CHOICE OF THE RIDGE FACTOR FROM THE CORRELATION MATRIX DETERMINANT

C. García
Doctoral Program in Economics and Business Sciences
University of Granada (Spain)
garciaclaudia@ugr.es

R. Salmerón and C.B. García
Department of Quantitative Methods for Economics and Business
Faculty of Economics and Business
University of Granada (Spain)

Keywords: multicollinearity, ridge regression, variance inflation factor, mean square error.

ABSTRACT

Ridge regression is the alternative method to ordinary least squares, which is mostly applied when a multiple linear regression model presents a worrying degree of collinearity. A relevant topic in ridge regression is the selection of the ridge parameter, and different proposals have been presented in the scientific literature. Since the ridge estimator is biased, its estimation is normally based on the calculation of the mean square error (MSE) without considering (to the best of our knowledge) whether the proposed value for the ridge parameter really mitigates the collinearity. With this goal and different simulations, this paper proposes to estimate the ridge parameter from the determinant of the matrix of correlation of the data, which verifies that the variance inflation factor (VIF) is lower than the traditionally established threshold. **The possible relation between the VIF and the determinant of the matrix of correlation is also analyzed.** Finally, the contribution is illustrated with three real examples.

1. INTRODUCTION

To study the linear relationships among different variables, there is the ordinary least squares (OLS) estimator, which is the best linear unbiased estimator (BLUE), Duzan and Shariff (2015). According to Kibria and Banik (2016), there can be model specification errors such as the omission of relevant variables, inclusion of unnecessary explanatory variables or highly inter-correlated explanatory variables. The existence of multicollinearity is a serious issue because the estimation may be unstable, and the variance of the estimators may be large compared to the values of the estimated parameters, which can be insignificant or have the wrong sign (Salmerón Gómez et al. (2016)).

The problem of multicollinearity can be mitigated using numerous methods such as ridge regression that is one of the most applied methodology to estimate models with collinearity. This method, which was developed by Hoerl and Kennard (1970a,b), introduces a parameter $k \geq 0$, known as the ridge parameter, in the diagonal of the matrix $\mathbf{X}^t\mathbf{X}$ to avoid the singularity of this matrix (García et al. (2017)). According to Kibria and Banik (2016), Hoerl and Kennard found that there is a nonzero value of k for which the mean squared error (MSE) of the ridge regression estimator is smaller than the variance of the OLS estimator. Many authors have proposed different algorithms to obtain the biasing parameter k (see Kibria and Banik (2016) for a detailed list), but the traditional k value in the literature is the one proposed by Hoerl et al. (1975), see for example Halawa and El Bassiouni (2000). However, these k values do not always mitigate the collinearity ignoring the indications of Marquardt (1970) that the maximum variance inflation factor (VIF) must be lower than 10. The main goal of this paper is to propose a value of the ridge parameter that mitigates the collinearity in a specified model. To that end, we will follow the work of García et al. (2017), who analyze the relation of the ridge factor (k) with the squared coefficient of correlation (ρ^2) in a model of two standardized independent variables, and conclude that this relation is linear. In that case, it is verified that $\rho^2 = 1 - det(\mathbf{R})$, where $det(\mathbf{R})$ denotes the determinant of the matrix of correlations, $\mathbf{R}$, of the independent variables. To obtain results in line with this particular case, this paper will work with $1 - det(\mathbf{R})$.

To detect the multicollinearity in a specific model, the variance inflation factor (VIF) is widely used. Thus, it is known that VIF> 10 implies collinearity. However, there are other options such as finding the value of the determinant of the correlation matrix.

Thus, taking into account that the VIFs can be found on the main diagonal of the inverse of the matrix of correlations, $\mathbf{R}$, it is possible to obtain the following relation between the VIF and the determinant of the correlation matrix (see, for example, Fox and Monette (1992)):

$$VIF_i = \frac{det(\mathbf{R}_{ii})}{det(\mathbf{R})}, \quad \forall i, \tag{1}$$

where $\mathbf{R}_{ii}$ is the matrix obtained after eliminating the file and column i of $\mathbf{R}$. In this case, from (1) is obtained that $det(\mathbf{R}) = det(\mathbf{R}_{ii}) \cdot VIF_i^{-1}$, and consequently, a $VIF_{max} > 10$ can be expressed as $det(\mathbf{R}) < 0.1 \cdot det(\mathbf{R}_{ii})$.

Finally, since the determinant of a matrix of correlation is between 0 and 1, it is verified that:

$$VIF_{max} > 10 \Rightarrow det(\mathbf{R}) < 0.1. \tag{2}$$

For three explanatory variables (two independent variables and a constant), it can be demonstrated that $det(\mathbf{R}) = 1 - \rho^2 = \frac{1}{VIF}$. Then, it is evident that if $VIF > 10$ then $det(\mathbf{R}) < 0.1$. That is, if the determinant of the correlation matrix is less than 0.1, multicollinearity is likely present in the specified model.

The second principal objective of this work is to generalize the expression (2) taking into account the number of variables and observations of the multiple linear regression. Before presenting the contribution of this paper, we first present the ridge regression (RR) in Section 2. Here, we also briefly specify the MSE and VIF for the RR, which have different formulas from those in the traditional OLS estimation (subsections 2.1 and 2.2). In subsection 2.3, we present some criteria to select the ridge factor. Then, Section 3 presents a Monte Carlo simulation that helps us to obtain our k values (subsection 3.1). **The relation between bias and variance for the different values of k is analyzed in subsections 3.2 and 3.3**. Finally, the full contribution is shown with three empirical examples in Section 4, and the conclusions and future lines of research are presented in Section 5.

2. RIDGE REGRESSION

Hoerl and Kennard (1970b) proposed an estimation process to mitigate the collinearity problem, which appears in a multiple regression model when the explanatory or independent variables are not orthogonal. This process consists of modifying the matrix $\mathbf{X}^t\mathbf{X}$ by adding a positive and tiny value in its diagonal. In other terms, suppose the following multiple linear regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad (3)$$

where $\mathbf{X}$ is nxp (range p), $\boldsymbol{\beta}$ is $px1$, and the random perturbance $\mathbf{u}$ is $nx1$ and spherical ($E[\mathbf{u}] = \mathbf{0}$; $Var(\mathbf{u}) = \sigma^2\mathbf{I}$, $\mathbf{0}$ a zero vector and $\mathbf{I}$ being the identity matrix).

To avoid the instability of the ordinary least squares (OLS) estimator, which is determined by:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y},$$

they suggest to use the ridge estimator instead of the OLS estimator[1]. The ridge estimator can be expressed as:

$$\widehat{\boldsymbol{\beta}}(k) = (\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Y}, \qquad (4)$$

where k is known as the ridge factor, and $k \geq 0$.

Considering $\mathbf{Z}_k = (\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{X}$, we can verify that:

- $\widehat{\boldsymbol{\beta}}(k) = \mathbf{Z}_k\widehat{\boldsymbol{\beta}}$, $\forall k \geq 0$.
- $\widehat{\boldsymbol{\beta}}(k)$ is a biased estimator of $\boldsymbol{\beta}$, except for $k = 0$: $E[\widehat{\boldsymbol{\beta}}(k)] = \mathbf{Z}_k E[\widehat{\boldsymbol{\beta}}] = \mathbf{Z}_k\boldsymbol{\beta} \neq \boldsymbol{\beta}$.
- $Var(\widehat{\boldsymbol{\beta}}(k)) = \mathbf{Z}_k Var(\widehat{\boldsymbol{\beta}})\mathbf{Z}_k^t = \sigma^2(\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^t\mathbf{X}(\mathbf{X}^t\mathbf{X} + k\mathbf{I})^{-1}$.

Marquardt (1970) said that (4) minimizes the sum of squared residuals (SSR), which is

[1] The condition number (CN) of the matrix $\mathbf{X}^t\mathbf{X}+k\mathbf{I}$ ($k > 0$) is less than the CN for matrix $\mathbf{X}^t\mathbf{X}$ (Casella (1985)). Its inverse is sounder.

expressed by:

$$\begin{aligned}\text{SSR}(k) &= \mathbf{e}(k)^t\mathbf{e}(k) = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(k))^t(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(k)) \\ &= \text{SSR} + k^2\widehat{\boldsymbol{\beta}}(k)^t(\mathbf{X}^t\mathbf{X})^{-1}\widehat{\boldsymbol{\beta}}(k),\end{aligned}$$

where SSR is the sum of squared residuals of model (3). SSR(k) increases monotonically in k, and because $\widehat{\boldsymbol{\beta}}(k)^t(\mathbf{X}^t\mathbf{X})^{-1}\widehat{\boldsymbol{\beta}}(k)$ is a positive-definite cuadratic form, it can be demonstrated that $\text{SSR}(k) > \text{SSR}$ for $k > 0$.

2.1. MEAN SQUARE ERROR IN RIDGE REGRESSION

The estimated parameters obtained from (4) are biased estimators of $\boldsymbol{\beta}$ if $k \neq 0$; thus, it is interesting to calculate their mean squared error (MSE). For an estimator $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, the MSE is obtained using the following expression:

$$\text{MSE}(\widetilde{\boldsymbol{\beta}}) = E[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^t(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})] = tr(Var(\widetilde{\boldsymbol{\beta}})) + (E[\widetilde{\boldsymbol{\beta}}] - \boldsymbol{\beta})^t(E[\widetilde{\boldsymbol{\beta}}] - \boldsymbol{\beta}). \quad (5)$$

It is always positive, and a smaller value is better.

