**Big data and precision**

**By Niti Mishra, Miquel Torrens and Bálint Ván (Group 3).**

The article *Big Data and Precision* by David Cox addresses an underlying issue of standard statistical theory in which the accuracy of sample estimates increases when working in Big Data frameworks. This lies in excessively acute significance of these estimates given the reduction in their variance when the number of observations grows sufficiently large. The variance of the estimation grows proportionally to 1/n in the standard statistical procedure. However, empirical evidence shows in many cases this relation is in fact weaker due to the effect of persistent autocorrelation, here related to the concept of self-similarity. Long-range dependence enhances variance of estimates even for large number of observations.

The contribution by the article is to extend this autocorrelation concept by thinking of big data in a time series fashion, as a piece of data that starts small and then grows, evolving on some sort of notional time-frame. With this idea, the concept of autocorrelation can apply and the power of sample size could be diminished in terms of asserting estimated model coefficients, thereby allowing for higher variance on these estimations.

A more realistic approach might be to investigate the sources of variability of the data and base the modelling on the insights gathered there. There are sound incentives to keep the amount of data as big as it is still understandable. In that sense working collaboratively with the experts in the field at stake is fundamental to be able to choose the best model and make practical inferences on the data.

The biggest danger of the Big Data approach perhaps lies in incorrectly inferring the results. The standard statistical procedures are not always accurate in estimating precision and so using predictors that are really not relevant. This leads to the misconception that having a lot more data produces better estimates.

This paradigm is both frightening and beautiful; it is a double edge sword. On the one hand if methodologies are not rigorously adapted and analyses are not performed carefully, it clears the path for big mistakes. But with the right statistical approaches it has the potential to uncover complex solutions to some of the biggest question marks in science.