Computational Economics

The multiColl package versus other existing packages in R to detect multicollinearity

--Manuscript Draft--

Manuscript Number:	CSEM-D-21-00259R1
Full Title:	The multiColl package versus other existing packages in R to detect multicollinearity
Article Type:	Software review
Keywords:	Multicollinearity; Detection; Intercept; Dummy; Software; R package
Corresponding Author:	catalina garcia Universidad de Granada Granada, Andalucia SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad de Granada
Corresponding Author's Secondary Institution:	
First Author:	Román Salmerón-Gómez
First Author Secondary Information:	
Order of Authors:	Román Salmerón-Gómez
	catalina garcia
	JOSE GARCÍA-PÉREZ
Order of Authors Secondary Information:	
Funding Information:	universidad de granada (PP2019-EI-02) — dr Román Salmerón-Gómez
Abstract:	The main goal of this paper is to show the advantages of the multiColl package in R, comparing its results with other existing packages in R for the treatment of multicollinearity.
Response to Reviewers:	The paper now fits the format of a software review (as part of R) with an extension of 10 pages. Thank you for giving us the opportunity to revise and resubmit this manuscript.

Noname manuscript No.
(will be inserted by the editor)

The *multiColl* package versus other existing packages in R to detect multicollinearity

Román Salmerón · Catalina B. García García · José García Pérez

Received: date / Accepted: date

Abstract The main goal of this paper is to show the advantages of the **multiColl** package in **R**, comparing its results with other existing packages in **R** for the treatment of multicollinearity.

Keywords Multicollinearity · Detection · Intercept · Dummy · Software · R

1 Introduction

As is known, linear regression is a widely applied statistical tool for analyzing the effect (if any) of a set of independent variables on a dependent variable. A model with n observations and k independent variables is specified as follows:

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times k} \cdot \boldsymbol{\beta}_{k\times 1} + \mathbf{u}_{n\times 1}, \quad (1)$$

where the first column of $\mathbf{X}$ is composed of ones representing the intercept (denoted by $\mathbf{1}_{n\times 1} = (1,\dots,1)^t$) and $\mathbf{u}$ represents the random disturbance assumed to be centered and spherical. The aim is to estimate the coefficients $\boldsymbol{\beta}$ of the independent variables and, from these values, establish the direction of relations (according to the signs) and quantify relations (with values).

The ordinary least squares (OLS) approach is the most frequently applied methodology for obtaining the estimates of coefficients using the well-known expression $\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y}$. Thus, the inverse of matrix $\mathbf{X}^t\mathbf{X}$ is required, and, consequently, it is assumed that the variables in matrix $\mathbf{X}$ are independent. If

Salmerón, R.
Quantitative Methods for the Economy and Business, University of Granada
E-mail: romansg@ugr.es

García, C.
Quantitative Methods for the Economy and Business, University of Granada
E-mail: cbgarcia@ugr.es

García-Pérez, J.
Department of Economics and Business, University of Almería
E-mail: jgarcia@ual.es

this condition does not hold, it is said that the model is characterized by near-multicollinearity.

When near-multicollinearity exists the conclusions obtained from the analysis will be questioned; such concerns justify development of an appropriate method to detect and treat collinearity. In relation to the detection, it is important the distinction presented by Marquardt and Snee (9) between non-essential multicollinearity (relationship between the intercept and the rest of the independent variables) and essential multicollinearity (relationship between the independent variables apart from the intercept) since not all measures are useful for detecting both types of approximate multicollinearity.

This paper presents a detailed analysis for a better understanding of all measures included in *multiColLM* function presented by Salmerón et al. (14) and an adequate application of these measures (for example, what measures are able to diagnose essential and non-essential multicollinearity). At the same time, these measures are compared to similar measures existing in other packages in **R**. The structure of the paper is as follows: Section 2 presents how to use the **multiColl** (16) package to calculate the measures most commonly applied in the scientific literature to detect the presence of multicollinearity by paying special consideration to qualitative explanatory variables which are usually ignored in the existing software. The different functions are compared with the functions included on the **car** (2), **mctest** (19), **mcvis** (8) and **rms** (5) packages in **R** which are also applied to obtain similar measures. Section 3 presents the particular case of the simple linear regression where non-essential multicollinearity can exist although it is ignored in others econometric software as is shown in this paper. Finally, Section 4 summarizes the main contributions of this paper.

An extended version of this paper can be consulted in Salmerón et al. (12), (15).