Starting from expression (3), the estimation in (4) can be expressed as $\widehat{\boldsymbol{\beta}}(k) = \mathbf{Z}_k \cdot \widehat{\boldsymbol{\beta}}$, thus following McDonald (2010)[2], the variance and its trace can be expressed as:

$$\begin{aligned}Var\left(\widehat{\boldsymbol{\beta}}(k)\right) &= \sigma^2 \cdot \boldsymbol{\Gamma} \cdot \mathbf{D}_{\frac{\lambda}{(\lambda+k)^2}} \cdot \boldsymbol{\Gamma}^t. \\ tr\left(Var\left(\widehat{\boldsymbol{\beta}}(k)\right)\right) &= tr(\sigma^2 \cdot \boldsymbol{\Gamma} \cdot \mathbf{D}_{\frac{\lambda}{(\lambda+k)^2}} \cdot \boldsymbol{\Gamma}^t) = \sigma^2 tr(\mathbf{D}_{\frac{\lambda}{(\lambda+k)^2}} \cdot \boldsymbol{\Gamma}^t \cdot \boldsymbol{\Gamma}) \\ &= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2}. \quad (6)\end{aligned}$$

Meanwhile, $E[\widehat{\boldsymbol{\beta}}(k)] = \mathbf{Z}_k \cdot \boldsymbol{\beta}$; then:

$$\begin{aligned}(E[\widehat{\boldsymbol{\beta}}(k)] - \boldsymbol{\beta})^t(E[\widehat{\boldsymbol{\beta}}(k)] - \boldsymbol{\beta}) &= \boldsymbol{\beta}^t(\mathbf{Z}_k - \mathbf{I})^t(\mathbf{Z}_k - \mathbf{I})\boldsymbol{\beta} \\ &= \boldsymbol{\beta}^t \cdot \boldsymbol{\Gamma} \cdot \mathbf{D}_{\frac{k^2}{(\lambda+k)^2}} \cdot \boldsymbol{\Gamma}^t \cdot \boldsymbol{\beta} = k^2 \cdot \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2}, \quad (7)\end{aligned}$$

[2]Note that $\mathbf{X}^t\mathbf{X} = \boldsymbol{\Gamma} \cdot \mathbf{D}_\lambda \cdot \boldsymbol{\Gamma}^t$, where $\boldsymbol{\Gamma}$ are the eigenvector matrix of $\mathbf{X}^t\mathbf{X}$; thus, $\boldsymbol{\Gamma}^t = \boldsymbol{\Gamma}^{-1}$, and $\lambda_1, \ldots, \lambda_p$ are its eigenvalues, being $\mathbf{D}_\lambda = diag(\lambda_1, \ldots, \lambda_p)$.

where α_j, for $j = 1, \ldots, p$ are the elements of vector $\boldsymbol{\alpha} = \boldsymbol{\Gamma}^t \cdot \boldsymbol{\beta}$.

From (6) and (7), the MSE in the ridge regression is:

$$\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) = \sigma^2 \cdot \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \cdot \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2} = \sum_{j=1}^{p} \frac{\sigma^2 \cdot \lambda_j + k^2 \cdot \alpha_j^2}{(\lambda_j + k)^2}. \quad (8)$$

Then, we can obtain the value of MSE from the estimated values of σ^2 and $\boldsymbol{\beta}$.

In expression (8), Hoerl and Kennard (1970b) distinguished between the variance of the estimate $\widehat{\boldsymbol{\beta}}(k)$, denoted as $\gamma_1(k)$, and the squared bias of the estimated $\widehat{\boldsymbol{\beta}}(k)$, denoted as $\gamma_2(k)$. That is to say:

$$\textbf{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) = \gamma_1(k) + \gamma_2(k),$$

It is verified that $\gamma_1(k)$ is decreasing in k (converges to zero) and $\gamma_2(k)$ is increasing. Thus, MSE $\left(\widehat{\boldsymbol{\beta}}(k)\right)$ presents a horizonal asymptote given by:

$$\lim_{k \to +\infty} \textbf{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) = \sum_{j=1}^{p} \alpha_j^2 = \boldsymbol{\alpha}^t \boldsymbol{\alpha} = \boldsymbol{\beta}^t \boldsymbol{\beta}.$$

Indeed, the asymptotic behaviour of $\gamma_1(k)$ leads to a value of k from which the MSE $\left(\widehat{\boldsymbol{\beta}}(k)\right)$ depends on $\gamma_2(k)$ in a high percentage.

Finally, by deriving the expression (8) in terms of k, we obtain:

$$\frac{\partial}{\partial k} \text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) = 2 \cdot \sum_{j=1}^{p} \frac{k \cdot \alpha_j - \sigma^2}{(\lambda_j + k)^3} \cdot \lambda_j. \quad (9)$$

From (9), we obtain that MSE $\left(\widehat{\boldsymbol{\beta}}(k)\right)$ is increasing if $k \geq \max_{j=1,\ldots,p} \left\{\frac{\sigma^2}{\alpha_j^2}\right\}$, and decreasing if $k \leq \min_{j=1,\ldots,p} \left\{\frac{\sigma^2}{\alpha_j^2}\right\}$. However, it is not possible to algebraically obtain the turning points; thus, it is not possible to analyze the monotony of MSE $\left(\widehat{\boldsymbol{\beta}}(k)\right)$ **due to the different monotony of $\gamma_1(k)$ and $\gamma_2(k)$**. Despite these difficulties, this expression has been used to estimate the ridge parameter (e.g., Hoerl and Kennard (1970b) or Hoerl et al. (1975)).

2.1.1 BIAS IN RIDGE REGRESSION

From expression (8), it is verified that

$$\textbf{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) = \sigma^2 \cdot \sum_{j=1}^{p} \frac{1}{\lambda_j} + k \cdot \sum_{j=1}^{p} \frac{(\alpha_j^2 \cdot \lambda_j - \sigma^2) \cdot k - 2 \cdot \lambda_j \cdot \sigma^2}{(\lambda_j + k)^2 \cdot \lambda_j}$$
$$= \textbf{MSE}\left(\widehat{\boldsymbol{\beta}}\right) + bias(k), \qquad (10)$$

where $bias(0) = 0$**,** $bias(k) < 0$ **if** $k < min_{j=1,\dots,p}\left\{\frac{2\cdot\lambda_j\cdot\sigma^2}{\alpha_j^2\cdot\lambda_j-\sigma^2}\right\}$ **and** $bias(k) > 0$ **for** $k > max_{j=1,\dots,p}\left\{\frac{2\cdot\lambda_j\cdot\sigma^2}{\alpha_j^2\cdot\lambda_j-\sigma^2}\right\}$**. For** $min_{j=1,\dots,p}\left\{\frac{2\cdot\lambda_j\cdot\sigma^2}{\alpha_j^2\cdot\lambda_j-\sigma^2}\right\} < k < max_{j=1,\dots,p}\left\{\frac{2\cdot\lambda_j\cdot\sigma^2}{\alpha_j^2\cdot\lambda_j-\sigma^2}\right\}$ **is not possible to conclude.**

Furthermore, if $bias(k) < 0$ then $\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) < \text{MSE}\left(\widehat{\boldsymbol{\beta}}\right)$ and if $bias(k) > 0$ then $\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right) > \text{MSE}\left(\widehat{\boldsymbol{\beta}}\right)$.

Note that $\frac{\partial}{\partial k} bias(k) = \frac{\partial}{\partial k}\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right)$.

2.2. VARIANCE INFLATION FACTOR IN RIDGE REGRESSION

After the ridge regression is applied, it is necessary to test whether the initial collinearity problem has been mitigated. As we have stated in the Introduction, to detect the problem, the variance inflation factor (VIF) is widely used.

García et al. (2015) show that many studies have some errors in the calculation of VIFs associated to ridge regression. García et al. (2016) proposed an alternative methodology to properly measure it. This proposal will be used in this work, considering that the multicollinearity has been mitigated when the value of the VIF **associated to ridge regression**, $\text{VIF}(k)$, is less than 10.

2.3. SOME CRITERIA FOR SELECTING THE RIDGE FACTOR

Hoerl and Kennard (1970b) proposed the use of the trace of ridge estimators, which consists of representing $\boldsymbol{\beta}_i(k)$ for $k \geq 0$. As $\boldsymbol{\beta}(k)^t\boldsymbol{\beta}(k)$ decreases in k (see Marquardt (1970)), it can be assumed that there is a set of values of ridge factor k, which stabilizes the estimation. The appropriate k is among these values.

In parallel, the most used criterion in choosing the k parameter is the following (Hoerl et al. (1975)):

$$k_{Hoerl} = p \cdot \frac{\widehat{\sigma}^2}{\widehat{\boldsymbol{\beta}}^t \widehat{\boldsymbol{\beta}}}. \tag{11}$$

The likelihood of giving estimators with smaller MSEs than the OLS estimators for this k is greater than 0.5.

However, this is not the only criterion in selecting k. In the literature, many authors have developed different expressions to estimate the ridge parameter. Starting from Kibria and Banik (2016), there are 25 criteria in addition to the previous one. Tables 1 and 2 present these different criteria.

