



News and narratives in financial systems: Exploiting big data for systemic risk assessment



Rickard Nyman^a, Sujit Kapadia^{b,1}, David Tuckett^{a,1,*}

^a University College London Centre for the Study of Decision-Making Uncertainty, United Kingdom

^b Bank of England, United Kingdom and European Central Bank, Germany

ARTICLE INFO

Article history:

Received 22 June 2020

Revised 12 March 2021

Accepted 21 March 2021

Available online 2 April 2021

JEL Classification: C53, D83, E32, G01, G17

Keywords:

Systemic risk

Text mining

Big data

Sentiment

Uncertainty

Narratives

Early warning indicators

ABSTRACT

This paper applies algorithmic analysis to financial market text-based data to assess how narratives and sentiment might drive financial system developments. We find changes in emotional content in narratives are highly correlated across data sources and show the formation (and subsequent collapse) of exuberance prior to the global financial crisis. Our metrics also have predictive power for other commonly used indicators of sentiment and appear to influence economic variables. A novel machine learning application also points towards increasing consensus around the strongly positive narrative prior to the crisis. Together, our metrics might help to warn about impending financial system distress.

© 2021 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0/>)

"We perceive the world before we react to it, and we react not to what we perceive, but always to what we infer." F.H. Knight (1921 p 118)

"Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as a result of animal spirits—of a spontaneous urge to action rather than inaction..." J.M. Keynes (1936 p 161)

"Significant market events generally only occur if there is similar thinking among large groups of people, and the news media are essential vehicles for the spread of ideas" (R.J. Shiller, 2000 p 71)

1. Introduction

Discussion of the role played by narrative sentiment in driving the economy and the financial system has a long history (Kindleberger and Aliber, 2005). The years preceding the global financial crisis are an example. Widespread exuberance took hold in the financial sector linked, as in many previous instances (Reinhart and Rogoff, 2009), to the idea of a new paradigm – the strong positive sentiment this time being 'justified' by the belief there was now greater efficiency of markets and distribution of risk around the system so that higher returns could be sustained. Acceptance of this narrative proved to have

* Corresponding author.

E-mail addresses: rickard.nyman.11@ucl.ac.uk (R. Nyman), sujit.kapadia@ecb.europa.eu (S. Kapadia), D.Tuckett@ucl.ac.uk (D. Tuckett).

¹ Sujit Kapadia started to work on this paper while at the Bank of England.

enormous costs, with sentiment reversing rapidly and fear pervading the financial system when the crash came during 2007 and 2008.

The crisis led to a renewed focus on the role of narrative sentiment (e.g. Akerlof and Shiller, 2009; Gennaioli and Shleifer, 2018; Shiller, 2019). This work identifies narrative and emotion as economically important, but typically interprets this as reflecting behavioural bias and framing at the micro-level, or as some kind of irrational herding at the macro level. By contrast, drawing on the insights of Knight (1921) and Keynes (1936) and joining a long tradition of social science thinking about narrative with contemporary psychology and neuroscience, we conceive narratives as an everyday means for perceiving and interpreting the world (see for example, Bruner, 1990), with actions necessarily driven by emotions of excitement and anxiety (Tuckett and Nikolic, 2017). Both conceptions of narrative identify textual analysis as a promising approach to help understand the impact of narratives, but the latter is able to ground the textual approach with a theoretical filter drawn from psychological microfoundations i.e. conviction narratives.

This paper uses a conviction narrative approach towards attempting to identify sentiment and systemic risk in financial systems using big data. Specifically, we apply theoretically-motivated algorithmic analysis to large amounts of unstructured text-based data spanning three sources of potential interest related to the financial system: internal Bank of England daily commentary on market news and events; Reuters' news wire articles in the United Kingdom; and broker research reports. We identify quantitative metrics that try to capture shifts in financial system sentiment along with the extent of narrative topic consensus in the market. The two new measures we develop suggest changes in narrative sentiment were a significant influence on the financial system in the period leading up to the global financial crisis and thereafter.

