Residualization: justification, properties and application

Catalina. B. García[a], Román Salmerón[a], Claudia García[b] and José García[c]

[a]Department of Quantitative Methods for Economics and Business, University of Granada, Spain
[b]Doctoral Program in Economics and Business Sciences, University of Granada, Spain
[c]Department of Applied Economics, University of Almería, Spain

ABSTRACT
Although it is usual to find collinearity in econometric models, it is commonly disregarded. An extended solution is to eliminate the variable causing the problem but, in some cases, this decision may affect to the goal of the research. Alternatively, the residualization not only allows to mitigate the collinearity, but it also provides an alternative interpretation of the coefficients isolating the effect of the residualized variable. This paper develops completely the *residualization* and justifies its application not only to deal with multicollinearity but also to separate the individual effects of the regressor variables. This contribution is illustrated by two econometric models with financial and ecological data, although it can also be extended to many different fields.

KEYWORDS
Collinearity; Econometric; Isolated effect

1. Introduction

Explanatory variables of an econometric model can present strong collinearity and consequently the variance of the ordinary least squares (OLS) estimators may be large compared to the values of the estimated parameters that can be insignificant or have the wrong sing. Even when collinearity diagnostic measures consider that the collinearity is not worrying, it is possible that the individual effects of the variables cannot be separated or displayed clearly. This idea resembles the objective of Shapley value regression, [57], which presents an entirely different strategy for assessing the contribution of regressor variables to the dependent variable. It owes its origin to the theory of cooperative games. The value of R^2 obtained by fitting a linear regression model is regarded as the value of a cooperative game played by the independent variables (each variable is a member) against the dependent variable (explaining it). The analyst does not have sufficient information to disentangle the contributions made by the individual members; only their joint contribution R^2 is known. The Shapley value decomposition imputes the most likely contribution of each individual member. On the other hand, [3] proposed an alternative methodology to OLS based on ordered variable regression (OVR), originally presented by [64], which entirely resolves the issue of related predictors by creating and using predictors that are perfectly unrelated.

These antecedents lead to residualization that is a procedure applied in numerous research published in relevant social science journals in many different fields, such as

CONTACT Catalina. B. García. Tel.:+34958248790. Email: cbgarcia@ugr.es

linguistics ([1, 8, 26, 34–36]), environmental issues ([28–30]) or economic development and policies ([4, 6, 32, 38, 62]), for example. This method has been also applied in previous research by the name of regression with orthogonal variables (see [44, 53]). However, this method has not been developed completely in prior works and we consider that this lack of specification leads to different criticisms as the one in [66]. [65] also concluded that "residualization of predictor variables is not the hoped-for panacea [to collinearity]". We consider that the key point not taken into consideration until now, is that this methodology provides an alternative interpretation for the estimated parameters, apart from the mitigation of collinearity. This could be seen as a limitation since the methodology is not always applicable but it can be also seen as an opportunity to obtain new interpretation not possible from the initial model. To briefly explain the general concept of the method, it might be say that by residualizing one of the explanatory variables, its effect is being isolated from the rest of the variables of the model, so we are including the part of this variable that has no relationship with the rest of independent variables, and thus we are introducing new interpretations of the residualized variable. Thus, this paper develops completely the residualization and justifies its application not only to deal with multicollinearity but also to separate the individual effects of the regressor variables. Main properties and inference are also presented together to the variance inflation factor and the condition number that allow to check if the collinearity has been mitigated after the application of residualization.

The structure of this paper is as follows: Section 2 presents the estimation and main properties of residualization showing that the estimation of the variance of the random disturbance, the global significance test, the individual significance test of the residualized variable and the goodness of fit obtained by the residualization will be similar to that of the original model. Section 3 analyzes how residualization mitigates collinearity. **Section 4 compares the residualization with OLS and others well known techniques, such as ridge regression, principal component regression or partial least squares regression**. Section 5 presents the successive residualization. Finally, Section 6 illustrates the contribution of this paper with two econometric models: the first one shows the application of the method when the main goal of the researcher is to mitigate collinearity, and the second one shows the application of the method when the purpose of the study is to obtain new interpretations of the variables. The two empirical examples belong to different fields: the first is a financial model, while the second is usually applied in ecological studies. Main conclusions are summarized in Section 7.

2. Estimation and properties

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

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

where the first column of $\mathbf{X}$ corresponds to the independent term and the random disturbance, $\mathbf{u}$, is spherical.

The first step is to define the following auxiliary regression:

$$\mathbf{X}_i = \mathbf{X}_{-i}\boldsymbol{\alpha} + \mathbf{v}, \quad i = 2, \ldots, p, \tag{2}$$

where $\mathbf{X}_{-i}$ is the result obtained after eliminating column (variable) i from matrix $\mathbf{X}$ and $\mathbf{X}_i$ represent the variable i. That is, $\mathbf{X} = (\mathbf{X}_{-i}\ \mathbf{X}_i)$.

By Ordinary Least Squares (OLS) estimation of (2), it will be obtained the correspondent estimated residuals, $\mathbf{e}_i$. They will represent the part of variable i that has no relation with any other exogenous variable of model (1) since the residuals $\mathbf{e}_i$ are orthogonal to $\mathbf{X}_{-i}$ (that is, $\mathbf{e}_i^t \mathbf{X}_{-i} = \mathbf{0}$, with $\mathbf{0}$ being a vector of zeros with appropriate dimensions).

Taking into account the previous, this method involves replacing the variable $\mathbf{X}_i$ by the estimated residuals of model (2), $\mathbf{e}_i$, in the original model (1). Thus, the residualization is obtained from the following expression:

$$\mathbf{Y} = \mathbf{X}_O \boldsymbol{\gamma} + \mathbf{w}, \tag{3}$$

where $\mathbf{X}_O = (\mathbf{X}_{-i}\ \mathbf{e}_i)$.

Once the basic procedure is explained, lets compare the results of model (1) and model (3).

2.1. Estimation

From $\mathbf{X} = (\mathbf{X}_{-i}\ \mathbf{X}_i)$, the OLS estimator for model (1), $\widehat{\boldsymbol{\beta}}$, will be:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y} = \begin{pmatrix} \mathbf{X}_{-i}^t\mathbf{X}_{-i} & \mathbf{X}_{-i}^t\mathbf{X}_i \\ \mathbf{X}_i^t\mathbf{X}_{-i} & \mathbf{X}_i^t\mathbf{X}_i \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{X}_{-i}^t\mathbf{Y} \\ \mathbf{X}_i^t\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^t & C \end{pmatrix} \cdot \begin{pmatrix} \mathbf{X}_{-i}^t\mathbf{Y} \\ \mathbf{X}_i^t\mathbf{Y} \end{pmatrix} \\ &= \begin{pmatrix} \left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}^t\mathbf{Y} - \widehat{\boldsymbol{\alpha}} \cdot \frac{\mathbf{e}_i^t\mathbf{Y}}{\mathbf{e}_i^t\mathbf{e}_i} \\ \frac{\mathbf{e}_i^t\mathbf{Y}}{\mathbf{e}_i^t\mathbf{e}_i} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{-i} \\ \widehat{\beta}_i \end{pmatrix}, \end{aligned} \tag{4}$$

taking into account that:

$$\begin{aligned} C &= \left(\mathbf{X}_i^t\mathbf{X}_i - \mathbf{X}_i^t\mathbf{X}_{-i}\left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}^t\mathbf{X}_i\right)^{-1} \\ &= \left(\mathbf{X}_i^t\left(\mathbf{I} - \mathbf{X}_{-i}\left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}^t\right)\mathbf{X}_i\right) = \left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1}, \\ \mathbf{B} &= -\left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}^t\mathbf{X}_i \cdot \left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1} = -\widehat{\boldsymbol{\alpha}} \cdot \left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1}, \\ \mathbf{A} &= \left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1} + \left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}^t\mathbf{X}_i \cdot \left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1}\mathbf{X}_i^t\mathbf{X}_{-i}\left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1} \\ &= \left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1} + \left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1} \cdot \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\alpha}}^t, \end{aligned}$$

where $\widehat{\boldsymbol{\alpha}}$ and $\mathbf{e}_i^t\mathbf{e}_i$ are, respectively, the OLS estimator and the sum of the square of the residuals of the auxiliary regression (2).