2 Multicollinearity detection

This section illustrates the application of the functions included in **multiColl** (16) package and similar functions existing in other packages in **R** by the following two examples: a) Henri Theil's textile consumption data (18) for the period from 1923 to 1939. We have included a dummy variable to distinguish between the 20s and the 30s and b) Klein and Goldberger (7) for consumption and wage incomes in the United States from 1936 to 1952.

2.1 **multiColl package**

Correlation matrix and its determinant: *RdetR* function. The correlation coefficient measures the linear relation between two quantitative variables, and for this reason, it should not be considered if one of the variables is non-quantitative (for example, a dummy variable) or is constant. Thus, this measure is appropriate for detection of near essential multicollinearity if there is a relation between two quantitative variables.

The following results are obtained for the considered examples:

```
> RdetR(theil.X, T, pos = 4)
$`Correlation matrix`
           income    relprice
income    1.0000000  0.1788467
relprice  0.1788467  1.0000000
$`Correlation matrix's determinant`
[1] 0.9680139
> RdetR(KG.X)
$`Correlation matrix`
                wage.income non.farm.income farm.income
wage.income       1.0000000   0.9431118       0.8106989
non.farm.income   0.9431118   1.0000000       0.7371272
farm.income       0.8106989   0.7371272       1.0000000
$`Correlation matrix's determinant`
[1] 0.03713592
```

In the first example, the presence of problematic near-multicollinearity is not detected with any measure. However, in the second example, although none of the correlation coefficients is higher than 0.9486833 (see García et al. (4)), the determinant is lower than $0.1013+0.00008626\cdot 14-0.01384\cdot 3 = 0.06098764$. Thus, in this case, a problematic essential near-multicollinearity has been detected.

Variance Inflation Factors: *VIF* function. García et al. (3), among others, shows that the main diagonal of the inverse of the correlation matrix is composed of the so-called variance inflation factors (VIFs). Thus, the VIFs are not appropriate for non-quantitative variables. Consequently, if a model includes a non-quantitative variable, the model is required to indicate such a variable's position in the design matrix $\mathbf{X}$. The first column of $\mathbf{X}$ corresponding to the intercept is also eliminated, and the model needs to include at least two quantitative variables.

The following results are obtained for the considered examples:

```
> VIF(theil.X, T, pos = 4)
  income  relprice
1.033043  1.033043
```

```
> VIF(KG.X)
wage.income non.farm.income farm.income
  12.296544        9.230073    2.976638
```

Considering that values of VIFs higher than 10 indicate problematic near-multicollinearity, it is possible to conclude that in the first example there is no multicollinearity, while the second example presents problematic multicollinearity.

Condition Number: *CNs* function. The condition number (CN) is obtained from the square root of the ratio between the maximum and the minimum eigenvalues of matrix $\mathbf{X}^t\mathbf{X}$ after the transformation of columns of matrix $\mathbf{X}$ to be of unit length. Values between 20 and 30 indicate a moderate near-multicollinearity, while values higher than 30 indicate problematic near-multicollinearity (see Belsley (1)). It could be interesting to calculate the CN with and without consideration of the intercept and to calculate the increase caused by a change from not accounting for the intercept to accounting for it (see Salmerón et al. (11)). The following results are obtained for the considered examples:

```
> CNs(theil.X)
$`CN without intercept`
[1] 24.15423
$`CN with intercept`
[1] 53.39671
$`Increase (in percentage)`
[1] 54.76458
```

```
> CNs(KG.X)
$`CN without intercept`
[1] 30.2987
$`CN with intercept`
[1] 35.88644
$`Increase (in percentage)`
[1] 15.57062
```

In the first case, the CN without the intercept indicates a moderate near-multicollinearity, and an increase of 54.76458% is observed when accounting for the intercept. In the second example, the CN without the intercept indicates a problematic near-multicollinearity. In addition, the increase when considering the intercept is 15.57062%, and consequently, it appears that the intercept does not have a relevant role in the existing problematic multicollinearity.

Stewart index: *ki* function. Stewart (17) presented the Stewart index, k_i^2, which measures the relation between the i-th column of $\mathbf{X}$ and the rest of the columns of $\mathbf{X}$. Although Stewart identified this index with the VIF, these measures only coincide if the variables to which the measures are applied have zero mean (see Salmerón et al. (10), (11) and (13) for more details).

