Abstract

Ridge estimation (RE) is an alternative method to ordinary least squares (OLS) when there exists a collinearity problem in a linear regression model. The variance inflator factor (VIF) is applied to test if the problem exists in the original model and is also necessary after applying the ridge estimate to check if the chosen value for parameter k **has mitigated the collinearity problem.** This paper shows that the application of the original data when working with the ridge estimate **leads to non-monotone VIF values**. **García et al. (2014) showed some problems with the traditional VIF used in RE. We propose an augmented VIF, $\mathbf{VIF}_R(j,k)$, associated with RE, which is obtained by standardizing the data before augmenting the model. The $\mathbf{VIF}_R(j,k)$ will coincide with the VIF associated with the OLS estimator when $k = 0$. The augmented VIF has the very desirable properties of being continuous, monotone in the ridge parameter and higher than one.**
Key words: Collinearity; linear regression; variance inflator factor; ridge regression; standardization

1 Introduction

One of the basic hypotheses of the ordinary least squares (OLS) method is that the exogenous variables must be linearly independent. **If there exists a perfect linear dependency – complete multicollinearity – the model has no unique solution, while if the dependency is approximate – near multicollinearity – the estimation will be unstable. Note that throughout this article, multicollinearity will mean near dependence.** Thus, when multicollinearity exists it is necessary to seek an alternative method to estimate the model. Ridge estimation (RE) (see e.g. Hoerl & Kennard, 1970a,b) is the most commonly applied method to analyze data by minimizing the effects of multicollinearity, without having to increase the sample, improve its quality and/or eliminate some of the variables of the model (options that are not feasible in many cases).

Once the ridge estimator is applied, it is necessary to verify whether the problem has disappeared or at least that the damage due to multicollineality has been sufficiently mitigated to not be decisive. To do so, it is important to have a measure to determine the presence of multicollinearity. A widely used measure in the literature is the variance inflator factor (VIF) (see e.g. Theil, 1971, Marquardt, 1970). If the VIF takes values below a generally accepted threshold (see O'Brien, 2007), the problem of multicollinearity is considered to have been overcome.

In addition to the VIF, the use of standardization to diagnose and mitigate collinearity of explanatory variables is another option that is questioned in the scientific literature. Marquardt & Snee (1975) agreed with standardizing variables to diagnose collinearity when stated that: "The ill conditioning that results from failure to standardize is all the more insidious because it is not due to any real defect in the data, but only the arbitrary origins of the scales on which the predictor variables are expressed". However, this paper received strong criticism from Smith & Campbell (1980) among others: "The essential problem with the VIF and similar measures (of collinearity) is that they ignore the parameters while trying to assess the information given by the data. Clearly, an evaluation of the strength of the data depend on the scale and nature of the parameters". **After noting Smith and Campbell's lack of understanding of the problem, Marquardt (1980) responded to their work,** concluding with the following recommendation: "In summary, when faced with the analysis of regression data whose quality may be good in all respects except the presence of multicollinearity, the statistician should remove the nonessential multicollinearity by standardizing the predictor variables and then use a biased estimator to reduce the effects of the remaining multicollinearity". Marquardt (1980) also added an important consideration: "The least squares objective function is mathematically independent of the scaling of the predictor variables (while the objective function in ridge regression is mathematically dependent on the scaling of the predictor variables)". This will be an essential issue in the aim of our work.

Several authors have contributed to the controversy surrounding the standardization of data. Belsley et al. (1980) indicated that: "Mean centering typically masks the role of the constant term in any underlying near dependencies and produces misreadingly favorable conditioning diagnostics. Especially for a ridge regression model". Belsley (1982), among others, criticized the prevailing practice of centering predictor variables (usually followed by scaling to unit length) prior to assessing the presence and effects of collinearity. On the other hand, Vinod & Ullah (1981) pointed out that: "The appearance of a ridge trace that does not plot standardized regression coefficients may be dramatically changed by a simple translation of the origin and scale transformation of the variables. In this case, there is the danger of naively misinterpreting the meaning of the plot". Gunst (1984) noted that "one of the problems of centering data is to carefully consider whether it is important to detect collinearity with the constant term". In Marquardt's comment to the paper presented by Stewart (1987), he stated: "I fully agree with Stewart that when there is a constant term in the model, the model should be centered before the important of the remaining variables is assessed and the centering simple shows the variable for what it is". Thus, most authors recommend standardizing the data as we will do, so that $\mathbf{X}'\mathbf{X}$ is in the form of a correlation matrix (King, 1986 and Stewart, 1987 among others).

On the other hand, Sardy (2008) proposed a new method to rescale the variable based on the diagonal elements of the covariance matrix of the maximum-likelihood estimator and stated that it has to be applied before ridge regression as an alternative to traditional standardizing. Dias & Castro (2011) showed that the real impact on variance can be overestimated by the traditional VIF when the explanatory variables contain no redundant information about the dependent variable and stated that a corrected version of this multicollinearity indicator becomes necessary. In their diagnostic of collinearity, Jensen & Ramirez (2013) made a distinction between centered and uncentered VIFs. Moreover, they pointed out the special relevance of uncentered VIF when the independent term plays an important role in the interpretation of the model.

In this paper we will provide some new ideas to clarify when it is recommendable to work with standardized or unstandardized explicative variables. Except in the case of the independent term, the VIFs will be generally equal to standardized or unstandardized variables. However, we will show a case where the standardization of the variables plays a key role: when applying the ridge estimator as a solution to collinearity. Remember that the ridge estimator (Hoerl & Kennard, 1970a,b) is biased, depends on the parameter k and is given by the following expression:

$$\widehat{\boldsymbol{\beta}}_R(k) = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{Y}. \tag{1}$$

Note that when $k = 0$, expression (1) coincides with the OLS estimator.

Following the analogy between multicollinearity and illness suggested by Marquardt & Snee (1975), it is necessary to calculate the VIF to *diagnose* the *disease* of the data (multicollinearity) and check whether or not it exceeds the threshold generally accepted in the literature. If the VIF exceeds the threshold, we can state that the data suffer from the *illness* and an adequate *treatment* will therefore be needed, including RE. The treatment will be more or less strong depending on the *doses*: the value of the parameter k.

However, the problem does not end there. After applying the treatment we must check that it was effective, and if not, increase the *dose* (increase k). That is, we need to recalculate the value of the VIF that will depend on the parameter k. This is the reason it is necessary to extend the concept of VIF to the case of RE, denoted by augmented VIF or $\text{VIF}_R(j,k)$. To the best of our knowledge, this topic has not yet been addressed in the literature. Even O'Brien (2007), who pointed out the controversy surrounding the standardization of variables in the first footnote of his excellent paper and cited the main papers about this issue, did not say anything about the need to correctly extend the concept of VIF to RE.

As we will see below, this extension presents a problem whose solution leads to the conclusion that data standardization in RE is not optional but required. The rationale behind this is simple and is based on the following observation: if the ridge estimator coincides with the OLS estimator when $k = 0$, then the $\text{VIF}_R(j,k)$ associated with RE must coincide with the VIF associated with the OLS estimator when $k = 0$. If this does not occur, it is evident that the extension of the VIF to RE is not correct.

Actually, the definition of the VIFs proposed by Marquardt (1970) as the diagonals of the inverse of the correlation matrix of centered and scaled regressors, does not verify this condition and cannot be applied when extending it to the RE. However, this definition has been accepted without further considerations by Marquardt & Snee (1975), Velleman & Welsch (1981), Montgomery & Peck (1982), Anderson (1985), Stewart (1987), Myers (1990) and Fox & Monette (1992), among others. The same occurs with the definition given by Fox & Monette (1992). We insist that the problem of these definitions arises when extending them to RE. Recently, García et al. (2014) established that the correct extension of the VIF to RE has to be done from the model $\mathbf{Y}^R = \mathbf{X}^R\boldsymbol{\beta} + \mathbf{u}^R$. The VIF obtained from this extension has some similarities to the one from the surrogate ridge regression presented by Jensen & Ramirez (2010b). Figure 1 summarizes this basic idea and the procedure to correctly extend the concept of VIF to RE.

In this work we explain why Marquardt (1980) was right when **he** pointed out that "The quality of the predictor variable (here regressors) structure of a data set can be assessed properly only in terms of a standardized scale. This applies to both, least squares and ridge estimation". We also note that two problems may arise if the data are not adequately standardized when applying the extension of the concept of the VIF in RE:

1. To obtain values of VIF less than 1 after applying the ridge estimator. This will be inconsistent with the theoretical definition of VIF (see e.g. Jamal & Rind, 2007, Irfan et al., 2013, Salmerón et al., 2013).