At this point, it is important to note that with all of these methods, the multicollinearity is not guaranteed to be mitigated. Thus, following the proposal from Marquardt (1970) and García et al. (2017) and considering the previous subsection, we have simulated a data set to obtain the smallest possible value of k that makes the VIFs of the ridge regression, VIF(k), smaller than 10.

3. MONTE CARLO SIMULATION

3.1. Relation between k and $det(\mathbf{R})$ and between $det(\mathbf{R})$ and VIF

In this subsection, we simulate values from the following:

$$\mathbf{X}_i = \sqrt{1-\xi^2} \cdot \mathbf{Z}_i + \xi \cdot \mathbf{Z}_p,$$

where $i = 2, \ldots, p$ with $p = 3, 4, 5$, $\mathbf{Z}_i \sim N(\mu, \sigma)$ with $\mu \in \{1, 2, 3, 4, 5\}$ and $\sigma^2 \in \{0.1, 0.2, 0.3, 0.4, \ldots, 3\}$, $\xi \in \{0.8, 0.81, 0.82, 0.83, \ldots, 0.99\}$ and $n \in \{15, 20, 25, 30, \ldots, 200\}$. This method to generate independent variables with different grades of collinearity (ξ is specified so that the correlation between any two independent variables is given by ξ^2) has been applied, e.g., by Gibbons (1981), McDonald and Galarneau (1975), Kibria (2003) or Salmerón et al. (2018).

The matrix $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \ \ldots \ \mathbf{X}_p]$ is constructed, where $\mathbf{X}_1$ is a vector with ones (which represents the independent term in model (3)). Then, for every variable, the minimum value

Table 1: Ridge parameter: different criteria for choosing k

Reference	Notation	Formula
Hoerl and Kennard (1970b)	k_{hk}	$\frac{\widehat{\sigma}^2}{\widehat{\beta}^2_{\max}}$
Hoerl et al. (1975)	k_{Hoerl}	$p\frac{\widehat{\sigma}^2}{\widehat{\beta}^t\widehat{\beta}}$
Hocking et al. (1976)	k_h	$\widehat{\sigma}^2\frac{\sum_{i=1}^p\left(\lambda_i\widehat{\beta}_i\right)^2}{\left(\sum_{i=1}^p\lambda_i\widehat{\beta}_i^2\right)^2}$
Lawless and Wang (1976)	k_{lw}	$\frac{p\widehat{\sigma}^2}{\widehat{\beta}^t\mathbf{X}^t\mathbf{X}\widehat{\beta}}$
Nomura (1988)	k_n	$\frac{p\widehat{\sigma}^2}{\sum_{i=1}^p\left(\left(\widehat{\beta}_i^2\right)\Big/\left(2+\lambda_i\left(\frac{\widehat{\beta}_i^2}{\widehat{\sigma}^2}\right)^{1/2}\right)\right)}$
Kibria (2003)	k_k^{am}	$\frac{1}{p}\sum_{i=1}^p\frac{\widehat{\sigma}^2}{\widehat{\beta}_i^2}$
	k_k^{gm}	$\frac{\widehat{\sigma}^2}{\left(\prod_{i=1}^p\widehat{\beta}_i^2\right)^{1/p}}$
	k_k^{med}	$\text{med}\left(\frac{\widehat{\sigma}^2}{\widehat{\beta}_i^2}\right)$
Khalaf and Shukur (2005)	k_{ks}	$\frac{\lambda_{\max}\widehat{\sigma}^2}{(n-p)\,\widehat{\sigma}^2+\lambda_{\max}\widehat{\beta}^2_{\max}}$
	k_{ks}^{am}	$\frac{1}{p}\sum_{i=1}^p\frac{\lambda_i\widehat{\sigma}^2}{(n-p)\,\widehat{\sigma}^2+\lambda_i\widehat{\beta}_i^2}$
	k_{ks}^{max}	$\max\left(\frac{\lambda_i\widehat{\sigma}^2}{(n-p)\,\widehat{\sigma}^2+\lambda_i\widehat{\beta}_i^2}\right)$
	k_{ks}^{med}	$\text{med}\left(\frac{\lambda_i\widehat{\sigma}^2}{(n-p)\,\widehat{\sigma}^2+\lambda_i\widehat{\beta}_i^2}\right)$

Table 2: Ridge parameter: different criteria for choosing k *(cont.)*

Reference	Notation	Formula
Muniz and Kibria (2009)	k_{mk_1}	$\max\left(\frac{1}{\sqrt{\widehat{\sigma}^2/\widehat{\beta}_i^2}}\right)$
	k_{mk_2}	$\max\left(\sqrt{\frac{\widehat{\sigma}^2}{\widehat{\beta}_i^2}}\right)$
	k_{mk_3}	$\left(\prod_{i=1}^{p}\frac{1}{\sqrt{\widehat{\sigma}^2/\widehat{\beta}_i^2}}\right)^{1/p}$
	k_{mk_4}	$\left(\prod_{i=1}^{p}\frac{\widehat{\sigma}^2}{\widehat{\beta}_i^2}\right)^{1/p}$
	k_{mk_5}	$\text{med}\left(\frac{1}{\sqrt{\widehat{\sigma}^2/\widehat{\beta}_i^2}}\right)$
	k_{mk_6}	$\text{med}\left(\sqrt{\frac{\widehat{\sigma}^2}{\widehat{\beta}_i^2}}\right)$
Dorugade and Kashid (2010)	k_{dk}	$\max\left(0, k_{Hoerl} - \frac{1}{n\left(\text{VIF}_i\right)_{\max}}\right)$
Khalaf (2012)	k_f	$k_{hk} + \frac{2}{\left(\lambda_{\max} + \lambda_{\min}\right)^t}$
Muniz et al. (2012)	k_{m_1}	$\max\left(\frac{(n-p)\,\widehat{\sigma}^2 + \lambda_{\max}\widehat{\beta}_i^2}{\lambda_{\max}\widehat{\sigma}^2}\right)$
	k_{m_2}	$\max\left(\frac{\lambda_{\max}\widehat{\sigma}^2}{(n-p)\,\widehat{\sigma}^2 + \lambda_{\max}\widehat{\beta}_i^2}\right)$
	k_{m_3}	$\left(\prod_{i=1}^{p}\frac{(n-p)\,\widehat{\sigma}^2 + \lambda_{\max}\widehat{\beta}_i^2}{\lambda_{\max}\widehat{\sigma}^2}\right)^{1/p}$
	k_{m_4}	$\left(\prod_{i=1}^{p}\frac{\lambda_{\max}\widehat{\sigma}^2}{(n-p)\,\widehat{\sigma}^2 + \lambda_{\max}\widehat{\beta}_i^2}\right)^{1/p}$
	k_{m_5}	$\text{med}\left(\frac{(n-p)\,\widehat{\sigma}^2 + \lambda_{\max}\widehat{\beta}_i^2}{\lambda_{\max}\widehat{\sigma}^2}\right)$

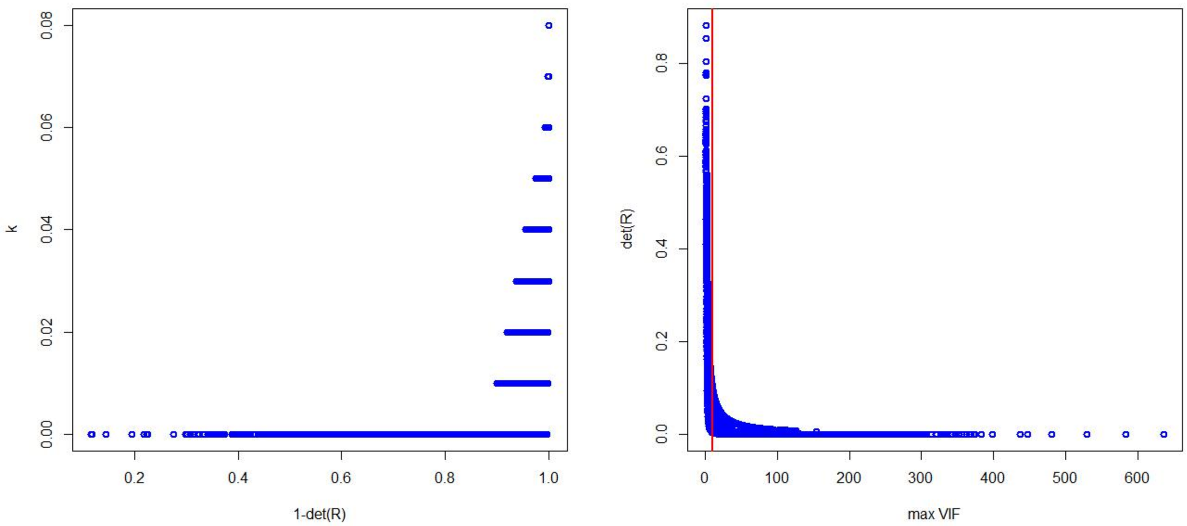

Figure 1: Graphic representation of k according to $1 - det(\mathbf{R})$ (right hand) and $det(\mathbf{R})$ according to VIF_{max} (left hand)

of k that makes the VIF associated with the ridge regression less than 10 and the determinant of the correlation matrix of $\mathbf{X}$ are calculated.