Because k_i^2 depends on the VIF, it is not appropriate to calculate this measure for dummy variables. In addition, as the VIF is only able to detect essential near-multicollinearity, the ratio $VIF(i)/k_i^2$ can be interpreted as the percentage of essential near-multicollinearity existing in the i-th variable of $\mathbf{X}$. The difference from 100% will be the percentage of non-essential near-multicollinearity. Another aspect that distinguishes this measure from the VIF is that it can be calculated for $i = 1$, that is, for the intercept. In this case, it measures the non-essential multicollinearity existing in the model.

The following results are obtained for the considered examples:

```
> ki(theil.X, T, pos = 4)
$`Stewart index`
[1] 403.20963 415.28266 23.50258
$`Proportion of essential collinearity in i-th independent variable
    (without intercept)`
[1] 0.2487566 4.3954455
$`Proportion of non-essential collinearity in i-th independent variable
    (without intercept)`
[1] 99.75124 95.60455
> ki(KG.X)
$`Stewart index`
[1] 17.86327 185.96422 156.50013 39.16836
$`Proportion of essential collinearity in i-th independent variable
    (without intercept)`
[1] 6.612317 5.897805 7.599598
$`Proportion of non-essential collinearity in i-th independent variable
    (without intercept)`
[1] 93.38768 94.10219 92.40040
```

In the first example, problematic near non-essential multicollinearity is detected. In addition, the second explanatory variable seems to be closely related to the intercept due to the 99.75124% of the near-multicollinearity caused by this variable is non-essential. The second example presents also a high percentage of non-essential multicollinearity, however the value of the Stewart index associated to the intercept is not very high. Anyway, the existence of a threshold for this measure will help to obtain conclusions.

2.2 Other packages in **R**

This subsection applies the functions existing in the **car** (2), **mctest** (19), **mcvis** (8), **rms** (5) and **perturb** (6) packages in **R** to the Theil data set to analyze how these packages treat the dummy variable *twentys* and if these packages are able to detect the non-essential multicollinearity existing due to the slight variability of variable *income*.

The car and rms packages. By using the *vif* command of the **car** (2) package it is possible to obtain the VIFs without any consideration:

```
>     reg.theil = lm(consume~income+relprice+twentys)
```

```
>       vif(reg.theil)
  income relprice  twentys
1.062760 6.007181 5.866333
```

However, the VIF associated to the variable *twentys* is obtained from the coefficient of determination of a regression whose dependent variables is dichotomous. As is well known, models are not linear in this kind of fit, and, for this reason, it is not appropriate to use the coefficient of determination in this case. Identical results are obtained by using the *vif* command of the **rms** (5) package. In addition, none of these packages indicate in their help instructions that the VIF is only suitable for explanatory quantitative variables.

Note that, to the best of our knowledge, the Stewart index is not calculated in any other package in **R**. It only can be obtained manipulating the *vif* command of the **rms** (5) package. This manipulation consists in introducing the intercept as an independent variable within the matrix **X** and indicating that the model does not have an intercept with the *lm* command:

```
> reg.0 = lm(theil.y~theil.X+0)
> vif(reg.0)
      theil.X    theil.Xincome  theil.Xrelprice   theil.Xtwentys
   427.445968       427.228985       136.668316         9.972766
```

As shown by (13), after this manipulation the model is considered to be non centered and, consequently, the *vif* command will be calculating the Stewart index instead of VIF. Thus, in this situation the researcher wrongly considers that is calculating the VIF when, in fact, the Stewart index was obtained.

The mctest package. The *mctest* command from the **mctest** (19) package allows us to obtain a great number of measures to detect multicollinearity from a joint point of view:

```
> mctest(reg.theil)
Call:
omcdiag(mod = mod, Inter = TRUE, detr = detr, red = red, conf = conf,
    theil = theil, cn = cn)
Overall Multicollinearity Diagnostics
                       MC Results detection
Determinant |X'X|:         0.1650         0
Farrar Chi-Square:        25.5246         1
Red Indicator:             0.5371         1
Sum of Lambda Inverse:    12.9363         0
Theil's Method:           -0.1837         0
Condition Number:         53.3967         1
1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

and from an individual point of view:

```
> mctest(reg.theil, type="i", corr=TRUE)
Call:
imcdiag(mod = mod, method = method, corr = TRUE, vif = vif, tol = tol,
    conf = conf, cvif = cvif, ind1 = ind1, ind2 = ind2, leamer = leamer,
    all = all)
All Individual Multicollinearity Diagnostics Result
            VIF    TOL      Wi      Fi Leamer    CVIF Klein   IND1
income   1.0628 0.9409  0.4393  0.9414 0.9700 -0.0718     0 0.1344
relprice 6.0072 0.1665 35.0503 75.1077 0.4080 -0.4060     0 0.0238
twentys  5.8663 0.1705 34.0643 72.9950 0.4129 -0.3964     0 0.0244
           IND2
income   0.1029
relprice 1.4520
twentys  1.4451
1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

```
twentys , coefficient(s) are non-significant may be due to multicollinearity
R-square of y on all x: 0.9529
* use method argument to check which regressors may be the reason of collinearity
===================================
Correlation Matrix
              income   relprice    twentys
income    1.00000000 0.1788467 0.09351197
relprice  0.17884669 1.0000000 0.90809254
twentys   0.09351197 0.9080925 1.00000000
=====================NOTE=====================
relprice and twentys may be collinear as |0.908093|>=0.7
```

Similar to packages previously analyzed, it is not possible to avoid the dichotomous variable and it is considered to calculate the matrix of simple correlations and its determinant. It even indicates that *relprice* and *twentys* may be collinear as their coefficient of simple correlation is higher than 0.7. Thus, it is not only calculating a correlation coefficient between a quantitative variable and a qualitative one but it is also using a threshold (0.7) very far away from the one ($\sqrt{0.9}$) proposed by García et al. (4). In addition, the VIF is also calculated for the dichotomous variable.

Finally, note that with this command the variable *twentys* is designated as the one which is responsible for the existing multicollinearity. However, from the results previously presented the multicollinearity is generated by the relation between the intercept and *income*.

The mcvis package. Contradictory results are obtained by using the *mcvis* command from the **mcvis** (8) package. When original data (without transformations) are used, it is not able to detect the existing non-essential multicollinearity, apart from obtaining numerous warnings:

```
> mcvis1 = mcvis(theil.X, standardise_method="none")
There were 50 or more warnings (use warnings() to see the first 50)
>   mcvis1
          income relprice twentys
tau4 0.02   0.26     0.36    0.36
```

According to the authors "larger values indicate the greater contribution of the variable explaining the observed severity of multicollinearity". Thus, it will be designating (similar to what occurred with the **mctest** package) *relprice* and *twentys* as the variables responsible for multicollinearity, leaving the intercept in the last place.

These results should not surprise us if we take into account that the values are related to the VIFs (which is unable to detect the non-essential multicollinearity) and the eigenvalues of matrix $\mathbf{X}^t\mathbf{X}$.

If data are transformed by the *Euclidean* method, centered by mean and divided by Euclidean length, the intercept is eliminated of the analysis and, as consequence, the non-essential multicollinearity is also ignored. In any case, the results designate the variable *income* as the one responsible for the multicollinearity although the reason will not be clear:

```
>   mcvis2 = mcvis(theil.X[,-1], standardise_method="euclidean")
>   mcvis2
     income relprice twentys
tau3   0.89     0.08    0.04
```

If data are transformed with the default method, *studentise*, centred by mean and divided by standard deviation, the intercept is also eliminated of the analysis and the variables *relprice* and *twentys* are designated as the ones responsible for the multicollinearity:

```
>    mcvis3 = mcvis(theil.X[,-1])
>    mcvis3
      income  relprice  twentys
tau3    0.06      0.39     0.55
```

The transformation of the variables is a common practice when multicollinearity is being treated and, for this reason, the *mcvis* package should provide robust results for the different transformations or, at least, should clarify in which situations it needs to be used.

The perturb package. The use of *colldiag* command from the **perturb** (6) package allows us to detect the relation between the intercept and the variable *income* providing results that are complementary to the ones obtained by the **multiColl** (16) package:

```
> colldiag(reg.theil)
Condition
Index     Variance Decomposition Proportions
            intercept income relprice twentys
1    1.000 0.000     0.000  0.001    0.005
2    2.758 0.001     0.001  0.000    0.168
3   25.967 0.053     0.055  0.999    0.826
4   53.397 0.946     0.944  0.000    0.001
```

Thus, from the different commands existing in the analyzed packages in **R** whose goal is to detect multicollinearity, this is the only one that allows detection of the existence of non-essential multicollinearity in the model.

3 A special case, the simple linear regression model: *SLM* function

The simple linear model (model (1) with $k = 2$) is a particular case where it being systematically ignored in many different statistical software packages to determine if the near-multicollinearity is problematic. However, a simple linear model can present non-essential collinearity (see Salmerón et al. (11)). In this case, only the CN and the Stewart index seem appropriate measures.