2. To consider that the problem of collinearity has been overcome when it has not.

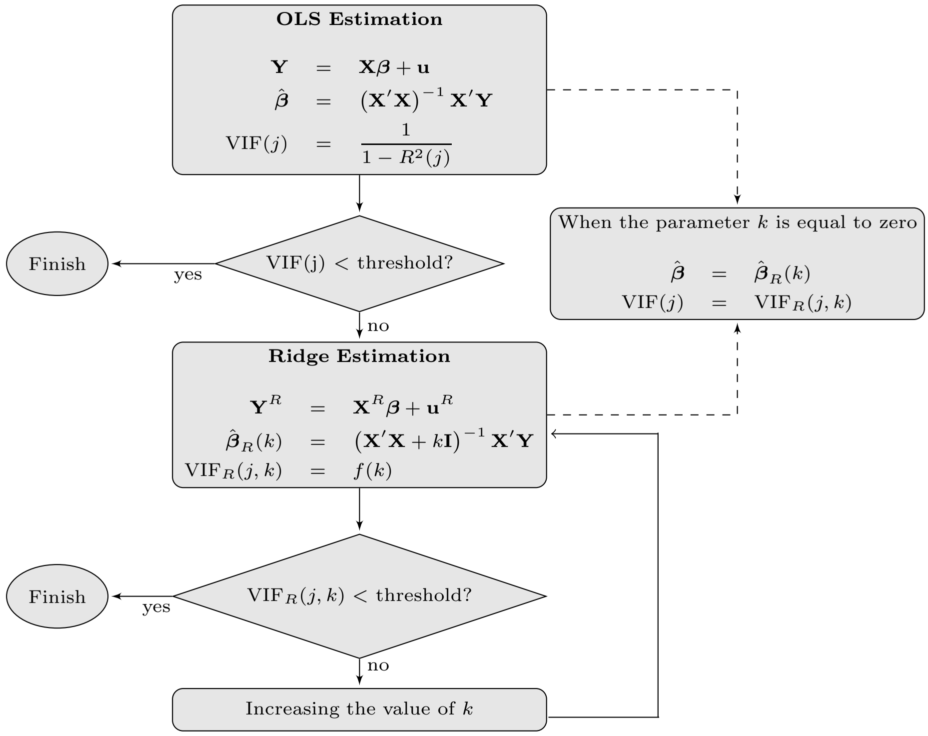

Figure 1: Diagnostic of multicollinearity in RE

In the core of this paper, we will show that the objective function in RE is not mathematically independent of the scaling of the data. We analyze how the use of original or standardized data in a linear model affects the calculation of the VIF, particularly after applying RE. Thus, the main goal of this paper is to show that standardization is not optional but required when analyzing the existence of multicollinearity in RE. The paper is structured as follows: in section 2 we include the notation used in the paper, in section 3 we calculate the VIF for standardized and unstandardized data after applying OLS estimation while in section 4 it is calculated after applying RE. In section 5 we study the effects of several factors on the variance of RE coefficients. Section 6 shows the results of the numerical example. Finally, section 7 presents the main consequences and conclusions of this work.

2 Notation

Vector and matrices are set in bold type. The transpose and inverse of matrix $\mathbf{A}$ are $\mathbf{A}'$ and $\mathbf{A}^{-1}$, respectively. Matrix $\mathbf{A}_{-j}$ is matrix $\mathbf{A}$ without column j. The special arrays are the identity matrix, $\mathbf{I}$, and the null vector, $\mathbf{0}$.
The original model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{Y}$ is the vector of order $n \times 1$; $\mathbf{X} = [\mathbf{1}_n, \mathbf{X}_2, \dots, \mathbf{X}_p]$ of dimension $n \times p$ `(note that the model contains a constant term)` and X_{ij} is the observation i of the variable j; $\boldsymbol{\beta}$ is the vector of parameters of order $p \times 1$, $\mathbf{u}$ is the vector of random disturbance for $n \times 1$ and R^2 is the coefficient of determination. The number of observations is n and the number of independent variables is p. The OLS estimators are $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. The auxiliary regression of $\mathbf{X}_j$ on the rest of the independent variables, $j = 2, \dots, p$, is $\mathbf{X}_j = \mathbf{X}_{-j}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. The transformed model is denoted in lowercase.
For the ridge regression we consider the model $\mathbf{Y}^R = \mathbf{X}^R\boldsymbol{\beta} + \mathbf{u}^R$ where $\mathbf{X}^R = \left(\frac{\mathbf{X}}{\sqrt{k}\mathbf{I}}\right)$ and $\mathbf{Y}^R = \left(\frac{\mathbf{Y}}{\mathbf{0}}\right)$, since its OLS estimator is equal to $\hat{\boldsymbol{\beta}}_R(k) = \left\{(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \quad k \geq 0\right\}$ (Marquardt, 1970; Zhang & Ibrahim, 2005) and $R^2(k)$ is the coefficient of determination. The auxiliary model will be denoted by $\mathbf{X}_j^R = \mathbf{X}_{-j}^R\boldsymbol{\delta} + \boldsymbol{\nu}$ and the transformed model in lowercase, with $j = 2, \dots, p$. **Note that $\mathbf{X}_{-j}$ and $\mathbf{X}_{-j}^R$ contain a constant column since $j \neq 1$.**

The explained sum of squares (ESS), the total sum of squares (TSS) and the coefficient of determination are denoted as:

- $\text{ESS}(j)$, $\text{TSS}(j)$ and $R^2(j)$ for the auxiliary regression of the dependent variable $\mathbf{X}_j$,
- $\text{ESS}_T(j)$, $\text{TSS}_T(j)$ and $R_T^2(j)$ for the transformed auxiliary regression of the dependent variable $\mathbf{x}_j$,
- $\text{ESS}(j,k)$, $\text{TSS}(j,k)$ and $R^2(j,k)$ for the auxiliary regression of the dependent variable $\mathbf{X}_j^R$,
- $\text{ESS}_T(j,k)$, $\text{TSS}_T(j,k)$ and $R_T^2(j,k)$ for the transformed auxiliary regression of the dependent variable $\mathbf{x}_j^R$.

Finally, the variance inflator factor (VIF) associated to the original model will be noted as $\text{VIF}(j)$ while the VIF associated to the RE will be noted as $\text{VIF}_R(j,k)$.

3 Obtaining the Variance Inflator Factor after application of OLS

Consider the following general linear model for p independent variables and n observations:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \tag{2}$$

The model is assumed to be correctly specified, that is, the random disturbance is centered, homoscedastic and uncorrelated. It also verifies that the random disturbance is uncorrelated with the independent variables.
The collinearity problem consists in the existence of linear dependency between the independent variables in a general linear model similar to (2). Collinearity prevents estimating the model (if the collinearity is perfect) or can imply unstable estimations (if the collinearity is approximate). Determining the existence of this problem in the model will be an essential issue.
VIF is the most widely applied measure to detect the existence of collinearity (see e.g. Theil, 1971), which is defined for $\mathbf{X}_j$ with $j = 2, \dots, p$, as:

$$\text{VIF}(j) = \frac{1}{1 - R^2(j)}, \tag{3}$$

where $R^2(j)$ is the coefficient of determination of the auxiliary regression of $\mathbf{X}_j$ on the rest of the independent variables, $j = 2, \dots, p$:

$$\mathbf{X}_j = \mathbf{X}_{-j}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{4}$$

To calculate the coefficient of determination of this model we will use the following expression:

$$R^2(j) = \frac{\sum\limits_{i=1}^{n} \left(\widehat{X}_{ij} - \overline{\mathbf{X}}_j\right)^2}{\sum\limits_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j\right)^2}. \tag{5}$$

If we consider the transformation (6):

$$\mathbf{x}_j = \frac{\mathbf{X}_j - a_j}{b_j}, \quad j = 1, \dots, p, \tag{6}$$

where $a_j \in \mathbb{R}$ and $b_j > 0$, model (4) will be given by:

$$\mathbf{x}_j = \mathbf{x}_{-j}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{7}$$

We can obtain the coefficient of determination of the transformed model using the following expression:

$$R_T^2(j) = \frac{\sum\limits_{i=1}^{n} \left(\widehat{x}_{ij} - \overline{\mathbf{x}}_j\right)^2}{\sum\limits_{i=1}^{n} \left(x_{ij} - \overline{\mathbf{x}}_j\right)^2}. \tag{8}$$

From transformation (6) it is easy to check that $\overline{\mathbf{x}}_j = \frac{1}{b_j}\left(\overline{\mathbf{X}}_j - a_j\right)$ and then:

$$\sum_{i=1}^{n} \left(\widehat{x}_{ij} - \overline{\mathbf{x}}_j\right)^2 = \frac{1}{b_j^2} \sum_{i=1}^{n} \left(\widehat{X}_{ij} - \overline{\mathbf{X}}_j\right)^2,$$
$$\sum_{i=1}^{n} \left(x_{ij} - \overline{\mathbf{x}}_j\right)^2 = \frac{1}{b_j^2} \sum_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j\right)^2.$$

Thus, we can conclude that $\text{ESS}_T(j) = \frac{1}{b_j^2}\text{ESS}(j)$ and $\text{TSS}_T(j) = \frac{1}{b_j^2}\text{TSS}(j)$, and it is verified that $R_T^2(j) = R^2(j)$. That is, the coefficient of determination of both models (original (4) and transformed (7)) are equal for any values of a_j **and** $b_j > 0$**,** $j = 2, \dots, p$, **as expected since standardizing data does not change the correlation.** Therefore, the VIF values obtained by using the auxiliary regressions will also be equal when working with original, centered and standardized data[1].

