Computational Statistics

Transformation of variables and the Condition Number in Ridge Estimation

--Manuscript Draft--

Manuscript Number:	COST-D-16-00163R2
Full Title:	Transformation of variables and the Condition Number in Ridge Estimation
Article Type:	Original Paper
Keywords:	Collinearity Ridge Regression Condition Number Econometric Models, Monte Carlo
Corresponding Author:	Catalina García Universidad de Granada SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad de Granada
Corresponding Author's Secondary Institution:	
First Author:	ROMAN SALMERON
First Author Secondary Information:	
Order of Authors:	ROMAN SALMERON
	JOSE GARCIA
	Catalina García
	MARIA DEL MAR LOPEZ
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Ridge estimation (RE) is an alternative method to ordinary least squares (OLS) estimation when collinearity is detected in a linear regression model. After applying RE, it is sensible to determine whether such collinearity has been mitigated. The condition number (CN) is a commonly applied measure to detect the presence of collinearity in econometric models, but to the best of our knowledge, it has not been extended to be applied after RE. In OLS estimation, Belsley et al (1980) established that the regressors must be of unit length and not centered to correctly calculate the CN. This paper reviews this requirement in the context of RE and presents an expression to calculate the CN in RE.

Computational Statistics manuscript No.
(will be inserted by the editor)

Transformation of variables and the Condition Number in Ridge Estimation

Román Salmerón · José García · Catalina García · María del Mar López

Received: date / Accepted: date

Abstract Ridge estimation (RE) is an alternative method to ordinary least squares (OLS) estimation when collinearity is detected in a linear regression model. After applying RE, it is sensible to determine whether such collinearity has been mitigated. The condition number (CN) is a commonly applied measure to detect the presence of collinearity in econometric models, but to the best of our knowledge, it has not been extended to be applied after RE. In OLS estimation, Belsley et al (1980) established that the regressors must be of unit length and not centered to correctly calculate the CN. This paper reviews this requirement in the context of RE and presents an expression to calculate the CN in RE.

Keywords Collinearity · Ridge Regression · Condition Number · Econometric Models

1 Introduction

Collinearity has been widely analyzed by econometricians from three research perspectives: analyzing its consequences, diagnosing its causes and identifying estimation methodologies to address it. Regarding the consequences of collinearity, many authors have stated that when there is near collinearity, the results of the estimation are unstable and present inflated variances, covariances, correlations, and estimated variance, instability in the estimated parameters and problems with individual significance tests (Farrar and Glauber, 1967; Gunst and Mason, 1977; Marquardt, 1970; Marquardt and Snee, 1975; Silvey, 1969; Willan and Watts, 1978). Given the relevant consequences of the existence of near collinearity, a good diagnosis will be essential.

Román Salmerón
Department of Quantitative Methods for Economics and Business
E-mail: romansg@ugr.es

José García
Department of Economic and Business, University of Almería, Spain
E-mail: jgarcia@ual.es

Catalina García
Department of Quantitative Methods for Economics and Business
E-mail: cbgarcia@ugr.es

María del Mar López
Department of Mathematics Didactic, University of Granada, Spain
E-mail: mariadelmarlopez@ugr.es

Despite its relevance, there is not single and general methodology to diagnose the existence of collinearity. Commonly applied measures include the Variance Inflation Factor (VIF) (Fox and Monette, 1992; García et al, 2015a; Marquardt, 1970; Theil, 1971), the eigenvalues obtained through the Condition Index (CI) (Belsley, 1982; Belsley et al, 1980), the Condition Number (CN) (Marshall and Olkin, 1965, 1969; Casella, 1980b; Belsley et al, 1980; Casella, 1980a; Besley, 1991; Lazaridis, 2007), the Variance Decomposition Proportions (VDP), used to analyze the correlations between different vectors and their angles (Rawlings et al, 1998; Wichers, 1975), and the biplot method, which when used to visually diagnose collinearity problems is called a collinearity biplot (Friendly and Kwan, 2009). Another much-discussed issue in the literature is that there is no statistical test to objectively determine the presence of collinearity but only 'rules of thumb'. For instance, it is generally accepted that when the VIF is higher than 10, there is collinearity. Other authors suggest that one should consider collinearity to be present given VIF values higher than 4. In the case of the CN, it is considered that the collinearity is moderate for values below 30, high for values between 30 and 100 and unacceptable for values higher than 100. The threshold can vary from author to author. Other collinearity measures, for example the *Red* indicator proposed by Kovacs et al (2005), do not even have an established threshold.

Regardless of the indicator used, once near collinearity is detected, it is necessary to select an appropriate methodology to estimate the model. It is not the goal of this paper to analyze the best estimation method in presence of collinearity, but ridge estimation (RE) (Hoerl and Kennard, 1970b,a) is certainly one of the most commonly applied methods. Figure 1 shows that extending RE to incorporate the CN , and collinearity diagnostic indicators in general, is motivated by the need to verify that after applying this estimation methodology from an appropriately chosen k value, the collinearity has been mitigated to a sufficient extent to not be a problem for the estimation. **Thus, the question is, what should one expect in terms of an appropriately chosen k value? For example, Hoerl et al (1975) proposed a value of $k = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}}$ that has a probability greater than 0.5 of producing estimations with a smaller mean square error (MSE) than ordinary least squares. However, this choice does not guarantee that the CN of $\mathbf{X}'\mathbf{X}+k\mathbf{I}$ will be less than the established threshold. Another example can be found in McDonald (2010), where different criteria for selecting k are proposed. One of these criteria leads to VIF values higher than 10, indicating that collinearity has not been mitigated. The same occurs if we apply the expression proposed by Hoerl et al (1975) to the example of McDonald (2010).**

Therefore, if our goal is to mitigate collinearity, we should not only seek to reduce the MSE but also to select a value of k that produces CN values that are less than the established threshold. Thus, to avoid that the CN of $\mathbf{X}'\mathbf{X} + k\mathbf{I}$ will be high after the application of RE, the next step could be to develop criteria to select k that consider not only the MSE but also collinearity diagnostic measures. Then, a possible criterion for selecting k could be to select the value of k that presents the lowest MSE and a CN value lower than the established threshold. Therefore, it is sensible to extend the CN to be applied after RE.

In this case, while the detection of collinearity after the application of ordinary least squares (OLS) is a widely discussed problem for which many different indicators (CN, VIF, CI, VDP, *Red*, etc.) have been proposed, there are few references regarding the application of these measures within RE. The extension of these measures to RE requests a special analysis, such as that in García et al (2015a), who presented an extension of the VIF to RE by applying the augmented model proposed by Marquardt (1970). Thus, García et al (2015a) showed that the expression of the VIF as traditionally applied in RE does not satisfy the following conditions: decreasing in k, continuous for $k = 0$ and always equal to or higher than one. Thus, the diagnosis generated by the expression traditionally applied in RE can be misleading or even erroneous.

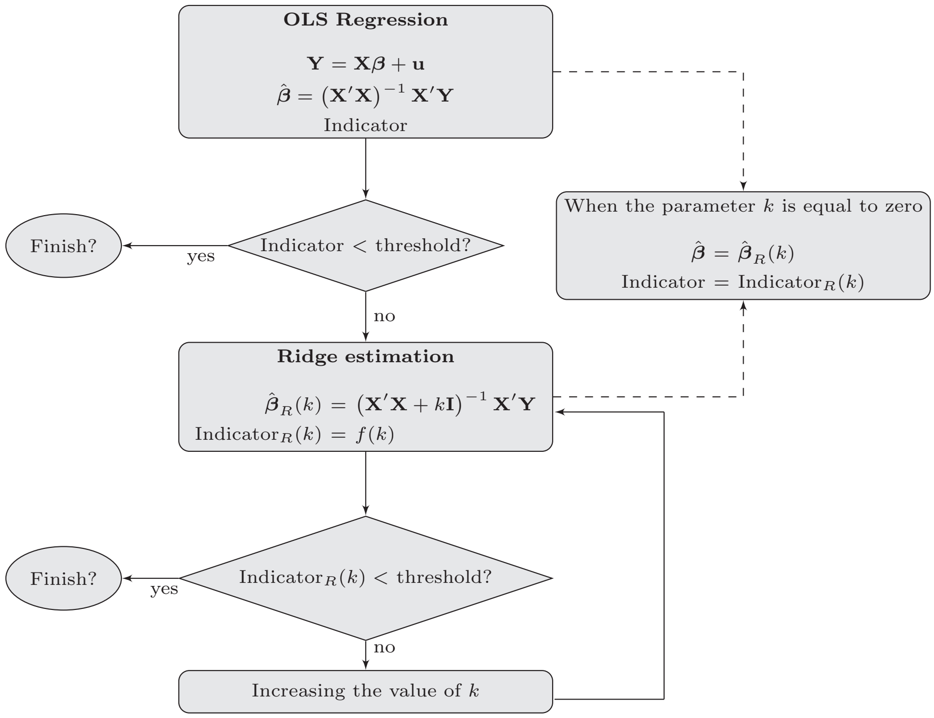

Fig. 1: Collinearity diagnostic in ridge estimation

Recall that Belsley et al (1980) established that the regressors must be of unit length and not centered to correctly calculate the CN in OLS. However, by following Belsley et al (1980), Lazaridis (2007, pp.130) showed that different scenarios for a given model lead to different values of the CN in OLS. The main goal of this paper is to analyze how different transformations of the original data (such as the standardization widely applied in data with collinearity or unit length recommended by Belsley et al (1980) for the calculation of the CN) affect the calculation of the CN in OLS and RE. **In this sense, Belsley et al (1980) highlighted that in OLS, "the general problem of optimal scaling-column scaling that results in a data matrix $\mathbf{X}$ with minimal condition number $k(\mathbf{X})$ remains unsolved. However, scaling for equal column lengths (which our unit column length is but a simple means for effecting) has known near-optimal properties in this regard (pp. 184)."** It is also is important to recognize that when RE is applied, the matrix $\mathbf{X}'\mathbf{X}$ is transformed into $\mathbf{X}'\mathbf{X} + k\mathbf{I}$, and then the transformations performed on the matrix $\mathbf{X}'\mathbf{X}$ are not retained in the matrix $\mathbf{X}'\mathbf{X} + k\mathbf{I}$.

Finally, although the most widely applied measure to detect collinearity is the VIF, its application is not always possible. For example, this is the case when any of the independent variables is qualitative because of the problems generated in the coefficient of determination of the auxiliary regression. **Another example could be a moderated regression when the interaction term is obtained from a dichotomous variable. Furthermore, the VIF is not robust to the presence of high leverage points (outliers) and can be affected by the sample size (the coefficient of determination and the size of the data tend to be inversely related, such that with a small number of observations, it is easy to obtain a high coefficient of determination). Furthermore, taking into account the expression for the estimated variance of the estimated parameters, high VIF values**

might not imply high estimated variance because it can be countered by the ratio of the variance of the error terms divided by the variation in the respective independent variable. In these cases, an alternative is the CN. For this reason, this paper focuses on the CN, presenting two expressions to extend it to RE and analyzing its properties. The structure of the paper is as follows: Section 2 presents the CN in OLS, while Sections 3 and 4 show, under different assumptions for the matrix of the augmented model proposed by Marquardt (1970), two extensions of the CN to RE. The results are illustrated with an empirical application presented in Section 5. Finally, section 6 provides the main conclusions.

