ANALYSING THE INTERACTION BETWEEN THE PIB AND THE EDUCATION TO EXPLAIN THE CO2 EMISSIONS

Catalina García García – Claudia García García – Román Salmerón Gómez-José García Pérez

Abstract

Different methodologies have been developed to analyze the environmental degradation that can be summarized in identity and econometrics techniques. Regardless of the methodology applied, the main factors that are traditionally used to explain the environmental impact are the population, affluence and technology. In this paper, we introduce the level of education (a dummy variable created from national learning achievement scores for PISA 2012) as a main factor that affects to the emissions of CO_2. Indeed, we analyze the interaction of this variable with the PIB. The introduction of interactions between independent variables in a multiple linear regression is a specific application called moderated regression analysis. As is expected, this kind of models usually presents collinearity. We propose to use the raise regression to mitigate collinearity while maintains the global and individual significance of the initial model. The contribution will be illustrated with an empirical approach using environmental data, relating CO2 emissions of 21 countries of the EU-28 for 2013.

Key words: collinearity, moderated regression, interaction, raised regression, CO2 emissions

JEL Code: C51, O52, Q53

1 Introduction

The United Nations Conference on Environment and Development, held in Stockholm on 5 to 16 June 1972, can be considered the first proof of the global dimension of climate problems. However, the climate change is not recognized until the United Nations Framework Convention on Climate Change, adopted in May 1992, signed by more than 150 countries at the Earth Summit in Rio de Janeiro. With this landmark, developed countries begin to implement internal policies trying to reduce their emissions to the atmosphere. To make the correct policies, it is necessary to know what aspects of the economy cause pollution problems, or in other words, what are the variables that affect the emissions to the atmosphere.

Different methodologies have been developed with this purpose that can be summarized in identity and econometrics techniques. As regards the first methods, one important and recognized formula is the IPAT identity, presented by Ehrlich and Holdren (1970). This identity considers that the environmental impact can be decomposed in population, affluence and technology. It is possible to find alternative versions of this identity in the scientific literature such as the IPBAT that includes the Behavior (Schulze, 2002) and the ImPACT that divides Technology in two components (Waggoner and Ausubel, 2002).

Another important methodology traditionally applied is the STIRPAT that is considered as the stochastic model of the IPAT identity, Dietz and Rosa (1994). From the original version, different works have proposed incorporating new variables to this methodology: Fan et al. (2006) incorporated the variables of the percentage of the population aged 15-64 and the proportion of the population living in the urban areas. This last variable is also applied by Jia et al. (2009) in conjunction with population, GDP per capita, the quadratic terms of GDP per capita and the percentage of the economy not included in the service sector and the share of the urban population in the total population.

This paper incorporates the education as variable to explain the emissions of CO2 by following many different works that note the variable education is a main factor to explain the attitudes towards preventing environmental damage (Bord and O'Connor, 1997; Engel and Pötschke, 1998; Torgler and Garcia-Valiñas, 2007) . For this purpose, it is elaborated a dichotomy variable that measures the level of education from PISA 2012 (Programme for International Student Assessment) taking the value 1 if the country is better or equal than the European Union, and 0 in other case.

In addition, this paper uses the variable gross domestic product (GDP) as independent variable. The goal is not only to know the influence of the level of education on the environmental impact but also to analyze if the GDP has a different influence depending on the level of education. Thus, the variable interaction between GDP and education is introduced in the model as independent variable. This fact leads to collinearity that will be mitigated by the raise regression, García et al. (2010), and the regression with orthogonal variables, Novales et al. (2015).

The structure of the paper is as follows: Section 2 presents the methodology that will be used to mitigate the collinearity. The contribution of this paper is illustrated in the empirical application presented in Section 3. Finally, main conclusions are summarized in Section 4.

2 Methodology

The inclusion of an interaction term is a specific application of multiple linear regression analysis known as Moderated Regression Analysis (MRA), Hartmann and Moers (1999). This methodology allows to analyze how the effect of one of the independent variables ($\mathbf{x}_1$) is moderated by a second independent variable ($\mathbf{x}_2$) by adding a cross product term between them ($\mathbf{x}_1 \times \mathbf{x}_2$), as an additional explanatory variable, , defining, for example, the following model:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_1 \times \mathbf{x}_2 + \mathbf{u}$$

The introduction of interactions between independent variables in a multiple linear regression can lead to problems of collinearity. Numerous techniques have been developed to estimate a model with collinearity. The ridge estimation (RE) (Hoerl and Kennard, 1970) is one of the most applied but there is still an open debate about the justification of its procedure and the interpretation of its estimators. Apart of this, what happens is that the experimental F diminishes and the model may become globally insignificant as the ridge factor (k) increases. Thus, although the collinearity is mitigated, the model is neither globally or individually significant, Salmerón et al (2017).