4 Obtaining the Variance Inflator Factor after application of RE

When the presence of multicollinearity is approximate, the estimations will be unstable. **Another problem is the tendency to not reject the null hypothesis in the individual significance test, since $\hat{\beta}_i$ will have large standard errors. Problems with the estimator due to the presence of multicollinearity among the independent variables are inflated variances and covariances, inflated correlations and inflated prediction variance (Farrar & Glauber, 1967; Silvey, 1969; Marquardt, 1970; Marquardt & Snee, 1975; Gunst & Mason, 1977; Willan & Watts, 1978).** Thus, it is interesting to apply an estimation method to amend the damage caused by multicollinearity. The ridge estimator (see Hoerl & Kennard, 1970a,b) is one of the best known and most applied methods in this situation. Given model (2), the ridge estimation defines a class of estimators depending on the parameter k:

$$\widehat{\boldsymbol{\beta}}_R(k) = \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{Y}, \quad (9)$$

where $k \geq 0$ and its covariance matrix is:

$$\text{var}\left(\widehat{\boldsymbol{\beta}}_R(k)\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}. \quad (10)$$

The estimator given in (9) is a biased estimator when $k > 0$. When $k = 0$, the estimation is equal to the one obtained from the OLS estimator of model (2).
Marquardt (1970) and Zhang & Ibrahim (2005) showed that the OLS estimation of the ridge model:

$$\mathbf{Y}^R = \mathbf{X}^R\boldsymbol{\beta} + \mathbf{u}^R, \quad (11)$$

is $\widehat{\boldsymbol{\beta}}_R(k)$ since $\left(\mathbf{X}^R\right)'\mathbf{X}^R = \mathbf{X}'\mathbf{X} + k\mathbf{I}$ and $\left(\mathbf{X}^R\right)'\mathbf{Y}^R = \mathbf{X}'\mathbf{Y}$.
To detect the possible existence of multicollinearity in model (11), it is necessary to extend the concept of VIF (augmented VIF hereafter), which is defined as:

$$\text{VIF}_R(j,k) = \left\{ \begin{array}{ll} \text{VIF}(j), & \text{if } k = 0 \\ f(k), & \text{if } k > 0 \end{array} \right., \quad (12)$$

that is, when $k = 0$ the VIF of the model (11) will be equal to the VIF of the original model. However, when $k > 0$ the $\text{VIF}_R(j,k)$ will depend on k in the form $f(k) = \dfrac{1}{1 - g(k)}$ obtained below. In addition, it has to be verified that the function (12) will be continuous at $k = 0$:

$$\lim_{k \to 0} \text{VIF}_R(j,k) = \text{VIF}(j), \quad \textbf{for all} j = 2, \dots, p, \quad (13)$$

monotonically decreasing (see Jensen & Ramirez, 2010b; Hadi, 2011) and higher than one for all k (García et al., 2014). An extension of the VIF to RE in which these conditions are not verified will be an incorrect extension in our opinion.
The relevance of the properties of continuity and monotonicity is that if they are not verified:

[1] When $a_j = \overline{\mathbf{X}}_j$ and $b_j = 1$, for $j = 2, \dots, p$, data are centered. When $a_j = \overline{\mathbf{X}}_j$ and $b_j = \sqrt{n-1} \cdot S_{n-1}$ where S_{n-1} is the quasi standard deviation of the independent variable $\mathbf{X}_j$, for $j = 2, \dots, p$, data are standardized.

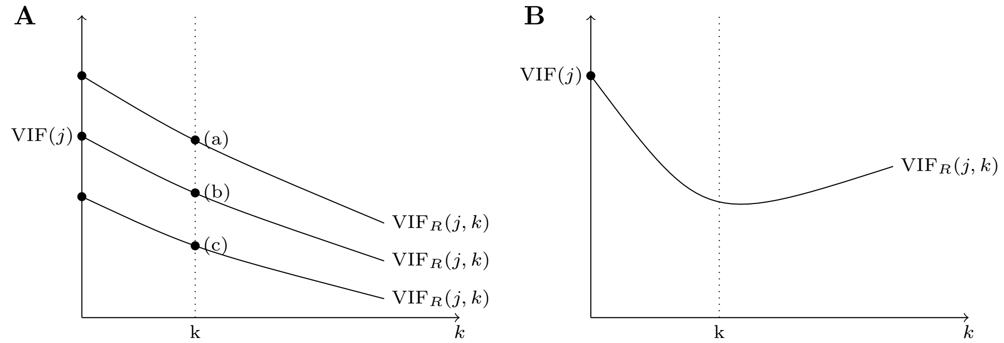

Figure 2: Continuity and monotonicity of VIF with the parameter k

- In scenario (a) of Figure 2A, the VIF$_R(j,k)$ obtained will be higher than **the values of the augmented VIF** (scenario (b)) and the conclusion will be that the problem of multicollinearity persists when it may actually have been solved.
- In scenario (c) of Figure 2A, the VIF$_R(j,k)$ obtained will be less than **the values of the augmented VIF** (scenario (b)) and the conclusion will be that the problem is solved when it actually still persists.
- In Figure 2B, the lack of monotonicity of VIF in RE allows the user to overshoot the "good" values and end up using a "bad" ridge parameter.

Note that an appropriate collinearity diagnostic will be obtained in scenario (b) of Figure 2A since the very desirable properties of continuity and monotonicity are verified.

4.1 Extensions of the VIF

To calculate the VIF in RE, we need to describe a new auxiliary regression as:

$$\mathbf{X}_j^R = \mathbf{X}_{-j}^R \boldsymbol{\delta} + \boldsymbol{\nu}, \tag{14}$$

where $\mathbf{X}_j^R = \left(X_{1j} \ X_{2j} \ \cdots \ X_{nj} \ 0 \ \cdots \ \sqrt{k} \ \cdots 0 \right)'$ and $\sqrt{k}$ is in the position $n+j$ of the vector $\mathbf{X}_j^R$, $j = 2, \dots, p$. Furthermore, the mean of the vector $\mathbf{X}_j^R$ is $\overline{\mathbf{X}}_j^R = \frac{n\overline{\mathbf{X}}_j + \sqrt{k}}{n+p}$.
Then, due to (note that $\widehat{\mathbf{X}}_j^R = \widehat{\mathbf{X}}_j$ for $j = 2, \dots, p$):

$$\begin{aligned} \text{ESS}(j,k) &= \sum_{i=1}^{n+p} \left(\widehat{X}_{ij}^R - \overline{\mathbf{X}}_j^R \right)^2 = \sum_{i=1}^{n} \left(\widehat{X}_{ij} - \overline{\mathbf{X}}_j^R \right)^2 + \sum_{i=n+1}^{n+p} \left(\widehat{X}_{ij}^R - \overline{\mathbf{X}}_j^R \right)^2, \\ \text{TSS}(j,k) &= \sum_{i=1}^{n+p} \left(X_{ij}^R - \overline{\mathbf{X}}_j^R \right)^2 = \sum_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j^R \right)^2 + (p-1) \left(\overline{\mathbf{X}}_j^R \right)^2 + \left(\sqrt{k} - \overline{\mathbf{X}}_j^R \right)^2, \end{aligned}$$

the coefficient of determination of the auxiliary regression (14) is:

$$R^2(j,k) = \frac{\sum\limits_{i=1}^{n} \left(\widehat{X}_{ij} - \overline{\mathbf{X}}_j^R \right)^2 + \sum\limits_{i=n+1}^{n+p} \left(\widehat{X}_{ij}^R - \overline{\mathbf{X}}_j^R \right)^2}{\sum\limits_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j^R \right)^2 + (p-1) \left(\overline{\mathbf{X}}_j^R \right)^2 + \left(\sqrt{k} - \overline{\mathbf{X}}_j^R \right)^2}. \tag{15}$$