Our sentiment measure appears, with the benefit of hindsight, to give early warning signs of significant financial events, with the key shifts identified by structural break tests. In particular, overall sentiment was at very high and stable levels in the mid-2000s, arguably indicative of exuberance in the financial system. From mid-2007, a surge in anxiety drove rapid falls in sentiment that continued until soon after the collapse of Lehman Brothers. And there were further falls in sentiment prior to the start of the euro area sovereign crisis in 2011–2. In a related exercise, we also illustrate how, when focused on particular topics such as 'property', the sentiment index may shed light on specific sectors of the economy. To gauge the robustness of our sentiment metrics, we compare them with both standard economic and financial measures of consumer confidence and market volatility, and with some commonly used, but more atheoretic, text-based measures of sentiment and uncertainty. Strikingly, we find that our sentiment metrics have some predictive power for other commonly used economic and financial measures, as indicated by Granger causality tests (Granger, 1969). They also appear to influence economic and financial variables in a simple vector autoregression (VAR) model.

Financial behaviour can often be homogenous. Our second, more experimental, measure therefore considers whether we can measure *structural* changes in the distribution of narratives, so as to assess whether sentiment shifts in narratives may also become associated with more or less consensus on narrative topics. Specifically, we develop a novel machine learning application using Latent Semantic Analysis (LSA) (Deerwester et al., 1988), X-means clustering (Pelleg and Moore, 2000) and entropy to measure 'narrative consensus / disagreement' in the distribution of narratives as they develop over time. The LSA algorithm turns text articles into numeric vectors, which allows us to group articles into topic clusters and to measure 'consensus' over topics. We use the X-means algorithm (an extension to the widely used K-means that uses the Akaike information criterion to decide when to stop iteratively dividing clusters) to discover automatically the number of clusters as well as cluster memberships, with the output of the discrete entropy measure on the number of articles in each cluster used as our measure of narrative consensus. Our novel measure is a possible way to reflect the extent to which some narratives have spread through social networks (see, for example, Bailey et al., 2018, Shiller, 2019 or Watts, 2002), possibly amplified by herding ("keeping up with the Goldmans" – Aikman et al., 2015), by the media (Shiller, 2000), or via social-psychological processes (*groupfeel* (Tuckett, 2011)).

Using this new measure of narrative dispersal, we find that consensus grew significantly over a period spanning several years prior to the global financial crisis. Taken with the sentiment series results, it confirms the view that in the pre-crisis period there was a growing and increasingly dominant narrative about a new paradigm in the financial system, in which excitement prevailed over anxiety and doubt about what might have been happening, so that avoidance emotions were substantially diminished. This could be interpreted as a possible signal that market attitudes had been captured by "phantastic object" thinking (Tuckett, et al., 2014) leading to greater risk-taking and, ultimately, impending distress. Based on all of these results, we contend that our metrics have potential value for gauging sentiment, consensus and systemic risk in financial systems. As such, they may serve as complements to more traditional indicators of systemic risk in financial systems, of the type discussed by, for example, Drehmann et al. (2011), Schularick and Taylor (2012), Bank of England (2014) or Giese et al (2014).

Our paper is related to a broad range of psychological and economic literature. Below, we briefly set out ideas about the role of narrative and sentiment in human decision-making applicable to both micro and macro levels, which inform our two novel measures. Specifically, we draw on ideas available in Knight (1921), Keynes (1936), and in modern brain and social science literatures, to explain how and why, to use Shiller's phrase quoted above, "similar thinking among large groups of people" comes to dominate market sentiment in the form of shared group narratives evoking either excitement or anxiety, which then function as coordinating influences on the evolution of economies and financial markets. After setting out this way of thinking, we review recent approaches to the text analysis of news and similar databases used in economics and finance to show how they differ. We then summarise our specific contribution to the literature.