This paper proposes to use the raise and the regression with orthogonal variables to mitigate collinearity while maintains the global and individual significance of the initial model. The raise regression was presented and developed by García et al. (2010) and Salmerón et al. (2017) and part from the following model:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u} \qquad (1)$$

Where the variables $\mathbf{x}_1$ and $\mathbf{x}_2$ presents collinearity between them. Geometrically, it can be represented as both vectors ($\mathbf{x}_1$ and $\mathbf{x}_2$) are geometrically close. It is to say, the angle that separate both vector is very small. See Figure 1.

The raise methodology proposes to use the following auxiliary regression:

$$\mathbf{x}_1 = \alpha_0 + \alpha_1 \boldsymbol{x}_2 + \boldsymbol{v} \tag{2}$$

The residuals $\boldsymbol{e}_1$ will be obtained from the estimation by ordinary least squares of expression (2) and the raised variable ($\tilde{\mathbf{x}}_1$) will be calculated as:

$$\tilde{\mathbf{x}}_1 = \boldsymbol{x}_1 + \lambda \boldsymbol{e}_1 \tag{3}$$

By substituting the variable $\mathbf{x}_1$ by the raised variable $\tilde{\mathbf{x}}_1$ in the original model (1), both variables will be separated geometrically and the collinearity will be mitigated.

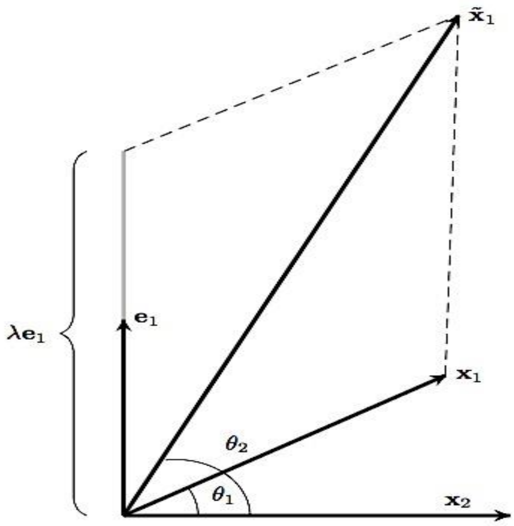

Fig. 1: Raise regression

On the other hand, the regression with orthogonal variables was presented for two variables by Novales et al. (2015). This methodology parts from the original model (1) and proposes to use an auxiliary regression similar to (2). Thus, the residuals of the auxiliary regression ($\boldsymbol{e}_1$) could be interpreted as the part of variable $\mathbf{x}_1$ not related to $\mathbf{x}_2$. The regression with orthogonal variables proposes to isolate the effect of variable $\mathbf{x}_1$ substituting it in model (1) by the residuals $\boldsymbol{e}_1$.

3. Empirical application

3.1. Data

We use a data set with information about the 28 countries of the European Union in 2013. The variables that we have considered important for our analysis are the following:

- Emissions (E): Greenhouse gas emissions in CO2 equivalent (tonnes). Source: European Environment Agency (EEA).

- GDP (G): Gross Domestic Product at current PPP (euros). Source: EUROSTAT.
- Education (Ed): This is a dummy variable created from PISA 2012 (Programme for International Student Assessment). Source: Own elaboration. We use the national learning achievement scores for PISA 2012. Note that Malta does not appear but we consider it like similar countries such as Spain, Greece, Italy or Portugal. To construct this variable, we have considered that the level of a country is better than the European Union average if:
  a) Its marks in two of the three subjects analysed are higher than in the EU-28.
  b) The mean of the three subjects analysed is higher than in the EU-28.

So, this dummy variable takes the value 1 if the country is better or equal than the European Union, and 0 in other case. We have the same values for each country by the two ways: a) and b) (see Table 1 in Appendix).