Since this calculation is repeated three times for each value of p, 1026000 values are obtained, which are presented in Figure 1. Figure 2 shows k according to $1 - det(\mathbf{R})$ after deleting the values of $k = 0$ (when the collinearity in the model it is not an issue of concern)[3].

This graphical representations suggest the following relationships:

$$k_{exp} = \beta \cdot e^{1-det(\mathbf{R})} + u_t, \tag{12}$$

$$k_{lineal} = \gamma \cdot (1 - det(\mathbf{R})) + u_t, \tag{13}$$

$$k_{sq} = \delta_1 \cdot (1 - det(\mathbf{R}))^2 + \delta_2 \cdot (1 - det(\mathbf{R})) + u_t, \tag{14}$$

$$det(\mathbf{R}) = \alpha \cdot \frac{1}{\text{VIF}_{\text{max}}} + v_t, \tag{15}$$

After adding the sample size (n) and number of variables (p), these estimations are presented

[3]As we have said in the Introduction, we use $1 - det(\mathbf{R})$ because for $p = 3$, this value coincides with the value of ρ^2. Then, the results are comparable to the findings in García et al. (2017)

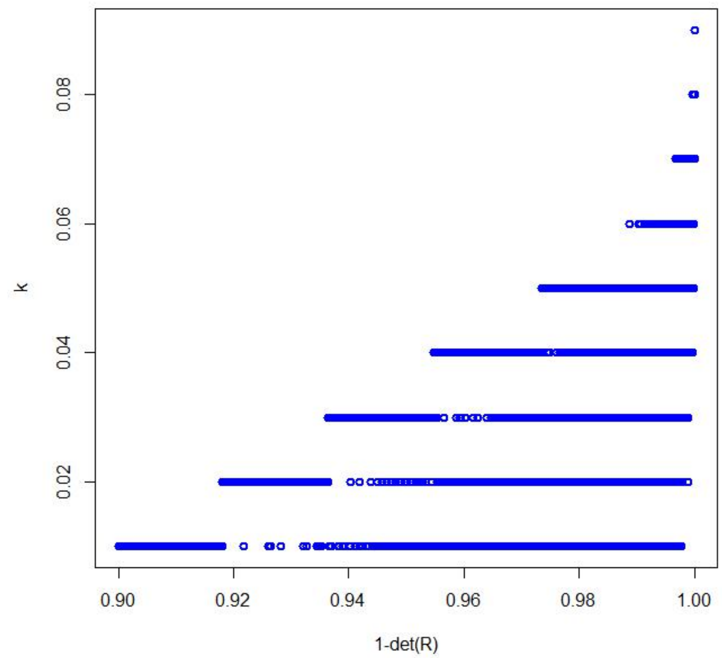

Figure 2: Graphic representation of k according to $1 - det(\mathbf{R})$ after deleting values $k = 0$

Table 3: Estimation of model (12) for $k \neq 0$

Variable	Estimation	Typical deviation	p-values
$e^{1-det(\mathbf{R})}$	0.006639	0.00007381	$< 2 \cdot 10^{-16}$
n	-0.00001241	0.0000005299	$< 2 \cdot 10^{-16}$
p	0.005745	0.00004296	$< 2 \cdot 10^{-16}$
R^2	0.807	$F_{3,479191}$	667700

Table 4: Estimation of model (13) for $k \neq 0$

Variable	Estimation	Typical deviation	p-values
$1 - det(\mathbf{R})$	0.01837	0.0002013	$< 2 \cdot 10^{-16}$
n	-0.00001262	0.0000005298	$< 2 \cdot 10^{-16}$
p	0.005678	0.00004309	$< 2 \cdot 10^{-16}$
R^2	0.8071	$F_{3,479191}$	668100

in Tables 3-6.

Using the information in these tables, Figure 3 shows the estimations obtained for k considering that $det(R) \in \{0, 0.1, 0.2, \dots, 1\}$, $n \in \{15, 20, 25, \dots, 200\}$ and $p \in \{3, 4, 5, \dots, 10\}$. Note that the estimations for k_{exp} and k_{lineal} are notably similar, and k_{sq} can take negative values. Analogously, Table 7 provides the maximum and minimum values obtained for k_{exp}, k_{lineal} and k_{sq}. These results seem to indicate that the estimations of k obtained from the quadratic relation may be not adequate in some cases.

Moreover, to test whether the coefficient of $\frac{1}{\text{VIF}_{\text{max}}}$ is significantly lower than 1, we obtain that the null hypothesis of equality is not rejected because

$$t_{exp} = \frac{|1.013 - 1|}{0.0003066} = 42.40052 > 1.959966 = t_{1025997}(0.975).$$

With these results, it is possible to assume that our initial assumption is maintained: VIF > 10 implies that $det(\mathbf{R}) < 0.1$. Furthermore, with the results in Table 6, if we consider the sample size and number of independent variables, we can obtain Table 8. This last table

Table 5: Estimation of model (14) for $k \neq 0$

Variable	Estimation	Typical deviation	p-values
$(1 - det(\mathbf{R}))^2$	0.7922	0.001482	$< 2 \cdot 10^{-16}$
$1 - det(\mathbf{R})$	-0.6901	0.001335	$< 2 \cdot 10^{-16}$
n	-0.000007567	0.0000004195	$< 2 \cdot 10^{-16}$
p	-0.01081	0.00004599	$< 2 \cdot 10^{-16}$
R^2	0.8791	$F_{4,479190}$	871200

Table 6: Estimation of model (15)

Variable	Estimation	Typical deviation	p-values
$\frac{1}{\text{VIF}_{\max}}$	1.013	0.0003066	$< 2 \cdot 10^{-16}$
n	0.00008626	0.000000578	$< 2 \cdot 10^{-16}$
p	-0.01384	0.00001732	$< 2 \cdot 10^{-16}$
R^2	0.9411	$F_{3,1025997}$	5463000

Table 7: Maximum and minimum values of k_{exp}, k_{lineal} and k_{sq} for $det(R) \in \{0, 0.1, 0.2, \ldots, 1\}$, $n \in \{15, 20, 25, \ldots, 200\}$ and $p \in \{3, 4, 5, \ldots, 10\}$

	Minimum	Maximum
k_{exp}	0.021392	0.07531
k_{lineal}	0.0145	0.07496
k_{sq}	-0.2598	0.0695

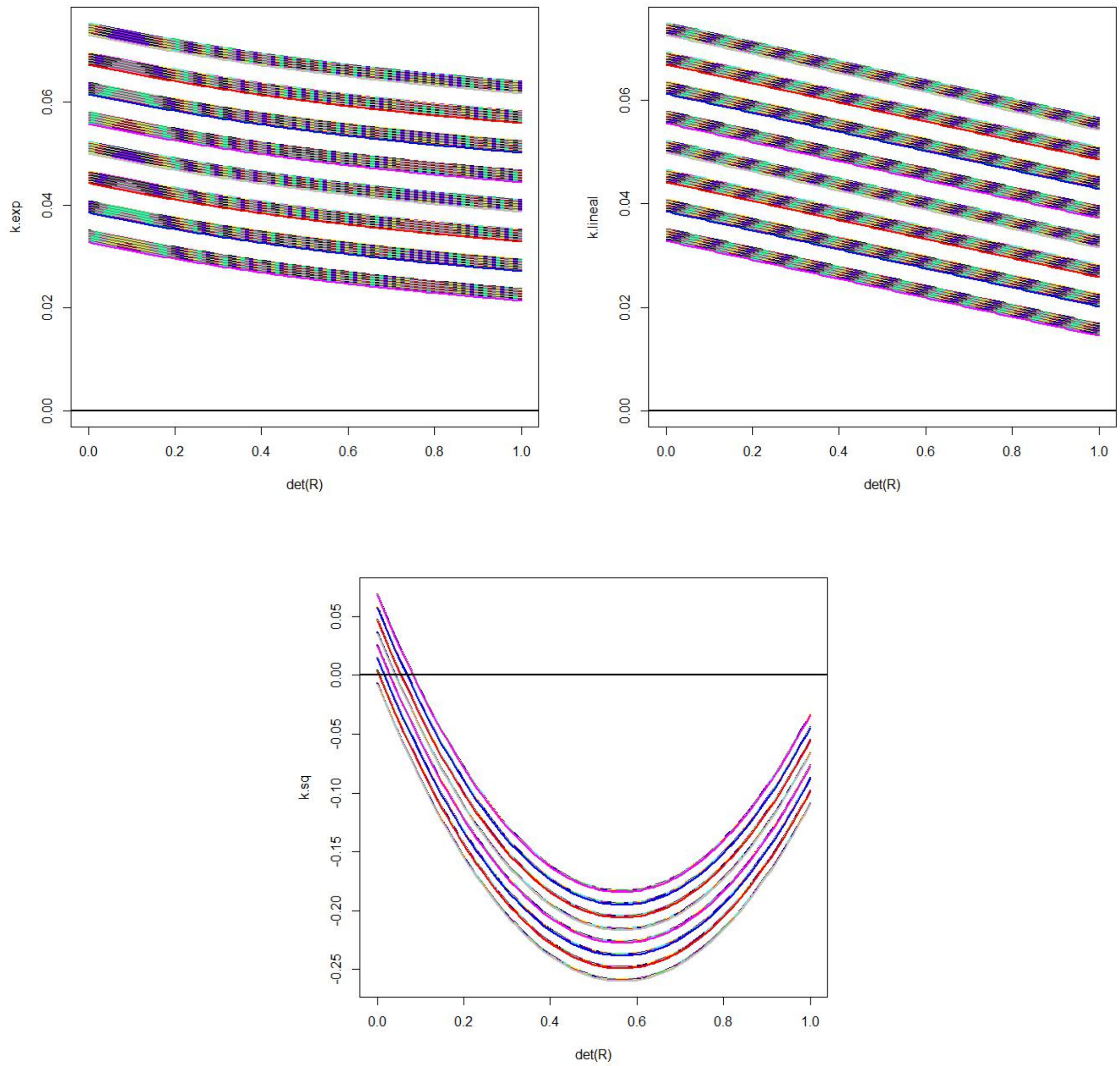

Figure 3: Graphic representation of k_{exp}, k_{lineal} and k_{sq} (top to bottom) for $det(R) \in \{0, 0.1, 0.2, \ldots, 1\}$, $n \in \{15, 20, 25, \ldots, 200\}$ and $p \in \{3, 4, 5, \ldots, 10\}$

Table 8: Values of $det(R)$ when VIF = 10 for different sample sizes and numbers of independent variables.

