

# Collinearity diagnostic applied in ridge estimation through the VIF

## Abstract

The Variance Inflation Factor (VIF) is used to detect the presence of linear relationships between two or more independent variables (that is, collinearity) in the multiple linear regression model. However, the traditionally used VIF definitions encounter some problems when extended to the case of the Ridge Estimation (RE). This paper presents an extension of the VIF in RE by providing two alternative VIF expressions that overcome these problems in the general case. Some characteristics of these expressions are also presented and compared with the traditional expression. The results are illustrated with an economic example in the case of three independent variables and with a Monte Carlo simulation for the general case.

*Keywords:* multiple Linear Regression; Collinearity; Ridge Regression; Multivariables

## 1 Introduction

The collinearity problem is the existence of linear relationships between two or more independent variables in model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

where  $E[\mathbf{u}] = \mathbf{0}$ ,  $E[\mathbf{u}\mathbf{u}'] = \sigma^2\mathbf{I}$ . It is considered that the variables in (1) are standardized, therefore  $\mathbf{X}'\mathbf{X}$  is the correlation matrix and  $\mathbf{X}'\mathbf{Y}$  is the vector of correlation coefficients of the response variable with each of the explanatory variables. Collinearity will be perfect or approximate depending on the kind of relation. In the first case, the model does not satisfy the full range condition and has infinite solutions, while, in the second one, although the condition is fulfilled, the estimation will be unstable **and the variance of the estimators may be large compared to the values of the estimated parameters that can be insignificant or have the wrong sign**. Obviously, the second case is troubling. Remember that the collinearity is a data problem belonging to numerical analysis research area that can be found in any field. The Variance Inflation Factor (VIF) (Marquardt, 1970; Theil, 1971; Fox & Monette, 1992) has been widely applied in scientific literature to diagnose the existence of collinearity, although it is possible to find other measures such as: the eigenvalues through the Condition Index (CI), (Belsley et al., 1980; Belsley, 1982), the Condition Number (CN), (Belsley et al., 1980; Besley, 1991), the Variance Decomposition Proportions (VDP) to analyze the correlations between different vectors and their angles, (Wichers, 1975; Rawlings et al., 1998), the Red indicator (Kovacs et al., 2005), the corrected VIF (Dias & Castro, 2007), or the use of biplot method in the visualization diagnostic of multicollinearity problems called collinearity biplot, (Friendly & Kwan, 2009).

Focusing on the VIF, the definition provided by Theil (1971) allows to measure the impact of collinearity of the variable  $\mathbf{X}_i$ ,  $i = 1, \dots, p$ , with the rest of the independent variables:

$$\text{VIF}(i) = \frac{1}{1 - R_i^2}, \quad i = 1, \dots, p. \quad (2)$$

where  $R_i^2$  is the coefficient of determination of  $\mathbf{X}_i$  on the rest of independent variables. Marquardt (1970) defined the VIF as the elements of the principal diagonal of the inverse correlation matrix. Thus, the VIF will be the elements of the diagonal of the inverse of  $\mathbf{X}'\mathbf{X}$  since the correlation matrix of the independent variables is the matrix  $\mathbf{X}'\mathbf{X}$  when the data are standardized. This measure has well-known weaknesses that should be noted. Firstly, the controversy about the use of centered or not centered data that has been recently revised by García et al. (2015b). Secondly, the fact that there is no measure to know how closely  $R_i^2$  must be to 1 to imply collinearity (Besley, 1991). Furthermore, the VIF is not resistant to the presence of high leverage

points (outliers). Finally, it is generally accepted that values of VIF higher than 10 indicate severe collinearity (Kennedy, 1992) but this rule of thumb lacks a theoretical basis. **Indeed, taking into account the expression of the estimated variance of the estimated parameters**

$$\hat{\text{var}}(\hat{\beta}_i) = \frac{\hat{\sigma}_u^2}{\text{nvar}(X_i)} \cdot \text{VIF}(i), \quad i = 1, \dots, p \quad (3)$$

where  $n$  is the number of observations, high values of VIF could not imply high estimated variance since it can be countered by the ratio of the variance of the error terms divided by the variation in the respective independent variable. Thus, the variance of the error is also an important factor to get high variance and, for this reason, practitioners should apply statistical skills for model modification to minimise it and thereby the variance of the estimated parameters.

**Focusing on the diagnose, and not in the solution of the collinearity**, the VIF is widely applied as can be noted from the paper of O'Brien (2007) with more than 1700 cites in a great variety of fields. The concept of VIF was generalized in Ordinary Least Squares (OLS) by Fox & Monette (1992), who defined the Generalized Variance Inflation Factor (GVIF) as the measure of the impact of collinearity on the square of the length of the joint confidence region (two or more coefficients)

$$\text{GVIF} = \frac{|\mathbf{R}_1||\mathbf{R}_2|}{|\mathbf{R}|}, \quad (4)$$

where  $|\cdot|$  is the matrix determinant,  $\mathbf{R}_1$  is the correlation matrix of a particular set of regressors,  $\mathbf{X}_1$ ,  $\mathbf{R}_2$  is the correlation matrix of the rest of the regressors,  $\mathbf{X}_2$ , and  $\mathbf{R}$  is the correlation matrix of all the regressors. Note that when the number of variables in  $\mathbf{X}_1$  is equal to one, the initial expression is simplified and the GVIF coincides with the VIF (see, for example, Berk, 1977).

On the other hand, the Ridge Estimation (RE) is a widespread method to overcome the problem of collinearity defining a class of estimators depending on the non-negative scalar parameter  $\lambda$

$$\hat{\beta}(\lambda) = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{Y}. \quad (5)$$

Its covariance matrix is

$$\text{var}(\hat{\beta}(\lambda)) = \sigma^2 (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}. \quad (6)$$

The estimator given in (5) is a biased estimator when  $\lambda > 0$  and when  $\lambda = 0$  coincides with the OLS estimator. Despite earlier usage in numerical analysis (Levenberg, 1944; Riley, 1955), the ridge estimator is an interesting topic usually labeled in statistic and econometric and with applications in many different fields such as medicine, physics and chemistry. Indeed, McDonald (2009) conducted a detailed study of the scientific literature on the ridge estimator since its presentation in the seminal papers of Hoerl & Kennard (1970a,b) and concluded that more than 240 works have been published (see McDonald, 2009) in prestigious journals.

Once the parameter  $\lambda$  is selected and RE is applied, it is necessary to calculate again the value of a diagnostic measure to check if collinearity has been mitigated enough. This fact justifies the extension of the collinearity indicators to be applied after RE. García et al. (2015a) extended the VIF for the case  $p = 2$  and the condition number was extended by García et al. (2015c). The purpose of this article is to show the deficiencies obtained when applying in RE the definitions of VIF originally created to OLS estimation and propose an alternative expression for  $p > 2$  that verifies some desirable properties.

The structure is as follows: Section 2 presents the problems when applying the traditional VIF definitions in RE by using an example. In Section 3 the VIF of surrogate ridge model and the VIF expression from the vector that generate the matrix of the ridge estimators for  $p = 3$  are calculated. Some properties of both VIFs are also shown and is calculated an explicit expression based on the correlation coefficients for the GVIF when  $p = 3$  in RE. These expressions are used to analyze the presence of multicollinearity in real economic data in Section 4. The

results are compared with the extension of the condition number in RE. Due to the difficulty of obtaining expressions for these VIFs for  $p > 3$ , Section 5 shows how to study the presence of multicollinearity in a model with  $p > 3$  obtaining the above expressions computationally for any value of  $p$ . It also highlights some limitations of using generalized VIF in RE. Finally, Section 6 resumes the main contributions of the work.