Likewise, since $\mathbf{e}_i^t\mathbf{X}_{-i} = \mathbf{0}$, the OLS estimation of model (3), $\widehat{\boldsymbol{\gamma}}$, is given by:

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} &= \left(\mathbf{X}_O^t\mathbf{X}_O\right)^{-1}\mathbf{X}_O^t\mathbf{Y} = \begin{pmatrix} \mathbf{X}_{-i}^t\mathbf{X}_{-i} & \mathbf{X}_{-i}^t\mathbf{e}_i \\ \mathbf{e}_i^t\mathbf{X}_{-i} & \mathbf{e}_i^t\mathbf{e}_i \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{X}_{-i}^t\mathbf{Y} \\ \mathbf{e}_i^t\mathbf{Y} \end{pmatrix} \\ &= \begin{pmatrix} \left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}^t\mathbf{Y} \\ \frac{\mathbf{e}_i^t\mathbf{Y}}{\mathbf{e}_i^t\mathbf{e}_i} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\gamma}}_{-i} \\ \widehat{\gamma}_i \end{pmatrix}. \end{aligned} \tag{5}$$

Thus, it is possible to compare the OLS estimators, expression (5), of the residualized model (3) with the OLS estimators of model (1), expression (4), obtaining the following conclusions:

- The estimate of the coefficient of the residualized variable does not change in model (3), that is, $\widehat{\beta}_i = \widehat{\gamma}_i$. However, the interpretation is different: the variation produced in dependent variable, $\mathbf{Y}$, given an increase in $\mathbf{e}_i$, that is to say, the part of the independent variable $\mathbf{X}_i$ that is not related to the rest of independent variables, $\mathbf{X}_{-i}$. Hence, due to the new interpretation of the residualized variable, the residualization could be applied to obtain conclusions that would not otherwise be possible.
- The orthogonality of $\mathbf{e}_i$ with $\mathbf{X}_{-i}$ verifies the *ceteris paribus* assumption, that is to say, when this variable increases, the other variables remain constant.
- The estimation of the non-residualized variables in model (3) differs:

$$\widehat{\boldsymbol{\beta}}_{-i} = \widehat{\boldsymbol{\gamma}}_{-i} - \widehat{\boldsymbol{\alpha}} \cdot \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i}. \tag{6}$$

  However, the interpretation will be the same as in model (1).

In addition, it is interesting to take into consideration that:

- For convenience purposes, see Section 3, all of the independent variables in model (1) have been included in the auxiliary regression (2). However, it will be possible to include only some of the independent variables, depending on the interest of the researcher (for example, trying to obtain interpretable residuals). In this case, the estimations of explanatory variables which are not included in the auxiliary regression will not change their value also. Furthermore, if the constant is included in the auxiliary regression, nonessential collinearity will be mitigated because the residuals will be orthogonal to the constant. See Section 3 for details on distinguishing among the different types of collinearity.
- The estimate of the non-residualized variables in model (3) coincides with the estimate obtained from model $\mathbf{Y} = \mathbf{X}_{-i}\boldsymbol{\delta} + \boldsymbol{\vartheta}$. That is, the estimation and interpretation of the non-residualized variables will be the same as that obtained in a regression in which the residualized variable is eliminated. Nevertheless, this coincidence only occurs when we introduce in the auxiliary regression all the rest of explanatory variables of the original model. Furthermore, since the two models have different residuals, the inference associated with these coefficients will be different.

Remark 1. *Another interesting question is how to select the variable to residualize. Throughout the work, we present different criteria that can be applied, or a combination of them, depending to the goal of the research.*

If the goal is to look for new interpretations, the variable to residualize will be the one that leads to the new interpretation desired by the researcher since the only interpretation that changes is that of the residualized variable. In this case, in the auxiliary regression (2), it will be possible to use all the independent variables or only some of them.

It is also interesting to rank the independent variables in model (1) according to their relevance to avoid residualizing variables considered to be relevant to avoid changing the interpretation of their coefficients. This fact was already been proposed in [3] with the use of ordered variable regression models.

2.2. Goodness of fit, estimation of the variance of the random disturbance and joint significance

The estimated residuals of original model (1) will be given by:

$$\begin{aligned}
\mathbf{e} &= \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X} \cdot \widehat{\boldsymbol{\beta}} = \mathbf{Y} - (\mathbf{X}_{-i} \ \mathbf{X}_i) \cdot \begin{pmatrix} \left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1} \mathbf{X}_{-i}^t \mathbf{Y} - \widehat{\boldsymbol{\alpha}} \cdot \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i} \\ \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i} \end{pmatrix} \\
&= \mathbf{Y} - \mathbf{X}_{-i} \left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1} \mathbf{X}_{-i}^t \mathbf{Y} + \mathbf{X}_{-i} \widehat{\boldsymbol{\alpha}} \cdot \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i} - \mathbf{X}_i \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i} \\
&= \mathbf{Y} - \mathbf{X}_{-i} \left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1} \mathbf{X}_{-i}^t \mathbf{Y} - \mathbf{e}_i \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i},
\end{aligned} \tag{7}$$

since $\mathbf{e}_i$ are the residuals of the auxiliary regression (2), it is verified that $\mathbf{e}_i = \mathbf{X}_i - \mathbf{X}_{-i}\widehat{\boldsymbol{\alpha}}$.

And those of the residualized model (3) are:

$$\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}_O = \mathbf{Y} - \mathbf{X}_O \cdot \widehat{\boldsymbol{\gamma}} = \mathbf{Y} - \mathbf{X}_{-i} \left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1} \mathbf{X}_{-i}^t \mathbf{Y} - \mathbf{e}_i \frac{\mathbf{e}_i^t \mathbf{Y}}{\mathbf{e}_i^t \mathbf{e}_i}. \tag{8}$$

It is evident that expression (8) coincides with (7), that is to say, the residuals of the original (1) and residualized (3) models coincide. Therefore, it is possible to conclude the following:

- It is evident that the squared sums of the residuals of the two models coincide and, consequently, that both models yield the same estimate of the variance of the random disturbance.
- Since the two models employ the same dependent variable, the total sum of squares will be also the same, and consequently, the coefficients of determination will also coincide.
- Since the F statistic of the global significance test can be expressed as a function of the coefficient of determination, it is evident that the global significance tests of both models will also be the same.
- It is clear that $\widehat{\mathbf{Y}} = \widehat{\mathbf{Y}}_O$, it is to say, that the original model and the residualized one provide the same prediction. Recall that if the goal is simply to predict $\mathbf{Y}$ from a set of variables $\mathbf{X}$, then multicollinearity is not a problem because the predictions will still be accurate [25]. Thus, residualization mitigates multicollinearity but maintains the same prediction.

2.3. Individual inference

Since the random disturbances are spherical, the individual inference will be given by the main diagonal of matrix $\left(\mathbf{X}^t \mathbf{X}\right)^{-1}$, that is to say (see expression (4)), by:

$$\begin{pmatrix} \left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1} + \left(\mathbf{e}_i^t \mathbf{e}_i\right)^{-1} \cdot \widehat{\boldsymbol{\alpha}} \widehat{\boldsymbol{\alpha}}^t & -\widehat{\boldsymbol{\alpha}} \cdot \left(\mathbf{e}_i^t \mathbf{e}_i\right)^{-1} \\ -\widehat{\boldsymbol{\alpha}}^t \cdot \left(\mathbf{e}_i^t \mathbf{e}_i\right)^{-1} & \left(\mathbf{e}_i^t \mathbf{e}_i\right)^{-1} \end{pmatrix}. \tag{9}$$

Taking into account the following expression:

$$\left(\mathbf{X}_O^t \mathbf{X}_O\right)^{-1} = \begin{pmatrix} \left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{e}_i^t \mathbf{e}_i\right)^{-1} \end{pmatrix}, \tag{10}$$

it is evident that the main diagonal of both matrices are different, except for the i element. Since the estimation of the variance of the random disturbance is the same, considering the estimation of the coefficients, it is possible to conclude the following:

- The inference related to the individual significance (t-Student test) of the unchanged variables differs between models (1) and (3).
- The inference related to the individual significance (t-Student test) of the residualized variable coincides in models (1) and (3).

Consequently, the residualization of the initial model does not affect the estimation of the variance of the random disturbance, the coefficient of determination, the global significance test, or the individual significance test of the residualized variable. It only changes the individual significance of unaltered variables.

Remark 2. *Another option is to residualize a variable with a coefficient that is significantly different from zero in the original model since the individual significance test of the residualized variable is maintained in the residualized model.*

3. Collinearity

As mentioned in the introductory Section, multicollinearity consists of the presence of interdependency between explanatory variables ([19]), distinguishing between two principal types of multicollinearity: perfect collinearity and near-collinearity [43, 47, 58]. The first type occurs when the interdependency between variables is exact, and the second occurs when it is approximate. Near-collinearity, also known as imperfect or approximate collinearity, may be divided into essential and nonessential collinearity ([39, 40, 59]). The former concerns the relationship between explanatory variables, excluding the intercept, while the latter involves the relationship between the intercept and at least one of the remaining independent variables of the model.

In addition to the new interpretation of the coefficient of the residualized variable, another result of interest in the residualized model is the effect on the linear relationship between the independent variables of the initial model. To verify that collinearity is mitigated after the residualization of the initial model, in the residualized model we analyze, in the residualized model, the estimated variances of the estimated coefficients, the Variance Inflation Factor (VIF) and the Condition Number (CN).

3.1. Decrease in estimated standard variance

Considering that the estimation of the random disturbance variance is the same in the original and the residualized model, the estimated variances of the coefficients will be determined by the main diagonal of the matrices $\left(\mathbf{X}^t\mathbf{X}\right)^{-1}$ and $\left(\mathbf{X}_O^t\mathbf{X}_O\right)^{-1}$. As noted above, the element corresponding to the residualized variable is the same in both matrices, and thus, the estimated variance will be also the same. That is, $\widehat{Var\left(\widehat{\beta}_i\right)} = \widehat{Var\left(\widehat{\gamma}_i\right)}$.