3.2. Estimations

From the above data, the following model is estimated:

$$\ln \boldsymbol{E}_i = \beta_0 + \beta_1 \ln \boldsymbol{G}_i + \beta_2 \boldsymbol{Ed}_i + \beta_3 \boldsymbol{I}_i + \boldsymbol{u}_i \qquad (4)$$

where $\boldsymbol{I}_i$ is the interaction variable obtained as $\ln \boldsymbol{G}_i \times \boldsymbol{Ed}_i$. Table 2 presents the estimation of the model by ordinary least squares (OLS).

Tab. 2: Estimation of the original model

	Intercept	GDP ($\boldsymbol{G}_i$)	Education ($\boldsymbol{Ed}_i$)	Interaction ($\ln \boldsymbol{G}_i \times \boldsymbol{Ed}_i$)
Coefficient p-value VIF	−2.93815 0.3456	0.8018937 <0.0001*** 2.326	−0.383708 0.9296 304.174	0.0251406 0.8813 317.274
R^2	0.826687			
p-value F_{exp}	2.71e-09			

In relation to the interpretation of coefficients, a level of education lower than the mean leads to diminishing the emissions 0.383708%, the effect of an increment of 1% in the GDP is an increase of 0.8018937% in the emissions. For countries with a level of education higher than the mean, this effect is increased in 0.0251406%. However, note than only the variable GDP is individually significant although the coefficient of determination is high and the model is globally significant. These are the common symptoms of a model with collinearity. This fact is confirmed with the variance inflator factors that are higher than 10.

For this reason, the model is estimated with other alternative methodologies as the raise regression and the regression with orthogonal variables.

With this purpose, firstly it is estimated the following auxiliary regression:

$$\ln \boldsymbol{G}_i = \alpha_0 + \alpha_1 \boldsymbol{Ed}_i + \alpha_2 \boldsymbol{I}_i + \boldsymbol{v}_i \tag{5}$$

Then, the raise variable is obtained as:

$$\ln \tilde{\boldsymbol{G}}_i = \ln \boldsymbol{G}_i + \lambda \boldsymbol{v}_i \tag{6}$$

This variable will be used in the original model (4) instead of the original variable $\boldsymbol{G}_i$, otbained the following model that is estimated by OLS:

$$\ln \boldsymbol{E}_i = \beta_0 + \beta_1 \ln \tilde{\boldsymbol{G}}_i + \beta_2 \boldsymbol{Ed}_i + \beta_3 \boldsymbol{I}_i + \boldsymbol{u}_i \tag{7}$$

Table 3 shows the estimations of the raise regression for different values of λ. Note that from $\lambda = 0.5$ all the independent variables start to be individually significant. The raise regression presents the characteristics of keeping the coefficient of determination and the global significance of the original model. However, even for values of λ higher than 500, the VIFs are still higher than 10. Thus, although the collinearity is still presents, it has been mitigated to not affect to the significance of the independent variables.

For this reason, it is selected the value of λ equal to 0.9 since from this value all the indepent variables are significant in a level of 99%. Then, the interpretation of the coefficients is that an increment of 1% in GDP leads to an increment of 0.422049 and this increment is also duplicated (+ 0.404985) to countries with a lever of education higher than the mean. However the intercept is 6.67153 for countries with a level of education lower than the mean and becomes -3.3216 for countries with a level of education higher than the mean.

The regression with orthogonal variables is obtained directly substituting the variable $\boldsymbol{G}_i$ in the original model (4) by the residuals obtained in the auxiliary regression (5):

$$\ln \boldsymbol{E}_i = \beta_0 + \beta_1 \boldsymbol{v}_i + \beta_2 \boldsymbol{Ed}_i + \beta_3 \boldsymbol{I}_i + \boldsymbol{u}_i \tag{8}$$

