Real-Time Measurement of Uncertainty During Crises

Rickard Nyman[1] and Paul Ormerod[2]

June 2020

JEL classification: C45; C82; D89

Keywords: uncertainty, text as data, machine learning, Granger causality

Abstract *We analyse news feed data, using daily data from the late 1990s through May 2020, to construct a simple, general measure of uncertainty in both the United States and the United Kingdom using a highly cited machine learning methodology.*

The level of uncertainty rises in both countries as expected in periods such as the aftermath of the dot com boom and the financial crisis. However, there is only a modest increase observed during the Covid crisis.

The highly cited Economic Policy Uncertainty index gives similar results, except that it shows a massive rise in the US during Covid.

However, using data from the late 1990s through May 2020, we show that our series unequivocally Granger-causes the EPU in both the UK and the US, and there is no Granger-causality in the reverse direction.

[1] Centre for Decision Making Uncertainty, University College London, rickardnyman@gmail.com and Algorithmic Economics Ltd

[2] Department of Computer Science, University College London, p.ormerod@ucl.ac.uk Algorithmic Economics Ltd

1. Introduction

The vast increase in digital information combined with advances in machine learning techniques offers the opportunity both to create different kinds of data and to create them in real time. In a recent issue of the *Journal of Economic Literature*, for example, Gentzkow et al. (2019) point out that "New technologies have made available vast quantities of digital text, recording an ever-increasing share of human interaction, communication, and culture. For social scientists, the information encoded in text is a rich complement to the more structured kinds of data traditionally used in research" (p.535).

Here, we use text data from the Reuters newsfeed to construct a real-time measure of uncertainty in both the United States and the United Kingdom. We apply the unsupervised learning algorithm for obtaining vector representations for words developed by Pennington et al. (2014)[3].

The Economic Policy Uncertainty Index (EPU) developed by Baker at al. (2016) has considerable traction within economics. We show that the uncertainty measure which we develop Granger-causes the EPU in both the UK and the US and that there is no causality from the EPU to our measure.

Section 2 describes the construction of the uncertainty index (referred to below as UNCERT), and section 3 briefly describes the Granger causality tests. An Appendix giving full details of the tests is available on request.

2. Real time measurement of uncertainty

We use the Reuters newsfeed over the period 1 January 1996 through 31 May 2020 as the textual source. To construct the series for America, we analyse all stories published by the New York and Washington offices, amounting to a total of 2,540,233 articles. For the UK, we use the stories published by the London office, which gives 2,040,337 articles.

A very simple way of measuring uncertainty would be to count the number of times the word itself appears each day.

We note here that, regardless of the word(s) which are the focus of any search, the resulting raw data should be scaled by counting the number of articles that mention at least one of the words, divided by the total number of articles. The scale is therefore the proportion of articles that matches the keyword search.

Essentially, we do base the series on a count of the word "uncertainty" itself. However, we augment the list of words in the search using approach which has become standard in machine learning known as GloVe (Pennington et al. op.cit.). A clear overview, with a

[3] a paper which, incidentally, has over 13,000 citations

description of how to download and use the method, is given at https://nlp.stanford.edu/projects/glove/.

The authors assemble a very large corpus of words from various sources. We use the one described on the GloVe website as Common Crawl (glove.42B.300d.zip). A co-occurrence matrix is constructed, which describes how frequently pairs of words co-occur with each other in any given corpus.

The referenced webpage above states: "The training objective of GloVe is to learn word vectors such that their dot product equals the logarithm of the word's probability of co-occurrence. Owing to the fact that the logarithm of a ratio equals the difference of logarithms, this objective associates (the logarithm of) ratios of co-occurrence probabilities with vector differences in the word vector space".

The eventual output of the process is that every word in the corpus has a unique n-dimensional vector associated with it. The elements of each vector are real valued numbers which essentially describe the closeness of the word to all other words in the corpus. This description is perforce rather imprecise. It is only intended to give a broad non-technical indication of what is going on.