2 The Condition Number in OLS

Given the linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \tag{1}$$

where the well-known basic hypotheses are verified and the dimension of the matrix $\mathbf{X} = (\mathbf{X}_1 \dots \mathbf{X}_m)$ is $n \times m$ ($n > m$), where $\mathbf{X}_j$ is $n \times 1$ with $j = 1, \dots, m$, and the condition number is defined, following Belsley et al (1980); Rawlings et al (1998), as follows:

$$K(\mathbf{X}) = \frac{\mu_{max}}{\mu_{min}}, \tag{2}$$

where μ_j are the singular values of the matrix $\mathbf{X}$ with $j = 1, ..., m$. It is known that $K(\mathbf{X}) \geq 1$. If the columns of $\mathbf{X}$ are orthogonal, then $K(\mathbf{X}) = 1$, which means that this is the lower bound and $K(\mathbf{X})$ tends to infinity in the case of perfect multicollinearity; see Lazaridis (2007).

Since the eigenvalues of the matrix $\mathbf{X}'\mathbf{X}$, λ_j, with $j = 1, ..., m$, coincide[1] with the square of singular values of matrix $\mathbf{X}$, that is $\mu_j^2 = \lambda_j$, it is possible to define the condition number with the following expression:

$$K(\mathbf{X}) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}. \tag{3}$$

Belsley et al (1980) stated that values of $K(\mathbf{X})$ lower than 20 imply light collinearity, values between 20 and 30 imply moderate collinearity, and values higher than 30 imply strong collinearity.

In addition, Belsley et al (1980) established that the regressors must be of unit length and not centered. Based on this statement, we derive the following considerations are presented:

a) It is not clear that the data should not be centered. Belsley (1984) argues that mean-centering typically masks the role of the constant term in any underlying near dependencies. In contrast, Marquardt (1980) states that centering observations removes the nonessential ill conditioning. Further, we can suppose that the data are centered if we do not wish to study the influence of the independent term.
b) What does it mean that the data should have unit length? Belsley et al (1980), page 120, argued that having unit length is similar to transforming the cross-products matrix $\mathbf{X}'\mathbf{X}$ into a correlation matrix, except that the mean zero property is not needed. This is to say that each variable should be divided by its standard deviation multiplied by the square root of the number of observations. However, Draper and Smith (1998) and Greene (1993) held that the unit length should be obtained by dividing each variable by its length, it is to say, by the square root of the sum of every element squared.

[1] The singular value decomposition of $\mathbf{X}$ is $\mathbf{X} = \mathbf{UDV}'$ where $\mathbf{U}'\mathbf{U} = \mathbf{I} = \mathbf{V}'\mathbf{V}$. Thus, the matrix $\mathbf{X}'\mathbf{X} = \mathbf{VDU}'\mathbf{UDV}' = \mathbf{VD}^2\mathbf{V}'$.

Regarding the first question, we should highlight that the decision of whether to center the data is not a minor question. In fact, García et al (2015b) showed that the values of the CN in OLS are different and that transforming the data can lead to different conclusions about the existence of collinearity. These authors recommended centering the data, and the present paper will review this recommendation. Regarding the definition of the term 'unit length,' this paper will consider the definitions of Belsley et al (1980) and Draper and Smith (1998) and an additional transformation, namely, dividing every variable by its standard deviation. Thus, the extensions to RE proposed in the following sections are performed under the following scenarios:

S.1) original uncentered data.
S.2) original centered data.
S.3) original uncentered data divided by their standard deviation.
S.4) original centered data divided by their standard deviation (typified data[2]).
S.5) original data divided by their standard deviation multiplied by the square root of the number of observations.
S.6) original centered data divided by their standard deviation multiplied by the square root of the number of observations (standardized data[3]).
S.7) original uncentered data divided by the square root of the sum of every variable squared (we will consider this transformation to be the unit length).
S.8) original centered data divided by the square root of the sum of every variable squared[4].

From an algebraic perspective, it has not considered the first column of ones (intercept) in matrix $\mathbf{X}$ because of the following:

- When the variables are centered, the intercept is lost and the first column of the resulting matrix $\mathbf{X}$ contains zeros. Then, one of the eigenvalues will be zero, and consequently, the condition number will be infinite. For this reason, in scenarios S.2, S.4, S.6 and S.8, it is not possible to consider the intercept in matrix $\mathbf{X}$.
- If there is an intercept, it is not possible to divide by its variance because the latter is zero. For this reason, in situations S.3, S.4, S.5, S.6 and S.8, it is not possible to include the intercept in matrix $\mathbf{X}$.

However, from an econometric perspective, it is possible to contemplate the existence of an intercept in scenarios S.1 and S.7. Thus, two additional scenarios are considered:

S.9) original uncentered data (S.1), where $\mathbf{X}$ contents an intercept.
S.10) original uncentered data divided by the square root of the sum of every variable squared (S.7), where $\mathbf{X}$ contains an intercept.

Finally, to illustrate the influence of the transformation of the data on the calculation of the CN, we calculate it in OLS for the above scenarios and for the following sets of data previously considered in the literature:

Data 1: 14 observations with 365 days of aging (see Alkhamisi and MacNeill (2015)).
Data 2: 10 observations of the percentage of GNP spent on total national research and development expenditures by country from 1972 to 1986 (see Alkhamisi and MacNeill (2015)).

[2] If the data are typified, their mean is zero, variance is equal to 1 and the cross products matrix is equal to the correlation matrix multiplied by the number of observations.

[3] If the data are standardized, their mean is zero, variance is equal to 1 divided by the number of observations and the the cross products matrix is equal to the correlation matrix.

[4] Note that if data are centered, this transformation coincides with standardization since $\sum\limits_{i=1}^{n} x_{ij}^2 = n\left(Var(\mathbf{X}_j) - \overline{\mathbf{X}}_j\right) = n \cdot Var(\mathbf{X}_j)$ for $j = 1, \ldots, m$.

	Data 1	Data 2	Data 3	Data 4	Data 5
S.1	21.05	93.68	57.509	12235103125.707	29931285.78
S.2	52.15	12.33	5.25	10799690757.19	10968511.56
S.3	10.25	171.74	41.09	3109.65	3109.65
S.4	42.67	12.1	3.209	7085.38	7085.38
S.5	10.25	171.74	41.09	3109.65	3109.65
S.6	42.67	12.1	3.209	7085.38	7085.38
S.7	9.602	86.08	25.506	1590.76	1590.76
S.8	42.67	12.1	3.209	7085.38	7085.38
S.9	608.08	410.03	1811.52	87917026016.11	21854820807.84
S.10	251.901	232.32	50.09	92055.89	92055.89

Table 1: Calculation of the CN in OLS for different data sets

Data 3: 21 observations of the oxidation of ammonia to nitric acid (see Alkhamisi and MacNeill (2015)).
Data 4: 16 observations of the percentage of the conversion of n-heptane to acetylene (see Macedo (2015)).
Data 5: The above data set but after modifying the scale of one of the independent variables by multiplying it by 10,000 (see Macedo (2015)).

From Table 1 it is possible to conclude the following:

– In the first three data sets, whether one concludes that collinearity exists changes depending on the scenario considered given the decision threshold of CN$>$ 30.
– The results from all of data sets show the number of condition matches for situations S.3 and S.5 and for situations S.4 and S.6 (recall that S.6 and S.8 coincide in this case). This will be shown in Appendix A.
– From data sets 4 and 5, if the data are not transformed (scenarios S.1, S.2 and S.9), then the CN is affected by the changes in scale. Thus, as stated by Stewart (1987), *the remedy for this problem is to adopt a standard scaling of the columns before computing the condition number, but what the standard should be is by no means clear.*

Then, it seems evident that the transformation of the data is necessary, but it is unclear which transformation is more appropriate. Answering this question for RE is the main goal of this paper.

3 Condition Number in Ridge Estimation: extension 1

The augmented model proposed by Marquardt (1970) allows us to know the columns of matrix $\mathbf{X}$ associated with matrix $\mathbf{X}'\mathbf{X}+k\mathbf{I}$, thereby making it feasible to transform this matrix. For this reason, to extend the CN to be applied in RE, we begin with the augmented model given by

$$\mathbf{y}^R = \mathbf{X}^R\beta + \mathbf{u}^R, \tag{4}$$

where $\mathbf{X}^R_{(n+m)\times m} = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ and $\mathbf{Y}_{(n+m)\times 1} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{0}_m \end{pmatrix}$, with $\mathbf{I}_m$ being the identity matrix of order m and $\mathbf{0}_m$ a zero vector with m rows.

By denoting the CN in the RE as $K(\mathbf{X}^R, k)$ it can be obtained as follows:

$$K(\mathbf{X}^R, k) = \sqrt{\frac{\delta_{max}}{\delta_{min}}}, \tag{5}$$

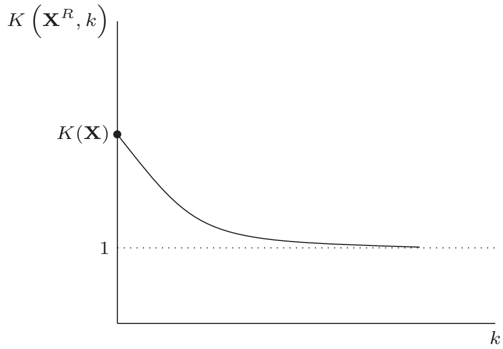

Fig. 2: The continuity, monotony and limit of the condition number

where δ_j, $j = 1, \ldots, m$, are the eigenvalues of the matrix $(\mathbf{X}^R)'\mathbf{X}^R$. Since $(\mathbf{X}^R)'\mathbf{X}^R = \mathbf{X}'\mathbf{X} + k\mathbf{I}$ and from the diagonalization $\mathbf{X}'\mathbf{X} = \mathbf{Q}\mathbf{D}_{\lambda_j}\mathbf{Q}'$, where $\mathbf{D}_{\lambda_j}$ is a diagonal matrix of eigenvalues and $\mathbf{Q}$ is orthogonal ($\mathbf{Q}\mathbf{Q}' = \mathbf{I}$), it is verified that $\mathbf{X}'\mathbf{X} + k\mathbf{I} = \mathbf{Q}\mathbf{D}_{\lambda_j}\mathbf{Q}' + k\mathbf{Q}\mathbf{Q}' = \mathbf{Q}(\mathbf{D}_{\lambda_j} + k\mathbf{I})\mathbf{Q}'$. Thus, we obtain that $\delta_j = \lambda_j + k$. Consequently, the CN in RE is given by

$$K(\mathbf{X}^R, k) = \sqrt{\frac{\lambda_{max} + k}{\lambda_{min} + k}}. \tag{6}$$

It is easy to show that $K(\mathbf{X}^R, k) \leq K(\mathbf{X})$ for all k. Furthermore, Appendix B shows that $K(\mathbf{X}^R, k)$ is decreasing in k, continuous for $k = 0$ (that is to say that it coincides with the CN obtained in OLS) and always equal to or higher than 1. From these properties, it is clear that $K(\mathbf{X}^R, k)$ will present a representation similar to Figure 2.