On the other hand, if we transform the variables of model (14) by means of $\mathbf{x}_j^R = \dfrac{\mathbf{X}_j^R - a_j}{b_j}$ with $j = 2, \dots, p$; $a_j \in \mathbb{R}$ and $b_j > 0$, that is

$$\mathbf{x}_j^R = \left(\frac{X_{1j} - a_j}{b_j} \ \frac{X_{2j} - a_j}{b_j} \ \cdots \ \frac{X_{nj} - a_j}{b_j} \ 0 \ \cdots \ \sqrt{k} \ \cdots 0 \right)', \tag{16}$$

we will obtain the following transformed model:

$$\mathbf{x}_j^R = \mathbf{x}_{-j}^R \boldsymbol{\delta} + \boldsymbol{\nu}. \tag{17}$$

To calculate the coefficient of determination of the auxiliary regression (17), from

$$\overline{\mathbf{x}}_j^R = \frac{n(\overline{\mathbf{X}}_j - a_j) + b_j\sqrt{k}}{b_j(n+p)}, \tag{18}$$

and (16), we obtain that:

$$\begin{aligned}
\widehat{x}_{ij}^R - \overline{\mathbf{x}}_j^R &= \left\{ \begin{array}{ll} \frac{1}{b_j}\left(\widehat{X}_{ij} - a_j - b_j\overline{\mathbf{x}}_j^R\right), & i = 1,\dots,n \\ \frac{1}{b_j}\left(b_j\widehat{X}_{ij}^R - b_j\overline{\mathbf{x}}_j^R\right), & i = n+1,\dots,n+p \end{array} \right. , \\
x_{ij}^R - \overline{\mathbf{x}}_j^R &= \left\{ \begin{array}{ll} \frac{1}{b_j}\left(X_{ij} - a_j - b_j\overline{\mathbf{x}}_j^R\right), & i = 1,\dots,n \\ -\overline{\mathbf{x}}_j^R, & i = n+1,\dots,n+p,\ i \neq n+j \\ \sqrt{k} - \overline{\mathbf{x}}_j^R, & i = n+j \end{array} \right. .
\end{aligned}$$

Taking into account that

$$\begin{aligned}
\mathrm{ESS}_T(j,k) &= \sum_{i=1}^{n+p} \left(\widehat{x}_{ij}^R - \overline{\mathbf{x}}_j^R\right)^2 = \frac{1}{b_j^2}\left[\sum_{i=1}^{n}\left(\widehat{X}_{ij} - a_j - b_j\overline{\mathbf{x}}_j^R\right)^2 + b_j^2 \sum_{i=n+1}^{n+p}\left(\widehat{X}_{ij}^R - \overline{\mathbf{x}}_j^R\right)^2\right], \\
\mathrm{TSS}_T(j,k) &= \sum_{i=1}^{n+p} \left(x_{ij}^R - \overline{\mathbf{x}}_j^R\right)^2 = \frac{1}{b_j^2}\left[\sum_{i=1}^{n}\left(X_{ij} - a_j - b_j\overline{\mathbf{x}}_j^R\right)^2 + (p-1)b_j^2\left(\overline{\mathbf{x}}_j^R\right)^2 + \left(\sqrt{k} - b_j\overline{\mathbf{x}}_j^R\right)^2\right],
\end{aligned}$$

we obtain that:

$$R_T^2(j,k) = \frac{\sum\limits_{i=1}^{n}\left(\widehat{X}_{ij} - a_j - b_j\overline{\mathbf{x}}_j^R\right)^2 + b_j^2\sum\limits_{i=n+1}^{n+p}\left(\widehat{X}_{ij}^R - \overline{\mathbf{x}}_j^R\right)^2}{\sum\limits_{i=1}^{n}\left(X_{ij} - a_j - b_j\overline{\mathbf{x}}_j^R\right)^2 + (p-1)b_j^2\left(\overline{\mathbf{x}}_j^R\right)^2 + \left(\sqrt{k} - b_j\overline{\mathbf{x}}_j^R\right)^2}. \tag{19}$$

It is evident that expressions (15) and (19) only coincide when $a_j = 0$ and $b_j = 1$, $j = 2,\dots,p$, that is, when no transformation is applied. Thus, the coefficient of determination of models (14) and (17) will be different and, by extension, the VIFs obtained from them will also be different.
The question is which coefficient of determination has to be applied in the definition of (12). We will answer this topic in the next subsection.

4.2 Selecting the correct extension of VIF

We have just seen that the coefficients of determination of the auxiliary regression in RE for the original and transformed data do not match. Therefore, the VIFs obtained from them will be different. At this point, we wonder whether or not it is necessary to standardize the data to calculate the VIF in the ridge estimation or, in another way, which of the two coefficients of determination is correct.
The RE coincides with the OLS estimation when $k = 0$. Thus, to determine the need to standardize the data we must prove which auxiliary regression, (14) or (17), provides a coefficient of determination that is equal to $R_T^2(j)$ and $R^2(j)$ when $k \rightarrow 0$. **Indeed, the selected option should lead to an augmented VIF higher than one and monotone.**
For $k = 0$, the coefficient of determination of the regression of model (14), that is, expression (15), will be given by:

$$R^2(j,0) = \frac{\sum\limits_{i=1}^{n}\left(\widehat{X}_{ij} - \frac{n}{n+p}\overline{\mathbf{X}}_j\right)^2 + \sum\limits_{i=n+1}^{n+p}\left(\widehat{X}_{ij}^R - \frac{n}{n+p}\overline{\mathbf{X}}_j\right)^2}{\sum\limits_{i=1}^{n}\left(X_{ij} - \frac{n}{n+p}\overline{\mathbf{X}}_j\right)^2 + p\left(\frac{n}{n+p}\overline{\mathbf{X}}_j\right)^2}. \tag{20}$$

This expression only coincides with the coefficient of determination $R_T^2(j)$ and $R^2(j)$ when $p = 0$, that is, when there are no independent variables in the model, **which is not possible.**
On the other hand, for $k = 0$, the coefficient of determination of the regression of the model (17), that is, expression (19), will be given by:

$$R_T^2(j,0) = \frac{\sum_{i=1}^{n}\left(\widehat{X}_{ij} - a_j - \frac{n(\overline{\mathbf{X}}_j - a_j)}{n+p}\right)^2 + \sum_{i=n+1}^{n+p}\left(b_j\widehat{X}_{ij}^R - \frac{n(\overline{\mathbf{X}}_j - a_j)}{(n+p)}\right)^2}{\sum_{i=1}^{n}\left(X_{ij} - a_j - \frac{n(\overline{\mathbf{X}}_j - a_j)}{n+p}\right)^2 + p\left(\frac{n(\overline{\mathbf{X}}_j - a_j)}{n+p}\right)^2}. \tag{21}$$

Taking into account the particular case when $a_j = \overline{\mathbf{X}}_j$ (centered data), $j = 2, \dots, p$, the expression (21) becomes:

$$R_T^2(j,0) = \frac{\sum_{i=1}^{n}\left(\widehat{X}_{ij} - \overline{\mathbf{X}}_j\right)^2 + b_j^2\sum_{i=n+1}^{n+p}\left(\widehat{X}_{ij}^R\right)^2}{\sum_{i=1}^{n}\left(X_{ij} - \overline{\mathbf{X}}_j\right)^2}. \tag{22}$$

We can conclude that $R_T^2(j,0) = R^2(j) = R_T^2(j)$ since:

$$\lim_{k \to 0} \widehat{X}_{ij}^R = 0, \quad i = n+1, \dots, n+p,$$

due to $X_{ij}^R = 0$ for $i = n+1, \dots, n+p$, $i \neq n+j$ and $X_{ij}^R = \sqrt{k}$ for $i = n+j$.
That is, for $k = 0$ the coefficient of determination of the auxiliary regression of the RE with standardized data (although centered data will be enough) coincides with the one obtained by OLS estimation. Thus, the VIFs obtained from these coefficients of determination will also be equal.
In consequence, the augmented VIF defined by:

$$\mathrm{VIF}_R(j,k) = \begin{cases} \mathrm{VIF}(j), & \text{if } k = 0 \\ \dfrac{1}{1 - R_T^2(j,k)}, & \text{if } k > 0 \end{cases} \tag{23}$$

is continuous in $k = 0$. In addition, $\mathrm{VIF}_R(j,k)$ will be a decreasing function on k (see Appendix A), and due to:

$$\lim_{k \to +\infty} R_T^2(j,k) = \frac{\frac{1}{n+p}}{1 - \frac{1}{n+p}} = \frac{1}{n+p-1},$$

it is evident that $\mathrm{VIF}_R(j,k)$ will always be higher than one since it is verified that:

$$\lim_{k \to +\infty} \frac{1}{1 - R_T^2(j,k)} = \frac{n+p-1}{n+p-2} > 1.$$

We therefore conclude that standardizing the data before augmenting the model leads to a definition of VIF which verifies the properties of continuity, monotonicity and provides values higher than one. Hence, we consider that standardizing is not optional but required. This fact explains why the surrogate ridge estimator obtained from a standardized model performs better than RE.

