for $k \geq 0$ by:

$$\begin{aligned} \hat{\boldsymbol{\beta}}(k) &= \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{y} = [(1+k)^2 - \rho^2]^{-1}\begin{pmatrix} (1+k)\gamma_1 - \rho\gamma_2 \\ (1+k)\gamma_2 - \rho\gamma_1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{(1+k)\gamma_1 - \rho\gamma_2}{(1+k)^2 - \rho^2} \\ \frac{(1+k)\gamma_2 - \rho\gamma_1}{(1+k)^2 - \rho^2} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1(k) \\ \hat{\beta}_2(k) \end{pmatrix}. \end{aligned} \tag{22}$$

Hoerl and Kennard (1970a,b) showed that the ridge and OLS estimators are related by the following expression:

$$\hat{\boldsymbol{\beta}}(k) = \mathbf{Z}^{-1}\hat{\boldsymbol{\beta}}, \tag{23}$$

where $\mathbf{Z} = \mathbf{I} + k\left(\mathbf{X}'\mathbf{X}\right)^{-1}$. Thus, taking into account (7) we obtain that:

$$\hat{\boldsymbol{\beta}}(\lambda) = \mathbf{H}\hat{\boldsymbol{\beta}}(k), \tag{24}$$

where

$$\mathbf{H} = \mathbf{M}^{-1} + k\left(\mathbf{X}'\mathbf{X}\mathbf{M}\right)^{-1} = \frac{1}{1+\lambda(1-\rho)-\rho^2}\begin{pmatrix} \frac{1+\lambda(1-\rho)-\rho^2+k(1+\lambda)}{1+\lambda} & -k\rho \\ \frac{\lambda\rho(1+\lambda(1-\rho)-\rho^2)-k\rho(1+\lambda)}{1+\lambda} & (1+\lambda(1-\rho))(k+1)-\rho^2 \end{pmatrix}.$$

A.1 Variance Inflation Factor

Taking into account that the variance-covariance matrix of (22) is:

$$\text{var}(\hat{\boldsymbol{\beta}}(k)) = \sigma^2[\mathbf{X}'\mathbf{X}+k\mathbf{I}]^{-1}\mathbf{X}'\mathbf{X}[\mathbf{X}'\mathbf{X}+k\mathbf{I}]^{-1}, \tag{25}$$

and by following Marquardt (1970) the element of the principal diagonal of the matrix $[\mathbf{X}'\mathbf{X}+k\mathbf{I}]^{-1}\mathbf{X}'\mathbf{X}[\mathbf{X}'\mathbf{X}+k\mathbf{I}]^{-1}$, it is to say the VIFs, will be:

$$\text{VIF}(k) = \frac{(1+k)^2 - 2(1+k)\rho^2 + \rho^2}{[(1+k)^2-\rho^2]^2}. \tag{26}$$

For $k=0$, $\text{VIF}(k)$ matches with VIF and with the raise regression VIF for $\lambda = 0$.
García et al (2015a) showed that expression (26) presents some anomalies that make its application inadvisable.

A.2 Individual significance test

The experimental t values for the individual significance test, t_{exp}, for ridge estimators can be obtained, for $i,j=1,2$ and $i \neq j$, by:

$$t_{exp}(\beta_i(k)) = \frac{\hat{\beta}_i(k)}{\sqrt{\hat{\sigma}^2\text{VIF}(k)}} = \frac{(1+k)\gamma_i - \rho\gamma_j}{\hat{\sigma}\sqrt{(1+k)^2-2(1+k)\rho^2+\rho^2}}. \tag{27}$$

Also, for $i=1,2$:

$$\lim_{k\to+\infty} t_{exp}(\beta_i(k)) = \frac{\gamma_i}{\hat{\sigma}}.$$

Note that for $i=2$:

$$\lim_{k\to+\infty} t_{exp}(\beta_2(k)) = \lim_{\lambda\to+\infty} t_{exp}(\beta_2(\lambda)).$$

It is evident that the t_{exp} of both ridge estimators depend on k and they can present values less than its critical value. We can therefore state that the fit may not be individually or globally significant when using the ridge estimator. However, if the initial model is globally significant and presents an individually significant raise variable, then we can apply the raise method and maintain these properties.
On the other hand, Halawa and El Bassiouni (2000) proposed to substitute in expression (27) the variance of the random perturbation given by:

$$\hat{\sigma}^2(k) = \frac{\text{RSS}(k)}{n-2}, \tag{28}$$

where $\text{RSS}(k) = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(k))'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(k))$ is the residual mean square of the ridge model. From the relation given by (23) it is verified that:

$$\text{RSS}(k) = \text{RSS} + k^2\hat{\boldsymbol{\beta}}(k)'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\hat{\boldsymbol{\beta}}(k).$$

Then, it is verified that $\text{RSS}(0) = \text{RSS}$ and, since $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ is positive defined, $\text{RSS}(k) > \text{RSS}$ for all $k>0$. In consequence, $\hat{\sigma}^2(k) > \hat{\sigma}^2$, it is to say, the experimental t values for the individual significance test obtained from $\hat{\sigma}^2(k)$ will always be less than the one obtained from $\hat{\sigma}^2$.