To construct the UNCERT series, we count each day the number of times "uncertainty" and a list of words identified by GloVe as being very close to it appear in the relevant Reuters news feed. To be clear, the closeness is identified using the general corpus of words in GloVe, and not in the specific Reuters text feed.

The closest word to "uncertainty" (except of course for the word itself) is "uncertainties". The Euclidean distance between the vector associated with "uncertainty" and the vector associated with "uncertainties" is 5.40. Of the 1.9 million words in the GloVe corpus, the Euclidean distance to the one furthest away is 17.70. The median is 8.64 and the standard deviation 0.68.

The "nearness" declines rather quickly, as Figure 1 shows.

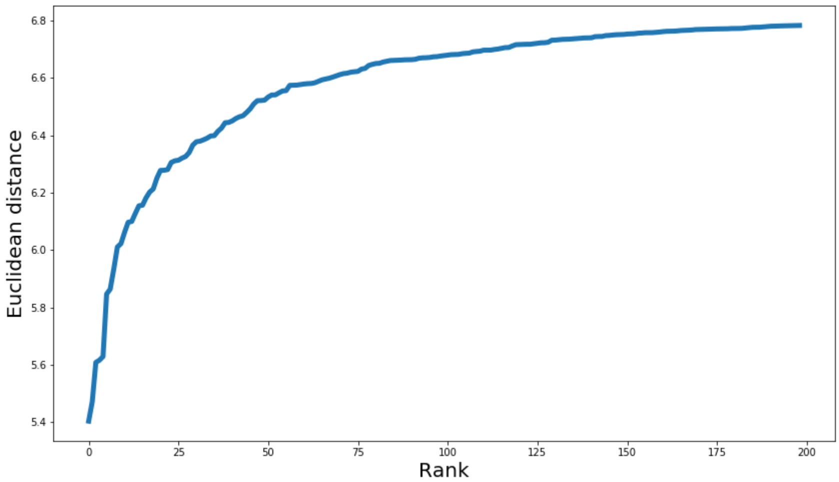

Figure 1 *Euclidean distance from the GloVe vector of "uncertainty" of the 200 words nearest to it*

The ten words closest to "uncertainty" are: uncertainty, uncertainties, uncertain, unpredictability, ambiguity, certainty, confusion, turmoil, expectation, instability. We exclude "certainty" and "expectation" on the grounds that their meaning is different and count the frequency with which the remaining 8 words appear.

Figures 2a and 2b plot, respectively, the UNCERT series for the US and the UK. We aggregate the data onto a monthly basis for presentational purposes, to screen out the noise observed at the daily frequency. We also index each series so that the average for January 1996 = 100.