5 The effects of several factors on the variance of the ridge regression coefficients

O'Brien (2007) analyzed the effects of several factors on the variance of the regression coefficients in OLS. We will develop this same analysis in the context of RE taking into account that the correct auxiliary regression to calculate the $\mathrm{VIF}_R(j,k)$ is (17).
For this purpose we start from the variance of the jth regression coefficient provided by Greene (1993):

$$\mathrm{var}\left(\widehat{\beta}_R(j,k)\right) = \sigma_{\mathbf{u}^R}^2\left[\left(\mathbf{x}_j^R\right)'\mathbf{M}_{-j}\mathbf{x}_j^R\right]^{-1}, \tag{24}$$

where $\mathbf{x}_j^R$ is the $n \times 1$ vector representing the values of the $n+p$ observations on the jth independent variable (see expression (16)) and:

$$\mathbf{M}_{-j} = \mathbf{I} - \mathbf{x}_{-j}^R \left[\left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_{-j}^R \right]^{-1} \mathbf{x}_{-j}^R = \mathbf{I} - \mathbf{N}_{-j}. \tag{25}$$

We consider that $\mathbf{x}_j^R$ is centered around its mean denoted by $\widetilde{\mathbf{x}}_j^R$, (see O'Brien, 2007). Then

$$\left(\widetilde{\mathbf{x}}_j^R\right)' \mathbf{M}_{-j} \widetilde{\mathbf{x}}_j^R = \left(\widetilde{\mathbf{x}}_j^R\right)' \widetilde{\mathbf{x}}_j^R - \left(\widetilde{\mathbf{x}}^R\right)' \mathbf{N}_{-j} \widetilde{\mathbf{x}}_j^R = \left(\widetilde{\mathbf{x}}_j^R\right)' \widetilde{\mathbf{x}}_j^R \left(1 - \frac{\left(\widetilde{\mathbf{x}}_j^R\right)' \mathbf{N}_{-j} \widetilde{\mathbf{x}}_j^R}{\left(\widetilde{\mathbf{x}}_j^R\right)' \widetilde{\mathbf{x}}_j^R} \right), \tag{26}$$

where it is verified that $\frac{\left(\widetilde{\mathbf{x}}_j^R\right)' \mathbf{N}_{-j} \widetilde{\mathbf{x}}_j^R}{\left(\widetilde{\mathbf{x}}_j^R\right)' \widetilde{\mathbf{x}}_j^R}$ is the coefficient of determination of the regression (17), $R_T^2(j,k)$.

Following O'Brien (2007), if we consider that $a_j = \overline{\mathbf{X}}_j$ and $b_j = 1$:

$$\left(\widetilde{\mathbf{x}}_j^R\right)' \widetilde{\mathbf{x}}_j^R = \sum_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j\right)^2 + \frac{(n+p-1)^2 + p}{(n+p)^2} \cdot k, \tag{27}$$

then expression (24) could be rewritten as:

$$\text{var}\left(\widehat{\beta}_R(j,k)\right) = \frac{\sigma_{\mathbf{u}^R}^2}{\left[\sum\limits_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j\right)^2 + \frac{(n+p-1)^2+p}{(n+p)^2} \cdot k \right] \left(1 - R_T^2(j,k)\right)}. \tag{28}$$

On the other hand, an unbiased estimation of $\sigma_{\mathbf{u}^R}^2$ can be obtained from:

$$\widehat{\sigma}_{\mathbf{u}^R}^2 = \frac{\left(1 - R^2(k)\right) \text{TSS}(k)}{(n+p) - p} = \frac{\left(1 - R^2(k)\right) \left[\sum\limits_{i=1}^{n} Y_i^2 - \frac{n^2}{n+p} \overline{\mathbf{Y}}^2 \right]}{n}, \tag{29}$$

where $\text{TSS}(k)$ is the total sum of squares of model (11). Thus, the unbiased estimator of the variance of the jth ridge regression coefficient is obtained by substituting expression (29) in expression (28):

$$\widehat{\text{var}}\left(\widehat{\beta}_R(j,k)\right) = \frac{\left(1 - R^2(k)\right) \left[\frac{1}{n} \sum\limits_{i=1}^{n} Y_i^2 - \frac{n}{n+p} \overline{\mathbf{Y}}^2 \right]}{\left[\sum\limits_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j\right)^2 + \frac{(n+p-1)^2+p}{(n+p)^2} \cdot k \right] \left(1 - R_T^2(j,k)\right)}. \tag{30}$$

From (30) note that:

- Since $R^2(k)$ is a decreasing function on k (see McDonald, 2010), it is evident that $1 - R^2(k)$ will be an increasing function on k. Hence, it is acting as a variance inflator factor.
- Since $R_T^2(j,k)$ is a decreasing function on k (see Appendix A), it is evident that $1 - R_T^2(j,k)$ is an increasing function on k. Since it is placed in the denominator it is acting as a variance deflator factor or, what it is the same, $\text{VIF}_R(j,k)$ is acting as a variance inflator factor.
- The parameter k associated to the ridge regression appears in the denominator of expression (30) and it will then be verified that

  $$\lim_{k \to +\infty} \widehat{\text{var}}\left(\widehat{\boldsymbol{\beta}}_R(j,k)\right) = 0,$$

  since $0 \leq R^2(k), R_T^2(j,k) \leq 1$.
- **By comparing expression (30) with the expression of O'Brien (2007) for model (2) we can see that $1 - R^2$ changes from acting as a variance deflator factor in OLS to $1 - R^2(k)$ acting as a variance inflator factor in RE. Note that $R^2(k)$ is decreasing in k.**

6 Numerical example

We conclude this paper with an example to illustrate the above results. We analyze the data previously used by Kunugi et al. (1961), Himmelblau (1970) and Marquardt & Snee (1975), among others. They studied the relation between the conversion of n-heptane to acetylene (%), $\mathbf{Y}$, and the reactor temperature (°C), $\mathbf{X}_1$, the ratio of H_2 to n-heptane (mole ratio), $\mathbf{X}_2$, and the contact time (sec), $\mathbf{X}_3$, with a quadratic model considering standardized and unstandardized data.

They obtained the VIF from the definition given by Marquardt (1970), who considered that the VIFs are the diagonal elements of the inverse of the simple correlation matrix. Since the correlation matrix of a data set coincides with the one of the same -standardized- data, then the VIFs obtained from this definition must be the same in both cases. Thus, the question is why the VIFs presented in Table 1 are not equal.

The problem is that they do not standardize all the variables but only $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$. From this standardization they obtain the rest of the variables $\mathbf{X}_4$, $\mathbf{X}_5$, $\mathbf{X}_6$, $\mathbf{X}_7$, $\mathbf{X}_8$ and $\mathbf{X}_9$ that may not be standardized. This causes the matrix $\mathbf{X}'\mathbf{X}$ to not coincide with the correlation matrix. In this case, it is not correct to take the diagonal elements of this matrix as the VIFs (as many authors have done). In addition, the correlation matrix of these new variables is different from the original one, thus leading to different results for the calculation of the VIFs.

In contrast, as shown in section 3, the VIFs will coincide if all the variables are standardized and we calculate the VIF from expression (3) or from Marquardt's proposal. Note that when estimating standardized data by OLS, both expressions are similar and are presented in the first row of Table 1. From the obtained VIFs, we can clearly conclude that the model has a serious problem of collinearity. Thus, it will be appropriate to consider the ridge estimation as the adequate estimation method. Since the estimation of this model will depend on k, it will be necessary to obtain the value of the VIF associated to each value of k to prove whether the problem of multicollinearity was solved.

Table 2 shows the VIFs associated to the RE, obtained from the original data and its standardization. As shown in section 4, the VIFs obtained from these standardized data in the ridge estimation when $k = 0$ will coincide with the one obtained by using OLS. Thus, these VIFs have to be considered when applying RE. If we represent the values of Table 2 we will obtain a figure similar to 2A. As an example, in Figure 3 we represent the VIFs obtained from original and standardized data for variables $\mathbf{X}_2$, $\mathbf{X}_4$, $\mathbf{X}_6$ and $\mathbf{X}_8$. In addition, we have highlight the value of VIF in OLS. It is evident that the VIFs obtained from standardized data verify all the required conditions presented in section 4.