Taking into account the scenarios established in section 2, it is possible to distinguish the following situations:

1) Scenario S.1 for $\mathbf{X}$, then expression (6) will be denoted as $K_U(\mathbf{X}^R, k)$.
2) Scenario S.2 for $\mathbf{X}$, then expression (6) will be denoted as $K_C(\mathbf{X}^R, k)$.
3) Scenario S.3 for $\mathbf{X}$, then expression (6) will be denoted as $K_{UT}(\mathbf{X}^R, k)$.
4) Scenario S.4 for $\mathbf{X}$, then expression (6) will be denoted as $K_T(\mathbf{X}^R, k)$.
5) Scenario S.5 for $\mathbf{X}$, then expression (6) will be denoted as $K_{US}(\mathbf{X}^R, k)$.
6) Scenario S.6 for $\mathbf{X}$, then expression (6) will be denoted as $K_S(\mathbf{X}^R, k)$.
7) Scenario S.7 for $\mathbf{X}$, then expression (6) will be denoted as $K_{UL}(\mathbf{X}^R, k)$.
8) Scenario S.8 for $\mathbf{X}$, then expression (6) will be denoted as $K_{CUL}(\mathbf{X}^R, k)$.
9) Scenario S.9 for $\mathbf{X}$, then expression (6) will be denoted as $K_{Uit}(\mathbf{X}^R, k)$.
10) Scenario S.10 for $\mathbf{X}$, then expression (6) will be denoted as $K_{ULit}(\mathbf{X}^R, k)$.

4 Condition Number in Ridge Estimation: extension 2

Although the previously presented extension verifies the properties highlighted in the introduction (continuity, monotony and being equal to or higher than one), the matrix $\mathbf{X}^R = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ does not have vectors of unit length. Therefore, this section proposes an alternative extension for calculating the CN in RE, distinguishing the same ten scenarios.

4.1 Centered or uncentered variables

Beginning with the matrix $\mathbf{X}$ in S.1 or S.2, we will encounter a problem because the matrix $\mathbf{X}^R = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is not in S.1 or S.2. Then, we argue that to correctly extend the CN to RE, the matrix $\mathbf{X}^R$ has to be in S.1 (uncentered columns) or S.2 (centered columns).

Scenario 1

If $\mathbf{X}$ is in S.1, then $\mathbf{X}^R = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is in S.1, and hence, the CN in RE will be obtained from the eigenvalues $\left(\mathbf{X}^R\right)' \mathbf{X}^R = \mathbf{X}'\mathbf{X} + k\mathbf{I}$. Thus, the resulting eigenvalues should coincide with the eigenvalues obtained from $K_U(\mathbf{X}^R, k)$.

Scenario 2

However, if $\mathbf{X}$ is not in S.2, we should transform it in the following way:

$$\widetilde{\mathbf{X}}_C = \mathbf{X} - \mathbf{i}\overline{\mathbf{X}} = \begin{pmatrix} X_{11} - \overline{X}_1 & \cdots & X_{1m} - \overline{X}_m \\ \vdots & \ddots & \vdots \\ X_{n1} - \overline{X}_1 & \cdots & X_{nm} - \overline{X}_m \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1}, \quad \overline{\mathbf{X}} = \left(\overline{X}_1 \dots \overline{X}_m\right)_{1\times m}.$$

However, $\mathbf{X}_C^R = \begin{pmatrix} \widetilde{\mathbf{X}}_C \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is not in S.2, and the following transformation is needed $\mathbf{x}_C^R = \mathbf{X}_C^R - \mathbf{i}^R\overline{\mathbf{X}}_C^R$. Then, [5]:

$$\begin{aligned} \left(\mathbf{x}_C^R\right)' \mathbf{x}_C^R &= \left(\mathbf{X}_C^R - \mathbf{i}^R\overline{\mathbf{X}}_C^R\right)' \left(\mathbf{X}_C^R - \mathbf{i}^R\overline{\mathbf{X}}_C^R\right) = \widetilde{\mathbf{X}}_C'\widetilde{\mathbf{X}}_C + k\mathbf{I}_m - \frac{k}{n+m}\boldsymbol{\Pi} \\ &= \mathbf{X}'\mathbf{X} + k\mathbf{I}_m - n\overline{\mathbf{X}}'\overline{\mathbf{X}} - \frac{k}{n+m}\boldsymbol{\Pi}, \end{aligned} \tag{7}$$

where:

$$\mathbf{i}^R_{(n+m)\times 1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \overline{\mathbf{X}}_C^R = \mathbf{i}'\frac{\sqrt{k}}{n+m}, \quad \boldsymbol{\Pi} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}_{m\times m}.$$

Expression (7) allows us to define the CN associated with matrix $\mathbf{x}_C^R$, which we will denote as $K_C(\mathbf{x}^R, k)$. Note that there is not a closed expression, but it is easily programmable.

4.2 Typified variables

When transforming matrix $\mathbf{X}$ in S.3 or S.4, that is,

$$\widetilde{\mathbf{X}}_{UT} = \mathbf{X}\cdot\boldsymbol{\Omega} = \begin{pmatrix} \frac{X_{11}}{\sqrt{Var(X_1)}} & \cdots & \frac{X_{1m}}{\sqrt{Var(X_m)}} \\ \vdots & \ddots & \vdots \\ \frac{X_{n1}}{\sqrt{Var(X_1)}} & \cdots & \frac{X_{nm}}{\sqrt{Var(X_m)}} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{\sqrt{Var(\mathbf{X}_1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{Var(\mathbf{X}_m)}} \end{pmatrix},$$

[5] If $\mathbf{X}^R = \begin{pmatrix} \widetilde{\mathbf{X}} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ where the variables in $\widetilde{\mathbf{X}}$ have zero mean, and thus, $\overline{X}_j^R = \frac{\sqrt{k}}{n+m}$ for $j = 1, \dots, m$.

or

$$\widetilde{\mathbf{X}}_T = \left(\mathbf{X} - \mathbf{i}\overline{\mathbf{X}}\right) \cdot \boldsymbol{\Omega} = \widetilde{\mathbf{X}}_C \cdot \boldsymbol{\Omega} = \begin{pmatrix} \frac{X_{11}-\overline{\mathbf{X}}_1}{\sqrt{Var(X_1)}} & \cdots & \frac{X_{1m}-\overline{\mathbf{X}}_m}{\sqrt{Var(X_m)}} \\ \vdots & \ddots & \vdots \\ \frac{X_{n1}-\overline{\mathbf{X}}_1}{\sqrt{Var(X_1)}} & \cdots & \frac{X_{nm}-\overline{\mathbf{X}}_m}{\sqrt{Var(X_m)}} \end{pmatrix},$$

we will encounter a problem because matrix $\mathbf{X}^R = \begin{pmatrix} \widetilde{\mathbf{X}} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is not in S.3 (uncentered columns) or S.4 (centered columns) and the CN will not be correctly extended to RE.

Scenario 3

Beginning from matrix $\widetilde{\mathbf{X}}$ in S.3 (the variance of each column is 1), to obtain that $\mathbf{X}^R$ is in S.3, the following transformation is required $\mathbf{x}^R_{UT} = \mathbf{X}^R_{UT} \cdot \boldsymbol{\Omega}_{UT}$, where:

$$\mathbf{X}^R_{UT} = \begin{pmatrix} \widetilde{\mathbf{X}}_{UT} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix},$$

$$\boldsymbol{\Omega}_{UT} = \begin{pmatrix} \frac{1}{\sqrt{Var\left(\mathbf{X}^R_{1,UT}\right)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{Var\left(\mathbf{X}^R_{2,UT}\right)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{Var\left(\mathbf{X}^R_{m,UT}\right)}} \end{pmatrix},$$

where $\mathbf{X}^R_{j,UT}$ is the j-th column of $\mathbf{X}^R_{UT}$ and $j = 1, \dots, m$. In this case:

$$\begin{aligned} \left(\mathbf{x}^R_{UT}\right)' \mathbf{x}^R_{UT} &= \boldsymbol{\Omega}_{UT} \cdot \left(\mathbf{X}^R_{UT}\right)' \mathbf{X}^R_{UT} \cdot \boldsymbol{\Omega}_{UT} = \boldsymbol{\Omega}_{UT} \cdot \left(\widetilde{\mathbf{X}}'_{UT}\widetilde{\mathbf{X}}_{UT} + k\mathbf{I}_m\right) \cdot \boldsymbol{\Omega}_{UT} \\ &= \boldsymbol{\Omega}_{UT} \cdot \left(\boldsymbol{\Omega}\mathbf{X}'\mathbf{X}\boldsymbol{\Omega} + k\mathbf{I}_m\right) \cdot \boldsymbol{\Omega}_{UT}. \end{aligned} \tag{8}$$

Scenario 4

When beginning from matrix $\widetilde{\mathbf{X}}$ in S.4 (the mean and variance of each column are zero and one, respectively), to obtain that $\mathbf{X}^R$ is in S.4, the following transformation is required: $\mathbf{x}^R_T = \left(\mathbf{X}^R_T - \mathbf{i}^R\overline{\mathbf{X}}^R_T\right) \cdot \boldsymbol{\Omega}_T$, where[6]:

$$\mathbf{X}^R_T = \begin{pmatrix} \widetilde{\mathbf{X}}_T \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}, \quad \overline{\mathbf{X}}^R_T = \mathbf{i}'\frac{\sqrt{k}}{n+m}, \quad \boldsymbol{\Omega}_T = \left(\sqrt{\frac{n(n+m) + (n+m-1)\cdot k}{(n+m)^2}}\right)^{-1} \cdot \mathbf{I}_m.$$

In this case,

$$\begin{aligned} \left(\mathbf{x}^R_T\right)' \mathbf{x}^R_T &= \Omega_T \left(\mathbf{X}^R_T - \mathbf{i}^R\overline{\mathbf{X}}^R_T\right)' \left(\mathbf{X}^R_T - \mathbf{i}^R\overline{\mathbf{X}}^R_T\right) \cdot \boldsymbol{\Omega}_T \\ &= \Omega_T \left(\Omega\left(\mathbf{X}'\mathbf{X} - n\overline{\mathbf{X}}'\overline{\mathbf{X}}\right)\Omega + k\mathbf{I}_m - \frac{k}{n+m}\boldsymbol{\Pi}\right)\Omega_T. \end{aligned} \tag{9}$$

---

[6] If $\mathbf{X}^R = \begin{pmatrix} \widetilde{\mathbf{X}} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$, where the variables contained in $\widetilde{\mathbf{X}}$ have zero mean and variance equal to one, and it is verified that $\overline{X}^R_j = \frac{\sqrt{k}}{n+m}$ and $Var\left(X^R_j\right) = \frac{n(n+m)+(n+m-1)k}{(n+m)^2}$ for $j = 1, \dots, m$ (see Appendix C).