## 2 VIF extension to ridge estimation

The first extension of the VIF associated with the ridge estimator was given by Marquardt (1970) who proposed detecting the presence of collinearity by using the diagonal elements of the matrix  $(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}$  as the VIF. Note that Marquardt automatically extended the definition of the VIF in OLS regression to RE. Kutner et al. (2005) proposed a definition of the VIF in RE that coincides with the extension of the Marquardt's VIF definition in RE (see Appendix A). However, the expression obtained for the VIF( $\lambda$ ) (see Marquardt, 1970 for the case of two variables) does not satisfy the condition of being larger than one (see Salmerón et al., 2013; García et al., 2013, 2015c). This is the first problem presented by the expression proposed by Marquardt. Furthermore, the expression given by Theil (1971) cannot be calculated since initially we do not know the matrix  $\mathbf{Z}$  from which to obtain  $\mathbf{Z}'\mathbf{Z} = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$ . The same occurs to the one proposed by Fox & Monette (1992) although in this case Friendly (2012) provides the correlation matrix of the data by conveniently transforming the covariance matrix (6). By using the R statistical environment, the *genridge* package (see Friendly, 2012) allows us to calculate the GVIF in RE by using the following expression

$$\text{GVIF}(i) = \frac{|\mathbf{R}_{-i}|}{|\mathbf{R}|}, \quad i = 1, \dots, p, \quad (7)$$

where  $\mathbf{R}$  denotes the correlation matrix among all the columns of  $\mathbf{X}$  and  $\mathbf{R}_{-i}$  is the resulting matrix by eliminating the  $i$ th row and  $i$ th column in matrix  $\mathbf{R}$ . Remember that  $|\mathbf{R}| = 1$  for orthogonal data and  $|\mathbf{R}| = 0$  for perfectly collinear data (see R Development Core Team, 2012). To illustrate all these affirmations, we will use the model previously used by Wissel (2009) relative to credit in American people with the following variables: the **total** mortgage dept outstanding,  $\mathbf{Y}$ , personal consumption,  $\mathbf{X}_1$ , personal incomes,  $\mathbf{X}_2$ , and consumer credit outstanding,  $\mathbf{X}_3$ . All variables are measured in billions of dollars. We have enlarged the sample by using data from 1995 to 2011 obtained from Economic reports of the President (2011). The values shown in Table 1 are the extension to RE of the VIF definition given by Marquardt (1970), denoted as REMVIF, and the extension of the general VIF definition given by Fox & Monette (1992), denoted as REGVIF.

Table 1: Values of VIFs of Marquardt (REMVIF( $\lambda, i$ )), Fox and Monette (REGVIF( $\lambda, i$ )) and Theil (VIF( $i$ )) of the  $i$ th variable for  $\lambda = 0$  and  $\lambda = 0.1$

	REMVIF(0, $i$ )	REGVIF(0, $i$ )	VIF( $i$ )	REMVIF(0.1, $i$ )	REGVIF(0.1, $i$ )
$i = 1$	154.9487	154.9487	154.9487	0.6272	1.724408
$i = 2$	37.2681	37.2681	37.2681	1.0804	1.699081
$i = 3$	222.8143	222.8143	222.8143	0.3658	1.254078

Note that all the definitions presented lead to the same result when the OLS estimation is applied ( $\lambda = 0$ ). However, when these definitions are extended and applied in RE ( $\lambda > 0$ ) they provide different results. The source of the problem may be that when we use the OLS estimator, the matrix of independent variables,  $\mathbf{X}$ , and the correlation matrix of the regressors,  $\mathbf{R}$ , are known. This situation is very different when we work with the ridge estimator: the data which generate the matrix  $\mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$  are initially unknown and consequently we do not

have the information to obtain the determination coefficient to calculate the corresponding VIFs (expression proposed by Theil, 1971) or the correlation matrix of the data (expression proposed by Fox & Monette, 1992). These limitations may be the cause to that numerous authors have followed the proposal by Marquardt (1970) to extend the concept of VIF to RE. This fact leads not only to values of VIF lesser than one but also they do not have the desirable property of being monotonic. Furthermore, as shown in the example, every measure to diagnostic collinearity leads to a different solution and, in some cases, to a different conclusion. In the next section, we present an alternative expression for the VIF in RE with  $p > 2$  and analyze its properties. We will use this same example to illustrate empirically the application of the proposed methodology.

### 3 RESVIF, REVIF and REGVIF in models with three exogenous variables

As mentioned above, the main problem in extending the definitions used in the OLS estimation to calculate the VIF in RE is that the matrix  $\mathbf{Z}$  that verifies  $\mathbf{Z}'\mathbf{Z} = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$  is initially unknown. This question was initially solved by García et al. (2015c) for the case  $p = 2$ . In this work we present the extension to the general case beginning for the case  $p = 3$ .

This problem can be solved from the matrix  $\mathbf{X}_S$ , verifying  $\mathbf{X}'_S\mathbf{X}_S = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$ , which was provided by Jensen & Ramirez (2010) in the presentation of the surrogate ridge estimation. Note that the matrix  $\mathbf{X}_S$  keeps the dimension of the matrix  $\mathbf{X}$  and it is orthonormalized. The VIF obtained from  $\mathbf{X}_S$  will be named as RESVIF.

Within the ridge estimator methodology, we can use the augmented model proposal of Marquardt (1970) to obtain by OLS the ridge estimator and use the matrix  $\mathbf{X}_Z = \left( \frac{\mathbf{X}}{\sqrt{\lambda\mathbf{I}}} \right)$  instead of the matrix  $\mathbf{X}_S$ . We will call the VIF calculated from this matrix REVIF.

In Appendix B we obtain explicit expressions for the RESVIF and REVIF when  $p = 3$ . Both expressions are obtained from the coefficient of determination of an auxiliary regression. Some properties of both VIFs are also shown. To finish, we obtain an explicit expression based on the correlation coefficients for REGVIF when  $p = 3$ . Follow the original expressions to calculate the RESVIF, REVIF and REGVIF are presented as a major contribution of this paper:

**Ridge Estimator Surrogate VIF (RESVIF):** for the  $i$ th variable is:

$$\text{RESVIF}(\lambda, i) = \frac{(1 + \lambda) \left[ (1 + \lambda)^2 - \rho_{jk}^2 \right]}{(1 + \lambda) \left[ (1 + \lambda)^2 - \rho_{ij}^2 - \rho_{ik}^2 - \rho_{jk}^2 \right] + 2\rho_{ij}\rho_{jk}\rho_{ik}}. \quad (8)$$

It is verified that  $\text{RESVIF}(\lambda, i) \geq 1$ ,  $\lim_{\lambda \rightarrow \infty} \text{RESVIF}(\lambda, i) = 1$  and the RESVIF is monotone decreasing with increasing  $\lambda$ .

**Ridge Estimator VIF (REVIF):** for the  $i$ th variable is:

$$\text{REVIF}(\lambda, i) = \frac{n + 3 + \lambda(n + 2)}{(n + 3) \left[ \frac{\lambda(2\rho_{ij} + 2\rho_{ik} - (\lambda + \rho_{jk} - n)) + (n + 3)(\rho_{jk} + 1)}{B} + \frac{2\rho_{ij}\rho_{ik}D - (\rho_{ij}^2 + \rho_{ik}^2)C}{A} \right] + \lambda(n + 2)}. \quad (9)$$

It is verified that  $\text{REVIF}(\lambda, i) \geq 1$ ,  $\lim_{\lambda \rightarrow \infty} \text{REVIF}(\lambda, i) = 1$  and the REVIF is monotone decreasing with increasing  $\lambda$ .