1.1. Narratives

As in much social and psychological science, we conceive narratives as a means, perhaps the primary means, for perceiving and making sense of the world as well as for sharing our perceptions and understanding with others (see, for example, Bruner, 1990 and Tuckett and Nikolic, 2017). Because the world and what is happening around us (and can be expected to happen as the result of our actions) is often complex, and the forces driving it are the dynamic and changing outcome of numerous interactions, we think economic agents must necessarily turn whatever data they find into “information”. That “information” is then transformed into “actionable intelligence” or “knowledge” underpinning expectations about the outcome of different actions (see Tuckett et al., 2020). Constructing knowledge and making plans based on such narrative expectations is what enables agents to act both “rationally” and with conviction even when situations are uncertain or complex – a point made by Knight (1921) as part of his self-labelled “brief excursion” into the “field of the theory of knowledge”, when his analysis obliged him to drop the classical assumption of “practical omnipotence” (Knight, 1921: p. 116). His point, quoted at the head of this paper, is that actionable intelligence – informed knowledge of what it is rational to do – is not given to economic agents as if they are Gods but has to be inferred by them and then believed in by them. Our social-psychological view is that narratives, devices for “constructing” the world and its evolution via selections and interpretations of data within assumptions of causal dynamics, provide this intelligence. Narratives can be created by an individual agent *de novo* or picked up or tested within an agent’s social networks (Collier and Tuckett, 2021). It is not a question of any narrative will do – actionable narratives can be assessed for their accuracy and the relevance of their core assumptions and organizational schemas. But, by definition, they are partial knowledge – systems for making the “facts” fit when there are potentially different interpretations and models. So, while one narrative will appear to serve well at one moment, like other explanations of the world, even scientific ones, it will always be open to being constructed differently should events cause facts to be reinterpreted or causal dynamics to be newly understood. New competing narratives may then flourish. In short, narratives in this view are constitutive of knowledge.

1.2. Sentiment

By sentiment we refer, precisely, to subjectively evoked and measurable emotional states – “states elicited by rewards and punishers” (Rolls, 2014 p14) which are triggered and evaluated within brain networks as we engage the world and respond to it. Emotions, specifically avoidance or approach emotions, therefore, motivate attention and action – creating conviction about how desirable the outcome of any planned actions and engagements will be, or not. Of particular significance in this context, is what is called the “as-if body loop” in contemporary neuroscience (Damasio and Damasio, 2006). It means that imagined stories about the world, like experience of the world itself, evoke similar neural processes to actual experience and so provide a means for cognitive and affective evaluation. In this way a “conviction narrative”, a narrative about how a plan will work, evokes *felt* and *thought* confidence about the outcome of action (Tuckett and Nikolic, 2017).

Although writing before brain scanning equipment allowed contemporary neuroscientific insights, Keynes’s concept of animal spirits showed that he recognised that, to act in the face of uncertainty, economic agents had to *feel* convinced about their “knowledge” and this meant they had to overcome doubt. In his view, *the level* of animal spirits, a measure of the extent to which the human urge to spontaneous action was sufficiently widespread within an economy, necessarily underpinned macroeconomic conditions. The more positive the sentiment in prevailing narratives, the more urge to risk-taking and vice-versa.

1.3. News texts, economic approaches to text analysis and our contribution

News texts, like most human communications, necessarily contain narrative *content* and *sentiment*. They, therefore, have the potential to capture shifts in the narratives circulating in the economy and so to capture both shifts in sentiment and what Shiller (2000) described as *changes* that may be occurring in the “similar thinking among large groups of people” that can drive market events. Here, we focus our attention on measuring narrative sentiment shifts in news and complementary text sources by exploiting a psychologically-grounded dictionary of ordinary English words designed to indicate whether narratives evoke approach or excitement rather than avoidance or anxiety.

In this vein we add to a growing literature exploiting text-based data sources to quantify sentiment and uncertainty in the economy and the financial system, sometimes with a view to forecasting macroeconomic and financial developments.¹ Economically or financially orientated word lists have been applied to content analysis of different text databases to try to predict various aspects of asset prices (e.g., Loughran and McDonald, 2011; Tetlock, 2007; Tetlock et al., 2008; Tetlock, 2011; Soo, 2013), to forecast macroeconomic variables (Thorsrud, 2018; Larsen and Thorsrud, 2019; Ardia et al., 2019; Rambaccussing and Kwiatkowski, 2020; Kalamara et al., 2020), and to capture both economic policy uncertainty (Baker et al., 2016) and other components of uncertainty (Larsen, 2020). The literature on measuring consensus is more limited but some

¹ In turn, this text-based literature is also related to a wider literature on how sentiment, as captured via surveys, market proxies or events, may affect financial markets and related opinion dynamics (e.g., Baker and Stein, 2004; Baker and Wurgler, 2006, 2007; Barsky and Sims, 2012; Brown and Cliff, 2005; Edmans et al., 2007; Lux, 2011; Greenwood and Nagel, 2009). But one advantage of text-based approaches is that they are available at high frequency and rapidly in real time.

recent work has attempted to use topic models such as latent dirichlet allocation to explore these issues (e.g., [Nimark and Pitschner, 2019](#)).