Expressions (8) and (9) allow us to define the CN associated with matrix $\mathbf{x}_{UT}^R$ or $\mathbf{x}_T^R$, which we will denote as $K_{UT}(\mathbf{x}^R, k)$ and $K_T(\mathbf{x}^R, k)$, respectively. Note that there is not a closed expression, but it is easily programmable.

4.3 Standardized variables

By transforming matrix $\mathbf{X}$ in S.5 or S.6, we obtain the following:

$$\widetilde{\mathbf{X}}_{US} = \frac{1}{\sqrt{n}} \cdot \mathbf{X} \cdot \boldsymbol{\Omega} = \frac{1}{\sqrt{n}} \widetilde{\mathbf{X}}_{UT} \begin{pmatrix} \frac{X_{11}}{\sqrt{n \cdot Var(X_1)}} & \cdots & \frac{X_{1m}}{\sqrt{n \cdot Var(X_m)}} \\ \vdots & \ddots & \vdots \\ \frac{X_{n1}}{\sqrt{n \cdot Var(X_1)}} & \cdots & \frac{X_{nm}}{\sqrt{n \cdot Var(X_m)}} \end{pmatrix},$$

or

$$\widetilde{\mathbf{X}}_S = \frac{1}{\sqrt{n}} \cdot \left(\mathbf{X} - \mathbf{i}\overline{\mathbf{X}}\right) \cdot \boldsymbol{\Omega} = \frac{1}{\sqrt{n}} \cdot \widetilde{\mathbf{X}}_C \cdot \boldsymbol{\Omega} = \frac{1}{\sqrt{n}} \cdot \widetilde{\mathbf{X}}_T = \begin{pmatrix} \frac{X_{11} - \overline{\mathbf{X}}_1}{\sqrt{n \cdot Var(X_1)}} & \cdots & \frac{X_{1m} - \overline{\mathbf{X}}_m}{\sqrt{n \cdot Var(X_m)}} \\ \vdots & \ddots & \vdots \\ \frac{X_{n1} - \overline{\mathbf{X}}_1}{\sqrt{n \cdot Var(X_1)}} & \cdots & \frac{X_{nm} - \overline{\mathbf{X}}_m}{\sqrt{n \cdot Var(X_m)}} \end{pmatrix},$$

we will encounter a problem because matrix $\mathbf{X}^R = \begin{pmatrix} \widetilde{\mathbf{X}} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is not in S.5 (uncentered columns) or S.6 (centered columns), and the CN will not be correctly extended to RE.

Scenario 5

Beginning from matrix $\widetilde{\mathbf{X}}$ in S.5 (the variance of each column is $\frac{1}{n}$), to obtain that $\mathbf{X}^R$ is in S.5, the following transformation is required $\mathbf{x}_{US}^R = \mathbf{X}_{US}^R \cdot \boldsymbol{\Omega}_{US}$, where:

$$\mathbf{X}_{US}^R = \begin{pmatrix} \widetilde{\mathbf{X}}_{US} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix},$$

$$\boldsymbol{\Omega}_{US} = \frac{1}{\sqrt{n+m}} \cdot \boldsymbol{\Omega}_{UT} = \begin{pmatrix} \frac{1}{\sqrt{(n+m) \cdot Var\left(\mathbf{X}_{1,US}^R\right)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{(n+m) \cdot Var\left(\mathbf{X}_{2,US}^R\right)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{(n+m) \cdot Var\left(\mathbf{X}_{m,US}^R\right)}} \end{pmatrix},$$

where $\mathbf{X}_{j,US}^R$ is the j-the column of $\mathbf{X}_{US}^R$ with $j = 1, \dots, m$. In this case,

$$\begin{aligned} \left(\mathbf{x}_{US}^R\right)' \mathbf{x}_{US}^R &= \boldsymbol{\Omega}_{US} \cdot \left(\mathbf{X}_{US}^R\right)' \mathbf{X}_{US}^R \cdot \boldsymbol{\Omega}_{US} = \boldsymbol{\Omega}_{US} \cdot \left(\widetilde{\mathbf{X}}_{US}' \widetilde{\mathbf{X}}_{US} + k\mathbf{I}\right) \cdot \boldsymbol{\Omega}_{US} \\ &= \boldsymbol{\Omega}_{US} \cdot \left(\frac{1}{n} \boldsymbol{\Omega} \mathbf{X}' \mathbf{X} \boldsymbol{\Omega} + k\mathbf{I}\right) \cdot \boldsymbol{\Omega}_{US} \\ &= \frac{1}{n+m} \boldsymbol{\Omega}_{UT} \cdot \left(\frac{1}{n} \boldsymbol{\Omega} \mathbf{X}' \mathbf{X} \boldsymbol{\Omega} + k\mathbf{I}\right) \cdot \boldsymbol{\Omega}_{UT}. \end{aligned} \quad (10)$$

Scenario 6

However, beginning from matrix $\widetilde{\mathbf{X}}$ in S.6 (the mean and variance of each column is zero and $\frac{1}{n}$, respectively), to obtain that $\mathbf{X}^R$ is in S.6, the following transformation is required $\mathbf{x}_S^R = \left(\mathbf{X}_S^R - \mathbf{i}^R\overline{\mathbf{X}}_S^R\right) \cdot \boldsymbol{\Omega}_S$, where:[7]:

$$\mathbf{X}_S^R = \begin{pmatrix} \widetilde{\mathbf{X}}_S \\ \sqrt{k}\mathbf{I}_m \end{pmatrix} \quad \overline{\mathbf{X}}_S^R = \mathbf{i}' \frac{\sqrt{k}}{n+m}, \quad \boldsymbol{\Omega}_S = \left(\sqrt{\frac{(n+m)+(n+m-1)\cdot k}{n+m}}\right)^{-1} \cdot \mathbf{I}_m.$$

In this case,

$$\begin{aligned} \left(\mathbf{x}_S^R\right)' \mathbf{x}_S^R &= \Omega_S \left(\mathbf{X}_S^R - \mathbf{i}^R\overline{\mathbf{X}}_S^R\right)' \left(\mathbf{X}_S^R - \mathbf{i}^R\overline{\mathbf{X}}_S^R\right) \cdot \boldsymbol{\Omega}_S \\ &= \Omega_S \left(\frac{1}{n}\Omega\left(\mathbf{X}'\mathbf{X} - n\overline{\mathbf{X}}'\overline{\mathbf{X}}\right)\Omega + k\mathbf{I}_m - \frac{k}{n+m}\boldsymbol{\Pi}\right)\Omega_S. \end{aligned} \quad (11)$$

Then, expressions (10) and (11) allow us to define the CN associated with matrix $\mathbf{x}_{US}^R$ or $\mathbf{x}_S^R$, which we will denote as $K_{US}(\mathbf{x}^R, k)$ and $K_S(\mathbf{x}^R, k)$, respectively. Note that there is not a closed expression, but it is easily programmable.

4.4 Unit length variables

Scenario 7

In this section, we assume that $\mathbf{X}^R = \begin{pmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ allows $\mathbf{X}$ to be is in S.7. In the case in which $\mathbf{X}$ is not in S.7, it should be transformed in the following way:

$$\widetilde{\mathbf{X}}_{UL} = \mathbf{X} \cdot \boldsymbol{\Omega}_{UL}, \quad \boldsymbol{\Omega}_{UL} = \begin{pmatrix} \frac{1}{||\mathbf{X}_1||} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{||\mathbf{X}_m||} \end{pmatrix},$$

where $||\mathbf{X}_j|| = \sqrt{\sum_{i=1}^{n} X_{ij}^2}$ with $j = 1, \ldots, m$. That is to say, $\widetilde{\mathbf{X}}_{j,UL} = \frac{\mathbf{X}_j}{||X_j||}$ such that $||\widetilde{\mathbf{X}}_{j,UL}|| = 1$. However, $\mathbf{X}_{UL}^R = \begin{pmatrix} \widetilde{\mathbf{X}}_{UL} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is not in S.7, and to have $\mathbf{X}^R$ be in S.7, the following transformation is required: $\mathbf{x}_{UL}^R = \mathbf{X}_{UL}^R \boldsymbol{\Omega}_{UL}^R$, where

$$\boldsymbol{\Omega}_{UL}^R = \begin{pmatrix} \frac{1}{||\mathbf{X}_{1,UL}^R||} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{||\mathbf{X}_{m,UL}^R||} \end{pmatrix},$$

[7] If $\mathbf{X}^R = \begin{pmatrix} \widetilde{\mathbf{X}} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$, where the variables contained in $\widetilde{\mathbf{X}}$ present zero mean and variance equal to $\frac{1}{n}$, and it is verified that $\overline{X}_j^R = \frac{\sqrt{k}}{n+m}$ y $Var\left(X_j^R\right) = \frac{(n+m)+(n+m-1)k}{(n+m)^2}$ for $j = 1, \ldots, m$ (see Appendix C).

being $\mathbf{X}^R_{j,UL}$ the j-th column of $\mathbf{X}^R_{UL}$ with $j = 1, \dots, m$ and with $||\mathbf{X}^R_{j,UL}|| = \sqrt{\sum\limits_{i=1}^{n} \widetilde{X}^2_{ij,UL} + k} = \sqrt{||\widetilde{\mathbf{X}}_{j,UL}||^2 + k} = \sqrt{1+k}$ for $j = 1, \dots, m$.
In this case,

$$\begin{aligned}\left(\mathbf{x}^R_{UL}\right)' \mathbf{x}^R_{UL} &= \left(\boldsymbol{\Omega}^R_{UL}\right)' \left(\mathbf{X}^R_{UL}\right)' \mathbf{X}^R_{UL} \boldsymbol{\Omega}^R_{UL} = \boldsymbol{\Omega}^R_{UL} \left(\widetilde{\mathbf{X}}'_{UL} \widetilde{\mathbf{X}}_{UL} + k\mathbf{I}_m\right) \boldsymbol{\Omega}^R_{UL} \\ &= \boldsymbol{\Omega}^R_{UL} \left(\boldsymbol{\Omega}_{UL} \mathbf{X}'\mathbf{X} \boldsymbol{\Omega}_{UL} + k\mathbf{I}_m\right) \boldsymbol{\Omega}^R_{UL}. \end{aligned} \tag{12}$$

Expression (12) allows us to define the CN associated with matrix $\mathbf{x}^R_{UL}$, which we will denote as $K_{UL}(\mathbf{x}^R, k)$. Note that, similar to the cases above, there is not a closed expression, but it is easily programmable.