**Ridge Estimator Generalized VIF (REGVIF):** for the  $i$ th variable is:

$$\text{REGVIF}(i) = \frac{1 - \rho_{jk}^2}{1 - \rho_{jk}^2 - \rho_{ij}^2 - \rho_{ik}^2 + 2\rho_{jk}\rho_{ij}\rho_{ik}}. \quad (10)$$

Note that  $\rho_{ij}$  represents the correlation between the variables  $\mathbf{X}(i)$  and  $\mathbf{X}(j)$  with  $i, j, k \in \{1, 2, 3\}$ ,  $i \neq j$ ,  $i \neq k$  and  $j \neq k$ .

It can be easily demonstrated for  $\lambda = 0$  that:

$$\text{RESVIF}(0, i) = \text{REVIF}(0, i) = \text{REGVIF}(i) = \frac{1 - \rho_{jk}^2}{1 - \rho_{jk}^2 - \rho_{ij}^2 - \rho_{ik}^2 + 2\rho_{jk}\rho_{ij}\rho_{ik}} = \text{VIF}(i). \quad (11)$$

## 4 Obtaining the VIF in economic data

In section 2, we showed the deficiencies obtained when applying in RE the definitions of VIF originally created to OLS estimation by using the model relative to credit in American people previously applied by Wissel (2009) but enlarging the sample with data from 1995 to 2011 obtained from Economic reports of the President (2011). Now we will use this same example to illustrate how to calculate the REVIF and compare it with the RESVIF and the REMVIF.

From this data, the multiple linear estimation of model (1) leads to the following results:

$$\hat{\mathbf{Y}} = \begin{matrix} -920.9 & -1.396 \cdot \mathbf{X}_1 & +0.8886 \cdot \mathbf{X}_2 & +0.00667 \cdot \mathbf{X}_3, \\ (8995) & (2.399) & (0.6057) & (0.00639) \end{matrix}$$

with  $R^2 = 0.9583$  and  $F_{exp} = 99.65$ . Note that none parameter is statistically significant but they are not simultaneously zero (joint significance). We also observe that the coefficient of determination is very high. These results are typical of models with collinearity. **As was noted in the introduction, a good specification of the model is a first relevant step to solve these problems. However, the collinearity may persist.**

In this case, the correlation matrix shows very high correlations between all independent variables:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.9795893 & 0.9966152 \\ 0.9795893 & 1 & 0.9858508 \\ 0.9966152 & 0.9858508 & 1 \end{pmatrix}. \quad (12)$$

This high correlation was expected since the independent variables (personal consumption,  $\mathbf{X}_1$ , personal incomes,  $\mathbf{X}_2$ , and consumer credit outstanding,  $\mathbf{X}_3$ ) seem to be very related from an economic point of view. Indeed, Wissel (2009) recognised that the data were chosen to obtain multicollinearity. In addition, by calculating the  $R_i^2$ ,  $i = 1, 2, 3$ , we obtain the VIF of each variable which is always over 10 indicating the existence of collinearity (see Table 2).

Table 2: Values of VIF

	$i = 1$	$i = 2$	$i = 3$
$R_i^2$	0.9935	0.9731	0.9955
$\text{VIF}(i) = \frac{1}{1-R_i^2}$	154.9487	37.2681	222.8143

According to the interpretation given by Fox (1997) about the VIF, we can state, for example, that the confidence interval for  $\beta_1$  and  $\beta_3$  is  $\sqrt{154.9487} = 12.4478$  and  $\sqrt{222.8143} = 14.9269$  times greater respectively than if there were no multicollinearity. All this suggests the existence of collinearity in the proposed model which seems ideal to use the expression obtained in Section 3.

The VIFs of each variable, using the different expressions REMVIF, RESVIF, REVIF, REGVIF (from standardized variables) and REGVIF (from typified<sup>1</sup> variables) are represented in Figure 1 for  $\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ .

<sup>1</sup>Note that a standardized variable is the value of the variable minus its mean, divided by the square root of the number of observations multiplied by its variance while a typified variable is the value of the variable minus its mean, divided by its standard deviation.

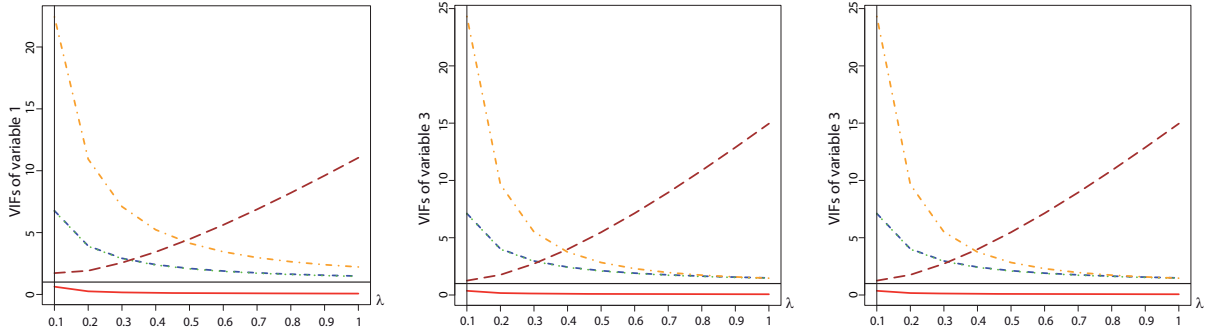


Figure 1: REMVIF(red-solid); RESVIF(blue-dashed); REVIF(green-dotted); REGVIF from standardized variables (brown-dotdash); REGVIF from typified variables (orange-dotdash) join the line  $y = 1$ .

Note that for  $\lambda$  varying between 0.1 and 1, the REMVIF of the variables  $\mathbf{X}_1$  and  $\mathbf{X}_3$  are always less than 1 while for variable  $\mathbf{X}_2$  only takes values higher than one when  $\lambda = 0.1$ . These results contradict the consequences obtained from expression (2). On the other hand, the values of REGVIF are different from the one obtained for the RESVIF and REVIF. This difference is very significative when variables are standardized.

In subsection B.2, expression (32) shows that the RESVIF and the REVIF coincide when  $n \rightarrow +\infty$ . However, in Figure 1 we can see that both VIFs graphs overlap and in this case the value of  $n$  is only 17. Furthermore, in Figure 2 shows the difference between RESVIF and REVIF for  $\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ . Note that the differences are of the order of  $10^{-2}$ . So a large sample size is not necessary to allow the property to be verified.

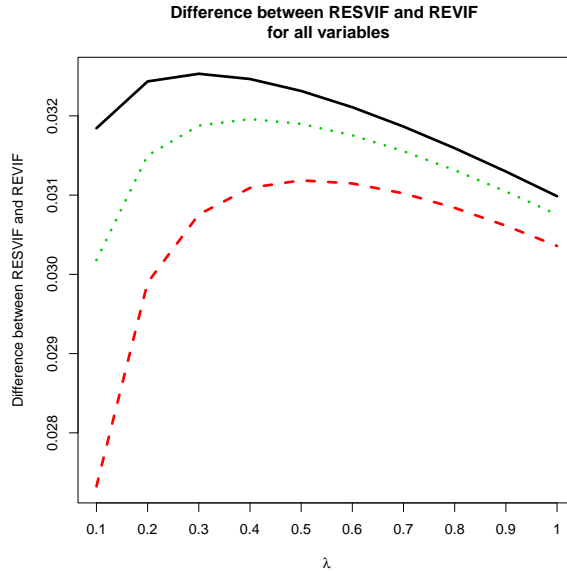


Figure 2: Representing the difference between RESVIF and REVIF for  $\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  for each variables.