$p \setminus n$	15	45	75	105	135	165	195
2	0.0749139	0.0775017	0.0800895	0.0826773	0.0852651	0.0878529	0.0904407
3	0.0610739	0.0636617	0.0662495	0.0688373	0.0714251	0.0740129	0.0766007
4	0.0472339	0.0498217	0.0524095	0.0549973	0.0575851	0.0601729	0.0627607
5	0.0333939	0.0359817	0.0385695	0.0411573	0.0437451	0.0463329	0.0489207
6	0.0195539	0.0221417	0.0247295	0.0273173	0.0299051	0.0324929	0.0350807

shows that the values of $det(R)$ diminish when we have more variables and fewer observations, which is consistent with Table 6, where the estimated parameter of n is positive, and the sign of p parameter is negative.

The second goal of this paper is achieved from this table. Thus, it will be possible to conclude that the collinearity is worrying if the model presents a $det(\mathbf{R})$ lower that the values presented in this table (values of $det(\mathbf{R})$ that are not collected in Table 8 can be easily obtained from estimations provided in Table 6).

3.2. Analysis of the squared bias of the estimated $\widehat{\boldsymbol{\beta}}(k)$

In this subsection, 15200 simulations have been carried out by following subsection 3.1 but considering that $\xi \in \{0.965, 0.967, 0.968, \dots, 0.999\}$ to ensure worrying collinearity. Indeed, the following model has been generated:

$$\mathbf{Y} = \beta_0 + \beta_1 \cdot \mathbf{Z}_1 + \cdots + \beta_p \cdot \mathbf{Z}_p + \mathbf{u},$$

where $\mathbf{u} \sim N(0, 1)$ and β_i with $i = 0, 1, \dots, p$ is obtained randomly from the set $\{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$.

The percentage of MSE caused by the component $\gamma_2(k)$ has been calculated for each simulation for the estimators of k proposed in this paper (k_{exp}, k_{lineal} and

Table 9: Percentage of cases where $a(k^*) - a(k_{Hoerl}) < 0$

k^*	p=3	p=4	p=5
k_{exp}	21.483%	5.178%	1.339%
k_{lineal}	21.459%	5.172%	1.3408%
k_{sq}	25.723%	3.467%	1.263%

k_{sq}) in the following way:

$$\delta(k) = \frac{\gamma_2(k)}{\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right)} \cdot 100, \tag{16}$$

The value proposed by Hoerl et al. (1975), k_{Hoerl} will be considered the reference value to analyze the behaviour in relative terms of the squared bias of the estimated $\widehat{\boldsymbol{\beta}}(k)$ for the estimations of k given by k_{exp}, k_{lineal} and k_{sq}.

Table 9 presents for different values of p, the percentage of cases where the difference $a(k^*) - a(k_{Hoerl})$ is negative, it is to say, where the percentage of variability of the MSE caused by $\gamma_2(k)$ is lower for values k_{exp}, k_{lineal} and k_{sq}. Note that these percentages diminish as p increases, which suggests that as the number of independent variables in the model increases, so does the distance between these estimates.

Table 10 presents the mean and the standard deviation for the difference when it is positive, it is to say, when the percentage of MSE caused by $\gamma_2(k)$ is higher in the estimations proposed in this paper. In this case, the percentage increases between 32% and 46% approximately depending on the value of p. It is to say, looking for an estimator of k to mitigate the approximate multicollinearity supposes an increase in relative terms of the squared bias of the estimated $\widehat{\boldsymbol{\beta}}(k)$ of approximately 46% in the worst case analyzed.

Table 10: Expected value of $a(k^*) - a(k_{Hoerl})$ when the difference is positive (the standard deviation in parentheses)

k^*	p=3	p=4	p=5
k_{exp}	32.2909% (31.416%)	38083% (32.886%)	45.177% (33.3522%)
k_{lineal}	32.88% (31.41%)	38.10% (32.889%)	45.158% (33.522%)
k_{sq}	39.245% (35.383%)	46.534% (33.512%)	46.109% (33.515%)

Table 11: Analysis of $bias(k)$ for $p = 3$

k^*	$k^* < min_{j=1,\dots,p} a_j$	$min_{j=1,\dots,p} a_j < k^* < max_{j=1,\dots,p} a_j$	$max_{j=1,\dots,p} a_j < k^*$
k_{exp}	16.835%	82.48%	0.68%
k_{lineal}	16.822%	82.493%	0.684%
k_{sq}	25.907%	72.598%	1.49%

3.3. Analysis of $bias(k)$

In this subsection, 15200 simulations have been carried out by following subsection 3.2. Values of k_{exp}, k_{lineal}, k_{sq} and $a_j = \frac{2 \cdot \lambda_j \cdot \sigma^2}{\alpha_j^2 \cdot \lambda_j - \sigma^2}$ (where σ^2 is replaced by its estimates) have been calculated for each simulation. Tables 11 to 13 show the results for $p = 3, 4, 5$, respectively. Note that it is not possible to conclude in a high percentage of cases and that this percentage increases as p increases.

It was also calculated the value of k, $k_{critical}$, from which MSE $\left(\widehat{\boldsymbol{\beta}}(k)\right)$ changes from being lower than MSE $\left(\widehat{\boldsymbol{\beta}}\right)$ to being higher. Table 14 shows the percentage of cases where k_{exp}, k_{lineal} and k_{sq} are lower than $k_{critical}$, it is to say, the

Table 12: Analysis of $bias(k)$ for $p = 4$

k^*	$k^* < min_{j=1,\dots,p} a_j$	$min_{j=1,\dots,p} a_j < k^* < max_{j=1,\dots,p} a_j$	$max_{j=1,\dots,p} a_j < k^*$
k_{exp}	2.48%	96.532%	0.986%
k_{lineal}	2.467%	96.546%	0.986%
k_{sq}	1.611%	95.96%	2.427%

Table 13: Analysis of $bias(k)$ for $p = 5$

k^*	$k^* < min_{j=1,\dots,p} a_j$	$min_{j=1,\dots,p} a_j < k^* < max_{j=1,\dots,p} a_j$	$max_{j=1,\dots,p} a_j < k^*$
k_{exp}	0.519%	97.986%	1.49%
k_{lineal}	0.519%	97.986%	1.493%
k_{sq}	0.473%	97.861%	1.664%

Table 14: Percentage of cases where $k^* < k_{critical}$, it is to say, $\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k^*)\right) < \text{MSE}\left(\widehat{\boldsymbol{\beta}}\right)$

k^*	p=3	p=4	p=5
k_{exp}	27.355%	19.69%	14.75%
k_{lineal}	27.355%	19.657%	14.75%
k_{sq}	37.934%	14.151%	14.75%

percentage of cases where $bias(k^*) < 0$ and then $\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k^*)\right) < \text{MSE}\left(\widehat{\boldsymbol{\beta}}\right)$ with $k^* = k_{exp}, k_{lineal}, k_{sq}$.

Since values k_{exp}, k_{lineal} and k_{sq} are calculated to obtain an associated VIF lower than 10, these results have to be interpreted as the percentage of cases where the MSE associated to k_{exp}, k_{lineal} and k_{sq} is lower than the MSE obtained by OLS when the collinearity is mitigated.

By comparing these values with the values provided by Hoerl et al. (1975), who stated that *the use of the ridge estimator with biasing parameter* $k_a = p \cdot \frac{\widehat{\sigma}^2}{\widehat{\boldsymbol{\beta}}^t \widehat{\boldsymbol{\beta}}}$ *has a probability greater than 0.5 of producing estimates with a smaller mean square error than least squares*, it is noted that the mentioned probability diminishes when the mitigation of collinearity is imposed as a condition. Further, Table 14 shows that the percentages diminish as p increases. This conclusion contrasts with the results obtained by Hoerl et al. (1975), who stated that *the probability of a smaller MSE using* k_a *increases as* p *increases.*

4. EXAMPLE

The contribution of this paper is illustrated in this section with the application of the proposed methodology to three different real datasets with different numbers of independent variables. Then, the application of the following three estimations proposed in this paper

$$k_{exp} = 0.006639 \cdot e^{1-det(\mathbf{R})} - 0.00001241 \cdot n + 0.005745 \cdot p, \tag{17}$$

$$k_{lineal} = 0.01837 \cdot (1 - det(\mathbf{R})) - 0.00001262 \cdot n + 0.005678 \cdot p, \tag{18}$$

$$k_{sq} = 0.7922 \cdot (1 - det(\mathbf{R}))^2 - 0.6901 \cdot (1 - det(\mathbf{R})) - 0.000007567 \cdot n - 0.01081 \cdot p, \tag{19}$$

are compared with other proposals in the literature and summarized in Tables 1 and 2.