Scenario 8

Now, we will see that if matrix $\mathbf{X}$ is centered, this transformation (scenario S.8) coincides with scenario S.6. Indeed, Appendix D shows that working with centered data, the transformation

$$\widetilde{\mathbf{X}}_{CUL} = \left(\mathbf{X} - \mathbf{i}\overline{\mathbf{X}}\right) \boldsymbol{\Omega}_{CUL}, \quad \boldsymbol{\Omega}_{CUL} = \begin{pmatrix} \frac{1}{||\mathbf{X}_1||} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{||\mathbf{X}_m||} \end{pmatrix},$$

is inadequate since the columns of matrix $\widetilde{\mathbf{X}}_{CUL}$ do not have length equal to 1. In that case, the correct transformation will be given by

$$\widetilde{\mathbf{X}}_{CUL} = \left(\mathbf{X} - \mathbf{i}\overline{\mathbf{X}}\right) \boldsymbol{\Omega}_{CUL} = \widetilde{\mathbf{X}}_C \boldsymbol{\Omega}_{CUL}, \quad \boldsymbol{\Omega}_{CUL} = \begin{pmatrix} \frac{1}{||\widetilde{\mathbf{X}}_{1,C}||} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{||\widetilde{\mathbf{X}}_{m,C}||} \end{pmatrix},$$

where $\widetilde{\mathbf{X}}_{j,C}$ is the j-th column of $\widetilde{\mathbf{X}}_C$ with $j = 1, \dots, m$. That is to say,

$$||\widetilde{\mathbf{X}}_{j,C}||^2 = \sum_{i=1}^{n} \widetilde{X}^2_{ij,C} = \sum_{i=1}^{n} \left(X_{ij} - \overline{\mathbf{X}}_j\right)^2 = n \cdot Var\left(\mathbf{X}_j\right), \quad j = 1, \dots, m,$$

and then $\boldsymbol{\Omega}_{CUL} = \frac{1}{\sqrt{n}} \boldsymbol{\Omega}$.
However, again, $\mathbf{X}^R_{CUL} = \begin{pmatrix} \widetilde{\mathbf{X}}_{CUL} \\ \sqrt{k}\mathbf{I}_m \end{pmatrix}$ is not in S.8, and the following transformation is required: $\mathbf{x}^R_{CUL} = \left(\mathbf{X}^R_{CUL} - \mathbf{i}^R \overline{\mathbf{X}}^R_{CUL}\right) \boldsymbol{\Omega}^R_{CUL} = \widetilde{\mathbf{X}}^R_C \boldsymbol{\Omega}_{CUL}$, where

$$\overline{\mathbf{X}}^R_{CUL} = \mathbf{i}' \frac{\sqrt{k}}{n+m}, \quad \boldsymbol{\Omega}^R_{CUL} = \begin{pmatrix} \frac{1}{||\widetilde{\mathbf{X}}_{1,C}||} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{||\widetilde{\mathbf{X}}_{m,C}||} \end{pmatrix},$$

Table 2: VIF values in ridge regression for $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$

k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	k	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
0	154.948757	37.268158	222.814314	0	154.9488	37.2682	222.8143
0.01	39.399287	21.955079	49.459193	0.1	6.7362	6.1326	7.0842
0.02	24.114188	16.601705	28.446546	0.2	3.8902	3.713	3.9924
0.03	17.718249	13.469864	20.168229	0.3	2.8876	2.802	2.937
0.04	14.135007	11.385386	15.720666	0.4	2.3768	2.3257	2.4063
0.05	11.824723	9.892666	12.938903	0.5	2.0684	2.034	2.0882
0.06	10.204944	8.769506	11.032729	0.6	1.8626	1.8377	1.8769
0.07	9.003767	7.893256	9.644171	0.7	1.716	1.6971	1.7269
0.08	8.076331	7.190354	8.587250	0.8	1.6067	1.5918	1.6153
0.09	7.338069	6.613907	7.755671	0.9	1.5223	1.5101	1.5293
0.1	6.736164	6.132574	7.084235	1	1.4553	1.4452	1.4611

and where $\widetilde{\mathbf{X}}^R_{j,C}$ is the j-th column of $\widetilde{\mathbf{X}}^R_C$ with $j = 1, \dots, m$. That is (see Appendix C),

$$||\widetilde{\mathbf{X}}^R_{j,C}||^2 = \sum_{i=1}^{n+m} \left(\widetilde{X}^R_{ij,C} \right)^2 = \sum_{i=1}^{n+m} \left(X^R_{ij} - \overline{\mathbf{X}}^R_j \right)^2$$
$$= (n+m) \cdot Var\left(\mathbf{X}^R_{S,j} \right) = \frac{(n+m) + (n+m-1)k}{n+m}, \quad j = 1, \dots, m,$$

and then $\boldsymbol{\Omega}^R_{CUL} = \boldsymbol{\Omega}_S$.
In that case,

$$\left(\mathbf{x}^R_{CUL} \right)' \mathbf{x}^R_{CUL} = \boldsymbol{\Omega}^R_{CUL} \left(\boldsymbol{\Omega}_{CUL} \left(\mathbf{X}'\mathbf{X} - n\overline{\mathbf{X}}'\overline{\mathbf{X}} \right) \boldsymbol{\Omega}_{CUL} - k\mathbf{I}_m - \frac{k}{n+m} \boldsymbol{\Pi} \right) \boldsymbol{\Omega}^R_{CUL}$$
$$= \boldsymbol{\Omega}_S \left(\frac{1}{n} \boldsymbol{\Omega} \left(\mathbf{X}'\mathbf{X} - n\overline{\mathbf{X}}'\overline{\mathbf{X}} \right) \boldsymbol{\Omega} - k\mathbf{I}_m - \frac{k}{n+m} \boldsymbol{\Pi} \right) \boldsymbol{\Omega}_S = \left(\mathbf{x}^R_S \right)' \mathbf{x}^R_S.$$

5 Empirical application

To illustrate the contribution of this paper, we analyze the data collected by Wissel (2009) but extend the sample with data from 1995 to 2011 obtained from Economic reports of the President Economic reports of the President (2011). The variables considered are mortgage debt outstanding (in trillions of dollars), $\mathbf{Y}$, personal consumption (in trillions of dollars), $\mathbf{X}_1$, personal income (in trillions of dollars), $\mathbf{X}_2$, and consumer credit outstanding (in trillions of dollars), $\mathbf{X}_3$. This data set was previously analyzed by Salmerón et al (2016) via RE, which yielded the VIF values reported in Table 2. Note that the VIF value obtained via OLS ($k = 0$) is always higher than 10, indicating the presence of collinearity.
Table 3 shows the CN in OLS obtained from expression (3) from data in scenarios S.1-S.10. Note that different transformations of the original data lead to different CN values, and in all cases, the values are higher than 30, which is the threshold above which the model is considered to exhibit severe collinearity. Indeed, as anticipated, the results for scenarios S.3 and S.5 coincide, as do the results for scenarios S.4, S.6 and S.8.
We present the CN values in RE from extensions 1 (see section 3) and 2 (see section 4) for different values of k in Tables 4 and 5, respectively. We can establish the following:

- The CN also provides different values in RE in scenarios S.1-S.10. Consequently, we can affirm that different transformations of the data lead to different conclusions regarding whether collinearity has been mitigated.

Table 3: CN values in OLS from data in scenarios S.1-S.10

Scenario	S.1	S.2	S.3	S.4	S.5
CN	5942.57595218	5609.76501621	75.16946674	33.19428616	75.16946674

Scenario	S.6	S.7	S.8	S.9	S.10
CN	33.19428616	60.09226704	33.19428616	87249657.88750964	233.88678067

- In scenarios S.1, S.2 and S.9, the results are excessively high and inconsistent. For $k = 1$, it is established that there is a high degree of collinearity.
- For scenarios S.3 and S.4, the CN decreases slowly, while for S.5, S.6, S.7, S.8 and S.10 it decreases substantially more rapidly.
- For extension 1, the CN in RE is always continuous for $k = 0$. For extension 2, it is also continuous for $k = 0$ except in S.3 and S.5.
- The CN is always decreasing in k and higher than 1.
- In OLS, the CN coincides in scenarios S.4, S.6 and S.8, while in RE it coincides in scenarios S.6 and S.8.
- Establishing 20 as the threshold above which collinearity is a concern, we conclude the following:
  - In scenarios S.1, S.2, S.3 and S.9, collinearity is not mitigated for the considered values of k (the CN is always higher than 30).
  - In scenario S.5, the CN is lower than 20 for $k > 0.2$.
  - In scenario S.4, the CN is lower than 20 for $k \geq 0.08$.
  - In scenarios S.6, S.7, S.8 and S.10, the CN is lower than 20 for $k \geq 0.01$.

  If we compare these conclusions with those obtained from the VIF in Table 2, and because the collinearity is mitigated for values of the $VIF < 10$ (which is the case for values of $k \geq 0.07$), we conclude that the results are consistent only in scenario S.4.

From these considerations, we recommend calculating the CN in RE based on scenario S.4, that is, from typified data (taking the mean values of the original data and dividing them by the standard deviation). Thus, the CN in OLS should also be calculated is the same way. In this case, the recommendation of Belsley et al (1980) should not be followed: having unit length is similar to transforming the cross-products matrix $\mathbf{X}'\mathbf{X}$ into a correlation matrix, except that the mean zero property is not needed (scenario S.5). Note that in this case, the results will be consistent with those obtained from the VIF with the threshold VIF< 4.

Table 6 reports the difference between the values obtained for the CN in extensions 1 and 2 in each proposed scenario. Note the following:

- The two extensions nearly coincide in scenarios S.1, S.7, S.9 and S.10 (there is a difference of less than 10^{-7}).
- Both extensions present values that are very similar in scenarios S.2 and S.4 (there is a difference of less than 10^{-3}) and in S.6 and S.8 (there is a difference of less than 10^{-2}).
- The two extensions present different values for scenarios S.3 and S.5.

To confirm these findings, we offer a simulation analysis for the following matrix in scenarios S.1 to S.10:

$$\mathbf{X} = [\mathbf{Z}_1,\ \mathbf{Z}_1 + \gamma \cdot \mathbf{W}], \quad \mathbf{X} = [\mathbf{Z}_1,\ \mathbf{Z}_2,\ \mathbf{Z}_1 + \mathbf{Z}_2 + \gamma \cdot \mathbf{W}],$$
$$\mathbf{X} = [\mathbf{Z}_1,\ \mathbf{Z}_2,\ \mathbf{Z}_3,\ \mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_3 + \gamma \cdot \mathbf{W}],$$

where $\mathbf{Z}_i$, $i = 1, 2, 3$, and $\mathbf{W}$ have been independently generated from a normally typified matrix; that is, we consider matrices with 2, 3 and 4 independent variables with different degrees of linear relationships as a function of the parameter γ. We have considered the following:

Table 4: Values of CN in RE from extension 1 and data in scenarios S.1-S.10

k	$K_U(\mathbf{X}^R, k)$	$K_C(\mathbf{X}^R, k)$	$K_{UT}(\mathbf{X}^R, k)$	$K_T(\mathbf{X}^R, k)$	$K_{US}(\mathbf{X}^R, k)$
0	5942.57595218	5609.76501621	75.16946674	33.19428616	75.16946674
0.1	5942.57580567	5609.76288895	64.05068737	18.25046713	28.34847761
0.2	5942.57565916	5609.76076170	56.74835869	14.01895889	20.80880563
0.3	5942.57551265	5609.75863446	51.48335761	11.81344632	17.22026098
0.4	5942.57536614	5609.75650721	47.45615300	10.40645118	15.01966267
0.5	5942.57521964	5609.75437997	44.24716843	9.40966605	13.49489584
0.6	5942.57507313	5609.75225273	41.61226706	8.65629049	12.35870496
0.7	5942.57492662	5609.75012549	39.39846174	8.06124049	11.47003716
0.8	5942.57478011	5609.74799826	37.50436341	7.57597326	10.75046568
0.9	5942.57463360	56097.4587102	35.85976659	7.17050835	10.15244992
1	5942.57448710	5609.74374379	34.41431134	6.82519811	9.64527315

k	$K_S(\mathbf{X}^R, k)$	$K_{UL}(\mathbf{X}^R, k)$	$K_{CUL}(\mathbf{X}^R, k)$	$K_{Uit}(\mathbf{X}^R, k)$	$K_{ULit}(\mathbf{X}^R, k)$
0	33.19428616	60.09226704	33.19428616	87249657.88750964	233.88678067
0.1	5.47164817	5.53831084	5.47164817	25585350.55549748	6.37231440
0.2	3.95754764	3.98718809	3.95754764	18493513.77817653	4.56183814
0.3	3.28913038	3.30839233	3.28913038	15214243.27781145	3.76941719
0.4	2.89486438	2.90933595	2.89486438	13226288.51196591	3.30257135
0.5	2.62909088	2.64080648	2.62909088	11857231.28391437	2.98761930
0.6	2.43540130	2.44531983	2.43540130	10840819.35944916	2.75772559
0.7	2.28679508	2.29544341	2.28679508	10047734.50294692	2.58100163
0.8	2.16853196	2.17623071	2.16853196	9406595.12800260	2.44006913
0.9	2.07180330	2.07876240	2.07180330	8874355.41049992	2.32455247
1	1.99098430	1.99734895	1.99098430	8423311.00126480	2.22782891

k	$K_U(\mathbf{X}^R, k)$	$K_C(\mathbf{X}^R, k)$	$K_{UT}(\mathbf{X}^R, k)$	$K_T(\mathbf{X}^R, k)$	$K_{US}(\mathbf{X}^R, k)$
0	5942.57595218	5609.76501621	75.16946674	33.19428616	75.16946674
0.01	5942.57593753	5609.76480348	73.79014391	29.91512663	59.35826733
0.02	5942.57592288	5609.76459076	72.48408367	27.44927546	50.59592564
0.03	5942.57590822	5609.76437803	71.24502226	25.50805913	44.83688777
0.04	5942.57589357	5609.76416531	70.06742088	23.92849097	40.68276961
0.05	5942.57587892	5609.76395258	68.94636128	22.61066650	37.50436341
0.06	5942.57586427	5609.76373986	67.87745917	21.48948698	34.97121496
0.07	5942.57584962	5609.76352713	66.85679187	20.52047573	32.89093934
0.08	5942.57583497	5609.76331440	65.88083770	19.67207403	31.14293085
0.09	5942.57582032	5609.76310168	64.94642473	18.92118581	29.64723577
0.1	5942.57580567	5609.76288895	64.05068737	18.25046713	28.34847761

k	$K_S(\mathbf{X}^R, k)$	$K_{UL}(\mathbf{X}^R, k)$	$K_{CUL}(\mathbf{X}^R, k)$	$K_{Uit}(\mathbf{X}^R, k)$	$K_{ULit}(\mathbf{X}^R, k)$
0	33.19428616	60.09226704	33.19428616	87249657.88750964	233.88678067
0.01	15.33044156	16.65203518	15.33044156	60746880.67926855	19.86201719
0.02	11.48597416	12.02671221	11.48597416	49349444.05888925	14.08758600
0.03	9.58583554	9.90193716	9.58583554	42630764.91315606	11.52382919
0.04	8.40254741	8.61851209	8.40254741	38072980.07538167	9.99542679
0.05	7.57597326	7.73705117	7.57597326	34721075.27531780	8.95295873
0.06	6.95707226	7.08415797	6.95707226	32122619.62489100	8.18405792
0.07	6.47153515	6.57579191	6.47153515	30031958.63986814	7.58702885
0.08	6.07762206	6.16563442	6.07762206	28302685.60632128	7.10626687
0.09	5.74984578	5.82578676	5.74984578	26841402.82737018	6.70846934
0.1	5.47164817	5.53831084	5.47164817	25585350.55549748	6.37231440

Table 5: Values of CN in RE from extension 2 and data in scenarios S.1-S.10

k	$K_U(\mathbf{x}^R, k)$	$K_C(\mathbf{x}^R, k)$	$K_{UT}(\mathbf{x}^R, k)$	$K_T(\mathbf{x}^R, k)$	$K_{US}(\mathbf{x}^R, k)$
0	5942.57595218	5609.76501621	61.52013310	33.19428616	61.52013310
0.1	5942.57580567	5609.76300686	52.35231857	18.24759858	23.27996638
0.2	5942.57565916	5609.76099752	46.37328099	14.01455992	17.34775137
0.3	5942.57551265	5609.75898817	42.06955841	11.80789664	14.70218266
0.4	5942.57536614	5609.75697883	38.78179575	10.39994565	13.15178250
0.5	5942.57521964	5609.75496950	36.16503228	9.40232755	12.07006872
0.6	5942.57507313	5609.75296016	34.01882866	8.64820536	11.23871754
0.7	5942.57492662	5609.75095083	32.21768889	8.05247364	10.56553411
0.8	5942.57478011	5609.74894150	30.67849460	7.56657577	10.00266999
0.9	5942.57463360	5609.74693217	29.34371999	7.16052176	9.52151901
1	5942.57448710	5609.74492284	28.17211789	6.81465712	9.10337373

k	$K_S(\mathbf{x}^R, k)$	$K_{UL}(\mathbf{x}^R, k)$	$K_{CUL}(\mathbf{x}^R, k)$	$K_{Uit}(\mathbf{x}^R, k)$	$K_{ULit}(\mathbf{x}^R, k)$
0	33.19428616	60.09226704	33.19428616	87249657.88750964	233.88678067
0.1	5.45828586	5.53831084	5.45828586	25585350.55549821	6.37231440
0.2	3.93880510	3.98718809	3.93880510	18493513.77817661	4.56183814
0.3	3.26645371	3.30839233	3.26645371	15214243.27781127	3.76941719
0.4	2.86901519	2.90933595	2.86901519	13226288.51196609	3.30257135
0.5	2.60056281	2.64080648	2.60056281	11857231.28391432	2.98761930
0.6	2.40454854	2.44531983	2.40454854	10840819.35944930	2.75772559
0.7	2.25388778	2.29544341	2.25388778	10047734.50294686	2.58100163
0.8	2.13378475	2.17623071	2.13378475	9406595.12800262	2.44006913
0.9	2.03539171	2.07876240	2.03539171	8874355.41050002	2.32455247
1	1.95305509	1.99734895	1.95305509	8423311.00126478	2.22782891

k	$K_U(\mathbf{x}^R, k)$	$K_C(\mathbf{x}^R, k)$	$K_{UT}(\mathbf{x}^R, k)$	$K_T(\mathbf{x}^R, k)$	$K_{US}(\mathbf{x}^R, k)$
0	5942.57595218	56097.6501621	61.52013310	33.19428616	61.52013310
0.01	5942.57593753	56097.6481527	60.35997779	29.91465600	48.50884067
0.02	5942.57592287	56097.6461434	59.28058218	27.44841176	41.34474191
0.03	5942.57590822	56097.6441340	58.25956994	25.50685524	36.64569279
0.04	5942.57589357	56097.6421247	57.29071241	23.92698531	33.26234083
0.05	5942.57587892	56097.6401153	56.36931403	22.60888827	30.67849460
0.06	5942.57586427	56097.6381060	55.49143704	21.48745917	28.62332213
0.07	5942.57584962	56097.6360966	54.65366073	20.51821694	26.93926130
0.08	5942.57583497	56097.6340873	53.85296454	19.66959965	25.52759643
0.09	5942.57582032	56097.6320780	53.08665579	18.91850880	24.32292279
0.1	5942.57580567	56097.6300686	52.35231857	18.24759858	23.27996638

k	$K_S(\mathbf{x}^R, k)$	$K_{UL}(\mathbf{x}^R, k)$	$K_{CUL}(\mathbf{x}^R, k)$	$K_{Uit}(\mathbf{x}^R, k)$	$K_{ULit}(\mathbf{x}^R, k)$
0	33.19428616	60.09226704	33.19428616	87249657.88750964	233.88678067
0.01	15.32659029	16.65203518	15.32659029	60746880.67926394	19.86201719
0.02	11.48022083	12.02671221	11.48022083	49349444.05888682	14.08758600
0.03	9.57865585	9.90193716	9.57865585	42630764.91315262	11.52382919
0.04	8.39418270	8.61851209	8.39418270	38072980.07537316	9.99542679
0.05	7.56657577	7.73705117	7.56657577	34721075.27531861	8.95295873
0.06	6.94674923	7.08415797	6.94674923	32122619.62489043	8.18405792
0.07	6.46036747	6.57579191	6.46036747	30031958.63986744	7.58702885
0.08	6.06567357	6.16563442	6.06567357	28302685.60631992	7.10626687
0.09	5.73716859	5.82578676	5.73716859	26841402.82737157	6.70846934
0.1	5.45828586	5.53831084	5.45828586	25585350.55549821	6.37231440

Table 6: Difference in the CN between extensions 1 and 2 in scenarios S.1-S.10

k	$K_U(\mathbf{X}^R,k) - K_U(\mathbf{x}^R,k)$	$K_C(\mathbf{X}^R,k) - K_C(\mathbf{x}^R,k)$	$K_{UT}(\mathbf{X}^R,k) - K_{UT}(\mathbf{x}^R,k)$	$K_T(\mathbf{X}^R,k) - K_T(\mathbf{x}^R,k)$	$K_{US}(\mathbf{X}^R,k) - K_{US}(\mathbf{x}^R,k)$
0	0	0.000000000698492	13.649333642410191	0.000000000000021	13.649333642401885
0.1	0.000000000345608	-0.000117905260595	11.698368798444051	0.002868546422597	5.068511226762542
0.2	0.000000002045454	-0.000235810801314	10.375077693109887	0.004398973136713	3.461054267910853
0.3	0.000000000415639	-0.000353716176505	9.413799208495242	0.005549687080659	2.518078317581106
0.4	0.000000000805812	-0.000471621212455	8.674357249916184	0.006505528040746	1.867880169418458
0.5	0	-0.000589526155636	8.082136150093902	0.007338500037211	1.424827111790952
0.6	0.000000000344698	-0.000707430749571	7.593438399573735	0.008085127748556	1.119987426864531
0.7	0.000000000914952	-0.000825335258924	7.180772855413402	0.008766852448858	0.904503048668674
0.8	-0.000000000070031	-0.000943239037952	6.825868804236002	0.009397498975352	0.747795685048988
0.9	0.000000000804903	-0.001061143238985	6.516046599397331	0.009986585747389	0.630930907751742
1	-0.000000000780346	-0.001179046411380	6.242193442116371	0.010540992172928	0.541899423723818

k	$K_S(\mathbf{X}^R,k) - K_S(\mathbf{x}^R,k)$	$K_{UL}(\mathbf{X}^R,k) - K_{UL}(\mathbf{x}^R,k)$	$K_{CUL}(\mathbf{X}^R,k) - K_{CUL}(\mathbf{x}^R,k)$	$K_{Uit}(\mathbf{X}^R,k) - K_{Uit}(\mathbf{x}^R,k)$	$K_{ULit}(\mathbf{X}^R,k) - K_{ULit}(\mathbf{x}^R,k)$
0	-0.000000000000433	0.000000000017003	0.000000000002665	0	-0.000000000335774
0.1	0.013362309679123	-0.000000000000002	0.013362309679123	0.000002995133400	-0.000000000000008
0.2	0.018742537050687	0.000000000000007	0.018742537050690	-0.000000048428774	-0.000000000000004
0.3	0.022676666212001	0	0.022676666212001	0.000000383704901	-0.000000000000002
0.4	0.025849180542294	0.000000000000002	0.025849180542295	-0.000000178813934	-0.000000000000000
0.5	0.028528072641608	0.000000000000002	0.028528072641608	0.000000085681677	0.000000000000000
0.6	0.030852762408646	0.000000000000002	0.030852762408647	0.000000141561031	0.000000000000000
0.7	0.032907296629715	0.000000000000001	0.032907296629716	-0.000000054016709	0.000000000000000
0.8	0.034747211454716	0.000000000000001	0.034747211454717	0.000000035390258	-0.000000000000001
0.9	0.036411586310429	0.000000000000000	0.036411586310429	-0.000000039115548	-0.000000000000000
1	0.037929209538493	-0.000000000000001	0.037929209538493	0.000000031664968	0.000000000000000

– $\gamma \in [0.1,\ 10]$ with intervals of 0.01 considering 991 values. Note that the collinearity decreases as γ increases.
– The sizes of the generated sample are 20, 50, 100, 150 and 200.