Table 3 shows that, when using standardized data, all variables have a VIF less than 10 for $k \geq 0.09$. Therefore, they are within the established threshold and we can conclude that the problem of multicollinearity has been solved. On the other hand, note that when working with original data, the VIF values of the variables $\mathbf{X}_3$ and $\mathbf{X}_9$ increase from values of k equal to 0.7 and 0.1, respectively. This means that the VIFs are not monotonically decreasing as a function of the ridge parameter k when working with original data, contrary to what happens when working with standardized data. This inconsistency was previously shown by Jensen & Ramirez (2008) and summarized by Hadi (2011): "they observe that (a) the condition of the variance of $||\hat{\boldsymbol{\beta}}_S(k)||$ is monotonically increasing in k and (b) the maximum variance inflation factor is monotonically decreasing in k. These properties do not hold for the classical ridge estimator, $\hat{\boldsymbol{\beta}}(k)$". In relation to the first statement, Marquardt (1963, Theorem 2) showed that $||\hat{\boldsymbol{\beta}}_R(k)||^2$ decreases monotonically to 0 as $k \to \infty$. This is also stated in Jensen & Ramirez (2008). However, their example seems to contradict this fact. For their example however, Jensen & Ramirez (2008) followed a common rescaling convention of first transforming the data from their original units into standard "correlation form" units where the ridge regression was conducted and then transforming from standard units back to the original units. Under this convention $||\hat{\boldsymbol{\beta}}_S(k)||^2$ was not monotone. **The correct formulation is given in Kapat & Goel (2010) and acknowledged by Jensen & Ramirez (2010a).** As regards the second statement, which is verified in the surrogate ridge model, it should be noted that if the VIF in RE is a nondecreasing function in k, it is due to an incorrect extension of the VIF from OLS to RE. This work shows that with a correct extension, the VIF will be monotonically decreasing in k.

Finally, we have calculated the VIFs from the R statistical environment (see R Development Core Team, 2012) using the *genridge* package (see Table 4). Although for $k = 0$, we obtained the VIFs corresponding to OLS, clearly they are not monotonically decreasing in k and therefore do not verify one of the conditions required in definition (12). Thus, this package is not recommended as it presents an incorrect way to calculate the VIF in RE.

	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$
Unstandardized	2856748.93	10956.14	2017162.52	9802.9	1428091.88	240.36	2501944.59	65.73	12667.1
Standardizing only $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$	375.25	1.74	680.28	31.04	6563.35	35.61	1762.58	3.16	1156.77

Table 1: Acetylene data VIF by Marquardt & Snee (1975) ($\mathbf{X}_4 = \mathbf{X}_1\mathbf{X}_2$, $\mathbf{X}_5 = \mathbf{X}_1\mathbf{X}_3$, $\mathbf{X}_6 = \mathbf{X}_2\mathbf{X}_3$, $\mathbf{X}_7 = \mathbf{X}_1^2$, $\mathbf{X}_8 = \mathbf{X}_2^2$, $\mathbf{X}_9 = \mathbf{X}_3^2$)

VIF for original data

k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$
0	46241.194	31155.843	21174.411	26632.669	12092.525	363.281	39045.694	90.128	738.840
0.1	3330.439	2886.990	1.195	2824.681	40.478	26.380	3037.278	52.216	1.094
0.2	2580.023	1755.559	1.112	1823.559	29.539	14.179	2378.925	52.160	1.098
0.3	2298.633	1339.299	1.092	1451.323	25.633	9.860	2130.471	52.112	1.101
0.4	2145.607	1117.430	1.084	1250.385	23.626	7.656	1994.161	52.068	1.103
0.5	2046.476	976.546	1.082	1121.034	22.403	6.322	1904.979	52.027	1.105
0.6	1975.356	877.385	1.081	1028.720	21.579	5.428	1840.343	51.990	1.107
0.7	1920.838	802.712	1.081	958.260	20.987	4.788	1790.302	51.954	1.108
0.8	1877.084	743.752	1.082	901.907	20.540	4.307	1749.762	51.922	1.110
0.9	1840.783	695.550	1.083	855.279	20.190	3.934	1715.831	51.891	1.111
1	1809.902	655.089	1.084	815.702	19.910	3.634	1686.733	51.862	1.112

VIF for standardized data

k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$
0	2856748.930	10956.136	2017162.520	9802.903	1428091.880	240.359	2501944.590	65.734	12667.100
0.1	7.525	7.311	8.638	7.315	8.661	4.065	7.135	6.117	5.941
0.2	4.448	4.111	4.832	4.136	4.858	3.014	4.288	3.722	3.763
0.3	3.359	3.021	3.541	3.046	3.558	2.485	3.267	2.821	2.936
0.4	2.792	2.472	2.888	2.497	2.901	2.166	2.731	2.349	2.489
0.5	2.442	2.143	2.495	2.166	2.504	1.952	2.398	2.058	2.206
0.6	2.204	1.925	2.231	1.946	2.239	1.799	2.170	1.861	2.011
0.7	2.031	1.770	2.043	1.789	2.049	1.685	2.004	1.720	1.867
0.8	1.900	1.654	1.901	1.672	1.906	1.596	1.878	1.614	1.757
0.9	1.798	1.565	1.791	1.582	1.796	1.525	1.779	1.532	1.670
1	1.715	1.494	1.703	1.510	1.707	1.468	1.699	1.466	1.600

Table 2: VIFs values for unstandardized and standardized data in ridge estimation for $k = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$.

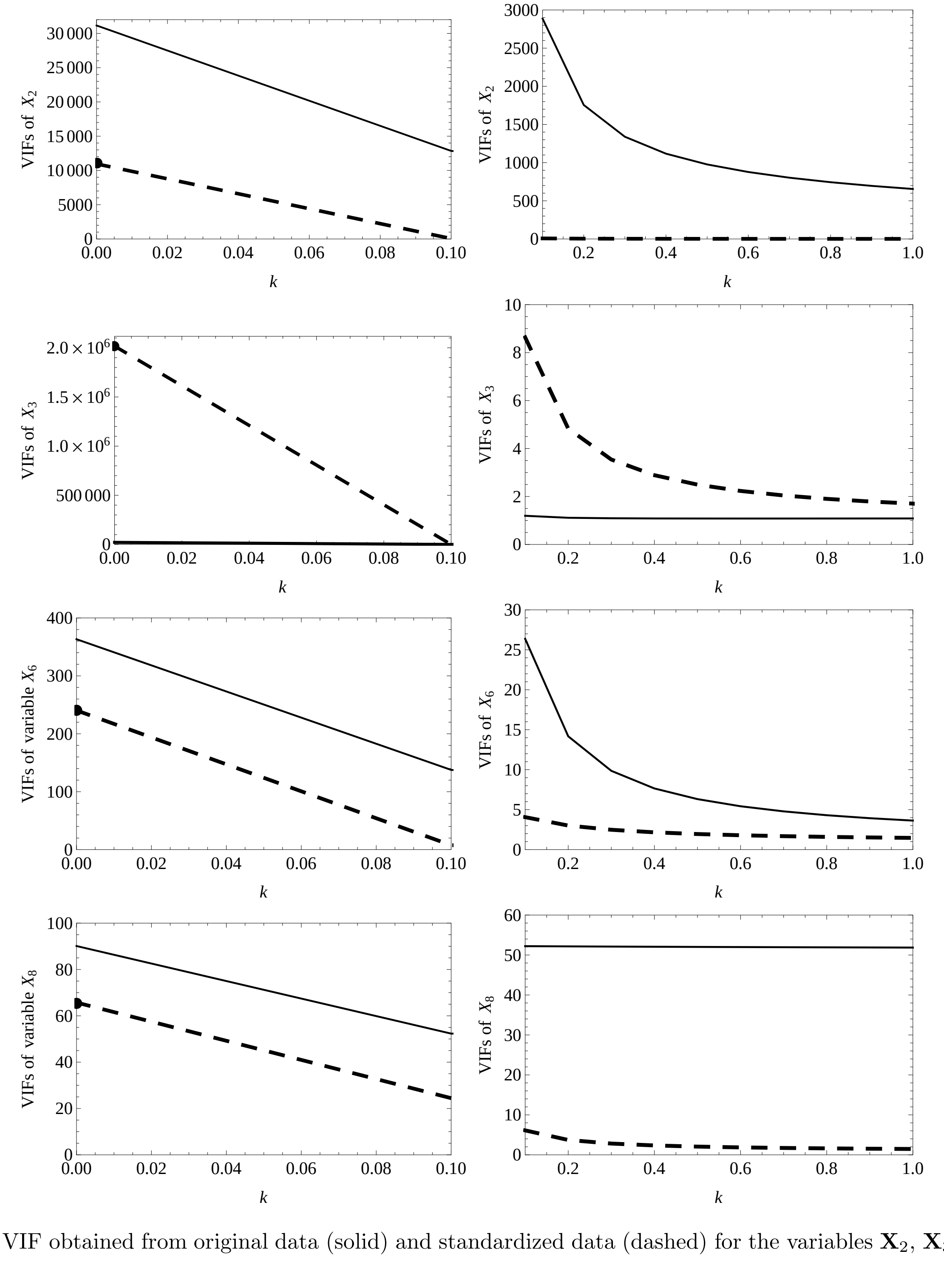

Figure 3: VIF obtained from original data (solid) and standardized data (dashed) for the variables $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_6$ and $\mathbf{X}_8$.