In Table 3 we study in more detail the results obtained when  $\lambda$  varies between 0 and 0.1. Note that from  $\lambda = 0.02$ , all values of REMVIF are less than 10 while this does not occur in RESVIF

Table 4: Values of RESVIF for  $\lambda = 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1$ 

$\lambda$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
0	154.948757	37.268158	222.814314
0.01	39.418727	21.965852	49.483674
0.02	24.137728	16.617822	28.474403
0.03	17.743912	13.489267	20.197535
0.04	14.162010	11.407017	15.750795
0.05	11.852657	9.915907	12.969567
0.06	10.233564	8.793964	11.063770
0.07	9.032913	7.918666	9.675491
0.08	8.105895	7.216529	8.618786
0.09	7.367971	6.640709	7.787377
0.1	6.766345	6.159898	7.116079

and REVIF until  $\lambda = 0.07$  (see Tables 4 and 5). Following REMVIF this fact would imply that, for  $\lambda = 0.02, 0.03, 0.04, 0.05, 0.06$ , the problem of collinearity has been mitigated when in fact it has not. Note that when the parameter  $\lambda$  is equal to zero the values of VIFs are equal since in this situation the original model has been estimated by OLS. Table 6 presents the Condition Number (CN) extended from OLS to RE without further considerations and the  $CN(\lambda)$  extended to RE taking into considerations the problems of this issue as shown by García et al. (2015c).

Table 3: Values of REMVIF for  $\lambda = 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1$ 

$\lambda$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
0	154.9487572	37.268158	222.8143143
0.01	12.5436403	14.231998	11.5699764
0.02	5.6365410	8.271826	4.1167910
0.03	3.3836467	5.440672	2.1973748
0.04	2.3043321	3.865205	1.4041879
0.05	1.6909349	2.897617	0.9950503
0.06	1.3055132	2.260648	0.7546938
0.07	1.0463587	1.818998	0.6007831
0.08	0.8632570	1.500161	0.4959592
0.09	0.7288631	1.262416	0.4211669
0.1	0.6271743	1.080372	0.3658185

From these results we can highlight the following comments:

- The concept of VIF in OLS is based on expression (2) and it has to be kept when extending to RE.
- The REMVIF presents values lesser than one unrespecting the concept of VIF. Furthermore, the REMVIF goes rapidly down the rules of thumb of VIF and it can lead to think that the multicollinearity is mitigated when it is not.
- The REVIF coincides asymptotically with the RESVIF, and for values of  $n$  equal to 17, as in the numerical example, the difference is depreciable.

Table 5: Values of REVIF for  $\lambda = 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1$

$\lambda$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
0	154.948757	37.268158	222.814314
0.01	39.399287	21.955079	49.459193
0.02	24.114188	16.601705	28.446546
0.03	17.718249	13.469864	20.168229
0.04	14.135007	11.385386	15.720666
0.05	11.824723	9.892666	12.938903
0.06	10.204944	8.769506	11.032729
0.07	9.003767	7.893256	9.644171
0.08	8.076331	7.190354	8.587250
0.09	7.338069	6.613907	7.755671
0.1	6.736164	6.132574	7.084235

Table 6: Values of REVIF and Condition Number for  $\lambda=0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$

$\lambda$	REVIF(1)	REVIF(2)	REVIF(3)	CN	CN( $\lambda$ )
0	154.9488	37.2682	222.8143	75.1695	33.1943
0.1	6.7362	6.1326	7.0842	64.5731	18.6334
0.2	3.8902	3.713	3.9924	57.48	14.3649
0.3	2.8876	2.802	2.937	52.3065	12.1218
0.4	2.3768	2.3257	2.4063	48.3183	10.6853
0.5	2.0684	2.034	2.0882	45.1227	9.6652
0.6	1.8626	1.8377	1.8769	42.4876	8.8929
0.7	1.716	1.6971	1.7269	40.2664	8.2823
0.8	1.6067	1.5918	1.6153	38.3609	7.7838
0.9	1.5223	1.5101	1.5293	36.7028	7.3669
1	1.4553	1.4452	1.4611	35.2428	7.0117



- The values obtained to the REGVIF (with original, typified or standardized data) do not coincide with the values obtained for the RESVIF or the REVIF.
- From Table 6 we see that the CN and  $CN(\lambda)$  are decreasing. While CN presents all values higher than 30, the  $CN(\lambda)$  takes values lesser than 10 from  $\lambda > 0.4$ . Thus, the  $CN(\lambda)$  and the REVIF lead to similar conclusions in the diagnostic of collinearity. However, the CN extended without further considerations indicates that the collinearity has not been mitigated even for  $\lambda = 1$ .

Finally, we would like to emphasize that in the case of the ridge estimator, the calculation of the VIFs proposed by Theil (1971) (see expression (2)), Marquardt (1970) (elements of the principal diagonal of the inverse correlation matrix) and Fox & Monette (1992) (see expression (7)) leads us to obtain the same results obtained in the calculation of REVIF. The REVIF is monotonically decreasing in  $\lambda$ , is higher than 1 and it verifies that  $\lim_{\lambda \rightarrow \infty} \text{REVIF}(\lambda, i) = 1$  and  $\text{REVIF}(0, i) = \text{VIF}(i)$ .

Then, we propose to use the REVIF instead of traditionally used REMVIF and REGVIF. The RESVIF, associated to surrogate ridge model, can be also recommended due to its simple calculation and its asymptotic equivalence with the REVIF but taking into consideration that it is associated to OLS surrogate ridge model.

## 5 A Monte Carlo Simulation

In the last section we have presented an economic example with three exogenous variables and standardized data obtaining the REMVIF, RESVIF, REVIF and REGVIF with standardized data finding certain anomalies in the results obtained by the REGVIF since it increases when  $\lambda$  increases (see Figure 1), which does not make sense. Due to this fact, we calculate the REGVIF from the original data obtaining different results even for  $\lambda = 0$ . We also calculate the REGVIF from typified data and we have included this case in the work since the obtained results coincide with the results provided by the package *genridge* in R software. In this last case we obtain values of REGVIF different from the one obtained for RESVIF and REVIF. In this section, we will studied in more detail this case and use a Monte Carlo simulation to present algorithms to obtain computationally the different VIFs for models with multiple independent variables. Following McDonald & Galarneau (1975); Wichern & Churchill (1978); Gibbons (1981) we generate five independent variables using:

$$\mathbf{X}(i) = \sqrt{1 - \gamma^2} \mathbf{Z}(i) + \gamma \mathbf{Z}(5), \quad i = 1, 2, 3, 4, 5, \quad (13)$$

where  $\mathbf{Z}(i)$  are independent pseudo-random numbers distributed as  $N(0, 100)$  and  $\gamma$  is specified so that the correlation between two any independent variable is given by  $\gamma^2$ . The dependent variable is generated as:

$$\mathbf{Y} = \mathbf{X}(1) + \mathbf{X}(2) + \mathbf{X}(3) + \mathbf{X}(4) + \mathbf{X}(5) + \mathbf{u}, \quad (14)$$

where  $\mathbf{u}$  are independent pseudo-random numbers distributed as  $N(0, 1)$ . Finally, we used three sample sizes: 100, 500 and 1000.

Since there are five independent variables it is necessary to use the algorithms proposed in Appendix C. The results are shown in Figure 3 for  $\lambda = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ . It is observed that:

- The REGVIF from standardized variables, as already mentioned, increases when  $\lambda$  increases.
- The REGVIF from typified variables increases when  $n$  increases.
- The REGVIF from typified variables is never lesser than the established limit 10. This means that the collinearity has not been mitigated.