4.1. Mortality rate ($p = 3$)

Since this work is the natural extension of García et al. (2017), the goal of this first example is to compare the estimations of k in García et al. (2017) with this new analysis. García et al. (2017) presented the example used by McDonald and Schwing (1973), where the mortality rate is related to the nitrogen oxide pollution potential and the hydrocarbon pollution potential ($p = 3$) for 60 cities ($n = 60$). Both independent variables present a coefficient of correlation ρ of 0.984. This model presents a VIF of 31.502, which denotes severe multicollinearity.

In this case, the estimation of k was 0.0372 with a VIF of 9.9999, whereas from expressions (17)-(19), the following values are obtained:

$$k_{exp} = 0.0339732, \quad k_{lineal} = 0.03406366, \quad k_{sq} = 0.04162561,$$

whose VIFs are 10.59599, 10.57821 and 9.289162, respectively.

A better behavior for the estimation of k was provided by García et al. (2017), which is expected because the simulation was exclusively performed for $p = 3$.

Furthermore, considering the results in Table 6, we obtain the following:

$$\widehat{det(\mathbf{R})} = 1.013 \cdot 0.1 + 0.00008626 \cdot 60 - 0.01384 \cdot 3 = 0.0649556.$$

In other words, in the case of 60 observations, 3 explanatory variables and VIF less than 10, we will have the determinant of the correlation matrix less than 0.0649556. Considering only $det(\mathbf{R}) = 0.031744$, we can conclude that the collinearity in this case is worrisome.

4.2. CO_2 emissions in China ($p = 4$) and number of people employed ($p = 5$)

With the following two examples, we attempt to compare the estimations of k proposed in this work with other estimations in the literature (summarized in Tables 1 and 2). On the one hand, the VIF will be used to check whether the collinearity has been mitigated; on the other hand, a preference criterion based on the MSE will be established.

The two following datasets will be applied:

- Data from China (1990-2014)[4] for the CO_2 emissions (dependent variable), population, per capita GDP and industrialization (% of GDP).
- Data from Longley (1967) (1947-1962)[5] for the number of employed people (dependent variable), number of unemployed people, number of people in the armed forces, non-institutionalized population aged 14 years or over and GNP implicit price deflator.

For the China dataset, we have 25 observations and 4 explanatory variables ($n = 25$ and $p = 4$) and the following estimations for the variance of the random disturbance, coefficient of the regressor, eigenvalues of the matrix $\mathbf{X}^t\mathbf{X}$ and determinant of the matrix of correlations:

$$\widehat{\sigma}^2 = 0.00192444 \ , \ \widehat{\beta}_i = \begin{pmatrix} 157.833822 \\ -7.675747 \\ 1.421072 \\ 1.864854 \end{pmatrix} \ , \ \lambda_i = \begin{pmatrix} 13082.34 \\ 8.329007 \\ 0.01033608 \\ 0.000005995894 \end{pmatrix} \ , \ det(\mathbf{R}) = 0.0219826.$$

These values will be applied to calculate different estimations in Table 15. In addition, it is possible to note that

$$\widehat{det(\mathbf{R})} = 1.013 \cdot 0.1 + 0.00008626 \cdot 25 - 0.01384 \cdot 4 = 0.0480965.$$

[4]Dataset extracted from the World Bank website. All data are expressed in logarithms.

[5]Dataset available in `R-project` (`longley` data).

Because $det(\mathbf{R}) = 0.0219826 < 0.0480965 = \widehat{det(\mathbf{R})}$, we can conclude that there is worrisome collinearity. This conclusion is supported by the values of the VIFs: 28.62464, 28.37358, and 1.606269.

For the dataset from Longley (1967), we have 16 observations and 5 explanatory variables ($n = 16$ and $p = 5$) and the following values applied to obtain different estimations in Table 15:

$$\widehat{\sigma}^2 = 0.3215902 \ , \ \widehat{\beta}_i = \begin{pmatrix} 13,781314004 \\ -0,012412375 \\ -0,005968381 \\ 0,306601491 \\ 0,207046452 \end{pmatrix} \ , \ \lambda_i = \begin{pmatrix} 3192925 \\ 113270.6 \\ 4559.912 \\ 132.2974 \\ 0.006778464 \end{pmatrix} \ , \ det(\mathbf{R}) = 0.007661552.$$

As before, we obtain:

$$\widehat{det(\mathbf{R})} = 1.013 \cdot 0.1 + 0.00008626 \cdot 16 - 0.01384 \cdot 5 = 0.03348016,$$

so $det(\mathbf{R}) = 0.007661552 < 0.03348016 = \widehat{det(\mathbf{R})}$; thus, the collinearity is worrying and is also supported by the values of the VIFs: 3.1476, 2.497795, 34.5883, and 35.97075.

Table 15 shows the k estimation values, as well as the maximum value of VIF and the value of MSE for both datasets. The values of VIF_{max} and MSE that correspond to the values of k out of the interval $[0, 1]$ have been omitted. The values of k that correspond to VIF_{max} less than 10 have been highlighted. Note the following:

- For the CO_2 emissions in the China dataset: k_{exp}, k_{lineal} and k_{sq} have a higher MSE than the one of OLS, but they are the estimations of k corresponding to a maximum VIF below 10. Furthermore, considering the results in Table 16, for $k > 0.007899581$, $\text{MSE}\left(\widehat{\boldsymbol{\beta}}(k)\right)$ is increasing; thus, it easily exceeds the value given by OLS.

- For the dataset of the number of employed people: the estimations of k that correspond to a VIF below 10 are k_{exp}, k_{lineal}, k_{ks}^{max}, k_{mk_3}, k_{mk_5}, k_{m_3} and k_{m_5}. All of these estimations present a MSE higher than that of OLS, but, among them, the proposed estimations in this paper (k_{exp} and k_{lineal}) have the lowest MSE.

Table 15: Values of $\text{VIF}_{\max}$ and MSE for different k

	China			Longley (1967)		
	k value	$\text{VIF}_{\max}(k)$	$\text{MSE}(k)$	k value	$\text{VIF}_{\max}(k)$	$\text{MSE}(k)$
k_{OLS}	0	28.624639	321.146	0	35.970754	47.44544
k_{Hoerl}	0.0000003082071	28.624159	350.2103	0.008460154	23.16867	67.87734
k_{exp}	**0.04032404**	**9.384603**	**24965.3757**	**0.04643538**	**9.454627**	**145.26514**
k_{lineal}	**0.04036268**	**9.378959**	**24965.3844**	**0.04641734**	**9.4571**	**145.25136**
k_{sq}	**0.0393946**	**9.522679**	**24965.1605**	0.0411237	10.251297	140.80546
k_{hk}	0.00000007725101	28.624518	317.0726	0.001693252	32.337479	37.95626
k_h	0.00000007725078	28.624518	317.0726	0.001693252	32.337479	37.95626
k_{lw}	0.000001299878	28.622614	1009.5692	0.00002349348	35.914541	47.12054
k_n	0.002910918	24.727075	24866.4623	70.86813	-	-
k_k^{am}	0.000384766	28.037886	24208.464	2225.245	-	-
k_k^{gm}	0.00003396358	28.571827	18045.1072	15.22808	-	-
k_k^{med}	0.000293016	28.175535	23977.5617	7.50183	-	-
k_{ks}	0.00000007725101	28.624518	317.0726	0.001693252	32.337479	37.95626
k_{ks}^{am}	0.00003064842	28.576973	17474.9074	-17.38358	-	-
k_{ks}^{max}	0.0003609909	28.073421	24159.6197	**0.9647501**	**1.432148**	**187.12299**
k_{ks}^{med}	0.000006239131	28.614922	6570.0691	0.0006159796	34.554742	41.18718
k_{mk_1}	3597.89	-	-	24.30184	-	-
k_{mk_2}	0.03086997	11.034913	24962.6248	95.01559	-	-
k_{mk_3}	171.5905	-	-	**0.256258**	**2.874257**	**180.13681**
k_{mk_4}	0.00003396358	28.571827	18045.1072	15.22808	-	-
k_{mk_5}	108.7411	-	-	**0.3651038**	**2.304299**	**182.92062**
k_{mk_6}	0.01461949	16.11944	24950.4202	2.738947	-	-
k_{dk}	0	28.624639	321.146	0.006722631	24.965223	59.00921
k_f	0.0001529552	28.388376	23121.0258	0.001693878	32.336276	37.95626
k_{m_1}	12944810	-	-	590.5796	-	-
k_{m_2}	0.0009529533	27.215456	24657.5261	8755.64	-	-
k_{m_3}	29443.32	-	-	**0.06616691**	**7.406175**	**156.54123**
k_{m_4}	0.00003396356	28.571827	18045.1038	15.1133	-	-
k_{m_5}	16211.16	-	-	**0.1333042**	**4.497771**	**171.95049**

