Measuring unexplained variation in data: a non-parametric approach based on global sensitivity indices

Giovanni Rabitti *

Department of Business, Economics, Mathematics and Statistics "Bruno de Finetti",
University of Trieste, Italy

Abstract

In many insurance datasets, it often happens that some risk factors are not included because they are unobservable or unreported. Then, it becomes essential to quantify the extent to which the observed explanatory risk factors provide relevant information on the response.

In this work, we provide a non-parametric approach to quantify the importance of unobserved variables for the variation in the response. We consider the total importance indices derived from sensitivity analysis literature, adapting them for the first time to address this particular problem. We discuss appealing theoretical properties of this importance measure in this context. We connect it to the mean squared error of a predictive model and propose a non-parametric numerical estimation strategy.

To illustrate the versatility of our approach, we apply it to insurance datasets with different response types, including medical malpractice costs, cyber data breaches, country life expectancy, and conduct a comparative analysis of four car insurance datasets. In this latter case, we show that telematic variables explain a significant portion of the variance in policyholders' claim frequency.

Keywords: Global sensitivity analysis, unobserved risk factors, unexplained variance.

*E-mail address: giovanni.rabitti@deams.units.it