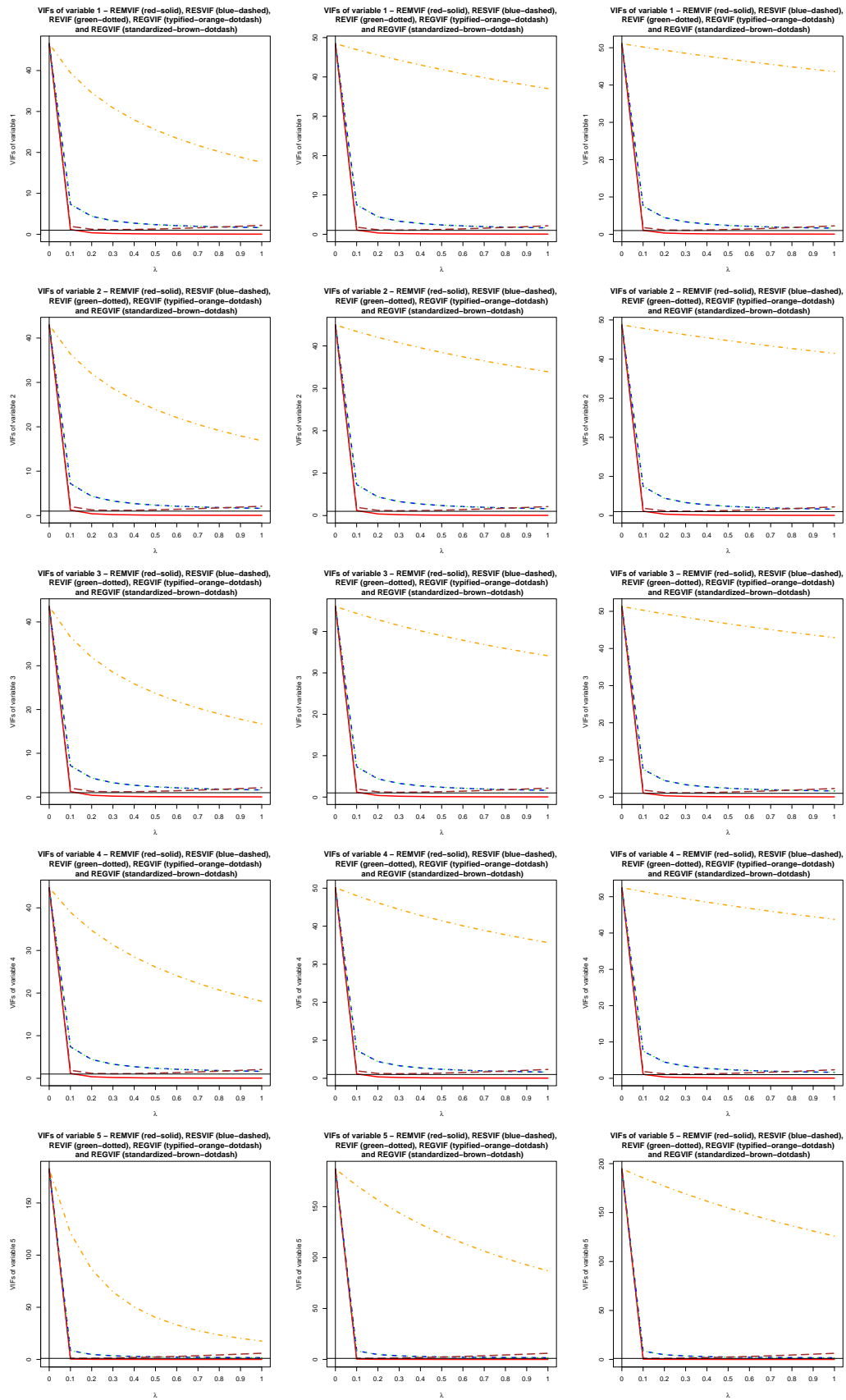


Figure 3: VIFs for all variables (first, second and third column corresponds to  $n = 100, 500, 1000$ , respectively).

Table 7: Difference between RESVIF and REVIF depending on sample size

Sample size	Minimum difference	Maximum difference
100	0.0064	0.0074
500	0.0013	0.0015
1000	0.00068	0.00078

- The other VIFs show the same behavior: a) RESVIF and REVIF graphs overlap and b) REMVIF takes values less than 1.
- The difference between RESVIF and REVIF decreases when  $n$  increases (see Table 7).

The first three points leads us to believe that the application of REGVIF (from typified or standardized data) is not appropriate in RE. This is not a minor issue since the package *genridge* in R software calculated the VIF in RE following this definition.

## 6 Conclusions

If all definitions of VIF (Marquardt, 1970; Theil, 1971; Fox & Monette, 1992; Kutner et al., 2005) lead to the same results under OLS estimation, its extension to RE should also lead to the same result. However, we have obtained different results and conclusions up to the point of considering that the collinearity problem is solved when it still persists.

The problem is that both Marquardt (1970) and Fox & Monette (1992) use the covariance matrix of RE by extending its initial definition (in the case of Marquardt) or obtaining the correlation matrix (in the case of Fox and Monette). All extensions lead to the same results if instead of selecting this matrix as a starting point we choose the ridge regressors matrix (see expressions in Marquardt, 1970), which establishes the equivalence between RE and an alternative OLS estimation. Thus, we can conclude that the expression proposed in this paper to calculate the VIF in RE is the right one. That is, the VIF associated with the ridge estimator should be calculated by the REVIF expression.

Furthermore, it has been shown that RESVIF (associated to ridge surrogate estimator) is easier to calculate than the REVIF and it can be its substitute since the differences are small, at least they are quoted to one hundredth.

When extending the  $VIF(i)$  to ridge regression, the  $VIF(\lambda, i)$  should be monotonically decreasing in  $\lambda$ , higher than one and it has to verify that  $\lim_{\lambda \rightarrow \infty} REVIF(\lambda, i) = 1$  and  $VIF(0, i) = VIF(i)$ .

These conditions are only verified by the REVIF (associated to ridge estimator) and the RESVIF (associated to surrogate ridge estimator). Although the RESVIF verifies these conditions and also presents desirable monotone properties, it is possible to get these same conditions within RE if the concept of VIF is appropriately extended. Thus, within RE the suitable extension should be the REVIF.

It was shown that the use of REGVIF in RE should be reviewed: a) from standardized variables major anomalies are obtained and b) from typified variables (methodology used in R software) are obtained different values to those obtained by RESVIF and REVIF, and as shown in the simulation section can lead to erroneous conclusions.

To sum up, the main contribution of this paper is to present the expressions and algorithms necessary to diagnose correctly the collinearity through the VIF after the application of the ridge estimation for  $p$  independent variables tending to make life simpler for those who come across collinearity issues in their regression model.