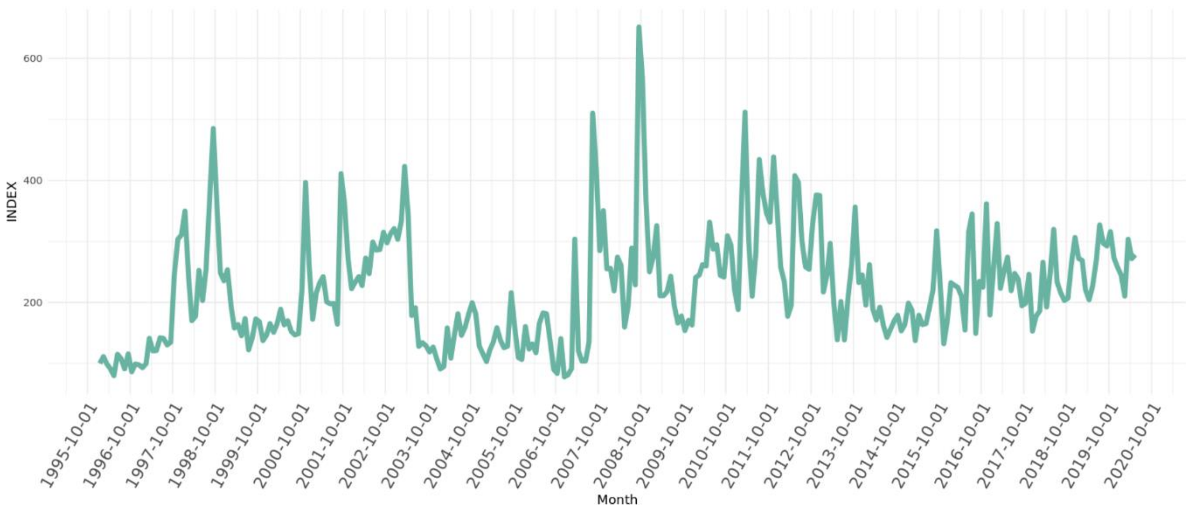

Figure 2a *The Uncertainty Index derived from the Reuters newsfeed for the United States, monthly averages January 1996 – May 2020, January 1996 = 100*

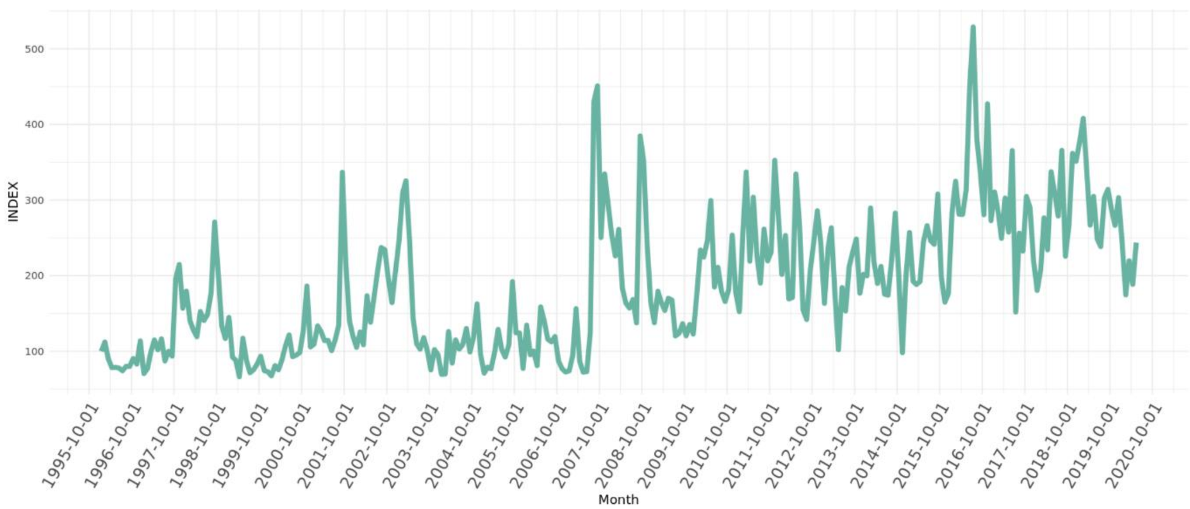

Figure 2b *The Uncertainty Index derived from the Reuters newsfeed for the United Kingdom, monthly averages January 1996 – May 2020, January 1996 = 100*

The two series have a great deal in common. The correlation between them is 0.69 and between the first differences it is 0.70.

Both series show increases in uncertainty at times when one would expect to see it:

- The Asian crisis of 1997/98
- The collapse of the dot com boom in the early 2000s
- The financial crisis of the late 2000s

In terms of the UK, we also observe a temporary but large rise in uncertainty in the immediate aftermath of the referendum in June 2016 and the vote to leave the EU.

There are two periods where the series give results which are perhaps less immediately obvious.

After an initial fall at the very end of the 2000s, once the financial crisis began to abate and the economies began to recover, uncertainty levels rose in the early years of the 2010s.

Throughout the West in general, economic growth was considerably lower in the recovery phase than has usually been the case after recessions. The corporate sector in particular retained large amounts of cash, and investment remained relatively subdued. The levels of uncertainty in the series is certainly consistent with this behaviour and the weak nature of the recovery.

