

Turning Big Data into knowledge for policy

Rickard Nyman

Financial Computing & Analytics
University College London
rickard.nyman.11@ucl.ac.uk

The use of Big Data to measure and predict variables of interest to policymakers has in recent years attracted a lot of attention, especially in central banks. Most of the discussion surrounding Big Data has focused on its promise to give accurate and immediate (i.e., near real time) measures of economic and social variables. For example, a recent study by engineers at Google showed the promise of estimating the number of influenza cases based on the number of searches for flu related terms on Google. Other similar studies have shown the potential to forecast a range of economic variables, such as consumer confidence and economic growth.

This study presents an example, considering the case of flu predictions using Google trends data, of how such forecasts can at times be misleading, and that a deeper understanding of the underlying processes generating the data is necessary. When people are searching independently, i.e. they genuinely need to know the information, search engine data is a reliable predictor. When social influence is important, i.e. many people are searching simply because many other people are, the forecasts can be highly misleading.

We argue that, as a consequence of this, a pure data mining approach to prediction using Big Data can be greatly misleading. Searches need to be guided by theory from the social sciences. This is of fundamental importance. These ideas extend to the analysis of Big Data in general.

A high priority for research is therefore the development of techniques and tools which enable very early identification of which motive is the more important in any particular search. This is essential in order to be able to convey accurate information to policy makers.