Figure 3 displays the Mean Absolute Error (MAE) of the differences in each scenario between extensions 1 and 2 in the 991 simulations, that is,

$$MAE = \frac{1}{991} \sum_{r=1}^{991} \sum_{k} \left| K_l \left(\mathbf{X}^R, k\right)_r - K_l \left(\mathbf{x}^R, k\right)_r \right|,$$

where $l = U, UT, US, C, T, S, UL, CUL, Uit, ULit$. The results are depicted from left to right for the 2, 3 and 4 variables. Every trend line corresponds to one of the considered sample sizes: 20, blue circles; 50, green diamonds; 100, red pluses; 150, cyan stars; and 200, magenta squares. Note the following:

– The MAE decreases as the sample size increases.
– In all cases, the highest errors correspond to scenarios S.2, S.6 and S.8.
– The highest MAE is lower than 0.4.

Thus, it has been shown that the two extensions present similar results. Therefore, we recommend obtaining the CN from the first extension, which is easier to calculate. Moreover, since the first extension has a closed expression, it was possible to show that the three desirable properties established in García et al (2015a) are verified. That is, the CN in OLS and RE should be calculated from $K_T(\mathbf{X}^R, k)$.

6 Conclusions

Ridge estimation is applied to estimate a linear model in which collinearity has been detected. Then, it is necessary to determine whether the collinearity has been mitigated after the application of this estimation approach. This fact motivates extending collinearity diagnostic indicators to ridge estimation (see Figure 1). In this paper, we have presented an extension of the condition number to ridge estimation under two approaches.

First, we extended it by requiring the presence of matrix $\mathbf{X}$ in different scenarios. This extension verifies the properties of being continuous in $k = 0$, monotonous decreasing in k and higher than 1. However, in this case, matrix $(\mathbf{X}^R)'\mathbf{X}^R$ does not, for example, have unit length or is not typified. For this reason, we present a second extension by requiring matrix $(\mathbf{X}^R)'\mathbf{X}^R$ to be present in various scenarios. In this second case, it is not possible to obtain an algebraic closed-form expression for the CN, and consequently, is not possible to analyze whether the desirable properties are verified.

Given the results, we recommend calculating the CN from typified data, as in this case, the conclusions are similar to those obtained with the VIF. **Further, although we are aware that, given the definition of the condition number, the second extension is the natural one, as both extensions lead to similar results, with very small differences, we consider it appropriate to use the first one because it is easier to apply.** Thanks to the closed-form expression of the CN, it was possible to show that it verifies the properties established in García et al (2015a).

Finally, since the calculation of the CN in OLS leads to different results depending on the data transformations, we also recommend calculating the CN in OLS using typified data to be consistent with the option recommended in RE. Note that typified data coincide with unit length centered data, in contrast to the statement of Belsley et al (1980).

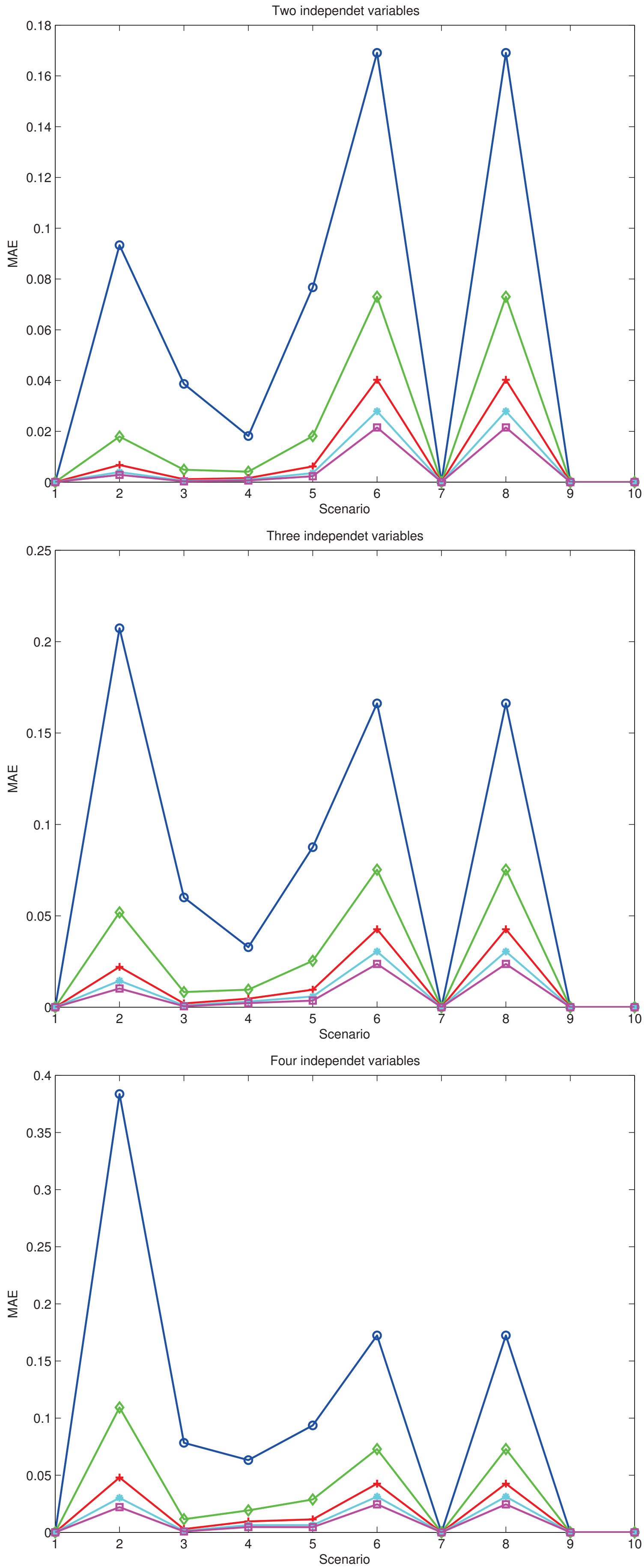

Fig. 3: MAE in scenarios S.1-S.10 for 2, 3 and 4 variables

References

Alkhamisi M, MacNeill I (2015) Recent results in ridge regression methods. METRON 73(3):359–376

Belsley DA (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 DA (1984) Demeaning conditioning diagnostics through centering. The American Statistician 38(2):73–77

Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons, New York

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

Casella G (1980a) Condition numbers and minimax ridge-regression estimators. Journal of American Statistical Association 80:753–758

Casella G (1980b) Minima ridge regression estimation. Ann Statist 8:1036–1056

Draper NR, Smith H (1998) Applied Regression Analysis. Wiley, NY

Economic reports of the President (2011) url:http://www.gpo.gov/fdsys/browse

Farrar DE, Glauber RR (1967) Multicollinearity in regression analysis: the problem revisited. The Review of Economic and Statistics pp 92–107

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

Friendly M, Kwan E (2009) Where's waldo? visualizing collinearity diagnostics. The American Statistician 63(1)

García J, Salmerón R, García C, López MDM (2015a) Standardization of variables and diagnostic of collinearity in the ridge regression. International Statistical Review DOI doi10.1111/insr.12099

García J, Salmerón R, López MDM, García C (2015b) Revisiting the condition number and red indicator in ridge regression. IProceedings of the International Work Conference on Time Series pp 271–22

Greene WH (1993) Econometric analysis, 2nd edn. Macmillan, New York

Gunst RF, Mason RL (1977) Biased estimation in regression: An evaluation using mean squared error. Journal of the American Statistical Association 72(359):616–628

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

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

Hoerl AE, Kennard RW, Baldwin KF (1975) Ridge regression: Some simulations. Communications in Statistics 4(2):105–123

Kovacs P, Petres T, Toth L (2005) A new measure of multicollinearity in linear regression models. International statistical review 73(3):405–412

Lazaridis A (2007) A note regarding the condition number: the case of spurious and latent multicollinearity. Quality and Quantity 41:123–135

Macedo P (2015) Ridge regression and generalized maximum entropy: An improved version of the ridge-gme parameter estimator. Communications in Statistics-Simulation and Computation (just-accepted)

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

Marquardt DW (1980) A critique of some ridge regression methods: Comment. Journal of the American Statistical Association 75(369):87–91

Marquardt DW, Snee SR (1975) Ridge regression in practice. The American Statistician 29(1):3–20
Marshall A, Olkin I (1965) Norms and inequalities for condition numbers. Pacific Journal of Mathematics 15(1):241–247
Marshall AW, Olkin I (1969) Norms and inequalities for condition numbers, ii. Linear Algebra and its Applications 2(2):167–172
McDonald GC (2010) Tracing ridge regression coefficients. Wiley Interdisciplinary Reviews: Computational Statistics 2:695–703
Rawlings JO, Pantula SG, Dickey DA (1998) Applied regression analysis: a research tool. Springer Science & Business Media
Riley JD (1955) Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix. Mathematical Tables and Other Aids to Computation pp 96–101
Salmerón R, García J, López MDM, García C (2016) Collinearity diagnostic applied in ridge estimation through the vif. Journal of Applied Statistic p In press
Silvey S (1969) Multicollinearity and imprecise estimation. Journal of the Royal Statistical Society Series B (Methodological) pp 539–552
Stewart GW (1987) Collinearity and least squares regression. Statistical Science 2(1):68–84
Theil H (1971) Principles of econometrics. Wiley, New York
Wichers CR (1975) The detection of multicollinearity: a comment. The Review of Economics and Statistics pp 366–368
Willan AR, Watts DG (1978) Meaningful multicollinearity measures. Technometrics 20(4):407–412
Wissel J (2009) A new biased estimator for multivariate regression models with highly collinear variables