In addition, the independent variable should present very little variability to be related to the intercept. For this reason, the calculation of its coefficient of variation (CV) can be very useful. Recently, Salmerón et al. (11) proposed to conclude that there is non-essential problematic collinearity if the coefficient of variation of an independent variable is lower than 0.1002506.

In the case of a dummy independent variable, the proportion of ones could be a good approximation to measure the relationship of this variable with the intercept (see Appendix A of Salmerón et al. (15)).

Considering the following simple linear regressions for Theil data, $consumption = f(income)$, $consumption = f(relprice)$ and $consumption = f(twentys)$, is obtained:

```
> SLM(theil.X[,1:2])
$`Coefficient of Variation`
[1] 0.04993766
$`Variance Inflation Factor`
[1] 1
$`Condition Number`
[1] 40.07489
$`Stewart index`
[1] 401.9994 401.9994
```

```
> SLM(theil.X[,c(1,3)])
$`Coefficient of Variation`
[1] 0.2144185
$`Variance Inflation Factor`
[1] 1
$`Condition Number`
[1] 9.43356
$`Stewart index`
[1] 22.75082 22.75082
```

```
> SLM(theil.X[,c(1,4)], T)
$`Proportion of ones in the dummy variable`
[1] 41.17647
$`Condition Number`
[1] 2.140501
```

In the first model, the value of CN indicates that the near non-essential multicollinearity is problematic. This value is accompanied by a value of CV that is very close to zero and very high values of the Stewart index. In the second model, the values of CV and the Stewart index are moderate. In addition, the value of CN is less than the established threshold. Finally, the third model is also characterized by a value of CN less than the established threshold, and the proportion of ones existing in the dummy variable also indicates a slight linear relation with the intercept.

3.1 Other packages in **R**

This subsection applies the functions existing in the **car** (2), **mctest** (19), **mcvis** (8), **rms** (5) and **perturb** (6) packages in **R** to the simple linear model commented above $consumption = f(income)$.

The car and mctest package. When the *vif* command of the **car** (2) package is applied, this eliminates the possibility that multicollinearity exists in this kind of model, since the following error is obtained: *Error in vif.default(reg.mls): model contains fewer than 2 terms.* A similar result is obtained when the *mctest* command of **mctest** (19) package is applied: *Error in if (ncol(x) < 2) stop("X matrix must contain more than one variable"): argument is of length zero.*

The mcvis package. The problem of data transformation appears again when the *mcvis* command of the **mcvis** (8) package is applied, since when the intercept is transformed a column of zeros is obtained. If the transformation is not applied, the command provides messages with numerous warnings and a result that could be interpreted as an indication of the existence of problematic near-multicollinearity:

```
> mcvis = mcvis(theil.X[,1:2], standardise_method="none")
There were 50 or more warnings (use warnings() to see the first 50)
> mcvis
        income
tau2 0.5     0.5
```

However, the same result is obtained for the model $consumption = f(relprice)$ where it was previously established that the existence of multicollinearity is not worrying. This fact leads us to consider that this measure is not adequate for detecting multicollinearity in the simple linear model.

The rms package. The following results are obtained by using the *vif* command of the **rms** (5) package:

```
> vif(lm(theil.y~theil.X[,2]))
theil.X[, 2]
           1
```

```
> vif(lm(theil.y~theil.X[,1:2]+0))
theil.X[, 1:2] theil.X[, 1:2]income
      401.9994             401.9994
```

In this case, it is possible to calculate the VIF being equal to 1 (which is in line with the results, previously commented, obtained by (11)). In addition, it is possible to calculate the VIF if the intercept is included in the design matrix and it is indicated that the model does not have an intercept. However, as previously commented, these results are not the VIFs but the Stewart indices. Thus, it is important to take special care when interpreting this information due to the fact, for lstlisting, that it will not be appropriate to use the thresholds traditionally considered for the VIF.

The perturb package. Finally, the *colldiag* command of the **perturb** (6) package allows us to calculate the condition index and variance decomposition proportions without any considerations:

```
> colldiag(reg.mls)
Index    Variance Decomposition Proportions
            intercept X
1    1.000 0.001       0.001
2   40.075 0.999       0.999
```

Thus, this command is the only one, except for the *SLM* proposed by the **multiColl** (16) package, that allows us to establish whether the degree of near-multicollinearity existing in the simple linear model is worrying.

