Regression Estimation in Small Samples: Problems and Solutions

By
Abd El Rahman Mahmoud Abd El Salam Khedr

SUPERVISED BY

Prof. Ahmed Mahmoud Gad
Professor of Statistics, Department of Statistics
Faculty of Economics and Political Science
Cairo University

Ass. Prof. Zohdy Mohammed Nofal
Associate Professor and chairman, Department of Statistics, Mathematics and Insurance - Faculty of Commerce, Benha University

A THESIS SUBMITTED TO
The Department of Statistics, Mathematics and Insurance
Benha University
In partial fulfillment of the requirement for
the Master degree in applied Statistics

2017
Abstract

Regression analysis is one of the most commonly statistical techniques used for analyzing data in different fields. Parametric methods used to fit the relation between the dependent variable and the independent variables require strong assumption to be met in the model.

When analyzing small samples, several problems encounter us such as Bias, reduced accuracy, increased standard deviations, and decreased statistical power. The parametric assumption of normality is particularly worrisome and the precise distribution of these estimators could not be known in small samples.

Nonparametric regression models do not assume a prespecified form of the regression function. Nonparametric regression models let the data decide which function fits them best without the restrictions assumed in parametric regression models.

A simulation study has been conducted to compare between parametric regression models and nonparametric regression models. 5000 replicates are simulated with sample size 10, 20, and 30. Several cases for the distribution of the errors are used. The criteria used to compare between them are Bias, MSE, and MAE.

Parametric estimation methods are Ordinary Least Squares (OLS), Maximum Likelihood (ML), and Least Absolute Deviations (LAD). Nonparametric estimation methods are Local Constant estimator (LC) which is the Nadaraya-Watson Estimator, k-nearest neighbor (KNN), Local linear estimator (LL), and Spline Smoothing (SS). The nonparametric methods used with two common Kernel densities Epanechnikov and Gaussian.

By examining the simulation results, it is shown that nonparametric method is better than parametric method. Spline Smoothing is considered the best method due to its good performance according to all criteria for sample size 10. LL.g also has good results according to MSE and MAE for sample size 10. For sample sizes 20 and 30, K-NN.ep, K-NN.g, SS and LL.g have the best results for most cases according to MSE and MAE. OLS and SS have the best results according to Bias.

The applications results confirmed that nonparametric method is the superior to parametric methods according to MSE and MAE. According to Bias, OLS and SS have the best results. ML is more efficient than OLS and LAD.