VIF for original data

k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$
0	46241.194	31155.843	21174.411	26632.669	12092.525	363.281	39045.694	90.128	738.840
0.01	9824.357	12847.914	3.297	11576.140	139.951	137.681	8726.421	52.323	1.086
0.02	7266.746	8913.011	2.078	8122.192	100.092	93.394	6484.656	52.292	1.086
0.03	5946.033	6884.999	1.690	6341.108	79.882	70.689	5327.651	52.275	1.087
0.04	5138.868	5647.617	1.504	5253.869	67.623	56.891	4620.758	52.263	1.089
0.05	4593.985	4813.607	1.396	4520.704	59.389	47.622	4143.653	52.253	1.090
0.06	4201.079	4213.124	1.326	3992.560	53.476	40.970	3799.650	52.245	1.091
0.07	3904.109	3759.941	1.278	3593.754	49.023	35.966	3539.643	52.237	1.092
0.08	3671.579	3405.631	1.243	3281.777	45.549	32.065	3336.042	52.230	1.092
0.09	3484.428	3120.902	1.216	3030.910	42.762	28.940	3172.152	52.223	1.093
0.1	3330.439	2886.990	1.195	2824.681	40.478	26.380	3037.278	52.216	1.094

VIF for standardized data

k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$
0	2856748.930	10956.136	2017162.520	9802.903	1428091.880	240.359	2501944.590	65.734	12667.100
0.01	57.206	57.586	72.766	56.382	67.312	7.646	52.898	24.459	31.437
0.02	30.099	30.636	37.809	30.233	36.455	6.538	27.841	17.404	19.060
0.03	20.870	21.249	25.834	21.046	25.313	5.936	19.365	13.783	14.114
0.04	16.190	16.412	19.763	16.295	19.530	5.504	15.077	11.512	11.413
0.05	13.348	13.445	16.088	13.374	15.980	5.161	12.475	9.938	9.696
0.06	11.433	11.434	13.622	11.391	13.576	4.877	10.721	8.778	8.501
0.07	10.052	9.978	11.851	9.953	11.838	4.633	9.456	7.885	7.617
0.08	9.007	8.875	10.516	8.863	10.522	4.420	8.497	7.176	6.934
0.09	8.186	8.009	9.475	8.006	9.491	4.233	7.743	6.598	6.388
0.1	7.525	7.311	8.638	7.315	8.661	4.065	7.135	6.117	5.941

Table 3: VIFs values for unstandardized and standardized data in ridge estimation for $k = 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1$.

k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_7$	$\mathbf{X}_8$	$\mathbf{X}_9$
0	2856748.9693	10956.136110	2017162.5388	9802.90279	1428091.896	240.359383	2501944.6311	65.733586	12667.09955
0.1	457.3444	8.570092	457.3636	10.12427	821.516	6.377054	312.2327	18.503457	22.47650
0.2	1098.6515	9.645159	743.6048	10.21393	1265.642	5.821969	891.6320	11.683160	21.68637
0.3	1925.9476	11.981106	1009.6772	12.17705	1589.838	5.521331	1698.2650	8.424357	26.74608
0.4	2904.0413	14.440581	1282.4596	14.46739	1867.126	5.361812	2681.7054	6.516822	33.80814
0.5	4002.1154	16.798451	1567.3391	16.77719	2121.404	5.298490	3803.8289	5.286679	42.03165
0.6	5194.8071	19.015065	1864.4789	19.02933	2362.830	5.305343	5034.5976	4.443063	51.06403
0.7	6461.3244	21.101708	2172.5606	21.21076	2596.428	5.364978	6349.8687	3.838568	60.69493
0.8	7784.6074	23.082649	2489.9029	23.32803	2824.912	5.464927	7730.0531	3.390796	70.77507
0.9	9150.6505	24.983173	2814.8474	25.39295	3049.836	5.595893	9159.1803	3.050399	81.18970
1	10547.9554	26.825937	3145.8946	27.41780	3272.125	5.750775	10624.1975	2.786225	91.84724

Table 4: Regression ridge VIF from *genridge* package

7 Conclusions

As shown, model (11) has been applied to obtain the VIFs from the matrix $\mathbf{X}^R$ following Theil (1971). In this paper we have obtained the following conclusions:

- If we use the original data without standardization, the VIFs will not verify the conditions of continuity and monotonic **decreasing** in k, which are required for the classical definition of VIF. Recall the comments of Hadi (2011) about the surrogate ridge model, which does verify the condition of monotonicity, contrary to what happens to the ridge estimator applied until now.
- The lack of monotonicity of the VIF makes the selection of the parameter k difficult since we may select a parameter k when the VIF is increasing, in contrast to what is intended.
- The lack of continuity of VIF for $k = 0$ can result in a misleading conclusion in multicollinearity diagnostic since for any value of $k > 0$ the behavior of VIF is uncontrolled, see Figure 2.

Summarizing, in multicollinearity diagnostic in RE the standardization of data is not optional. For a correct diagnostic of multicollinearity using the VIF, the data have to be standardized. This is in line with Marquardt (1980), who stated that "The quality of the predictor variable (here regressors) structure of a data set can be assessed properly only in terms of a standardized scale". Although Marquardt (1980) extends this conclusion to OLS and RE, we humbly think that the standardization is only required in RE.

On the other hand, Jensen & Ramirez (2008, p. 102, Table 8) analyzed the behavior of the VIFs associated to the surrogate ridge model that verify the condition of continuity and monotonicity, but this is mainly due to the fact that they use standardized data in addition to other advantages of surrogate ridge regression.

The obligation to standardize data in order to calculate the VIF in RE is also confirmed when extending the variance proposed by O'Brien (2007) in OLS to RE. The similarity of both expressions only occurs when starting from standardized data in model (11), thus reinforcing the requirement to standardize data.

A $\mathbf{VIF}_R(j, k)$ is a decreasing function of k

McDonald (2010) showed that the coefficient of determination in Ridge Estimation and with standardized variables is a decreasing function in k. When estimating model (11) by OLS, the ridge estimator is obtained. The coefficient of determination obtained will then also be decreasing in k. We will demonstrate this issue below.

The coefficient of determination of model (17) is obtained from:

$$R_T^2(j,k) = \frac{\widehat{\boldsymbol{\delta}}(k)' \left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_j - (n+p)\left(\overline{\mathbf{x}}_j^R\right)^2}{\left(\mathbf{x}_j^R\right)' \mathbf{x}_j^R - (n+p)\left(\overline{\mathbf{x}}_j^R\right)^2}, \tag{31}$$

where $\overline{\mathbf{x}}_j = \frac{n(\mathbf{X}_j - a_j) + b_j\sqrt{k}}{b_j(n+p)}$. Since it is verified that:

- $\widehat{\boldsymbol{\delta}}(k) = \left(\left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_{-j}^R\right)^{-1} \left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_j^R \Rightarrow \left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_j^R = \left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_{-j}^R \widehat{\boldsymbol{\delta}}(k)$.
- $\left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_{-j}^R = \mathbf{x}_{-j}'\mathbf{x}_{-j} + k\mathbf{I}$, $\left(\mathbf{x}_{-j}^R\right)' \mathbf{x}_j^R = \mathbf{x}_{-j}'\mathbf{x}_j$ and $\left(\mathbf{x}_j^R\right)' \mathbf{x}_j^R = \mathbf{x}_j'\mathbf{x}_j + k$.
- $a_j = \overline{\mathbf{X}}_j,\ \forall j \Rightarrow \overline{\mathbf{x}}_j^R = \frac{\sqrt{k}}{n+p} \Rightarrow (n+p)\left(\overline{\mathbf{x}}_j^R\right)^2 = \frac{k}{n+p}$.

we obtain that expression (31) for centered data is given by:

$$R_T^2(j,k) = \frac{\widehat{\boldsymbol{\delta}}(k)'\mathbf{x}_{-j}'\mathbf{x}_{-j}\widehat{\boldsymbol{\delta}}(k) + k\widehat{\boldsymbol{\delta}}(k)'\widehat{\boldsymbol{\delta}}(k) - \frac{k}{n+p}}{\mathbf{x}_j'\mathbf{x}_j + \frac{n+p-1}{n+p}k}, \tag{32}$$

where $\widehat{\boldsymbol{\delta}}(k) = \left(\mathbf{x}_{-j}'\mathbf{x}_{-j} + k\mathbf{I}\right)^{-1} \mathbf{x}_{-j}'\mathbf{x}_j$.