A Equivalence for transformations in OLS

Taking into account the notation given in Section 4, we now show that the CN in OLS coincides in scenarios S.3 and S.5, and in S.4 and S.6.
Since

$$\widetilde{\mathbf{X}}'_{UT}\widetilde{\mathbf{X}}_{UT} = \boldsymbol{\Omega}\mathbf{X}'\mathbf{X}\boldsymbol{\Omega}, \qquad \widetilde{\mathbf{X}}'_{US}\widetilde{\mathbf{X}}_{US} = \frac{1}{n}\boldsymbol{\Omega}\mathbf{X}'\mathbf{X}\boldsymbol{\Omega},$$

it is verified that if δ is an eigenvalue of $\widetilde{\mathbf{X}}'_{UT}\widetilde{\mathbf{X}}_{UT}$, then $\alpha = \frac{1}{n}\delta$ is an eigenvalue of $\widetilde{\mathbf{X}}'_{US}\widetilde{\mathbf{X}}_{US}$. In this case,

$$K\left(\widetilde{\mathbf{X}}_{US}\right) = \sqrt{\frac{\alpha_{max}}{\alpha_{min}}} = \sqrt{\frac{\frac{1}{n}\delta_{max}}{\frac{1}{n}\delta_{min}}} = \sqrt{\frac{\delta_{max}}{\delta_{min}}} = K\left(\widetilde{\mathbf{X}}_{UT}\right).$$

However, as

$$\widetilde{\mathbf{X}}'_{T}\widetilde{\mathbf{X}}_{T} = \boldsymbol{\Omega}\left(\mathbf{X}'\mathbf{X} - n\overline{\mathbf{X}}'\overline{\mathbf{X}}\right)\boldsymbol{\Omega}, \qquad \widetilde{\mathbf{X}}'_{S}\widetilde{\mathbf{X}}_{S} = \frac{1}{n}\boldsymbol{\Omega}\left(\mathbf{X}'\mathbf{X} - n\overline{\mathbf{X}}'\overline{\mathbf{X}}\right)\boldsymbol{\Omega},$$

it is verified that if δ is an eigenvalue of $\widetilde{\mathbf{X}}'_{T}\widetilde{\mathbf{X}}_{T}$, then $\alpha = \frac{1}{n}\delta$ is an eigenvalue of $\widetilde{\mathbf{X}}'_{S}\widetilde{\mathbf{X}}_{S}$. In this case,

$$K\left(\widetilde{\mathbf{X}}_{S}\right) = \sqrt{\frac{\alpha_{max}}{\alpha_{min}}} = \sqrt{\frac{\frac{1}{n}\delta_{max}}{\frac{1}{n}\delta_{min}}} = \sqrt{\frac{\delta_{max}}{\delta_{min}}} = K\left(\widetilde{\mathbf{X}}_{T}\right).$$

B Properties of the CN in RE: extension 1

The Condition Number (CN) was previously applied by Riley (1955) to show that the matrix $A + kI$ is better conditioned than the matrix A and to show that the CN of the matrix $A+kI$ is always lower than that of matrix A

for $k > 0$. Casella (1980a) developed some properties of the CN associated with ridge estimation. In this appendix, it is shown that $K(\mathbf{X}^R, k)$ is decreasing in k, continuous for $k = 0$ and always equal to or higher than one.

Decreasing in k: As $f(k) = \frac{\lambda_{max}+k}{\lambda_{min}+k}$, it is obtained that

$$\frac{\partial}{\partial k} f(k) = \frac{\lambda_{min} + k - \lambda_{max} - k}{(\lambda_{min} + k)^2} = \frac{\lambda_{min} - \lambda_{max}}{(\lambda_{min} + k)^2} \leq 0,$$

since $0 \leq \lambda_{min} \leq \lambda_{max}$ (because $\mathbf{X}'\mathbf{X}$ is positive defined, its eigenvalues are real positive numbers).
In this case, since $K(\mathbf{X}^R, k) = \sqrt{f(x)}$ and the positive square root is a strictly monotone function, it is verified that $K(\mathbf{X}^R, k)$ is decreasing in k.

Continuous for $k = 0$: It is evident that

$$K(\mathbf{X}^R, 0) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = K(\mathbf{X}),$$

as $K(\mathbf{X})$ is the condition number associated with model (1).

Always equal to or higher than 1: It is evident that

$$\lim_{k \to +\infty} K(\mathbf{X}^R, k) = \lim_{k \to +\infty} \sqrt{\frac{k}{k} \cdot \frac{\lambda_{max} + k}{\lambda_{min} + k}} = \lim_{k \to +\infty} \sqrt{\frac{\frac{\lambda_{max}}{k} + 1}{\frac{\lambda_{min}}{k} + 1}} = 1.$$

Then, since the first value is given by $K(\mathbf{X}) \geq 1$, is decreasing in k and its limit for $k \to +\infty$ is 1, it must be the case that it is always equal to or higher than one.

C Mean and variance of the independent variables of the augmented model

In this appendix, we obtain the mean and variance of the variables contained in matrix $\mathbf{X}^R$ given the model (4) and under the following assumptions:

- Matrix $\mathbf{X}$ has zero mean and variance equal to one (typified).
- Matriz $\mathbf{X}$ has zero mean and variance equal to one divided by the number of observations (standardized).

If the original matrix $\mathbf{X}$ does not satisfy these conditions, it has to be transformed in the following way:

- For each column of matrix $\mathbf{X}$, we subtract its mean and divide by its standard deviation. Thus, we have that the columns of $\mathbf{X}^R$ are given by

$$\mathbf{X}^R_{T,j} = \left(\frac{X_{1j} - \overline{\mathbf{X}}_j}{\sqrt{Var(\mathbf{X}_j)}} \ \cdots \ \frac{X_{nj} - \overline{\mathbf{X}}_j}{\sqrt{Var(\mathbf{X}_j)}} \ 0 \ \ldots \sqrt{k} \ \ldots 0 \right), \quad j = 1, \ldots, m.$$

In this case:

$$\begin{aligned}
\overline{\mathbf{X}}^R_{T,j} &= \frac{1}{n+m} \left(\sum_{i=1}^{n} \frac{X_{ij} - \overline{\mathbf{X}}_j}{\sqrt{Var(\mathbf{X}_j)}} + \sqrt{k} \right) = \frac{\sqrt{k}}{n+m}, \\
Var(\mathbf{X}^R_{T,j}) &= \frac{1}{n+m} \left(\sum_{i=1}^{n} \frac{(X_{ij} - \overline{\mathbf{X}}_j)^2}{Var(\mathbf{X}_j)} + k \right) - \left(\overline{\mathbf{X}}^R_{T,j} \right)^2 \\
&= \frac{1}{n+m} \left(\frac{n \cdot Var(\mathbf{X}_j)}{Var(\mathbf{X}_j)} + k \right) - \frac{k}{(n+m)^2} = \frac{n(n+m) + (n+m-1)k}{(n+m)^2}.
\end{aligned}$$

- For each column of $\mathbf{X}$, we subtract its mean and divide by its standard deviation multiplied by the square root of the number of observations. Thus, the columns of $\mathbf{X}^R$ are given by

$$\mathbf{X}^R_{S,j} = \left(\frac{X_{1j} - \overline{\mathbf{X}}_j}{\sqrt{n \cdot Var(\mathbf{X}_j)}} \ \cdots \ \frac{X_{nj} - \overline{\mathbf{X}}_j}{\sqrt{n \cdot Var(\mathbf{X}_j)}} \ 0 \ \ldots \sqrt{k} \ \ldots 0 \right), \quad j = 1, \ldots, m.$$

In this case,

$$\begin{aligned}
\overline{\mathbf{X}}^R_{S,j} &= \frac{1}{n+m} \left(\sum_{i=1}^{n} \frac{X_{ij} - \overline{\mathbf{X}}_j}{\sqrt{n \cdot Var(\mathbf{X}_j)}} + \sqrt{k} \right) = \frac{\sqrt{k}}{n+m}, \\
Var(\mathbf{X}^R_{S,j}) &= \frac{1}{n+m} \left(\sum_{i=1}^{n} \frac{(X_{ij} - \overline{\mathbf{X}}_j)^2}{n \cdot Var(\mathbf{X}_j)} + k \right) - \left(\overline{\mathbf{X}}^R_{S,j} \right)^2 \\
&= \frac{1}{n+m} \left(\frac{n \cdot Var(\mathbf{X}_j)}{n \cdot Var(\mathbf{X}_j)} + k \right) - \frac{k}{(n+m)^2} = \frac{(n+m) + (n+m-1)k}{(n+m)^2}.
\end{aligned}$$

D Order of operations to obtain unit length

Given a vector $\mathbf{X}$ of dimension $n \times 1$, the transformed vector $\mathbf{Z}$ given by

$$Z_i = \frac{X_i - \overline{\mathbf{X}}}{||\mathbf{X}||} = \frac{X_i - \overline{\mathbf{X}}}{\sqrt{\sum\limits_{i=1}^{n} X_i^2}}, \quad i = 1, \dots, n.$$

In this case,

$$\overline{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{1}{n} \frac{1}{\sqrt{\sum\limits_{i=1}^{n} X_i^2}} \sum_{i=1}^{n} \left(X_i - \overline{\mathbf{X}}\right) = 0,$$

$$Var(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - \overline{\mathbf{Z}}^2 = \frac{1}{n} \frac{1}{\sum\limits_{i=1}^{n} X_i^2} \sum_{i=1}^{n} \left(X_i - \overline{\mathbf{X}}\right)^2 = \frac{Var(\mathbf{X})}{\sum\limits_{i=1}^{n} X_i^2},$$

$$||\mathbf{Z}|| = \sqrt{\sum_{i=1}^{n} Z_i^2} = \sqrt{n \cdot Var(\mathbf{Z})} \neq 1. \tag{13}$$

That is, although the transformation leads to a centered vector (zero mean), such vectors do not have unit length. However, we then consider the transformed vector $\mathbf{Z}$ given by

$$Z_i = \frac{Y_i}{||\mathbf{Y}||} = \frac{Y_i}{\sqrt{\sum\limits_{i=1}^{n} Y_i^2}}, \quad Y_i = X_i - \overline{\mathbf{X}}, \quad i = 1, \dots, n.$$

In this case,

$$\overline{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{1}{n} \frac{1}{\sqrt{\sum\limits_{i=1}^{n} Y_i^2}} \sum_{i=1}^{n} \left(X_i - \overline{\mathbf{X}}\right) = 0,$$

$$Var(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - \overline{\mathbf{Z}}^2 = \frac{1}{n} \frac{1}{\sum\limits_{i=1}^{n} Y_i^2} \sum_{i=1}^{n} \left(X_i - \overline{\mathbf{X}}\right)^2 = \frac{1}{n},$$

$$||\mathbf{Z}|| = \sqrt{\sum_{i=1}^{n} Z_i^2} = \sqrt{n \cdot \frac{1}{n}} = 1. \tag{14}$$

That is, in this case, the transformation leads to a centered vector with variance $\frac{1}{n}$ (standardized data) and unit length.