4 Conclusions

This paper presents some of the functions of the **multiColl** (16) package to obtain the correlation matrix of the model's independent variables and its determinant (function *RdetR*), the variance inflation factors (*VIF* function), the condition number with and without intercept (*CN* and *CNs* functions) and the Stewart index (*ki* function). This package also allows the diagnosis of non-essential multicollinearity in a simple linear model (*SLM* function).

The main contributions of this paper can be summarized as follows: a) It was clarified that not all measures are adequate to detect all kinds of multicollinearity; b) When the model contains non-quantitative variables, unlike other packages existing in **R** analyzed in this paper, the **multiColl** (16) package allows us to obviate this kind of variable; and c) The superiority of the **multiColl** (16) package has been shown in comparison to other packages existing in **R** (except for the *colldiag* command of the **perturb** (6) package) to detect the degree of multicollinearity existing in a simple linear regression.

Funding

This work has been supported by project PP2019-EI-02 of the University of Granada, Spain.

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Belsley, D.: Conditioning diagnostics: collinearity and weak data in regression. John Wiley & Sons, New York (1991)

2. Fox, J., Weisberg, S., Price, B.: car: Companion to Applied Regression (2018). URL https://CRAN.R-project.org/package=car. R package version 3.0-2
3. García, C., García, J., López, M., Salmerón, R.: Collinearity: revisiting the variance inflation factor in ridge regression. Journal of Applied Statistics **42**(3), 648–661 (2018). URL https://www.tandfonline.com/doi/abs/10.1080/02664763.2014.980789
4. García, C., Salmerón, R., García, C.: A choice of the ridge factor from the correlation matrix determinant. Journal of Statistical Computation and Simulation **2**(89), 211–231 (2018). URL https://www.tandfonline.com/doi/abs/10.1080/00949655.2018.1543423?journalCode=gscs20
5. Harrell Jr, F.E.: rms: Regression Modeling Strategies (2020). URL https://CRAN.R-project.org/package=rms. R package version 6.0-1
6. Hendrickx, J.: perturb: Tools for evaluating collinearity (2012). URL https://CRAN.R-project.org/package=perturb. R package version 2.05
7. Klein, L., Goldberger, A.: An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan (1964)
8. Lin, C., Wang, K., Mueller, S.: mcvis: A new framework for collinearity discovery, diagnostic and visualization. Journal of Computational and Graphical Statistics (2020). DOI 10.1080/10618600.2020.1779729
9. Marquardt, D., Snee, R.: Ridge regression in practice. The American Statistician **1**(29), 3–20 (1975). URL https://www.tandfonline.com/doi/abs/10.1080/00031305.1975.10479105
10. Salmerón, R., García, C., García, J.: Comment on "a note on collinearity diagnostics and centering" by velilla (2018). The American Statistician (DOI: 10.1080/00031305.2019.1635527), 1–12 (2019)
11. Salmerón, R., Rodríguez, A., García, C.: Diagnosis and quantification of the non-essential collinearity. Computational Statistics (2019). URL https://doi.org/10.1007/s00180-019-00922-x
12. Salmerón, R., García, C., García, C.: multicoll: An r package to detect multicollinearity. arXiv (2019). URL https://arxiv.org/abs/1910.14590
13. Salmerón, R., García, C., García, C.: Detection of near-multicollinearity through centered and noncentered regression. Mathematics **6**(8), 931 (2020). URL https://doi.org/10.3390/math8060931
14. Salmerón, R., García, C., García, C.: A guide to using the r package "multicoll" for detecting multicollinearity. Computational Economics (2020). URL https://link.springer.com/article/10.1007/s10614-019-09967-y
15. Salmerón, R., García, C., García, C.: multicoll package and other packages to detect multicollinearity in r. arXiv (2021). URL https://arxiv.org/abs/2107.03077
16. Salmeron, R., Garcia, C., Garcia, J.: multiColl: Collinearity Detection in a Multiple Linear Regression Model (2019). URL https://CRAN.R-project.org/package=multiColl. R package version 1.0
17. Stewart, G.: Collinearity and least squares regression. Statistical Science **2**(1), 68–100 (1987)
18. Theil, H.: Principles of Econometrics. John Wiley & Sons, New York (1971)
19. Ullah, D.M.I., Aslam, D.M.: mctest: Multicollinearity Diagnostic Measures (2018). URL https://CRAN.R-project.org/package=mctest