For the rest of the variables, given expressions (9) and (10), we obtain the following:

$$\widehat{Var}\left(\widehat{\beta}_j\right) = \widehat{\sigma}^2 \cdot \left(w_{jj} + (\mathbf{e}_i^t \mathbf{e}_i)^{-1} \alpha_{jj}\right), \quad \widehat{Var}\left(\widehat{\gamma}_j\right) = \widehat{\sigma}^2 \cdot w_{jj}, \quad j = 1, \dots, p, \ j \neq i,$$

where w_{jj} y $\alpha_{jj} = \alpha_j^2$ are the elements (j, j) of the matrices $\left(\mathbf{X}_{-i}^t \mathbf{X}_{-i}\right)^{-1}$ and $\widehat{\boldsymbol{\alpha}} \widehat{\boldsymbol{\alpha}}^t$. Since $(\mathbf{e}_i^t \mathbf{e}_i)^{-1} \alpha_j^2 \geq 0$, it is verified that $\widehat{Var}\left(\widehat{\beta}_j\right) \geq \widehat{Var}\left(\widehat{\gamma}_j\right)$ for $j = 1, \dots, p$, with $j \neq i$. In consequence, the estimated variances of the residualized model will be always less than or equal to those in the original model.

This result is relevant since it demonstrates that the residualization implies a decrease in the estimated variances of the estimated coefficients (which are assumed to be inflated due to the presence of collinearity). Note that this result is contrary to the conclusions presented by [7].

Remark 3. *The linear relationship on the coefficients of the model (1) given in (6) can also be used to reduce the variance of the estimated coefficients only estimating model by restricted least-squares estimator. In this case the residualization could be used to mitigate one of the consequences of the existence of severe collinearity in the multiple linear regression model.*

3.2. Variance Inflation Factor

Each explanatory variable of model (1) has an associated VIF given by:

$$\text{VIF}_i = \frac{1}{1 - R_i^2}, \quad i = 2, \dots, p, \tag{11}$$

where R_i^2 is the coefficient of determination of model (2). It is generally accepted that values of VIF higher than 10 indicate severe collinearity, [31].

Applying this definition in the residualized model (3) and being $\mathbf{e}_i$ the dependent variable of the auxiliary regression, its coefficient of determination will be zero and the associated VIF will be one (the minimum value possible). In other case, R_j^2 will be obtained from the following auxiliary regression:

$$\mathbf{X}_j = \mathbf{X}_{O_{-j}} \ \boldsymbol{\xi} + \boldsymbol{\epsilon}, \quad j = 2, \dots, p, \ j \neq i, \tag{12}$$

where $\mathbf{X}_{O_{-j}}$ is the result obtained after eliminating column (variable) j from matrix $\mathbf{X}_O$.

Due to the orthogonality of $\mathbf{e}_i$ with $\mathbf{X}_{-i,-j}$ (matrix $\mathbf{X}$ without columns (variables) i and j), the residuals of (12) coincide[1] with the residuals of the following model:

$$\mathbf{X}_j = \mathbf{X}_{-i,-j} \ \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad i, j = 2, \dots, p, \ i \neq j. \tag{13}$$

[1]With $\mathbf{e}^1$ being the residuals of the auxiliary regression (12) and $\mathbf{e}^2$ being the residuals of regression (13) and given that $\mathbf{e}_i$ is orthogonal to $\mathbf{X}_j$ and $\mathbf{X}_{-i,-j}$, we obtain that

$$\begin{aligned} \mathbf{e}^1 &= \mathbf{X}_j - (\mathbf{X}_{-i,-j} \ \mathbf{e}_i) \widehat{\boldsymbol{\xi}} = \mathbf{X}_j - (\mathbf{X}_{-i,-j} \ \mathbf{e}_i) \cdot \begin{pmatrix} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j}\right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_j \\ \mathbf{0} \end{pmatrix} \\ &= \mathbf{X}_j - \mathbf{X}_{-i,-j} \left(\mathbf{X}_{-i,-j}^t \mathbf{X}_{-i,-j}\right)^{-1} \mathbf{X}_{-i,-j}^t \mathbf{X}_j = \mathbf{X}_j - \mathbf{X}_{-i,-j} \widehat{\boldsymbol{\eta}} = \mathbf{e}^2. \end{aligned}$$

Then, models (12) and (13) have the same coefficient of determination since the dependent variable is the same in both models.

However, the coefficient of determination of model (13) will be lower than that of model:

$$\mathbf{X}_j = \mathbf{X}_{-j}\boldsymbol{\theta} + \boldsymbol{\omega}, \quad i, j = 2, \dots, p, \ i \neq j, \tag{14}$$

since this latter model contains an additional independent variable, $\mathbf{X}_i$. Then, the coefficient of determination of model (12) is lower than that of model (14).

Thus, since the VIF associated with variable j in the original model (1) is obtained from the coefficient of determination of the auxiliary regression given by (14) and in the residualized model (3) from the coefficient of determination of model (12), it is clear that the VIF is decreased after residualizing the model. That is, the collinearity present in the model has been diminished.

Remark 4. *If the goal is to mitigate the collinearity in the model, we suggest residualizing the variable with the highest Variance Inflation Factor because, after the residualization, the VIF will be equal to 1. In this case, all independent variables have to be included in the auxiliary regression (2) to mitigate the collinearity in the most efficient way.*

3.3. Condition Number

Given the model (1), the condition number (CN) is given by:

$$K(\widetilde{\mathbf{X}}^t\widetilde{\mathbf{X}}) = \sqrt{\frac{\mu_{max}}{\mu_{min}}},$$

where μ_{min} and μ_{max} are, respectively, the minimum and maximum eigenvalue of $\mathbf{X}^t\mathbf{X}$. Note that previously the matrix $\mathbf{X}$ should be transformed to be unit length by columns, it is to say, data should be divided by the square root of the sum of its squared elements (see [5]). This author stated that values of CN between 20 and 30 indicate moderated collinearity and values higher than 30 indicate high collinearity.

Then, the CN associated to model (3) is obtained as follows:

$$K(\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}},$$

where λ_{min} and λ_{max} are, respectively, the minimum and maximum eigenvalue of $\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O$, where:

$$\widetilde{\mathbf{X}}_O = \left(\frac{\mathbf{X}_1}{||\mathbf{X}_1||} \cdots \frac{\mathbf{X}_{i-1}}{||\mathbf{X}_{i-1}||} \ \frac{\mathbf{X}_{i+1}}{||\mathbf{X}_{i+1}||} \cdots \frac{\mathbf{X}_p}{||\mathbf{X}_p||} \ \frac{\mathbf{e}_i}{||\mathbf{e}_i||} \right) = \left(\widetilde{\mathbf{X}}_{-i} \ \frac{\mathbf{e}_i}{||\mathbf{e}_i||} \right),$$

being $||\mathbf{X}_k|| = \sqrt{\sum\limits_{j=1}^{n} X_{kj}^2}$ for $k = 1, \dots, i-1, i+1, \dots, p$ and $||\mathbf{e}_i|| = \sqrt{\sum\limits_{j=1}^{n} e_{ij}^2}$. Then:

$$\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O = \begin{pmatrix} \widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}.$$

Thus, one of the p eigenvalues of $\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O$ will be equal to one and the rest will coincide with the eigenvalues of matrix $\widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i}$. Supposing that the eigenvalue equal to one is the first one, $\lambda_1 = 1$, it is verified that:

- If this is the minimum eigenvalue of $\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O$, the rest of eigenvalues will be equal or higher than one ($\lambda_i \geq 1, i = 2, \dots, p$), and, consequently, its sum will be equal or higher than $p-1$ $\left(\sum\limits_{i=2}^{p} \lambda_i \geq p-1\right)$. However, this sum will be equal to $p-1$ (since the trace of $\widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i}$ is equal to $p-1$). Then, all the eigenvalues will be equal to one ($\lambda_i = 1$ with $i = 1, 2, \dots, p$), it is to say, $\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O$ will be the identity matrix and, then, all the variables will be considered orthogonal between them.
- If this is the maximum eigenvalue of $\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O$, the rest of eigenvalues will be equal or lesser than one ($\lambda_i \leq 1, i = 2, \dots, p$), and, consequently, its sum will be equal or lesser than $p-1$ $\left(\sum\limits_{i=2}^{p} \lambda_i \leq p-1\right)$. However, it was justified that this sum is equal to $p-1$. Then, all the eigenvalues will be equal to one ($\lambda_i = 1$ with $i = 1, 2, \dots, p$). Thus, as before, all the variables will be considered orthogonal between them.

