

Multicollinearity and MRA in Sustainability: Application of the Raise Regression

Authors : Claudia García-García, Catalina B. García-García, Román Salmerón-Gómez

Abstract : Much economic-environmental research includes the analysis of possible interactions by using Moderated Regression Analysis (MRA), which is a specific application of multiple linear regression analysis. This methodology allows analyzing how the effect of one of the independent variables is moderated by a second independent variable by adding a cross-product term between them as an additional explanatory variable. Due to the very specification of the methodology, the moderated factor is often highly correlated with the constitutive terms. Thus, great multicollinearity problems arise. The appearance of strong multicollinearity in a model has important consequences. Inflated variances of the estimators may appear, there is a tendency to consider non-significant regressors that they probably are together with a very high coefficient of determination, incorrect signs of our coefficients may appear and also the high sensibility of the results to small changes in the dataset. Finally, the high relationship among explanatory variables implies difficulties in fixing the individual effects of each one on the model under study. These consequences shifted to the moderated analysis may imply that it is not worth including an interaction term that may be distorting the model. Thus, it is important to manage the problem with some methodology that allows for obtaining reliable results. After a review of those works that applied the MRA among the ten top journals of the field, it is clear that multicollinearity is mostly disregarded. Less than 15% of the reviewed works take into account potential multicollinearity problems. To overcome the issue, this work studies the possible application of recent methodologies to MRA. Particularly, the raised regression is analyzed. This methodology mitigates collinearity from a geometrical point of view: the collinearity problem arises because the variables under study are very close geometrically, so by separating both variables, the problem can be mitigated. Raise regression maintains the available information and modifies the problematic variables instead of deleting variables, for example. Furthermore, the global characteristics of the initial model are also maintained (sum of squared residuals, estimated variance, coefficient of determination, global significance test and prediction). The proposal is implemented to data from countries of the European Union during the last year available regarding greenhouse gas emissions, per capita GDP and a dummy variable that represents the topography of the country. The use of a dummy variable as the moderator is a special variant of MRA, sometimes called "subgroup regression analysis." The main conclusion of this work is that applying new techniques to the field can improve in a substantial way the results of the analysis. Particularly, the use of raised regression mitigates great multicollinearity problems, so the researcher is able to rely on the interaction term when interpreting the results of a particular study.

Keywords : multicollinearity, MRA, interaction, raise

Conference Title : ICSDS 2023 : International Conference on Statistics and Data Science

Conference Location : Toronto, Canada

Conference Dates : June 19-20, 2023