On the use of machine learning methods for the treatment of unit nonresponse in surveys

Khaled Larbi[1], John Tsang[2], David Haziza[3], Mehdi Dagdoug[4]

Abstract

In recent years, there has been a significant interest in machine learning in national statistical offices. Thanks to their flexibility, these methods may prove useful at the nonresponse treatment stage. In this article, we conduct an empirical investigation in order to compare several machine learning procedures in terms of bias and efficiency. In addition to the classical machine learning procedure, we assess the performance of ensemble approaches that make use of different machine learning procedures to produce a set of weights adjusted for nonresponse.

Keywords: Aggregation procedure, Efficiency, Nonresponse bias, Propensity score estimation.

[1] INSEE, khaled.larbi@insee.fr.

[2] University of Ottawa, john.tsang@uottawa.ca.

[3] University of Ottawa, dhaziza@uottawa.ca.

[4] McGill University, mehdi.dagdoug@mcgill.ca.