If the eigenvalue equal to one can not be the minimum or maximum eigenvalue of $\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O$, they have to be found on the rest of the eigenvalues of $\widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i}$. Thus, the CN of model (3) coincides with that of the auxiliary regression (2):

$$K(\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O) = K(\widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i}).$$

On the other hand, according to the Cauchy's Interlace Theorem for Eigenvalues of Hermitian Matrices[2], since $\widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i}$ is a submatrix of order $p-1$ of $\widetilde{\mathbf{X}}^t\widetilde{\mathbf{X}}$ is evident that it has to be verified that:

$$K(\widetilde{\mathbf{X}}_O^t\widetilde{\mathbf{X}}_O) = K(\widetilde{\mathbf{X}}_{-i}^t\widetilde{\mathbf{X}}_{-i}) \leq K(\widetilde{\mathbf{X}}^t\widetilde{\mathbf{X}}).$$

Thus, the condition number or the residualization (3) has to be equal or lesser than the condition number of the original model (1).

Remark 5. *If the goal is to mitigate the collinearity in the model, we suggest to residualize the variable i whose auxiliary regression (where the variable i is the dependent one) presents the lowest CN since it will coincides with the CN of the residualized model that will be always equal or lesser than the CN of the original model.*

4. Comparision of the residualization method with other existing methods

El objetivo de la presente sección es el de comparar mediante a Monte Carlo simulation the residualization method con otros métodos existentes para mitigar la multicolinealidad such as ridge regression (RR), principal component regression (PCR) and partial least squares regression (PLSR). Dicha comparación se realizará mediante el cálculo del mean square er-

[2]Given a matrix $\mathbf{A}$ with order p and eigenvalues $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_p$ and given its submatrix B with order $p-1$ and eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_{p-1}$ it is verified that $\xi_1 \leq \mu_1 \leq \xi_2 \leq \mu_2 \leq \xi_3 \leq \cdots \leq \xi_{p-1} \leq \mu_{p-1} \leq \xi_p$.

ror (MSE) y de los errores de predicción de las técnicas anteriores, del residualization method and OLS estimation.

Así, a continuación se muestra cómo calcular el MSE for the residualization method y cómo compararlo con el MSE de OLS y se presentarán las métricas que se van a usar para medir la capacidad predictiva de cada técnica.

4.1. Mean Square Error

Note that the original model is different from the residualized model, and for this reason, both models should be analyzed separately and the comparison may not be convenient. However, some publications have not considered this divergence (see, for example, [66]). For this reason, since $\widehat{\gamma}$ is a biased estimator of $\boldsymbol{\beta}$:

$$\begin{aligned} \widehat{\gamma} &= (\mathbf{X}_O\mathbf{X}_O)^{-1} \cdot \mathbf{X}_O^t\mathbf{Y} = (\mathbf{X}_O\mathbf{X}_O)^{-1} \cdot \mathbf{X}_O^t\mathbf{X} \cdot \boldsymbol{\beta} + (\mathbf{X}_O\mathbf{X}_O)^{-1} \cdot \mathbf{X}_O^t \cdot \mathbf{u} \\ E[\widehat{\gamma}] &= (\mathbf{X}_O\mathbf{X}_O)^{-1} \cdot \mathbf{X}_O^t\mathbf{X} \cdot \boldsymbol{\beta} \neq \boldsymbol{\beta} \text{ since } \mathbf{X}_O^t \neq \mathbf{X}, \end{aligned}$$

it could be interesting to calculate the MSE of residualization and to compare it with the MSE of the OLS estimator.

Given an estimator $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, its MSE is expressed as:

$$MSE\left(\widetilde{\boldsymbol{\beta}}\right) = trace\left(var\left(\widetilde{\boldsymbol{\beta}}\right)\right) + \left(E\left[\widetilde{\boldsymbol{\beta}}\right] - \boldsymbol{\beta}\right)^t \left(E\left[\widetilde{\boldsymbol{\beta}}\right] - \boldsymbol{\beta}\right).$$

In the case of the OLS estimator, $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator, where $E\left[\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$, and it is verified in the following:

$$\begin{aligned} MSE\left(\widehat{\boldsymbol{\beta}}\right) &= trace\left(var\left(\widehat{\boldsymbol{\beta}}\right)\right) \\ &= \sigma^2 \cdot \left[trace\left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1} + (\mathbf{e}_i^t\mathbf{e}_i)^{-1} \cdot trace\left(\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\alpha}}^t\right) + (\mathbf{e}_i^t\mathbf{e}_i)^{-1}\right], \quad (15) \end{aligned}$$

taking into account expression (9).

For the estimator $\widehat{\gamma}$, starting from (10), it is verified that:

$$\begin{aligned} MSE\left(\widehat{\gamma}\right) &= trace\left(var\left(\widehat{\gamma}\right)\right) + \left(E\left[\widehat{\gamma}\right] - \boldsymbol{\beta}\right)^t \left(E\left[\widehat{\gamma}\right] - \boldsymbol{\beta}\right) \\ &= \sigma^2 \cdot \left[trace\left(\mathbf{X}_{-i}^t\mathbf{X}_{-i}\right)^{-1} + (\mathbf{e}_i^t\mathbf{e}_i)^{-1}\right] + \beta_i \cdot \widehat{\boldsymbol{\alpha}}^t\widehat{\boldsymbol{\alpha}} \cdot \beta_i, \quad (16) \end{aligned}$$

Based on expressions (4), (5) and (6), we could say that $\widehat{\gamma} = \widehat{\boldsymbol{\beta}} + \mathbf{s}$, where:

$$\mathbf{s} = \begin{pmatrix} \widehat{\boldsymbol{\alpha}} \cdot \frac{\mathbf{e}_i^t\mathbf{Y}}{\mathbf{e}_i^t\mathbf{e}_i} \\ 0 \end{pmatrix}.$$

Thus, as:

$$\mathbf{e}_i^t\mathbf{Y} = \mathbf{e}_i^t\mathbf{X}\boldsymbol{\beta} + \mathbf{e}_i^t\mathbf{u} = [\mathbf{0}\ \mathbf{e}_i^t\mathbf{X}_i] \cdot \boldsymbol{\beta} + \mathbf{e}_i^t\mathbf{u} = \mathbf{e}_i^t\mathbf{X}_i\beta_i + \mathbf{e}_i^t\mathbf{u} = \mathbf{e}_i^t\mathbf{e}_i\beta_i + \mathbf{e}_i^t\mathbf{u},$$

we have that:

$$E\left[\widehat{\gamma}\right]=E\left[\widehat{\boldsymbol{\beta}}\right]+E\left[\mathbf{s}\right]=\boldsymbol{\beta}+\left(\begin{array}{c}\widehat{\boldsymbol{\alpha}}\cdot\beta_i\\0\end{array}\right)\Rightarrow\left(E\left[\widehat{\gamma}\right]-\boldsymbol{\beta}\right)^t\left(E\left[\widehat{\gamma}\right]-\boldsymbol{\beta}\right)=\beta_i^2\cdot\widehat{\boldsymbol{\alpha}}^t\widehat{\boldsymbol{\alpha}}.$$

From expressions (15) and (16), it is clear that:

$$MSE\left(\widehat{\gamma}\right)=MSE\left(\widehat{\boldsymbol{\beta}}\right)-\sigma^2\cdot\left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1}\cdot trace\left(\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\alpha}}^t\right)+\beta_i^2\cdot\widehat{\boldsymbol{\alpha}}^t\widehat{\boldsymbol{\alpha}},$$

so $\widehat{\gamma}$ has a lower MSE than $\widehat{\boldsymbol{\beta}}$ if:

$$\beta_i^2\cdot\widehat{\boldsymbol{\alpha}}^t\widehat{\boldsymbol{\alpha}}<\sigma^2\cdot\left(\mathbf{e}_i^t\mathbf{e}_i\right)^{-1}\cdot trace\left(\widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\alpha}}^t\right). \tag{17}$$

4.2. Metrics

Para medir la capacidad de ajuste de cada modelo se usarán the root mean squared error (RMSE) and the mean absolute error (MAE), mientras que para la capacidad predictiva the root mean squared prediction error (RMSPE) and the mean absolute prediction error (MAPE).

Dada una muestra de tamano n, supongamos que se divide en dos partes, una de tamano m y otra de tamano h de forma que $m+h=n$). Entonces, se usa la primera de ellas para aplicar cada una de los métodos de estimación comentados y calcular RMSE, mientras que en la segunda se evalúa su capacidad predictiva mediante la obtención de RMSPE y MAPE. Esto es:

$$RMSE=\sqrt{\frac{1}{m}\cdot\sum_{i=1}^{m}(Y_i-\widehat{Y}_i)^2},\qquad MAE=\frac{1}{m}\sum_{i=1}^{m}|Y_i-\widehat{Y}_i|,$$

$$RMSPE=\sqrt{\frac{1}{h}\cdot\sum_{i=m+1}^{n}(Y_i-\widehat{Y}_i)^2},\qquad MAPE=\frac{1}{h}\cdot\sum_{i=m+1}^{n}|Y_i-\widehat{Y}_i|.$$

4.3. Simulation

A continuación se describe la simulación realizada para conseguir los objetivos mencionados al inicio de la presente sección.