Table 16: Calculation of $\frac{\widehat{\sigma}^2}{\widehat{\alpha}_j^2}$ for CO_2 emissions in China and number of people employed data sets

$\frac{\widehat{\sigma}^2}{\widehat{\alpha}_j^2}$	CO_2 emissions in China	Number of people employed
j=1	0.004253279	15.27178
j=2	0.007899581	196.321
j=3	0.002887787	1.453364
j=4	0.00000007707998	5.288227
j=5	-	0.001694813

Figure 4 displays the MSE of both examples. The MSE as a function of k rapidly increases (in the example of the CO_2 emissions, the initial decrease is not appreciated[6]). Another symptom of the rapid increasing of MSE is that for CO_2 emission in China, the value of $\boldsymbol{\beta}^t\boldsymbol{\beta}$ [7] is 24975.93, while for the number of people employed, the value of $\boldsymbol{\beta}^t\boldsymbol{\beta}$ is 190.0617. This behavior favors the performance of the estimates provided by k_{exp}, k_{lineal} and k_{sq}; since they are the minimum values of k that make the VIF lower than 10, its MSE will be smaller than the rest of proposed estimations of k as long as the MSE exhibits an increasing behavior.

Remark 1. *To get a deeper analysis of the last question, following subsection 3.3, 15200 simulations have been carried. For each simulation we have calculated k_{exp}, k_{lineal}, k_{sq} and the turning point of MSE $\left(\widehat{\boldsymbol{\beta}}(k)\right)$ when it goes from being decreasing to growing, $k_{inflection}$. Table 17 summarizes the results obtaining the following conclusions:*

- *The turning point is lesser than the one proposed in this paper (between 82% and 93% for k_{exp} and k_{lineal}, and between 73% and 93% for k_{sq}). Thus, k_{exp}, k_{lineal} and k_{sq} are in the increasing part of the MSE in a high percentage of cases.*

- *Again, practically a identical behavior is observed between k_{exp} and k_{lineal}.*

[6]Considering Table 16, the MSE is decreasing if k is lower than 0.00000007707998.
[7]The limit of the MSE when k tends to infinity.

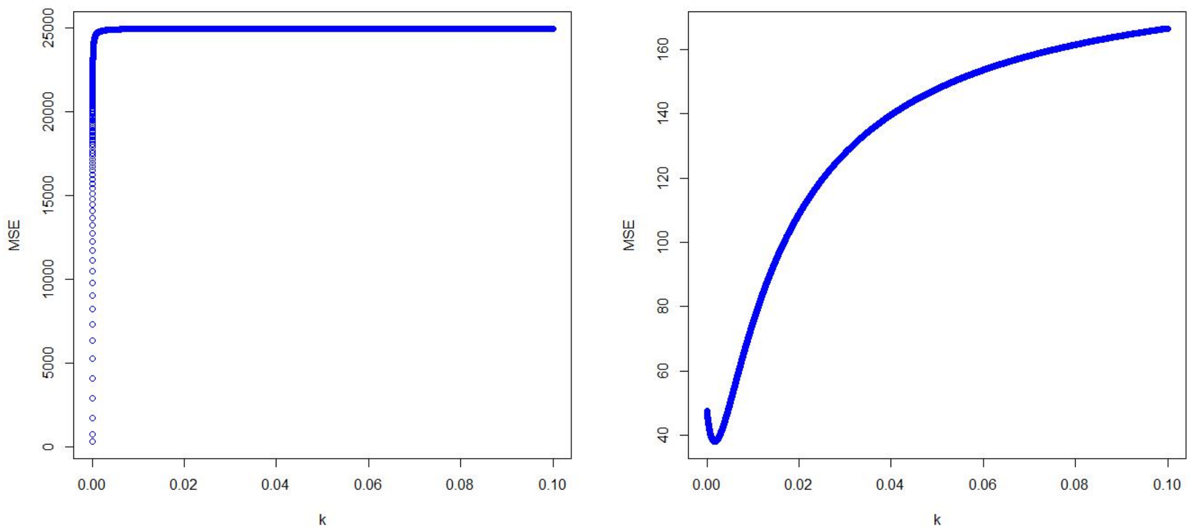

Figure 4: Graphic representation of MSE for CO_2 emissions in China and number of people employed datasets

Table 17: Percentage of cases in which k_{exp}, k_{lineal} and k_{sq} are higher than $k_{inflection}$

	$p = 3$	$p = 4$	$p = 5$
$k_{inlection} < k_{exp}$	82.177%	90.801%	93.868%
$k_{inlection} < k_{lineal}$	82.177%	90.927%	93.868%
$k_{inlection} < k_{lineal}$	73.737%	93.65%	93.868%

- *The percentage of cases in which k_{exp}, k_{lineal} and k_{sq} is higher than $k_{inflection}$ is higher as the number of independent variables of the model increases.*

The values of k displayed in Tables 1 and 2 have been calculated and compared to values k_{exp}, k_{lineal} and k_{sq} when the associated VIF is lower than 10. Tables 18 to 20 show the results and in parenthesis the percentage of cases where the VIF is lower than 10.

Note that the percentage where k_{exp}, k_{lineal} and k_{sq} are lower than the rest, is elevated (higher than 62%[8] *and mostly 90%), which supports the supposition. Furthermore, the*

[8] Situations where the number of cases is reduced (values of 50% and, even, 0%) are not considered.

number of cases where both conditions (k_{exp}, k_{lineal} and k_{sq} lower than the rest of values of k displayed in Tables 1 and 2 and VIF lower than 10) diminishes as p increases.

□

5. CONCLUSIONS

The first goal of this work is to provide an estimate of the ridge parameter k from the determinant of the matrix of correlations of the dataset, $det(\mathbf{R})$, which enables one to mitigate the worrying collinearity in a multiple linear regression. Thus, by following García et al. (2017), different datasets are simulated to calculate $det(\mathbf{R})$ and the first value of k that makes the VIF lower than the traditionally considered threshold of 10. Then, the estimation is obtained through a regression of k depending on $det(\mathbf{R})$. The VIF and MSE of these estimations are compared with the values obtained for other estimations in the literature.

Since the estimated values are obtained by considering the minimum values of k that makes the VIF lower than 10, these estimations also easily present a MSE higher than the value for OLS estimation but lower than other ridge estimations proposed in the literature if the MSE exhibits an increasing behavior. It is possible to conclude that k_{exp}, k_{lineal} and k_{sq} are higher than the turning point of monotony in between the 82% and the 93% of the cases depending on the number of independent variables of the econometric model; thus, k_{exp}, k_{lineal} and k_{sq} are in the growing part of MSE in a high percentage of cases. **Furthermore, in more than 62% of cases, the values of k_{exp}, k_{lineal} and k_{sq} are lower than the rest of values considered in the literature (except for k_{Hoerl}), and consequently the corresponding MSE will be also lower.**

It is important to highlight that this work has followed the indications provided by Marquardt (1970), who stated that *A rule of thumb for choosing the amount of bias to allow with ill conditioned data, whether by ridge or generalized inverse, is that the maximum variance inflation factor usually should be larger than 1 but certainly not as large as 10.* In addition, the indications provided by García et al. (2016) have been followed to avoid the problems indicated by García et al. (2015) and not obtain values of VIF below 1. **The**

Table 18: Percentage of case where k_{exp} is lower than values of k displayed in Tables 1 and 2 for $p = 3, 4, 5$ (minimum and maximum values are highlighted in bold)

	$k_{exp} < k^*$		
k^*	p=3	p=4	p=5
k_h	90.695% (6266)	91.269% (378)	92.307% (26)
k_{lw}	90.93% (2580)	94.696% (132)	88.888% (9)
k_n	99.39% (6926)	97.446% (470)	93.548% (31)
k_k^{am}	85.689% (5737)	91.798% (317)	88.461% (26)
k_k^{gm}	74.433% (4545)	89.603% (202)	73.333% (15)
k_k^{med}	77.106% (3180)	79.62% (211)	100% (9)
k_{ks}	71.81% (3180)	63.461% (52)	0% (1)
k_{ks}^{am}	96.088% (3656)	97.095% (241)	93.333% (15)
k_{ks}^{max}	91.113% (6392)	94.919% (433)	93.939% (33)
k_{ks}^{med}	73.631% (4183)	89.915% (238)	81.818% (11)
k_{mk1}	100% (3104)	100% (85)	100% (5)
k_{mk2}	100% (6956)	100% (461)	100% (38)
k_{mk3}	99.94% (6332)	100% (415)	100% (29)
k_{mk4}	74.43% (4545)	89.603% (202)	73.333% (15)
k_{mk5}	99.95% (4074)	100% (307)	100% (22)
k_{mk6}	99.93% (7269)	100% (489)	97.368% (38)
k_{dk}	**69.11%** (2791)	**62.068%** (58)	100% (1)
k_f	70.003% (3107)	64.788% (71)	50% (4)
k_{m_1}	100% (296)	100% (3)	(0)
k_{m_2}	97.35% (6650)	96.829% (410)	100% (32)
k_{m_3}	100% (1585)	100% (78)	100% (7)
k_{m_4}	74.24% (4570)	88.059% (201)	73.333% (15)
k_{m_5}	99.52% (1049)	100% (24)	100% (5)