Tab. 3: Estimation of the raise model for different values of λ

		Intercept	GDP ($\tilde{G}_i$)	Education (Ed_i)	Interaction ($\ln G_i \times Ed_i$)
$\lambda = 0.1$	Coefficient	−1.09387	0.728994	−2.22799	0.0980400
	p-value	0.6971	<0.0001***	0.5924	0.5427
	VIF		2.096	277.609	288.436
$\lambda = 0.2$	Coefficient	0.443034	0.668244	−3.76489	0.158789
	p-value	0.8633	<0.0001***	0.3504	0.3085
	VIF		1.921	257.404	266.502
$\lambda = 0.3$	Coefficient	1.74349	0.616841	−5.06534	0.210193
	p-value	0.4656	<0.0001***	0.1985	0.1674
	VIF		1.785	241.681	249.432
$\lambda = 0.4$	Coefficient	2.85816	0.572781	−6.18002	0.254253
	p-value	0.2031	<0.0001***	0.1106	0.0893*
	VIF		1.677	229.204	235.888
$\lambda = 0.5$	Coefficient	3.82421	0.534596	−7.14607	0.292438
	p-value	0.0731*	<0.0001***	0.0618*	0.0478**
	VIF		1.589	219.139	224.961
$\lambda = 0.6$	Coefficient	4.66951	0.501183	−7.99137	0.325851
	p-value	0.0224**	<0.0001***	0.0351**	0.0261**
	VIF		1.518	210.901	216.018
$\lambda = 0.7$	Coefficient	5.41536	0.471702	−8.73721	0.355332
	p-value	0.0061**	<0.0001***	0.0205**	0.0146**
	VIF		1.459	204.074	208.607
$\lambda = 0.8$	Coefficient	6.07834	0.445496	−9.40019	0.381538
	p-value	0.0015***	<0.0001***	0.0123**	0.0085**
	VIF		1.409	198.352	202.396
$\lambda = 0.9$	Coefficient	6.67153	0.422049	−9.99338	0.404985
	p-value	<0.0001***	<0.0001***	<0.0001***	<0.0001***
	VIF				
$\lambda = 1$	Coefficient	7.20540	0.400947	−10.5273	0.426087
	p-value	<0.0001***	<0.0001***	<0.0001***	<0.0001***
	VIF		1.332	189.377	192.652
$\lambda = 2$	Coefficient	10.5866	0.267298	−13.9084	0.559736
	p-value	<0.0001***	<0.0001***	<0.0001***	<0.0001***
	VIF		1.147	168.118	169.574
$\lambda = 20$	Coefficient	16.3829	0.0381854	−19.7047	0.788849
	p-value	<0.0001***	<0.0001***	<0.0001***	<0.0001***
	VIF		1.003	151.458	151.458

$\lambda = 500$	Coefficient	17.3084	0.00160059	−20.6303	0.825433
	p-value	<0.0001***	<0.0001***	<0.0001***	<0.0001***
	VIF		1	151.112	151.112

Tab. 4: Estimation of the orthogonal model

	Intercept	GDP ($\tilde{G}_i$)	Education (Ed_i)	Interaction ($\ln G_i \times Ed_i$)
Coefficient	17.3489	0.801893	−20.6708	0.827034
p-value	<0.0001***	<0.0001***	<0.0001***	<0.0001***
VIF		1	151.111	151.111

Table 4 presents the estimation of the model by the regression with orthogonal variables. This methodology also maintains the coefficient of determination and the global significance of the original regression. Note in this case all the independent variables are individually significant at a level of 99%. However, the VIFs are still higher than 10. This results are similar to the one obtained for the raise regression with $\lambda = 500$, except for the coefficient of the orthogonalized variable. In this case, the value of the orthogonalized variable is similar to the original model, but the interpretation is different. An increment of 1% in the part of GDP not related with the level of education leads to an increase of 0.801893% in the emissions. In relation to the level of education, countries with a level higher to the mean will diminish the emissions 20.6708%. The interpretation of the variable interaction is that the effect of the GDP in countries with a level of education higher than the mean is positive.

4 Conclusion

This work has analyzed the effect of the GDP, the level of education and its interaction on the emissions of CO2 in European countries. The estimations obtained by OLS were unstable due to the existence of collinearity in the model and other alternative methodologies have been applied. Thus, the raise regression and the regression with orthogonal variables were applied. It is not possible to say that the collinearity was overpassed since none of the methodologies gets values of VIF lower than 10. However, the collinearity is mitigated enough to not affect to the significance of the variables leading to models where all the variables are significant in a 99%. In addition, both methodologies maintain the coefficient of determination and global significance of the original model. It was showed, from an empirical point of view, that the regression with orthogonal variables presents results similar to the one obtained from the raise regression for high values of λ, except for the orthogonalized variable.

There are also important conclusions from the interpretation of the estimated coefficients. In fact, it was shown that the positive effect of the GDP on the emissions is also duplicated for countries with a level of education higher than the mean. On the other hand, the sign of the intercept for countries with a level of education lower than the mean is positive while it is negative for countries with a level of education higher than the mean. Thus, although a high level of education increases the effect of the GDP, it also helps to reduce the emissions.