Given the model $\mathbf{Y}=\beta_1+\beta_2\cdot\mathbf{X}_2+\beta_3\cdot\mathbf{X}_3+\mathbf{u}$, the following simulation is performed in order to establish the behavior of condition (17):

1. It is considered that $\boldsymbol{\mu}_{2\times1}=(\mu_1,\mu_2)^t$ with $\mu_1,\mu_2\in\{-10,-9,-8,\ldots,8,9,10\}$.
2. It is also considered that $a_1,a_2\in\{0,1,2,3,4\}$ and $b_1,b_2\in\{0.1,0.2,0.3,\ldots,1.9,2\}$, so $c^i_{5\times1}\sim N(a_i,b_i^2)$ is generated. Thus, given matrix $\mathbf{C}=[c^1\ c^2]$, a symmetric positive-definite matrix, $\Sigma_{2\times2}=\mathbf{C}^t\mathbf{C}$, is built.
3. $\mathbf{X}_2$ and $\mathbf{X}_3$ are generated from $N_2\left(\boldsymbol{\mu}_{2\times1},\Sigma_{2\times2}\right)$.

Table 1. Simulation results for MSE.

	n	25	50	75	100	125	150	Mean
Cond. (17)	Resid. variable: $\mathbf{X}_2$	8.13%	6.96%	7.15%	6.69%	6.96%	7.00%	7.159%
	Resid. variable: $\mathbf{X}_3$	8.42%	7.15%	7.12%	6.71%	6.77%	6.84%	
	$min\ cor(\mathbf{X}_2, \mathbf{X}_3)$	-0.9862	-0.9747	-0.9883	-0.9973	-0.9934	-0.9915	0.4877
	$max\ cor(\mathbf{X}_2, \mathbf{X}_3)$	0.9999	0.9999	0.9999	0.9999	0.9998	0.9999	
	min CV($\mathbf{X}_2$)	0.00958	0.00935	0.00911	0.0111296	0.00628	0.01072	6.6665
	max CV($\mathbf{X}_2$)	30222.48	12359.03	3249.84	3426.69	31001.38	1498.77	
	min CV($\mathbf{X}_3$)	0.0107	0.0116	0.00791	0.00682	0.0108	0.0114	5.7096
	max CV($\mathbf{X}_3$)	9763.29	37893.3	5199.45	2718.18	2655.24	3586.79	

4. The random perturbance, $\mathbf{u}$, is generated as $\mathbf{u} \sim N(0, d^2)$, where $d \in \{1, 2, 3, 4\}$, from which is calculated $\mathbf{Y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \beta_3 \cdot \mathbf{X}_3 + \mathbf{u}$, where $\beta_i \in \{-5, -4, \dots, 4, 5\}$.
5. Once the previous model and the corresponding auxiliary regressions are estimated, condition (17) is calculated from the obtained estimations of β_i, σ^2 and $\boldsymbol{\alpha}$.

A comparison of both models (OLS and residualization) is conducted with different sample sizes, $n \in \{25, 50, 75, 100, 125, 150\}$, such that 60000 simulations are performed in this experiment.

First, in Table 1, it can be observed that there are two types of situations: one where essential collinearity does not imply strong collinearity problems (the mean correlation is equal to 0.4877, which leads to a VIF value of 1.31208) and another where essential collinearity implies strong collinearity problems (the maximum and minimum correlations lead to VIF values of approximately 50.2512). It can also be observed that there are two types of situations in relation to nonessential collinearity: one where it is not worrisome (the mean value of the coefficient of variation (CV) is approximately 6, which implies the data have enough variability) and another where nonessential collinearity is worrisome (the minimum values of CV for each variable are close to zero, which implies slight variability of the data and indicates that the data may be considered almost constant and hence related to the intercept).

The first and second rows of Table 1 show the percentage of cases in which $MSE\left(\widehat{\gamma}\right) < MSE\left(\widehat{\boldsymbol{\beta}}\right)$ (condition (17) is verified), considering that variables $\mathbf{X}_2$ and $\mathbf{X}_3$ are residualized. Note that both results are similar and that there are no material differences for different sample sizes. The results (see Table 1) show that in only 7.159% of the cases, the condition $MSE\left(\widehat{\gamma}\right) < MSE\left(\widehat{\boldsymbol{\beta}}\right)$ is verified.

Second, si se realizan otras 60000 simulations dividiendo la muestra tal y como se indica en la subsection 4.2 teniendo en cuenta que $h = 0.15 \cdot n$ se obtienen los valores mostrados in Table 2. Para obtener los valores referentes a PCR y PLSR se ha usado the R's package *pls*. En el caso de la RR se ha obtenido el valor del estimador $\widehat{\boldsymbol{\beta}}(k)$ considerando aquel valor de k que ha mitigado el grado de multicolinealidad, entendiendo que esta situacin se alcanza cuando el valor del CN, calculado tal y como se muestra en [54], queda por debajo del umbral establecido como preocupante, 20. Esta forma de trabajar ya ha sido usada por [21] trabajando with the VIF en lugar del CN (trabajar en este caso con el VIF se debe a que en [52] se muestra que el VIF ignora la multicolinealidad aproximada del tipo no-esencial).

Table 2. Simulation results for RMSE, MAE, RMSPE and MAPE.

Metric	OLS	Resid. variable: $\mathbf{X}_2$	Resid. variable: $\mathbf{X}_3$	RR	PCR	PLSR
RMSE	2.635103	2.635103	2.635103	2.637032	8.810389	6.182757
MAE	1.960508	1.960508	1.960508	1.961239	8.507369	6.266333
RMSPE	2.725054	19.73536	19.64029	2.721985	2.725054	2.725054
MAPE	2.087758	17.56701	17.48688	2.084665	2.087758	2.087758

Atendiendo a los resultados obtenidos en la primera muestra, se tiene que the residualization method and OLS conducen a los mismos resultados para RMSE and MAE tal y como se adelantaba in subsection 2.2 al verificarse que $\widehat{\mathbf{Y}} = \widehat{\mathbf{Y}}_O$. Además, estos valores son inferiores a los de las demás tcnicas (levemente inferiores en el caso de RR).

Observando los resultados de RMSPE y MAPE en la segunda muestra, se tiene que the residualization method tiene peor capacidad predictiva que las dems técnicas. Sin embargo, el hecho de que las demás no mejoren sustancialmente los resultados proporcionados por OLS indican que en el caso de que el interés resida únicamente en la predicción, quizás, la mejor opción sea no hacer nada. Estos resultados concuerdan con las apreciaciones dadas por [25] ya comentadas: if the goal is simply to predict, then multicollinearity is not a problem because the predictions will still be accurate.

5. Successive residualization

It is possible that the goal of the researcher (to mitigate collinearity or obtain a new interpretation for the estimated coefficients) has not been achieved after residualizing the first variable. Then, it is necessary to residualize a second variable. In this case, the doubly residualized model will be given by

$$\mathbf{Y} = \mathbf{X}_{OO}\boldsymbol{\lambda} + \varsigma, \tag{18}$$

where $\mathbf{X}_{OO} = (\mathbf{X}_{-i,-j} \ \mathbf{e}_i \ \mathbf{e}_j)$ with $\mathbf{e}_j$ being the residuals of the auxiliary regression (13).

The goal of this Section is not to obtain the estimated inference of this model (18) but to highlight that successive residualization may be interesting, and in this case, the residuals $\mathbf{e}_i$ y $\mathbf{e}_j$ will be orthogonal since it is verified that:

$$\mathbf{e}_i^t \mathbf{e}_j = \mathbf{e}_i^t \left(\mathbf{X}_j - \mathbf{X}_{-i,-j} \widehat{\boldsymbol{\eta}} \right) = \mathbf{e}_i^t \mathbf{X}_j - \mathbf{e}_i^t \mathbf{X}_{-i,-j} \widehat{\boldsymbol{\eta}} = \mathbf{0}.$$

This relationship between the residuals will still hold if more variables are residualized; that is, the degree of multicollinearity will continue decreasing. Note that if the process is repeated $p-1$ times, all the variables of the model will be orthogonal between them. Interested readers should consult [53] for further information on successive residualization.

6. Empirical application

The methodology proposed herein is applied to two different models in relation to economic-financial and ecological data sets. The first example is focused on the application of residualization when the main goal of the researcher is mitigation of the collinearity and also presents a new interpretation of the modified variable. The second example shows the application of residualization when the major purpose of the study is to obtain new interpretations of the variables. The use of different models from different fields is motivated by the challenge of demonstrating the relevance and application of the method in real-world examples across a wide range of disciplines.

6.1. The economic-financial model

The first empirical application is based on a model developed by Wooldridge ([63]) with interest rate data from the market yields published by Salomon Brothers in *An Analytical Record of Yields and Yield Spreads* in the 1990s but modified in this paper for the period June 2008 to April 2019 (end-of-month data) by using a dataset from the U.S. Department of the Treasury:

$$\mathbf{c}_{52} = \beta_1 + \beta_2 \; \mathbf{c}_{13} + \beta_3 \; \mathbf{c}_{26} + \mathbf{u}, \tag{19}$$

where $\mathbf{u}$ is the random disturbance, which is spherical, and $\mathbf{c}_{13}$, $\mathbf{c}_{26}$ and $\mathbf{c}_{52}$ represent the coupon equivalents for different maturity tranches: 13 weeks, 26 weeks, and 52 weeks, respectively. The coupon equivalent, also called the bound equivalent or the investment yield, is the bill's yield based on the purchase price, discount, and a 365/366-day year. The coupon equivalent can be used to compare the yield on a discount bill to the yield on a nominal coupon bond that pays semiannual interest.

