

Believing by Seeing before Seeing by Believing: Visualizing the Gaussian Regression Model by the SIM.REG package for intuitive teaching

The Gaussian Regression Model is one of the fundamental techniques in econometric theory. There are several crucial assumptions of this model such as homoskedastic error variance, no-autocorrelation between error distributions, no high multicollinearity between independent variables, and identification of correct functional form. When the model is used for hypothesis testing the normality of error distributions too should be added to the list. These fundamental assumptions are of course always mentioned in the regression courses and several statistical tests are introduced to diagnose possible violations. Unfortunately, students, especially those with social science background, quite often fail to internalize the importance of these assumptions neither do they really appreciate the BLUE (Best Linear Unbiased Estimator) property of The Gaussian Regression Model. However, it is of crucial importance for the students to develop an intuitive understanding of the weaknesses and strengths of this basic approach in order to be able comprehend the premises of and need for more sophisticated modeling techniques. In other words students first need to ‘believe’ by ‘seeing’ to be able to ‘see’ more advanced theorems and models by ‘believing’.

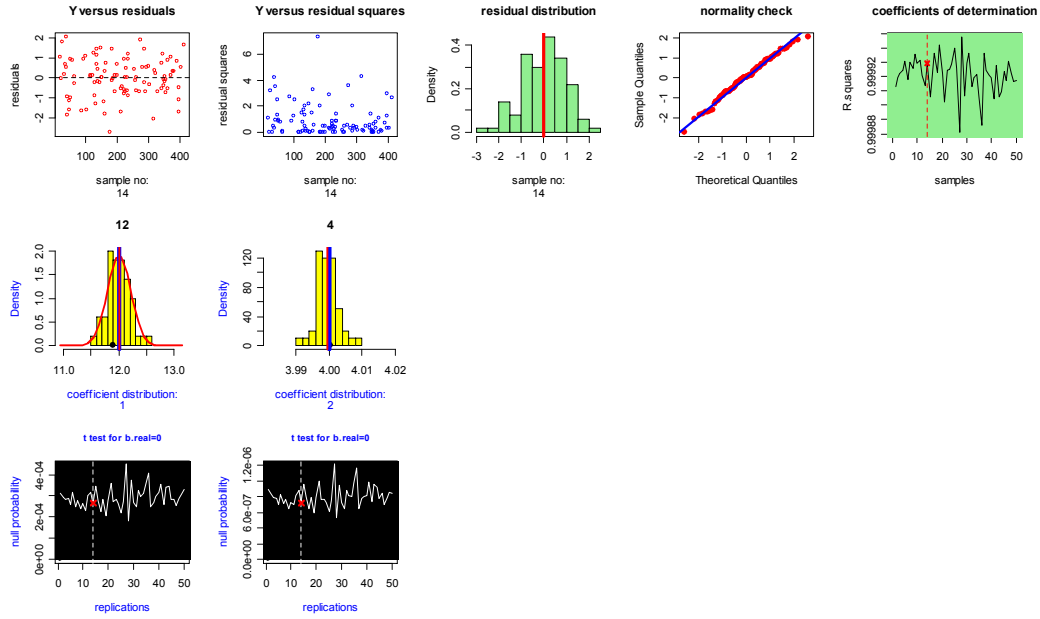
In this paper, a package programmed by the author in R language to conduct Monte Carlo simulations, the SIM.REG (Simulated Regression) is introduced and its visual and analytical strength in depicting ins and outs of The Gaussian Regression Model is demonstrated.

The SIM.REG allows creating, testing and visualizing data sets which contain desired degrees of heteroskedasticity, autocorrelation, multicollinearity and non-normality in order to reveal the isolated or simultaneous impact of these violations on the Gaussian regression model. Similarly, the SIM.REG also allows revealing the impact of wrong functional forms on the model. Indeed SIM.REG contains an option which allows seeing, for example, the impact of increasing level of autocorrelation on inference structure as an animation. Moreover under the SIM.REG one can also visually depict the BLUE property of the Gaussian Regression Model by making all assumptions hold. In this way students can be visually shown the degree to which violation of assumptions damage the capacity of the model to make reliable estimations and correct inference.

Moreover, it is also possible to use the SIM.REG to generate data sets which allow testing the sensitivity of statistical tests and indicators, such as rank correlation test, Durbin-Watson test, condition index etc... In other words by using the SIM.REG one can create data sets which contain an isolated problem such as heteroskedasticity and then scrutinize the ability of various statistical tests to identify the problem. In this way one can enable the students to develop a critical approach to various tests, that is, when and why to disregard the positive or negative test outcomes about particular problems. More importantly one can also create several problems simultaneously in order to reveal how these simultaneous violations may render some instruments of diagnosis ineffective.

The main purpose of the paper is, by sketching out various applications of the SIM.REG package, to show how it can be effectively used for teaching purposes in MA level regression courses.

Example 1: the SIM.REG output for simple linear regression with full compliance after 50 simulations



Example 2: the SIM.REG output for multiple linear regression with heteroskedasticity, severe autocorrelation and mild multicollinearity after 40 simulations