## Acknowledgements

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

## References

- 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, NY: John Wiley.
- Berk, K.N. (1977). Tolerance and condition in regression computations. *Journal of the American Statistical Association*, **72**, 863–866.
- Besley, D.A. (1991). Conditioning diagnostics: Collinearity and weak data in regression. New York: John Wiley.
- Dias, J. & Castro, J. (2007). New multicollinearity indicators in linear regression models. *International Statistical Review*, **75**, 114–121.
- Economic reports of the President (2011). url:<http://www.gpo.gov/fdsys/browse>.
- Fox, J. (1997). Applied regression analysis, linear models, and realted methods. CA: Sage: Thousand Oaks.
- Fox, J. & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, **87**, 178–183.
- Friendly, M. (2012). genridge: Generalized ridge trace plots for ridge regression. R package version 0.6-3.
- Friendly, M. & Kwan, E. (2009). Where’s Waldo? Visualizing collinearity diagnostics. *The American Statistician*, **63**.
- García, C., García, J., López, M.D.M. & Salmerón, R. (2015a). Collinearity: revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics*, **32**, 648–661. doi: <http://dx.doi.org/10.1080/02664763.2014.980789>.
- García, J., García, C., López, M.D.M. & Salmerón, R. (2013). Desarrollo del método de alzado para el tratamiento de la multicolinealidad. Determinación del factor de alzamiento. XXVII International Conference of Applied Economics. pp. 1848–1862.
- García, J., Salmerón, R., García, C. & López, M.D.M. (2015b). Standardization of variables and collinearity diagnostic in ridge regression. *International Statistical Review*. doi:DOI: 10.1111/insr.12099.
- García, J., Salmerón, R., López, M.D.M. & García, C. (2015c). Revisiting the condition number and red indicator in ridge regression. *Proceedings of the International Work Conference on Time Series*, 271–22.
- Gibbons, D.G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, **76**, 131–139.
- 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.
- Jensen, D.R. & Ramirez, D.E. (2010). Surrogate models in ill-conditioned systems. *Journal of Statistical Planning and Inference*, **140**, 2069–2077.
- Kennedy, P.A. (1992). A guide to Econometrics. 4th edition. MIT Press.
- Kovacs, P., Petres, T. & Toth, L. (2005). A new measure of multicollinearity in linear regression models. *International statistical review*, **73**, 405–412.
- Kutner, M.H., Nachatsheim, C.J., Neter, J. & Li, W. (2005). Applied linear statistical models. 5th edition. McGraw–Hill.

- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares.
- Marquardt, D.W. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics*, **12**, 591–612.
- McDonald, G.C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, **1**, 93–100.
- McDonald, G.C. (2010). Tracing ridge regression coefficients. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 695–703.
- McDonald, G.C. & Galarneau, D.I. (1975). A Monte Carlo evaluation of some ridge type estimators. *Journal of the American Statistical Association*, **70**, 407–416.
- 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.
- Rawlings, J.O., Pantula, S.G. & Dickey, D.A. (1998). Applied regression analysis: a research tool. Springer Science & Business Media.
- Riley, J. (1955). Solving systems of linear equations with a positive definite, symmetric but possibly ill-conditioned matrix. *Math. Tables Aids Comput.*, **9**, 96–101.
- 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.
- Theil, H. (1971). Principles of econometrics. New York: Wiley.
- Wichern, D.W. & Churchill, G.A. (1978). A comparison of ridge estimators. *Technometrics*, **20**, 301–311.
- Wichers, C.R. (1975). The detection of multicollinearity: a comment. *The Review of Economics and Statistics*, 366–368.
- Wissel, J. (2009). A new biased estimator for multivariate regression models with highly collinear variables. Ph.D. thesis, Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universität Würzburg.

## A Relationship between the definition of the VIF by Kutner et al. (2005) and the extension of the Marquardt’s VIF definition in RE

Kutner et al. (2005) considered that in RE the VIF is the  $i$ th element of the principal diagonal of the following matrix

$$\mathbf{V} \begin{pmatrix} \frac{\gamma_1}{(\gamma_1 + \lambda)^2} & 0 & \dots & 0 \\ 0 & \frac{\gamma_2}{(\gamma_2 + \lambda)^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\gamma_p}{(\gamma_p + \lambda)^2} \end{pmatrix} \mathbf{V}', \quad (15)$$

where  $\mathbf{V}$  is an orthogonal matrix of order  $p$  whose columns are the normalized eigenvectors of the correlation matrix of the independent variables,  $\mathbf{R}$ , and  $\gamma_i$  is the  $i$ th eigenvalue of the same matrix,  $i = 1, \dots, p$ . That is to say,  $\mathbf{R} = \mathbf{V} \cdot \mathbf{\Gamma} \cdot \mathbf{V}'$  with  $\mathbf{V}' \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{V}' = \mathbf{I}$  and  $\mathbf{\Gamma}$  being a diagonal matrix whose elements are the eigenvalues of  $\mathbf{R}$ . Since the data are standardized,  $\mathbf{R} = \mathbf{X}'\mathbf{X}$ , and then this proposal coincides with the one provided by Marquardt since

$$(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} = \mathbf{V}(\mathbf{\Gamma} + \lambda\mathbf{I})^{-1} \mathbf{\Gamma} (\mathbf{\Gamma} + \lambda\mathbf{I})^{-1} \mathbf{V}', \quad (16)$$

where

$$(\mathbf{\Gamma} + \lambda \mathbf{I})^{-1} \mathbf{\Gamma} (\mathbf{\Gamma} + \lambda \mathbf{I})^{-1} = \begin{pmatrix} \frac{\gamma_1}{(\gamma_1 + \lambda)^2} & 0 & \dots & 0 \\ 0 & \frac{\gamma_2}{(\gamma_2 + \lambda)^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\gamma_p}{(\gamma_p + \lambda)^2} \end{pmatrix}. \quad (17)$$

## B VIFs in RE

In this section we obtaining explicit expression for RESVIF, REVIF and REGVIF for  $p = 3$ .

### B.1 Ridge Estimator Surrogate VIF (RESVIF)

From surrogate ridge estimator presented by Jensen & Ramirez (2010), we can affirm that exists a matrix  $\mathbf{X}_S$  which verifies  $\mathbf{X}'_S \mathbf{X}_S = \mathbf{X}' \mathbf{X} + k \mathbf{I}$ . In the case of three variables ( $p=3$ ):

$$\mathbf{X}'_S \mathbf{X}_S = \mathbf{X}' \mathbf{X} + \lambda \mathbf{I} = \begin{pmatrix} 1 + \lambda & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 + \lambda & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 + \lambda \end{pmatrix}, \quad (18)$$

i.e.  $\sum_{i=1}^n x_{ij}^2 = 1 + \lambda$  for  $j = 1, 2, 3$ , and  $\sum_{i=1}^n x_{ij} x_{ik} = \rho_{jk}$  for  $j, k = 1, 2, 3$ ,  $j \neq k$ .

Through the regression of the standardized variable  $\mathbf{X}_S(i)$  on the standardized variables  $\mathbf{X}_S(j)$  and  $\mathbf{X}_S(k)$

$$\mathbf{X}_S(i) = \beta_j \mathbf{X}_S(j) + \beta_k \mathbf{X}_S(k) + \mathbf{v}, \quad (19)$$

with  $i, j, k \in \{1, 2, 3\}$ ,  $i \neq j$ ,  $i \neq k$  and  $j \neq k$ , the estimator of the parameters by OLS is

$$\widehat{\boldsymbol{\beta}}_S(i) = \begin{pmatrix} 1 + \lambda & \rho_{jk} \\ \rho_{jk} & 1 + \lambda \end{pmatrix}^{-1} \begin{pmatrix} \rho_{ij} \\ \rho_{ik} \end{pmatrix} = \frac{1}{(1 + \lambda)^2 - \rho_{jk}^2} \begin{pmatrix} (1 + \lambda)\rho_{ij} - \rho_{jk}\rho_{ik} \\ (1 + \lambda)\rho_{ik} - \rho_{jk}\rho_{ij} \end{pmatrix}. \quad (20)$$

In this case the Explained Sum of Squares (ESS) and the Total Sum of Squared (TSS) are

$$\text{ESS}_S(i) = \frac{1}{(1 + \lambda)^2 - \rho_{jk}^2} [(1 + \lambda)(\rho_{ij}^2 + \rho_{ik}^2) - 2\rho_{ij}\rho_{jk}\rho_{ik}], \quad (21)$$

$$\text{TSS}_S(i) = 1 + \lambda \quad (\text{by } \mathbf{X}_S \text{ definition}). \quad (22)$$