The following matrix represents the correlation between the variables:

$$\begin{array}{c} \\ \mathbf{c}_{52} \\ \mathbf{c}_{13} \\ \mathbf{c}_{26} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{c}_{52} & \mathbf{c}_{13} & \mathbf{c}_{26} \end{array} \\ \begin{pmatrix} 1.000 & & \\ 0.981 & 1.000 & \\ 0.995 & 0.993 & 1.000 \end{pmatrix} \end{array}.$$

From this matrix, it is observed that all explanatory variables are positively related to the dependent variable. In addition, the coefficient of correlation between explanatory variables $\mathbf{c}_{13}$ and $\mathbf{c}_{26}$ is equal to 0.993, which can serve as a first indication of the dependence between them. From this coefficient of correlation, it is determined that the VIF is equal to 71.516, while the condition number is equal to 23.233. Both measures confirm the existence of worrisome near-collinearity in this model. In relation to nonessential collinearity, the coefficients of variation ($CV(\mathbf{c}_{13}) = 1.486997$ and $CV(\mathbf{c}_{26}) = 1.280229$) are higher than the threshold established by [56], and consequently, it is possible to conclude that there is no relation between the intercept and any of the independent variables.

Table 3 presents the results of the estimation by ordinary least squares (OLS), ridge regression and residualization. As seen from the OLS results, the negative values of the estimated parameter $\widehat{\beta}_2$ do not make economic sense, particularly when taking into account the matrix of correlation previously presented that suggests the existence of a direct relation between the explanatory and the dependent variables.

Table 3. Results of Wooldridge model.

		OLS	Ridge	Residualization
Intercept	$\widehat{\beta}_1$	0.050 **	0.058	0.183 **
	(s.d.)	(0.008)	n/a	(0.007)
$\mathbf{c}_{13}$	$\widehat{\beta}_2$	-0.557 **	-0.464	1.067 **
	(s.d.)	(0.065)	n/a	(0.008)
$\mathbf{c}_{26}$	$\widehat{\beta}_3$	1.562 **	1.473	
	(s.d.)	(0.063)	n/a	
$\mathbf{e}_{\mathbf{c}_{26}}$	$\widehat{\beta}_{3_0}$			1.562 **
	(s.d.)			(0.063)
R^2		0.9935	n/a	0.9935
F statistic		9794	n/a	9794
p-value (of F)		$< 2.2 \cdot 10^{-16}$	n/a	$< 2.2 \cdot 10^{-16}$

**, * mean the coefficient is statistically significant at 0.01 (99% level of confidence) and at 0.05 (95% level of confidence), respectively.

Table 4. Results of Wooldridge model: Shapley values.

	$\mathbf{c}_{13}$	$\mathbf{c}_{26}$	R^2
Shapley value (OLS)	0.4828	0.5107	0.9935
Share (% of R^2)	48.594%	51.406%	100%

This nonexpected sign can be a consequence of the presence of collinearity. For this reason, other methodologies of estimation will be applied.

To apply residualization, the first step will be to select the variable to be residualized. In this case, it is possible to consider that the medium term, $\mathbf{c}_{26}$, includes in some way the short term, $\mathbf{c}_{13}$. From the Shapley values (Table 4), it can also be seen that it contributes more to the R^2 of the model. Additionally, the residuals obtained from the auxiliary regression between $\mathbf{c}_{26}$ and $\mathbf{c}_{13}$ will be the part of $\mathbf{c}_{26}$ that is not explained by $\mathbf{c}_{13}$. That is, the residuals will represent the second 13-week period. This fact could indicate that the variable selected to be residualized is $\mathbf{c}_{26}$. Note that residualization provides a new interpretation that is not possible to obtain from the initial model.

When residualization is applied, the value of the estimated parameter for $\mathbf{c}_{13}$ becomes positive. Additionally, all the estimated coefficients are individually significant with a level of confidence of 99%, the coefficient of determination of the original model is maintained, and the essential near-collinearity problem is mitigated (the lowest possible value for the VIF, 1.000, is obtained). In this case, the condition number is equal to 1.878, which indicates the relationship between $\mathbf{c}_{13}$ and the intercept, reflecting that this relationship is not worrisome. Finally, it is important to remark that the estimated variances of the estimators have been diminished, except in the case of the residualized variable, which remains constant.

It may be interesting to compare these results with the results obtained from ridge regression, which is widely applied to estimate models with collinearity. For this, a value of k was selected that results in a VIFvalue less than 10 (VIF = 9.948, calculated by following [55]), which is $k = 0.047$; see Table 3. Note that in this case, the signs obtained are not the expected ones. Furthermore, ridge regression does not allow drawing any conclusion about the global characteristics of the model, the individual significance, the inference or the analysis of the isolated effect of both variables, $\mathbf{c}_{13}$ and $\mathbf{c}_{26}$.

In addition, Figure 1 compares the estimations in terms of essential near-collinearity (the nonessential collinearity is not analyzed since it is not worrisome in this model), showing that the VIF value for ridge regression for different values of k is always lower than that of the OLS estimation but higher than the VIF obtained by residualization.

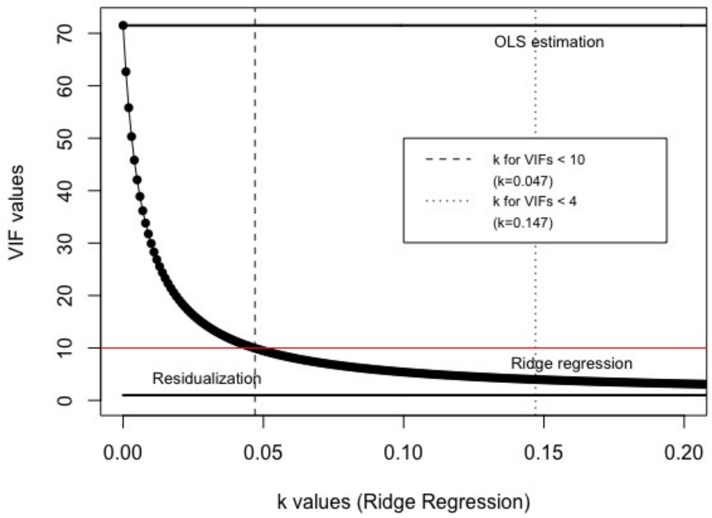

Figure 1. VIF values obtained by ridge regression of the Wooldridge model (monthly data). k in steps of 0.001.

6.2. The ecological model

The second example is based on the STIRPAT model that is usually applied in environmental economics to observe the influence of some social and economic variables on the atmospheric impact of a country or a group of countries. It can be defined as the stochastic version of the IPAT identity ([11, 12]), which identifies the impact on the atmosphere as a function of the population, the affluence measured from the gross domestic product (GDP), and some technology variable(s) (usually some variable(s) related to industries; see, for example, [41] or [42]). Ehrlich and Holdren, who were the first researchers to study the IPAT identity ([15–17]), were aware of the problem of the relationship between variables in this identity, and although collinearity is likely to appear, most STIRPAT applications have disregarded it (see, for example, [2, 9, 10, 13, 24, 33, 42, 45, 46, 48–50, 60]). However, there have been some efforts to address collinearity in STIRPAT models (see [14, 18, 20, 27, 37, 41, 51, 61]).

This paper applies the STIRPAT model to data from China (1990-2014), the most pollutant country in the world, as revealed by the World Bank, with a CO_2 emissions value of 10291926.878 kilotonnes (kt) in 2014. The traditional specification of the STIRPAT model is:

$$\mathbf{I} = \beta_1 + \beta_2 \ \mathbf{P} + \beta_3 \ \mathbf{A}_{pc} + \beta_4 \ \mathbf{T} + \mathbf{u}, \tag{20}$$

where $\mathbf{u}$ is the random disturbance, which is spherical; $\mathbf{I}$ represents CO_2 emissions; $\mathbf{P}$ the total of population (billions); variable $\mathbf{A}_{pc}$ is the per capita GDP (expressed in trillion constant 2010 US$); and finally, $\mathbf{T}$ is industrialization (% of GDP). The dataset has been extracted from the World Bank. However, in this paper, the following specification is proposed:

$$\mathbf{I} = \beta_1 + \beta_2 \ \mathbf{P} + \beta_3 \ \mathbf{e_A} + \beta_4 \ \mathbf{T} + \mathbf{u}, \tag{21}$$

where $\mathbf{A}$ is the GDP (expressed in trillion constant 2010 US$) and $\mathbf{e_A}$ are the residuals of the following auxiliary regression:

$$\mathbf{A} = \alpha_1 + \alpha_2 \ \mathbf{P} + \alpha_3 \ \mathbf{T} + \mathbf{v}. \tag{22}$$

The use of model (21) instead of model (20) intends to overcome the following disadvantages:

- Traditionally, per capita GDP has been used to avoid the existing dependency between the GDP and the population. However, as the reader will see with the following correlation matrix, the linear relationship between per capita GDP and population is higher than the relationship between GDP and population. This means that the linear relationship is not mitigated but increased.

$$\begin{array}{c} \\ \mathbf{I} \\ \mathbf{P} \\ \mathbf{A}_{pc} \\ \mathbf{A} \\ \mathbf{T} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{I} & \mathbf{P} & \mathbf{A}_{pc} & \mathbf{A} & \mathbf{T} \end{array} \\ \left(\begin{array}{ccccc} 1.0000 & & & & \\ 0.8896 & 1.0000 & & & \\ 0.9910 & 0.9111 & 1.0000 & & \\ 0.9905 & 0.9081 & 0.9999 & 1.0000 & \\ -0.5464 & -0.6194 & -0.6296 & -0.6320 & 1.0000 \end{array} \right) \end{array}$$

- In STIRPAT studies, $\mathbf{A}_{pc}$ (per capita GDP) is usually taken as the variable that represents affluence. However, the use of $\mathbf{A}_{pc}$ presents a disadvantage. From an interpretative point of view, a very important issue of the economy is ignored when using the per capita GDP (the ratio between GDP and population) since the distribution of income and the level of development of each region of the country are disregarded when all people are considered equal in terms of earnings. An increase in the per capita GDP does not necessarily mean the country is more developed; it may also indicate that the richest people in the country have increased their income.