A.3 Goodness of fit and global significance contrast

McDonald (2010) showed the following expression to calculate the coefficient of determination of the ridge estimator:

$$R^2(k) = \frac{\left(\hat{\boldsymbol{\beta}}(k)'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}(k) + k\hat{\boldsymbol{\beta}}(k)'\hat{\boldsymbol{\beta}}(k)\right)^2}{\hat{\boldsymbol{\beta}}(k)'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}(k)}, \quad (29)$$

which is decreasing in k. From this expression it is possible to obtain the experimental value for the global significance test:

$$F_{exp}(k) = \frac{R^2(k)}{\frac{1-R^2(k)}{n-2}}. \quad (30)$$

References

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

Batah FSM, Özkale MR, Gore SD (2009) Combining unbiased ridge and principal component regression estimators. Communications in Statistics-Theory and Methods 38(13):2201–2209

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

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

Chatterjee S, Hadi AS (2006) Regression Analysis by Example, vol 607. John Wiley & Sons

Crouse RH, Jin C, Hanumara RC (1995) Unbiased ridge estimation with prior information and ridge trace. Communications in Statistics-Theory and Methods 24(9):2341–2354

Efron B, Hastie T, Johnstone I, Tibshirani R, et al (2004) Least angle regression. The Annals of statistics 32(2):407–499

Farebrother RW (1984) The restricted least squares estimator and ridge regression. Communications in Statistics-Theory and Methods 13(2):191–196

Frank LE, Friedman JH (1993) A statistical view of some chemometrics regression tools. Technometrics 35(2):109–135

García C, García J, López MM, Salmerón R (2015a) Collinearity: revisiting the variance inflation factor in ridge regression. Journal of Applied Statistics 32(3):648–661

García C, García J, Salmerón R, López MM (2015b) Raise regression: Selection of the raise parameter. Proceeding of International Conference of Data Mining, Las Vegas, EEUU To appear

García CB, García J, Soto J (2010) The raise method: An alternative procedure to estimate the parameters in presence of collinearity. Quality and Quantity 45(2):403–423

Golan A (2008) Information and entropy econometrics: a review and synthesis, vol 2. Now Publishers Inc

Golan A, Judge GG, Miller D (1996) Maximum entropy econometrics: Robust estimation with limited data. Wiley New York

Groß J (2003) Restricted ridge estimation. Statistics & probability letters 65(1):57–64

Halawa A, El Bassiouni M (2000) Tests of regression coefficients under ridge regression models. Journal of Statistical Computation and Simulation 65(1-4):341–356

Hoerl AE, Kennard RW (1970a) Ridge regression: Applications to nonorthogonal problems. Technometrics 12:69–82

Hoerl AE, Kennard RW (1970b) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55–67

Jensen DR, Ramirez DE (2008) Anomalies in the foundations of ridge regression. International Statistical Review 76(1):89–105

Jensen DR, Ramirez DE (2010) Surrogate models in ill-conditioned systems. Journal of Statistical Planning and Inference 140(7):2069–2077

Kaciranlar S, Sakallioglu S, Akdeniz F (1998) Mean squared error comparisons of the modified ridge regression estimator and iiie restricted ridge regression estimator. Communications in Statistics-Theory and Methods 27(1):131–138

Kapat P, Goel PK (2010) Anomalies in the foundations of ridge regression: some clarifications. International Statistical Review 78(2):209–215

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 XQ, Gao F, Yu ZF (2013) Improved ridge estimators in a linear regression model. Journal of Applied Statistics 40(1):209–220

Marquardt DW (1970) Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. Technometrics 12(3):591–612

Massy WF (1965) Principal components regression in exploratory statistical research. Journal of the American Statistical Association 60(309):234–256

Mayer LS, Willke TA (1973) On biased estimation in linear models. Technometrics 15(3):497–508

McDonald GC (2009) Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics 1:93–100

McDonald GC (2010) Tracing ridge regression coefficients. Wiley Interdisciplinary Reviews: Computational Statistics 2:695–703

McDonald GC, Schwing RC (1973) Instabilities of regression estimates relating air pollution to mortality. Technometrics 15(3):463–481

Özkale MR (2009) Comment on ridge estimation to the restricted linear model. Communications in Statistics-Theory and Methods 38(7):1094–1097

Pati KD, Adnan R, Rasheed BA (2015) Using ridge least median squares to estimate the parameter by solving multicollinearity and outliers problems. Modern Applied Science 9(2):191–198

Rousseeuw PJ (1984) Least median of squares regression. Journal of the American Statistical Association 79(388):871–880

Rousseeuw PJ (1985) Multivariate estimation with high breakdown point. Mathematical statistics and applications 8:283–297

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

Salmerón R, García J, López MM, García C (2015) The correct extension of the variance inflator factor to the ridge estimation. Journal of Applied Statistics Under Review

Sarkar N (1992) A new estimator combining the ridge regression and the restricted least squares methods of estimation. Communications in Statistics-Theory and Methods 21(7):1987–2000

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, vol 1, pp 197–206

Stone M, Brooks RJ (1990) Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. Journal of the Royal Statistical Society Series B (Methodological) 52(2):237–269

Swindel BF (1976) Good ridge estimators based on prior information. Communications in Statistics-Theory and Methods 5(11):1065–1075

Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological) 58(1):267–288

Willan AR, Watts DG (1978) Meaningful multicollinearity measures. Technometrics 20(4):407–412

Zhang J, Ibrahim M (2005) A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data. Journal of Applied Statistics 32(6):571–588

Zhong Z, Yang H (2007) Ridge estimation to the restricted linear model. Communications in Statistics-Theory and Methods 36(11):2099–2115