Given that $\mathbf{x}'_{-j}\mathbf{x}_{-j}$ is a symmetric and positive defined matrix, we can affirm the existence of a matrix $\mathbf{\Gamma}$, which is orthogonal ($\mathbf{\Gamma}' = \mathbf{\Gamma}^{-1}$) and a diagonal matrix $\mathbf{D}_{\lambda_h}$, with $h = 1, \dots, p-1$, composed by the latent roots of $\mathbf{x}'_{-j}\mathbf{x}_{-j}$ verifying that $\mathbf{x}'_{-j}\mathbf{x}_{-j} = \mathbf{\Gamma}\mathbf{D}_{\lambda_h}\mathbf{\Gamma}'$. In this case, if we note $\boldsymbol{\gamma} = \mathbf{x}'_{-j}\mathbf{x}_j$, we obtain that:

$$\begin{aligned}
\widehat{\boldsymbol{\delta}}(k)'\mathbf{x}'_{-j}\mathbf{x}_{-j}\widehat{\boldsymbol{\delta}}(k) &= \boldsymbol{\gamma}'\mathbf{\Gamma}\mathbf{D}_{\frac{\lambda_h}{(\lambda_h+k)^2}}\mathbf{\Gamma}'\boldsymbol{\gamma}, \\
k\widehat{\boldsymbol{\delta}}(k)'\widehat{\boldsymbol{\delta}}(k) &= \boldsymbol{\gamma}'\mathbf{\Gamma}\mathbf{D}_{\frac{k}{(\lambda_h+k)^2}}\mathbf{\Gamma}'\boldsymbol{\gamma}.
\end{aligned}$$

Noting $\boldsymbol{\alpha}' = \boldsymbol{\gamma}'\mathbf{\Gamma}$ it is obtained that expression (32) can be rewritten as::

$$R_T^2(j,k) = \frac{\sum\limits_{h=1}^{p-1} \frac{\alpha_h^2}{\lambda_h+k} - \frac{k}{n+p}}{\sum\limits_{i=1}^{n} x_{ij}^2 + \frac{n+p-1}{n+p}k}. \tag{33}$$

Let $\lambda_0 = \min\{\lambda_1, \dots, \lambda_n\}$ then for $k \in (-\lambda_0, \infty)$ this expression is differentiable on (λ_0, ∞), **and in particular for** $k = 0$. Then:

$$\frac{\partial R_T^2(j,k)}{\partial k} = \frac{-\left(\sum\limits_{h=1}^{p-1} \frac{\alpha_h^2}{(\lambda_h+k)^2} + \frac{1}{n+p}\right)\left(\sum\limits_{i=1}^{n} x_{ij}^2 + \frac{n+p-1}{n+p}k\right) - \frac{n+p-1}{n+p}\left(\sum\limits_{h=1}^{p-1} \frac{\alpha_h^2}{\lambda_h+k} - \frac{k}{n+p}\right)}{\left(\sum\limits_{i=1}^{n} x_{ij}^2 + \frac{n+p-1}{n+p}k\right)^2}. \tag{34}$$

Since the numerator of expression (33) is clearly negative, $R_T^2(j,k)$ is a decreasing function on k.
Thus $\text{VIF}_R(j,k)$ is a decreasing function of k.

Acknowledgements

We would like to thank referees for their detailed, encouraging and constructive reviews of our paper which clearly contributed to improving both the structure and the content of this manuscript.
This paper has been partially supported by the project 'Valoración de proyectos gubernamentales a largo plazo: obtención de la tasa social de descuento', reference: P09-SEJ-05404, Proyectos de Excelencia de la Junta de Andalucía and Fondos FEDER.

References

Anderson, T.W. (1985). A introduction to multivariate statistical analysis. London: John Wiley & Sons.

Belsley, D.A. (1982). Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise. *Journal of Econometrics*, **20**, 211–253.

Belsley, D.A., Kuh, E. & Welsch, R.E. (1980). Regression diagnostics. New York, John Wiley.

Dias, J. & Castro, J. (2011). The corrected VIF. *Journal of Applied Statistics*, **38**, 1499–1507.

Farrar, D.E. & Glauber, R.R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92–107.

Fox, J. & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, **87**, 178– 183.

García, C., García, J., López, M.D.M. & Salmerón, R. (2014). Collinearity: Revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics*. doi:http://dx.doi.org/10.1080/02664763.2014.980789.

Greene, W.H. (1993). Econometric analysis. 2nd edition. New York: Macmillan.

Gunst, F.R. (1984). Toward a balanced assessment of collinearity diagnostics. *The American Statistician*, **38**, 79–82.

Gunst, R.F. & Mason, R.L. (1977). Advantages of examining multicollinearities in regression analysis. *Biometrics*, 249–260.

Hadi, A.S. (2011). Ridge and Surrogate Ridge Regressions. International Encyclopedia of Statistical Science. New York, NY: Springer.

Himmelblau, D.M. (1970). Process Analysis by Statistical Methods. New York, N.Y.: John Wiley & Sons.

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

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

Irfan, M., Javed, M. & Ali Raza, M. (2013). Comparison of shrinkage regression methods for remedy of multicollinearity problem. *Middle East Journal of Scientific Research*, **14**, 570–579.

Jamal, N. & Rind, M.Q. (2007). Ridge regression: A tool to forecast wheat area and production. *Pakistan Journal Statistical and Operation Research*, **3**, 125–134.

Jensen, D.R. & Ramirez, D.E. (2008). Anomalies in the foundations of ridge regression. *International Statistical Review*, **76**, 89–105.

Jensen, D.R. & Ramirez, D.E. (2010a). Anomalites in ridge regression: addendum. *International Statistical Review*, **78**, 215–217.

Jensen, D.R. & Ramirez, D.E. (2010b). Surrogate models in ill-conditioned systems. *Journal of Statistical Planning and Inference*, **140**, 2069–2077.

Jensen, D.R. & Ramirez, D.E. (2013). Revision: Variance inflation in regression. *Advances in Decision Sciences*, **2013**, 1–15. doi:http://dx.doi.org/10.1155/2013/671204.

Kapat, P. & Goel, P.K. (2010). Anomalies in the foundations of ridge regression: Some clarifications. *International Statistical Review*, **78**, 209–215.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, **30**, 666–687.

Kunugi, T., Tamura, T. & Naito, T. (1961). New acetylene process uses hydrogen dilution. *Chemical Engineering Progress*, **57**, 43–49.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, **11**, 431–441.

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

Marquardt, D.W. (1980). A critique of some ridge regression methods: Comment. *Journal of the American Statistical Association*, **75**, 87–91.

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

Marquardt, W.D. (1987). Collinearity and least squares regression: Comment. *Statistical Science*, **2**, 84–85.

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

Montgomery, D.C. & Peck, E.A. (1982). Introduction to linear regression analysis. New York: John Wiley & Sons.

Myers, R.H. (1990). Classical and modern regression with applications. 2nd edition. Boston, Mass, USA: PWS-Kent.

O'Brien, R.M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, **41**, 673–690.

R Development Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Salmerón, R., García, C., López, M.D.M. & García, J. (2013). A note about the variance inflation factor and the ridge regression. 2nd International Conference of Informatics and Management Sciences. pp. 197–199.

Sardy, S. (2008). On the practice of rescaling covariates. *International Statistical Review*, **76**, 285–297.

Silvey, S. (1969). Multicollinearity and imprecise estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 539–552.

Smith, G. & Campbell, F. (1980). A critique of some ridge regression methods. *Journal of the American Statistical Association*, **75**, 74–81.

Stewart, G.W. (1987). Collinearity and least squares regression. *Statistical Science*, **2**, 68–84.

Theil, H. (1971). Principles of econometrics. New York: Wiley.

Velleman, P.F. & Welsch, R.E. (1981). Efficient computing of regression diagnostics. *The American Statistician*, **35**, 234–242.

Vinod, H.D. & Ullah, A. (1981). Recent Advances in Regression Methods. New York: M. Dekker.

Willan, A.R. & Watts, D.G. (1978). Meaningful multicollinearity measures. *Technometrics*, **20**, 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**, 571–588.