In the residualization procedure, on the one hand, the relationship between GDP and population is deleted (in this case, the relationship between GDP and industrialization, variable $\mathbf{T}$, is also deleted), and, on the other hand, it is assumed that all people have the same income. Indeed, $\mathbf{e_A}$ coincides with the part of GDP that has no relationship with population and industrialization, as has been discussed. If $\mathbf{A}_{pc}$ could be interpreted as a tool to measure the enrichment of the people and not the enrichment of the country, $\mathbf{e_A}$ would be interpreted as a tool that measures whether the countries, and not the people, are richer in economic terms that are unrelated to industrialization.

Furthermore, in model (21), nonessential collinearity represents a formidable issue for this empirical example (see [56]) since $CV(\mathbf{P}) = 0.053$ and $CV(\mathbf{T}) = 0.026$. To mitigate the nonessential collinearity, the variables population and technology will be centered. Thus, the following model is proposed:

$$\mathbf{I} = \beta_1 + \beta_2 \ \mathbf{P}^* + \beta_3 \ \mathbf{e_A} + \beta_4 \ \mathbf{T}^* + \mathbf{u}, \tag{23}$$

Table 5. Results of STIRPAT models (20) and (23).

	Model (20)		Model (23)
Intercept	-10287191 *	Intercept	5405875 ***
	(3667784)		(53166)
$\mathbf{P}$	-1790259	$\mathbf{P}^*$	35883988 ***
	(1861596)		(1004251)
$\mathbf{A}_{pc}$	1837647	$\mathbf{e_A}$	1300875 ***
	(77813)		(57278)
$\mathbf{T}$	409211 ***	$\mathbf{T}^*$	24250
	(80784)		(82153)
R^2	0.9924		0.9918
F statistic	918.2		851.2
p-value (of F)	$< 2.2 \cdot 10^{-16}$		$< 2.2 \cdot 10^{-16}$

***, **, * mean the coefficient is statistically significant at 0.001 (99.9% level of confidence), at 0.01 (99% level of confidence) and at 0.05 (95% level of confidence), respectively.

where $\mathbf{P}^* = \mathbf{P} - \overline{\mathbf{P}}$ and $\mathbf{T}^* = \mathbf{T} - \overline{\mathbf{T}}$.

With this model, it is verified that the values of VIF ($\text{VIF}_{\mathbf{P}} = 1.622$, $\text{VIF}_{\mathbf{e_A}} = 1.000$ and $\text{VIF}_{\mathbf{T}} = 1.622$) are lower than the threshold of 10 (strong collinearity), and the same is true for CN, whose value is 2.063 (which is lower than 30). This means that the degree of the existing near-multicollinearity is not worrisome.

The results obtained by using OLS estimation of models (20) and (23) are shown in Table (5). The reader will observe the following:

- In model (20), the intercept has a coefficient that is significantly different from zero and has a negative value. This means that if population and GDP were null, the CO_2 emissions would be negative. This situation is corrected with model (23).
- In model (20), the estimated coefficient for population is not significantly different from zero; by contrast, in model (23), the coefficient is significant, and it has a positive value. This means that when the population increases, the CO_2 emissions also increase. This is in line with economic theory and the correlation matrix.
- In models (20) and (23), the GDP coefficient (obtained from $\mathbf{A}_{pc}$ and $\mathbf{e_A}$, respectively) is significantly different from zero and has a positive value. However, the interpretations of the two estimated coefficients are different. Thus, in model (23), the conclusion is that the increase in the wealth of the country when the production of goods and services is unrelated to industrialization supposes an increase in the CO_2 emissions.
- In model (20), the estimated coefficient for industrialization is significant and has a positive value. It is contrary to the sign expected by observing the correlation matrix; however, this situation is corrected by model (23), in which this coefficient is not significantly different from zero.

7. Conclusions

The estimation and inference of the multiple linear regression model estimated by residualization procedure is exhaustively developed in this paper and the results are compared with those of the original model. In this sense, it is important to point out that the application of residualization leads to conclusions about a model different to the original even though they have several identical characteristics (such as the variance estimation of the random perturbation, the coefficient of determination or the significance statistic).

The main contributions when applying this technique are:

- The new interpretations of the coefficients. The residualized model can answer questions that could not be answered with the initial model.
- The possibility of reducing the degree of collinearity in the initial model.

This paper proposes different criteria to select the variable to be residualized and it is also revealed the option of a successive residualization.

Finally, it is relevant to note that residualization is not always applicable because the interpretations of the new estimated coefficients are not always simple. For this reason, recommendations can be made to apply other **well known techniques, such as ridge regression, principal component regression or partial least squares regression. En este sentido se ha comprobado mediante a Monte Carlo simulation que dichas técnicas tienen una mejor capacidad predictiva que the residualization method, si bien hay que tener en cuenta que la interpretación de los coeficientes en estas técnicas alternativas es más controvertida si cabe que en the residualization method (PONER REFERENCIAS). Por tal motivo, se podría usar otra técnica alternativa como** raise regression (see [22] and [23]).

References

[1] B. Ambridge, J. Pine, and C. Rowland, *Semantics versus statistics in the retreat from locative overgeneralization errors*, Cognition 123 (2012), pp. 260–279.

[2] M. Azam and A. Khan, *Testing the environmental kuznets curve hypothesis: A comparative empirical study for low, lower middle, upper middle and high income countries*, Renewable and Sustainable Energy Reviews 63 (2016), pp. 556–567.

[3] G. Baird and S. Bieber, *The Goldilocks Dilemma: Impacts of Multicollinearity. A Comparison of Simple Linear Regression, Multiple Regression, and Ordered Variable Regression Models*, Journal of Modern Applied Statistical Methods 15 (2016), p. 18.

[4] N. Bandelj and M. Mahutga, *How socio-economic change shapes income inequality in post-socialist Europe*, Social Forces 88 (2010), pp. 2133–2161.

[5] D. Belsley, *Conditioning diagnostics: Collinearity and weak data in regression*, John Wiley, New York, 1991.

[6] Y. Bradshaw, *Urbanization and underdevelopment: A global study of modernization, urban bias, and economic dependency*, American Sociological Review 52 (1987), pp. 224–239.

[7] A. Buse, *Brickmaking and the collinear arts: a cautionary tale*, Canadian Journal of Economics (1994), pp. 408–414.

[8] A. Cohen-Goldberg, *Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production*, Journal of Memory and Language 67 (2012), pp. 184–198.

[9] D. Coondoo and S. Dinda, *Causality between income and emission: a country group-specific econometric analysis*, Ecological Economics 40 (2002), pp. 351–367.

[10] S. De Bruyn, J. Van Den Bergh, and J. Opschoor, *Economic growth and emissions: reconsidering the empirical basis of environmental Kuznets curves*, Ecological Economics 25 (1998), pp. 161–175.

[11] T. Dietz and E. Rosa, *Rethinking the environmental impacts of population, affluence and technology*, Human Ecology Review 1 (1994), pp. 277–300.

[12] T. Dietz and E. Rosa, *Effects of population and affluence on CO_2 emissions*, Proceedings of the National Academy of Sciences of the USA 94 (1997), pp. 175–179.