Table 19: Percentage of case where k_{lineal} is lower than values of k displayed in Tables 1 and 2 for $p = 3, 4, 5$ (minimum and maximum values are highlighted in bold)

$k_{lineal} < k^*$			
k^*	p=3	p=4	p=5
k_h	90.652% (6280)	91.315% (380)	92.307% (26)
k_{lw}	90.94% (2583)	94.736% (133)	88.888% (9)
k_n	99.39% (6945)	97.463% (473)	93.548% (31)
k_k^{am}	85.691% (5752)	91.798% (317)	88.461% (26)
k_k^{gm}	74.428% (4552)	89.655% (203)	73.333% (15)
k_k^{med}	77.142% (3185)	79.146% (211)	100% (9)
k_{ks}	71.75% (2662)	63.461% (52)	0% (1)
k_{ks}^{am}	96.093% (3661)	97.107% (242)	93.333% (15)
k_{ks}^{max}	91.131% (6405)	94.954% (436)	93.939% (33)
k_{ks}^{med}	73.584% (4187)	90.041% (241)	81.818% (11)
k_{mk1}	100% (3109)	100% (86)	100% (5)
k_{mk2}	100% (6977)	100% (464)	100% (38)
k_{mk3}	99.98% (6349)	100% (416)	100% (29)
k_{mk4}	74.42% (4552)	89.655% (203)	73.333% (115)
k_{mk5}	99.95% (4083)	100% (308)	100% (22)
k_{mk6}	99.91% (7290)	100% (493)	97.368% (38)
k_{dk}	**69.19%** (2795)	**62.068%** (58)	100% (1)
k_f	69.94% (3111)	64.788% (71)	50% (4)
k_{m_1}	100% (297)	100% (3)	(0)
k_{m_2}	97.36% (6670)	96.601% (412)	100% (32)
k_{m_3}	100% (1588)	100% (79)	100% (7)
k_{m_4}	74.24% (4577)	88.118% (202)	73.333% (15)
k_{m_5}	99.52% (1052)	100% (24)	100% (5)

Table 20: Percentage of case where k_{sq} is lower than values of k displayed in Tables 1 and 2 for $p = 3, 4, 5$ (minimum and maximum values are highlighted in bold)

$k_{sq} < k^*$			
k^*	p=3	p=4	p=5
k_h	92.149% (5439)	92.365% (4139)	93.75% (32)
k_{lw}	90.725% (2178)	91.276% (1387)	90.909% (11)
k_n	95.38% (7412)	98.878% (6239)	91.891% (37)
k_k^{am}	94.048% (5192)	94.664% (3711)	86.667% (30)
k_k^{gm}	92.715% (2567)	90.009% (2052)	72.222% (18)
k_k^{med}	92.723% (1704)	85.13% (1836)	100% (10)
k_{ks}	92.686% (1340)	82.037% (373)	0% (1)
k_{ks}^{am}	95.497% (4509)	97.262% (2959)	94.444% (18)
k_{ks}^{max}	97.777% (6210)	96.543% (5236)	97.368% (38)
k_{ks}^{med}	**89.258%** (1955)	86.598% (1955)	84.615% (13)
k_{mk1}	100% (2339)	100% (1128)	100% (6)
k_{mk2}	100% (8905)	100% (6225)	100% (45)
k_{mk3}	100% (5750)	100% (5486)	100% (34)
k_{mk4}	92.71% (2567)	90.009% (2052)	72.222% (18)
k_{mk5}	100% (3157)	100% (3975)	100% (27)
k_{mk6}	91.402% (6758)	98.549% (6411)	97.777% (45)
k_{dk}	91.32% (1372)	**80%** (390)	100% (1)
k_f	95.28% (1633)	81.213% (511)	50% (4)
k_{m_1}	99.68% (313)	100% (31)	(0)
k_{m_2}	94.69% (7672)	96.698% (5149)	100% (36)
k_{m_3}	99.92% (1283)	100% (942)	100% (7)
k_{m_4}	92.49% (2478)	88.309% (1976)	66.666% (18)
k_{m_5}	99.89% (921)	100% (285)	100% (6)

results obtained in the simulation to analyze the relation between bias and variance show that the values of k, k_{exp}, k_{lineal} and k_{sq}, estimated to obtain a VIF lower than 10, present a probability between 27.355% and 14.75% (for k and k_{exp}) and between 37.934% and 14.151% (for k_{lineal}) to obtain a MSE lower that the MSE obtained by OLS. This probability diminishes as p increases (at least for k and k_{exp}). By comparing these results with the results provided by Hoerl et al. (1975) for k_{Hoerl}, it is noted that the requirement to mitigate the collinearity has reduced the percentage of cases where the MSE is lower than in OLS.

The second goal of this work is to find a relation between the VIF and the $det(\mathbf{R})$ as the following rule of thumb: a VIF higher than 10 corresponds to $det(\mathbf{R})$ below 0.1. A table of equivalence is also provided to fit this relation to the number of variables and observations of a multiple linear regression.

Finally, due to the estimation of the variance of the random disturbance, σ^2, is present in most of the estimations of the ridge parameter, k, collected in Tables 1 and 2, a future research line could be to introduce $\widehat{\sigma}^2$ to the estimations of k proposed in expressions (17) to (19). It will be also interesting to analyze if providing a estimation of k for each value of p will be appropriate since it could be more efficient although some prediction ability could be lost for values of p not considered.

BIBLIOGRAPHY

Casella, G. (1985). Condition Numbers and Ridge Regression Estimators. *Journal of the American Statistical Association 80*(391), 753–758.

Dorugade, A. and D. Kashid (2010). Alternative method for choosing ridge parameter for regression. *Applied Mathematical Sciences 4*(9), 447–456.

Duzan, H. and N. Shariff (2015). Simulation Study to Verify the Appropriate k Value for Ridge Regression in Two-variable Regression Model. *Indian Journal of Science and Technology 8*(12), 1–8.

Fox, J. and G. Monette (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association 87*(417), 178–183.

García, C., J. García, M. López-Martín, and R. Salmerón (2015). Collinearity: Revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics 42*(3), 648–661.

García, C., R. Salmerón, J. García, and M. López-Martín (2017). On the selection of the ridge and raise factors. *Indian Journal of Science and Technology 10*(13), 1–8.

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

Gibbons, D. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association 76*(373), 131–139.

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

Hocking, R., F. Speed, and M. Lynn (1976). A class of biased estimators in linear regression. *Technometrics 18*(4), 55–67.

Hoerl, A. and R. Kennard (1970a). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics 12*(1), 69–82.

Hoerl, A. and R. Kennard (1970b). Ridge Regression: Biased Estimation for Northogonal Problems. *Technometrics 12*(1), 55–67.

Hoerl, A., R. Kennard, and K. Baldwin (1975). Ridge Regression: Some simulation. *Communications in Statistics-Theory and Methods 4*(2), 105–123.

Khalaf, G. (2012). A proposed ridge parameter to improve the least squares estimator. *Journal of Modern Applied Statistical Methods 11*(2), 443–449.

Khalaf, G. and G. Shukur (2005). Choosing ridge parameters for regression problems. *Communications in Statistics: Theory and Methods 34*(5), 1177–1182.

Kibria, B. (2003). Performance of some new ridge regression estimators. *Communications in Statistics: Simulation and Computation 32*(2), 419–435.

Kibria, B. and S. Banik (2016). Some Ridge Regression Estimators and Their Performances. *Journal of Modern Applied Statistical Methods 15*(1), 206–238.

Lawless, J. and P. Wang (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics: Theory and Methods 5*(4), 307–323.

Longley, J. (1967). An appraisal of least-squares programs for the electronic computer from the point of the user. *Journal of the American Statistical Association 62*(319), 819–841.

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

McDonald, G. (2010). Tracing ridge regression coefficients. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(6), 695–703.

McDonald, G. and D. Galarneau (1975). A Monte Carlo evaluation of some ridge type estimators. *Journal of the American Statistical Association 70*, 407–416.

McDonald, G. and R. Schwing (1973). Instabilities of regression estimates relating air pollution to mortality. *Technometrics 15*, 463â“–481.

Muniz, G. and B. Kibria (2009). On some ridge regression estimators: An empirical comparison. *Communications in Statistics: Simulation and Computation 38*(3), 621–630.

Muniz, G., B. Kibria, K. Mansson, and G. Shukur (2012). On developing ridge regression parameters: A graphical investigation. *Statistics and Operations Research Transactions 36*(2), 115–138.

Nomura, M. (1988). On the almost unbiased ridge regression estimation. *Communications in Statistics: Simulation and Computation 17*(3), 729–743.

Salmerón, R., C. García, and J. García (2018). Variance inflation factor and condition number in multiple linear regression. *Journal of Statistical Computation and Simulation 88*(12), 2365–2384.

Salmerón Gómez, R., J. García Pérez, M. López Martín, and C. García García (2016). Collinearity diagnostic applied in ridge estimation through the variance inflation factor. *Journal of Applied Statistics 43*(10), 1831–1849.