Using the expressions (21) and (22), the determination coefficient of the model (19) is

$$R_S^2(i) = \frac{\text{ESS}_S(i)}{\text{TSS}_S(i)} = \frac{(1 + \lambda)(\rho_{ij}^2 + \rho_{ik}^2) - 2\rho_{ij}\rho_{jk}\rho_{ik}}{(1 + \lambda)[(1 + \lambda)^2 - \rho_{jk}^2]}. \quad (23)$$

As a result, the  $i$ th variable RESVIF is

$$\text{RESVIF}(\lambda, i) = \frac{1}{1 - R_S^2(i)} = \frac{(1 + \lambda)[(1 + \lambda)^2 - \rho_{jk}^2]}{(1 + \lambda)[(1 + \lambda)^2 - \rho_{ij}^2 - \rho_{ik}^2 - \rho_{jk}^2] + 2\rho_{ij}\rho_{jk}\rho_{ik}}. \quad (24)$$

Because of  $\text{ESS}_S(i) \geq 0$  and  $(1 + \lambda)^2 - \rho_{jk}^2 \geq 0$  then  $2\rho_{ij}\rho_{jk}\rho_{ik} - (1 + \lambda)(\rho_{ij}^2 + \rho_{ik}^2) \leq 0$ . By adding  $(1 + \lambda)[(1 + \lambda)^2 - \rho_{jk}^2]$  on both sides we obtain

$$(1 + \lambda)[(1 + \lambda)^2 - \rho_{ij}^2 - \rho_{ik}^2 - \rho_{jk}^2] + 2\rho_{ij}\rho_{jk}\rho_{ik} \leq (1 + \lambda)[(1 + \lambda)^2 - \rho_{jk}^2].$$

Thus, we conclude that  $\text{RESVIF}(\lambda, i) \geq 1$ ,  $\lim_{\lambda \rightarrow \infty} \text{RESVIF}(\lambda, i) = 1$  and the RESVIF is monotone decreasing with increasing  $\lambda$  (see Jensen & Ramirez, 2010, Theorem 5, p. 2077).

## B.2 Ridge Estimator VIF (REVIF)

Ridge estimator VIF (REVIF) is the extension of the Theil (1971) definition (see expression (2)) to RE. In this case, it is necessary to know the matrix of regressors  $\mathbf{Z}$  so that  $\mathbf{Z}'\mathbf{Z} = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$ . Even though this matrix is unknown Marquardt (1970) showed that

$$\mathbf{X}_Z = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \\ \hline \sqrt{\lambda} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda}\mathbf{I} \end{pmatrix}, \quad \mathbf{Y}_Z = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \end{pmatrix},$$

where the bottom  $p \times 1$  submatrix of  $\mathbf{Y}_Z$  is not to be viewed as a random responses, the top  $n \times p$  submatrix of the  $X_Z$  matrix has to be standardized (see García et al. (2015b)) and RE in the model (1) is similar to the OLS regression in the following model:

$$\mathbf{Y}_Z = \mathbf{X}_Z\boldsymbol{\beta} + \mathbf{v}. \quad (25)$$

It is verifying that  $\hat{\boldsymbol{\beta}}(\lambda) = (\mathbf{X}'_Z\mathbf{X}_Z)^{-1}\mathbf{X}'_Z\mathbf{Y}_Z = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})\mathbf{X}'\mathbf{Y}$  since  $\mathbf{X}'_Z\mathbf{X}_Z = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$  and  $\mathbf{X}'_Z\mathbf{Y}_Z = \mathbf{X}'\mathbf{Y}$ . Then the matrix  $\mathbf{X}_Z$  can be used instead of the matrix  $\mathbf{Z}$ . The case  $p = 2$  is given in García et al. (2015a). For  $p = 3$ , the regression of the variable  $\mathbf{X}_Z(i)$  on the variables  $\mathbf{X}_Z(j)$  and  $\mathbf{X}_Z(k)$  is given by

$$\mathbf{X}_Z(i) = \beta_0 + \beta_j\mathbf{X}_Z(j) + \beta_k\mathbf{X}_Z(k) + \mathbf{w}, \quad (26)$$

with  $i, j, k \in \{1, 2, 3\}$ ,  $i \neq j$ ,  $i \neq k$  and  $j \neq k$ . In that case

$$\text{ESS}_Z(i) = \frac{\lambda[(\lambda + \rho_{jk} + 1) - 2\rho_{ij} - 2\rho_{ik}]}{B} - \frac{2\rho_{ij}\rho_{ik}D - (\rho_{ij}^2 + \rho_{ik}^2)C}{A}, \quad (27)$$

$$\text{TSS}_Z(i) = \frac{n + 3 + \lambda(n + 2)}{n + 3}, \quad (28)$$

where

$$\begin{aligned} A &= \lambda^2(n + 1) + \lambda(4 + 2n + 2\rho_{jk}) + (n + 3)(1 - \rho_{jk}^2), \\ B &= (n + 3)(1 + \rho_{jk}) + \lambda(n + 1), \\ C &= n + 3 + \lambda(n + 2), \\ D &= \rho_{jk}(n + 3) - \lambda. \end{aligned}$$

With this information, we conclude that the coefficient of determination is

$$R_Z^2(i) = \frac{\text{ESS}_Z(i)}{\text{TSS}_Z(i)} = \frac{n + 3}{n + 3 + \lambda(n + 2)}\text{ESS}_Z(i), \quad (29)$$

and then the  $i$ th variable REVIF is

$$\text{REVIF}(\lambda, i) = \frac{n + 3 + \lambda(n + 2)}{(n + 3) \left[ \frac{\lambda(2\rho_{ij} + 2\rho_{ik} - (\lambda + \rho_{jk} - n)) + (n + 3)(\rho_{jk} + 1)}{B} + \frac{2\rho_{ij}\rho_{ik}D - (\rho_{ij}^2 + \rho_{ik}^2)C}{A} \right] + \lambda(n + 2)}. \quad (30)$$

It can be easily demonstrated for  $\lambda = 0$  that

$$\text{RESVIF}(0, i) = \text{REVIF}(0, i) = \frac{1 - \rho_{jk}^2}{1 - \rho_{jk}^2 - \rho_{ij}^2 - \rho_{ik}^2 + 2\rho_{jk}\rho_{ij}\rho_{ik}} = \text{VIF}(i). \quad (31)$$

Thus, we conclude that  $\text{REVIF}(\lambda, i) \geq 1$ ,  $\lim_{\lambda \rightarrow \infty} \text{REVIF}(\lambda, i) = 1$  and the REVIF is monotone decreasing with increasing  $\lambda$  since  $R_Z^2(i)$  is decreasing in  $\lambda$  (see McDonald, 2010, p.696). On the other hand, with the help of symbolic computation software we can affirm that

$$\lim_{n \rightarrow \infty} (\text{REVIF}(\lambda, i) - \text{RESVIF}(\lambda, i)) = 0. \quad (32)$$

### B.3 Generalized VIF for $p = 3$ (REGVIF)

When  $p = 3$  the correlation matrix among all variables from (6) is

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}. \quad (33)$$

In this case the expression (7) would be written as

$$\text{REGVIF}(i) = \frac{1 - \rho_{jk}^2}{1 - \rho_{jk}^2 - \rho_{ij}^2 - \rho_{ik}^2 + 2\rho_{jk}\rho_{ij}\rho_{ik}}, \quad (34)$$

with  $i, j, k = 1, 2, 3$ ,  $i \neq j, i \neq k, j \neq k$ , since  $|\mathbf{R}| = 1 - \rho_{jk}^2 - \rho_{ij}^2 - \rho_{ik}^2 + 2\rho_{jk}\rho_{ij}\rho_{ik}$ ,  $\mathbf{R}_{-i} = \begin{pmatrix} 1 & \rho_{jk} \\ \rho_{jk} & 1 \end{pmatrix}$ , and  $|\mathbf{R}_{-i}| = 1 - \rho_{jk}^2$ .