[13] M. Disli, A. Ng, and H. Askari, *Culture, income, and CO_2 emission*, Renewable and

Sustainable Energy Reviews 62 (2016), pp. 418–428.
[14] J. Dong, C. Deng, R. Li, and J. Huang, *Moving Low-Carbon Transportation in Xinjiang: Evidence from STIRPAT and Rigid Regression Models*, Sustainability 9 (2016), p. 24.
[15] P. Ehrlich and J. Holdren, *The people problem*, Saturday Reveiw 4 (1970), pp. 42–43.
[16] P. Ehrlich and J. Holdren, *The Impact of Population Growth*, Science 171 (1971), pp. 1212–1217.
[17] P. Ehrlich and J. Holdren, *A bulletin dialogue on the "Closing Circle": Critique: One dimensional ecology*, Bulletin of the Atomic Scientists 28 (1972), pp. 16–27.
[18] Y. Fan, L. Liu, and Y. Wei, *Analyzing impact factors of CO_2 emissions using the STIRPAT model*, Environmental Impact Assessment Review 26 (2006), pp. 377–395.
[19] D. Farrar and R. Glauber, *Multicollinearity in regression analysis: the problem revisited*, The Review of Economic and Statistics (1967), pp. 92–107.
[20] Y. Fernández, M. Fernández, D. González, and B. Olmedillas, *El efecto regulador de los Planes Nacionales de Asignación sobre las emisiones de CO_2*, Revista de Economía Mundial 40 (2015), pp. 47–66.
[21] C. García, R. Salmerón, and C. García, *Choice of the ridge factor from the correlation matrix determinant*, Journal of Statistical Computation and Simulation 89 (2019), pp. 211–231.
[22] C. García, J. García, and J. Soto, *The raise method: An alternative procedure to estimate the parameters in presence of collinearity*, Quality & Quantity 45 (2011), pp. 403–423.
[23] J. García, R. Salmerón, C. García, and M. López-Martín, *The raise estimators. estimation, inference and properties*, Communications in Statistics-Theory and Methods 46 (2017), pp. 6446–6462.
[24] M. Gassebner, M. Lamla, and J. Sturm, *Determinants of pollution: what do we really know?*, Oxford Economic Papers 63 (2011), pp. 568–595.
[25] D. Gujarati, *Basic Econometrics*, 4th ed., McGraw-Hill, 2004.
[26] T. Jaeger, *Redundancy and reduction: Speakers manage syntactic information density*, Cognitive psychology 61 (2010), pp. 23–62.
[27] J. Jia, H. Deng, J. Duan, and J. Zhao, *Analysis of the major drivers of the ecological footprint using the STIRPAT model and the PLS method. A case study in Henan Province, China*, Ecological Economics 68 (2009), pp. 2818–2824.
[28] A. Jorgenson, *Global warming and the neglected greenhouse gas: A cross-national study of the social causes of methane emissions intensity, 1995*, Social Forces 84 (2006), pp. 1779–1798.
[29] A. Jorgenson and T. Burns, *The political-economic causes of change in the ecological footprints of nations, 1991–2001: a quantitative investigation*, Social Science Research 36 (2007), pp. 834–853.
[30] A. Jorgenson and B. Clark, *The economy, military, and ecologically unequal exchange relationships in comparative perspective: a panel study of the ecological footprints of nations, 1975?2000*, Social Problems 56 (2009), pp. 621–646.
[31] P. Kennedy, *A guide to Econometrics (3rd ed.)*, MIT Press, 1992.
[32] J. Kentor and E. Kick, *Bringing the military back in: Military expenditures and economic growth 1990 to 2003*, Journal of World-Systems Research 14 (2008), pp. 142–172.
[33] S. Kumar, *Environmentally sensitive productivity growth: A global analysis using Malmquist-Luenberger index*, Ecological Economics 56 (2006), pp. 280–293.
[34] V. Kuperman, R. Bertram, and R. Baayen, *Morphological dynamics in compound processing*, Language and Cognitive Processes 23 (2008), pp. 1089–1132.
[35] V. Kuperman, R. Bertram, and R. Baayen, *Processing trade-offs in the reading of Dutch derived words*, Journal of Memory and Language 62 (2010), pp. 83–97.
[36] K. Lemhöfer, T. Dijkstra, H. Schriefers, R. Baayen, J. Grainger, and P. Zwitserlood, *Native language influences on word recognition in a second language: A megastudy*, Journal of Experimental Psychology: Learning, Memory, and Cognition 34 (2008), p. 12.
[37] S. Lin, D. Zhao, and D. Marinova, *Analysis of the environmental impact of China based on STIRPAT model*, Environmental Impact Assessment Review 29 (2009), pp. 341–347.

[38] M. Mahutga and N. Bandelj, *Foreign investment and income inequality: The natural experiment of Central and Eastern Europe*, International Journal of Comparative Sociology 49 (2008), pp. 429–454.

[39] D. Marquardt, *A Critique of Some Ridge Regression Methods: Comment*, Journal of the American Statistical Association 75 (1980), pp. 87–91.

[40] D. Marquardt and S. Snee, *Ridge Regression in Practice*, Journal of the American Statistical Association 29 (1975), pp. 3–20.

[41] I. Martínez-Zarzoso, A. Bengochea-Morancho, and R. Morales-Lage, *The impact of population on CO_2 emissions: evidence from European countries*, Environmental and Resource Economics 38 (2007), pp. 497–512.

[42] I. Martínez-Zarzoso and A. Maruotti, *The impact of urbanization on CO_2 emissions: Evidence from developing countries*, Ecological Economics 70 (2011), pp. 1344–1353.

[43] A. Novales, *Econometría*, McGraw-Hill, Madrid, 1988.

[44] A. Novales, R. Salmerón, C. García, J. García, and M. López-Martín, *Tratamiento de la multicolinealidad aproximada mediante variables ortogonales*, Anales de Economía Aplicada. XXIX Congreso Internacional de Economía Aplicada (2015), pp. 1212–1227.

[45] M. Pablo-Romero and J. De Jesús, *Economic growth and energy consumption: The Energy-Environmental Kuznets Curve for Latin America and the Caribbean*, Renewable and Sustainable Energy Reviews 60 (2016), pp. 1343–1350.

[46] H. Pao and C. Tsai, *CO_2 emissions, energy consumption and economic growth in BRIC countries*, Energy Policy 38 (2010), pp. 7850–7860.

[47] R. Paul, *Multicollinearity: Causes, effects and remedies*, M. Sc. (Agricultural Statistics), Roll No. 4405 (2006). IASRI, New Delhi.

[48] A. Rafindadi, *Revisting the concept of environmental Kuznets curve in period of energy disaster and deteriorating income: Empirical evidence from Japan*, Energy Policy 94 (2016), pp. 274–284.

[49] J. Roberts and P. Grimes, *Carbon Intensity and Economic Development 1962-1991: A Brief Exploration of the Environmental Kuznets Curve*, World Development 25 (1997), pp. 191–198.

[50] J. Roca and E. Padilla, *Emisiones atmosféricas y crecimiento económico en España: la curva de Kuznets ambiental y el Protocolo de Kyoto*, Economía Industrial 351 (2003), pp. 73–86.

[51] M. Roy, S. Basu, and P. Pal, *Examining the driving forces in moving toward a low carbon society: an extended STIRPAT analysis for a fast growing vast economy*, Clean Technologies and Environmental Policy 19 (2017), pp. 2265–2276.

[52] R. Salmerón, C. García, and J. García, *Variance Inflation Factor and Condition Number in multiple linear regression*, Journal of Statistical Computation and Simulation 88 (2018), pp. 2365–2384.

[53] R. Salmerón, J. García, C. García, and C. García, *Treatment of collinearity through orthogonal regression: an economic application*, Boletín de Estadística e Investigación Operativa 32 (2016), pp. 184–202.

[54] R. Salmerón, J. García, C. García, and M. López-Martín, *Transformation of variables and the condition number in ridge estimation*, Computational Statistics 33 (2018), p. 14971524.

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

[56] R. Salmerón, A. Rodríguez, and C. García, *Diagnosis and quantifcation of the non-essential collinearity*, Computational Statistics (2019), p. In review.

[57] L. Shapley, *A value for n-person games*, Contributions to the Theory of Games (AM-28) 2 (2016), p. 307.

[58] S. Silvey, *Multicollinearity and imprecise estimation*, Journal of the Royal Statistical Society. Series B (Methodological) 31 (1969), pp. 539–552.

[59] R. Snee and D. Marquardt, *Comment: Collinearity diagnostics depend on the domain of*

prediction, the model, and the data, The American Statistician 38 (1984), pp. 83–87.

[60] M. Torras and J. Boyce, *Income, inequality, and pollution: a reassessment of the environmental Kuznets Curve*, Ecological Economics 25 (1998), pp. 147–160.

[61] G. Uddin, K. Alam, and J. Gow, *Estimating the major contributors to environmental impacts in Australia*, International Journal of Ecological Economics and Statistics 37 (2016), pp. 1–14.

[62] J. Walton and C. Ragin, *Global and national sources of political protest: Third world responses to the debt crisis*, American Sociological Review 55 (1990), pp. 876–890.

[63] J. Wooldridge, *Introduccin a la econometría. Un enfoque moderno*, 2nd ed., Thomson Paraninfo, Madrid, 2008.

[64] B. Woolf, *Computation and interpretation of multiple regressions*, Journal of the Royal Statistical Society. Series B (Methodological) (1951), pp. 100–119.

[65] L. Wurm and S. Fisicaro, *What residualizing predictors in regression analyses does (and what it does not do)*, Journal of Memory and Language 72 (2014), pp. 37–48.

[66] R. York, *Residualization is not the answer: Rethinking how to address multicollinearity*, Social science research 41 (2012), pp. 1379–1386.