Quantifying the Relationship between Financial News and Market Behaviour

Tobias Preis, *Warwick Business School* (Tobias.Preis@wbs.ac.uk)
Helen Susannah Moat, *University College London* (h.s.moat@ucl.ac.uk)
Nick Chater, *Warwick Business School* (Nick.Chater@wbs.ac.uk)

Mammoth amounts of data are now being generated through society's extensive interactions with technological systems, automatically documenting collective human behaviour in a previously unimaginable fashion (*1–3*). Analysis of such *big data* opens up new opportunities for a more precise and extensive quantification of real world social phenomena that was difficult to attain using complicated and expensive surveys and laboratory experiments alone.

A particularly fruitful area of research has focused on the analysis of Internet user search queries, as logged by search engines such as *Google*. Strong links have been found between changes in the information users are seeking online and events in the real world, ranging from reports of flu infections across the USA (*4*) to the trading volume in the US stock markets (*5*). A recent study has shown that Internet users from countries with a higher per capita GDP are significantly more likely to search for information about years in the future than years in the past (*6*).

The primary goal of this mini-project is to expand on findings quantifying relationships between "Big Data" sources and financial market behaviour (*5*) using historical news. For this mini-project, historical issues of the *Financial Times* are available to analyse the word frequency time series and their temporal relationship to financial market moves.

Previous preliminary experience in the following three areas is of advantage to complete this mini-project successfully:

> *(1) basic programming skills and experience with processing text corpora;*
> *(2) experience with statistical programming packages;*
> *(3) writing reports for an academic audience.*

This mini-project would be suitable as the basis for a PhD project in this area. Possible directions include incorporating other data sources such as *Google, Wikipedia*, *Flickr*, *Twitter* and others and developing agent-based models of this complex human behaviour.

Please do not hesitate to contact us if you have questions about this mini-project.

References:
1. King, G. Ensuring the Data-Rich Future of the Social Sciences. *Science* **331**, 719 (2011).
2. Lazer, D. *et al.* Computational Social Science. *Science* **323**, 721 (2009).
3. Vespignani, A. Predicting the Behavior of Techno-Social Systems. *Science* **325**, 425 (2009).
4. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012 (2009).
5. Preis, T., Reith, D. & Stanley, H. E. Complex dynamics of our economic life on different scales: insights from search engine query data. *Phil. Trans. R. Soc. A* **368**, 5707 (2010).
6. Preis, T., Moat, H. S., Stanley, H. E. & Bishop, S. R. Quantifying the Advantage of Looking Forward. *Sci. Rep.* **2**, 350 (2012).