References

1. Bord, R.J. & O'Connor, R.E., (1997). The gender gap in environmental attitudes: the case of perceived vulnerability to risk. *Social Science Quarterly*, 78, 830–840.
2. Dietz, T. & Rosa, E. A. (1994). Rethinking the environmental impacts of population, affluence and technology. *Human Ecology Review*, 1, 277-300.
3. Ehrlich, P. R. & Holdren, J. P. (1970). The people problem. *Saturday review*, 4(42), 42-43.
4. Engel, U., & Pötschke, M., (1998). Willingness to pay for the environment: social structure, value orientations and environmental behaviour in a multilevel perspective. *Innovation*, 11 (3), 315–332.
5. Fan, Y., Liu, L.-C., Wu, G., & Wei, Y.-M. (2006). Analyzing impact factors of CO2 emissions using the STIRPAT model. *Environmental Impact Assessment Review*, 26(4), 377-395.
6. Garcia, C., García, J., & Soto, J. (2011). The raise method. An alternative procedure to estimate the parameters in presence of collinearity. *Quality & Quantity*, 45(2), 403-423.
7. Hartmann, F. G., & Moers, F. (1999). Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis. *Accounting, Organizations and Society*, 24(4), 291-315.
8. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
9. Jia, J., Deng, H., Duan, J., & Zhao, J. (2009). 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(11), 2818-2824.

10. Novales, A., Salmerón, R., García, C., García, J., & López, M. (2015). Tratamiento de la multicolinealidad aproximada mediante variables ortogonales. *Anales de Economía Aplicada*, 1212-1227.
11. Salmeron, R., Garcia, C., Garcia, J., & Lopez, M. D. M. (2017). The raise estimator estimation, inference, and properties. *Communications in Statistics-Theory and Methods*, 46(13), 6446-6462.
12. Schulze, P. C. (2002). I = IPBAT. Ecological Economics, 40(2):149, 150-15 8
13. OECD (2012). PISA 2012. European Union, Paris.
14. Torgler, B., & Garcia-Valiñas, M. A. (2007). The determinants of individuals' attitudes towards preventing environmental damage. *Ecological economics*, 63(2), 536-552.
15. Waggoner, P. E. & Ausubel, J. H. (2002). A framework for sustainability science: a renovated
IPAT identity. *Proceedings of the National Academy of Sciences*, 99(12), 7860-7865.

Appendix

Tab. 1: Definition of variable Education for EU-28

COUNTRY	Math	Reading	Science	Mean	Value of the variable (a)	Value of the variable (b)
UE-28	489	489	497	492	-	-
Austria	506	490	506	501	1	1
Belgium	515	509	505	510	1	1
Bulgaria	439	436	446	440	0	0
Croatia	471	485	491	482	0	0
Cyprus	440	449	438	442	0	0
Czech Republic	499	493	508	500	1	1
Denmark	500	496	498	498	1	1
Estonia	521	516	541	526	1	1
Finland	519	524	545	529	1	1
France	495	505	499	500	1	1
Germany	514	508	524	515	1	1
Greece	453	477	467	466	0	0

Hungary	477	488	494	486	0	0
Irland	501	523	522	515	1	1
Italy	485	490	494	490	0	0
Latvia	491	489	502	494	1	1
Lithuania	479	477	496	484	0	0
Luxemburg	490	488	491	490	0	0
Malta	-	-	-	-	0	0
Netherlands	523	511	522	519	1	1
Poland	518	518	526	521	1	1
Portugal	487	488	489	488	0	0
Romania	445	438	439	441	0	0
Slovakia	482	463	471	472	0	0
Slovenia	501	481	514	516	1	1
Spain	484	488	496	489	0	0
Sweden	478	483	485	482	0	0
United Kingdom	494	499	514	502	1	1

Source: Own elaboration from PISA 2012.

Contact

Catalina Garcia Garcia
University of Granada
Campus Universitario de La Cartuja 18071 Granada
cbgarcia@ugr.es

Claudia Garcia Garcia
University of Granada
Campus Universitario de La Cartuja 18071 Granada
claugar@correo.ugr.es

José García Pérez
University of Almería
Ctra. Sacramento s/n, La Cañada de San Urbano, 04120 Almería
jgarcia@ual.es

Román Salmerón Gomez
University of Granada
Campus Universitario de La Cartuja 18071 Granada
romansg@ugr.es