PM2.5 Air Pollution Detection of Influential Observations based on the Linear Regression Model with Measurement Errors

Jia-Ren Tsai

Department of Statistics and Information Science, Fu Jen Catholic University

Abstract

This study assumes the observations in linear relationship between the Airbox sensor and the EPA air quality monitoring station, moreover, we adopts the adjusted least squared estimation to find unknown parameters and to develop detect abnormal influence points for the instrument on the linear measurement error model. In the data analysis of PM2.5 air pollution, the calculation of error variation by adjacent points and time axis are respectively used. The standard R^2 method of detecting influence points is developed and compared to other detection methods (GD, MD, GDF, MDF, GDB, MDB) under different methods of correcting measurement errors, and the results of abnormal values provided by these methods are investigated. When the independent variable contains measurement error, the detection method with measurement error should be used and the research results found that (1) The detection rules of the Cook's Distance series are more conservative; (2) When the measurement error is large, GDF, GDB and standard R^2 are more effective to detect possible influence points. Depend on diagnosing the impact of PM 2.5 air pollution data, this study can provide diagnostic results for practitioners as a reference, so that practitioners can assess whether appropriate adjustments and corrections to the instrument are necessary.

Keywords: Air pollution, Anomaly detection, Measurement error, Influential point