Therefore, the expression (34) clearly verifies that  $\text{RESVIF}(0, i) = \text{REVIF}(0, i) = \text{REGVIF}(i)$  (see expression (31)). Also  $\text{REGVIF}(i) \geq 1$ ,  $\forall i$ , since  $\text{REGVIF}(i) = \text{RESVIF}(0, i) \geq 1$ . Contrarily to RESVIF and REVIF, the REGVIF will not be decreasing in  $\lambda$ .

For  $\lambda = 0$  it is clear what  $\mathbf{R}(0) = R$  is, but the question is how to calculate  $\mathbf{R}(\lambda)$  for  $\lambda > 0$ . Friendly (2012) solves this problem by transforming the covariance matrix (6) of RE into a correlation matrix considering this last one as  $\mathbf{R}(\lambda)$ .

Note that if we consider the following matrix

$$\mathbf{R}(\lambda) = \begin{pmatrix} 1 + \lambda & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 + \lambda & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 + \lambda \end{pmatrix}, \quad (35)$$

and applying the definition (4) we obtain that

$$\text{REGVIF}(\lambda, i) = \frac{(1 + \lambda) \left[ (1 + \lambda)^2 - \rho_{jk}^2 \right]}{(1 + \lambda) \left[ (1 + \lambda)^2 - \rho_{ij}^2 - \rho_{ik}^2 - \rho_{jk}^2 \right] + 2\rho_{ij}\rho_{jk}\rho_{ik}}.$$

It is to say,  $\text{REGVIF}(\lambda, i) = \text{RESVIF}(\lambda, i)$  for all  $\lambda$  and  $i$ . However, in this case the election of  $\mathbf{R}(\lambda)$  will not correspond to a true correlation matrix since its mean diagonal is not constantly 1 for  $\lambda > 0$ .

## C Algorithms to obtain the VIFs in Ridge estimation in the general case

Consider that we have  $p$  standardized exogenous variables and an endogenous variable, if we consider the multiple linear regression model (1), the surrogate and ridge estimator VIFs for any value of  $p$  may be computed as shown in the Algorithms 1 and 2. In both cases, the matrices  $\mathbf{X}_S$  and  $\mathbf{X}_Z$  are calculated to obtain the regressions (19) and (26), respectively.

Finally, the coefficients of determination of these regressions are used to obtain the VIFs. Note that these algorithms simply reproduce the steps followed in subsections B.1 and B.2.



The generalized VIF is obtained from expression (7) as shown in Algorithm 3. We have also implemented an iterative procedure to calculate the VIF given by Marquardt (1970), although it is not the aim of the work. In Algorithm 4 the main diagonal elements of matrix  $(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}$  are obtained. Upon request, generate codes are available from the authors.

---

**Algorithm 1** Obtaining surrogate VIF (RESVIF)

---

**Require:** Calculate  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{I}$  and  $D(\delta)$  (discretization of the interval  $[0, 1]$  with  $\delta$  points)

- 1: **for**  $\lambda \in D(\delta)$  **do**
  - 2:   consider the surrogate matrix  $\mathbf{X}_S$  so that  $\mathbf{X}'_S \mathbf{X}_S = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$
  - 3:   **for**  $i \in \{1, 2, \dots, p\}$  **do**
  - 4:     the regression of column  $i$  of  $\mathbf{X}_S$  on the other variables
  - 5:     obtain the coefficient of determination,  $R_S^2(i)$ , of the regression
  - 6:     calculate surrogate VIF of  $i$ th variable as  $\text{RESVIF}(\lambda, i) = \frac{1}{1-R_S^2(i)}$
  - 7:   **end for**
  - 8: **end for**
- 

---

**Algorithm 2** Obtaining ridge estimator VIF (REVIF)

---

**Require:** Calculate  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{I}$  and  $D(\delta)$  (discretization of the interval  $[0, 1]$  with  $\delta$  points)

- 1: **for**  $\lambda \in D(\delta)$  **do**
  - 2:   calculate the matrix  $\mathbf{X}_Z$  that generates the matrix of the ridge estimators  $\mathbf{X}'_Z \mathbf{X}_Z = \mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$
  - 3:   **for**  $i \in \{1, 2, \dots, p\}$  **do**
  - 4:     the regression of column  $i$  of  $\mathbf{X}_Z$  on the other variables
  - 5:     obtain the coefficient of determination,  $R_Z^2(i)$ , of the regression
  - 6:     calculate surrogate VIF of  $i$ th variable as  $\text{REVIF}(\lambda, i) = \frac{1}{1-R_Z^2(i)}$
  - 7:   **end for**
  - 8: **end for**
- 

---

**Algorithm 3** Obtaining generalized VIF (REGVIF)

---

**Require:** Calculate  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{I}$ ,  $D(\delta)$  (discretization of the interval  $[0, 1]$  with  $\delta$  points) and  $\hat{\sigma}^2$  (the estimation of the variance of regression)

- 1: **for**  $\lambda \in D(\delta)$  **do**
  - 2:   calculate  $\text{var}(\hat{\boldsymbol{\beta}}(\lambda)) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}$
  - 3:   transform  $\text{var}(\hat{\boldsymbol{\beta}}(\lambda))$  into a correlation matrix  $\mathbf{R}$
  - 4:   **for**  $i \in \{1, 2, \dots, p\}$  **do**
  - 5:     calculate generalized VIF of  $i$ th variable as  $\text{REGVIF}(i) = \frac{|\mathbf{R}_{-i}|}{|\mathbf{R}|}$
  - 6:   **end for**
  - 7: **end for**
- 

---

**Algorithm 4** Obtaining Marquardt VIF (REMVIF)

---

**Require:** Calculate  $\mathbf{X}'\mathbf{X}$ ,  $\mathbf{I}$  and  $D(\delta)$  (discretization of the interval  $[0, 1]$  with  $\delta$  points)

- 1: **for**  $\lambda \in D(\delta)$  **do**
  - 2:   calculate  $\mathbf{M}(\lambda) = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}$
  - 3: **end for**
  - 4: consider the diagonal elements of  $\mathbf{M}(\lambda)$  as REMVIF
-

## D Data empirical application

Table 8: Data for empirical application. Source: Economic reports of the President (2011)

year	$Y$	$X_1$	$X_2$	$X_3$
1995	4524.80	6076.20	6200.90	1140.74
1996	4792.40	6288.30	6591.60	1253.43
1997	5104.80	6520.40	7000.70	1324.76
1998	5589.50	6862.30	7525.40	1420.99
1999	6195.10	7237.60	7910.80	1531.10
2000	6752.60	7604.60	8559.40	1716.96
2001	7460.40	7810.30	8883.30	1867.85
2002	8361.20	8018.30	9060.10	1972.11
2003	9376.20	8244.50	9378.10	2077.36
2004	10650.70	8515.80	9937.20	2192.24
2005	12097.70	8803.50	10485.90	2290.93
2006	13481.90	9054.50	11268.10	2384.96
2007	14566.00	9262.90	1112.30	2528.77
2008	14661.30	9211.70	12460.20	2548.86
2009	14370.00	9032.60	11867.00	2438.73
2010	13712.30	9196.20	12321.90	2545.28
2011	13383.80	9428.80	12947.30	2631.51