The most surprising feature of the data series is that uncertainty did not rise sharply during the Covid crisis in either country. The average of the index in January and February 2020 in the US was 227.6, in March 303.6 and the April and May average was 275.0. So, we observe some rise, but to put the levels in perspective in September 2008 the index was 651.4 and in October 2008 567.7. In the UK, we see hardly any increase at all. The January-February average was 209.1, in March 219.7 and the April-May average 216.2.

In general, the two series move closely with the Economic Policy Uncertainty news-based indices. The correlation between UNCERT and the EPU January 1996 to February 2020 was 0.72 in America, and 0.73 in the UK[4].

However, there is a marked contrast between the two during the Covid crisis, especially in the US. At the height of the financial crisis, in September and October 2008, the purely news based EPU index was 238 and 241 and the more broadly based EPU was 187 and 190. In May 2020, for example, the former was 504 and the latter 350.

3. Granger causality between UNCERT and the EPU

Here, we report the basic summary of Granger causality tests between UNCERT and the two versions of the EPU which are available for the United States January 1996-May 2020, and between UNCERT and the purely news based EPU for the UK January 1998-May 2020. A full description is available in an Appendix[5]. We follow the procedure set out in Toda and Yamamoto (1995).

The results are set out in Table 1.

Table 1 P-values in tests of Granger causality between UNCERT and the Economic Policy Uncertainty Indices

		United States	**United Kingdom**
From	**To**	**Jan 1996-May 2020**	**Jan1998-May 2020**
UNCERT	EPUNEWS	0.017	0.032
EPUNEWS	UNCERT	0.38	0.64
UNCERT	EPUGEN	0.010	n/a
EPUGEN	UNCERT	0.38	n/a

By way of explanation, the EPUNEWS[6] series is based on newspaper articles which contain the term 'uncertainty' or 'uncertain', the terms 'economic' or 'economy' and one or more of a list of "policy relevant terms". In the US, these include 'congress', 'legislation', 'white house', 'regulation', 'federal reserve', or 'deficit', and in the UK 'policy', 'tax', 'spending', 'regulation', 'Bank of England', 'budget', and 'deficit'.

In other words, both with our UNCERT series and the EPU news-based one the core word of each is the same, but the additional information used is different.

[4] The EPU is only available from January 1998 for the UK, so the correlation sample begins then
[5] rickard.nyman.11@ucl.ac.uk
[6] See the website https://www.policyuncertainty.com/index.html)

EPUGEN contains EPUNEWS as a component, but also incorporates information on tax code information and on disagreement amongst economic forecasters in the database published by the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters.

In each case, the UNCERT Granger-causes the EPU series, with no causality in the reverse direction.

4. Brief discussion

Advances in machine learning techniques during the past decade considerably advance the ability to extract meaningful quantitative time series from text-based data. Here, we develop a measure of uncertainty using the methodology developed by Pennington et al.(op.cit.).

Elsewhere (Nyman and Ormerod 2020), we apply the alternative machine learning technique of described in Mikolov et al. (2013) to construct a real-time index of well-being.

Both these seminal machine learning papers have over 10,000 citations, and their techniques can readily be applied using packages such as R.

The series we develop to measure uncertainty is in general very similar to the widely accepted Economic Policy Uncertainty index, though the two differ sharply in terms of the recent Covid crisis period. The EPU in both the UK and the US is Granger-caused by the machine learning based measure which we develop.

References

Baker, S.R., Bloom, N. and Davis, S.J., 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), pp.1593-1636.

Gentzkow, M., Kelly, B., and Taddy, M. 2019. Text as data. *Journal of Economic Literature*, 57(3), pp.535-574.

Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space.

Nyman, R. and Ormerod, P. 2020, Text as data: real time measurement of economic welfare, *arXiv preprint arXiv:2006.06457*

Pennington, J., Socher, R. and Manning, C.D., 2014, October. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Toda, H.Y. and Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, *66*(1-2), pp.225-250.