Our approach both draws on and departs from this existing literature in several ways. First, our sentiment measures follow those who use a dictionary approach. But we define the words in the dictionary with a theoretical determined filter, validated by psychological experiments ([Strauss, 2013](#)). The dictionary aims to capture solely pre-defined approach and avoidance words and is thus different to more general approaches as well as mitigating some of the difficulties associated with data mining, including the risk of obtaining seemingly significant correlations that do not generalize to new data or are highly context specific ([Lazer et al., 2014](#)). Second, and linked to the psychological microfoundations, our word dictionaries comprise solely common English language words and *do not contain* any economic or financial terms, even though the sources themselves are financial in nature. Third, our primary focus is specifically on gauging the systemic risk, rather than on movements in particular asset prices or broader macroeconomic developments. Fourth, much current research that applies some form of text-based sentiment analysis to study the economy or financial markets tends to exploit either newspaper or social media generated data. While we do exploit one news source, we also consider data sources more specifically connected to the financial system, including one source written within a central bank. Finally, we go beyond our dictionary-based sentiment metrics by applying machine learning techniques to develop a novel, but separate measure of the degree of topic consensus.

1.4. Organization of the paper

The remainder of the paper is structured as follows. [Section 2](#) explains the data and the methodology to construct our measure of sentiment. [Section 3](#) sets out our results and discusses relationships both with economic and financial variables and with some other atheoretic text-based based measures of sentiment and uncertainty. [Section 4](#) focuses on the measure of 'narrative consensus', explaining the methodology and results. [Section 5](#) discusses how both measures might complement more traditional indicators and analysis used in systemic risk assessment, and [Section 6](#) concludes.

2. Data and methodology

Based on the earlier discussion, we can suppose that in the case of a decision to invest or take risk, narratives evoking approach emotions must dominate the narrative, whereas in the case of a decision to disinvest or reduce risk-taking, avoidance narratives will dominate. Moving from a single investor to the economy as a whole, we may further suppose that at any given moment, there may be several competing narrative understandings of what is going on circulating among financial institutions and financial market participants, each evoking approach or avoidance emotions to produce an overall approach or avoidance tendency. Empirically, as and when dominant shared narratives shift, so that the emergence of one or other new narratives produces a change in overall approach or avoidance tendencies, these narrative shifts would be expected to manifest themselves in sentiment shifts and then in market movements or swings in risk-taking.

Our sentiment measure assumes that, as they emerge, some of these narratives, or pieces of them, will be contained within relevant text-based data sources, with common English language words evoking approach or avoidance used to describe them. Therefore, to measure narrative sentiment shifts, we select ordinary English words designed to indicate whether narratives evoke attraction or *approach* rather than repulsion or *avoidance*.²

In more ordinary language, the measure we introduce focuses on *excitement* about the potential gains from an action relative to *anxiety* about the potential losses. Thus, in the simplest case, the key variables of interest are the aggregate relative difference between *excitement* and *anxiety* and *shifts* in this difference over time.

We make use of a variety of data sources with a financial sector focus.

2.1. Bank of England internal market commentary

The Markets Directorate of the Bank of England produces a range of internal reports on financial markets and the financial system, some of which provide 'high-frequency' commentary on events and some of which provide deeper, or more thematic, analysis. For this study, we analysed some documents of the former kind, more specifically daily reports on the current state of markets, given that for the kind of analysis we employ here, the ideal type of data should remain as 'raw' as possible in order not to 'distort' the market emotions reflected within. These documents mainly cover financial news and how markets appear to respond to such news. We therefore expect these documents to correlate well with financial sentiment in the UK and potentially contain useful information on systemic risk.³

² Note, there is some difference between formulating distinctions between approach and avoidance (and anxiety and excitement) and positive and negative. Approach and avoidance have the advantage of a relatively unequivocal interpretation and expected brain response. Positive and negative are more complicated. For example, whilst anger is considered a negative emotion it is often associated with approach behaviours ([Carver & Harmon-Jones, 2009](#)), and whilst social approach motivation is associated with positive emotions at moderate levels, at high levels it may be associated with the negative emotions of separation distress ([Harmon-Jones et al., 2013](#)).

³ Note, the task of the staff of the directorate is to summarise news reaching them. Therefore, it may be important to note that one cannot conclude, that the content creators *themselves* had adopted as true any of the narratives portrayed in their documents.

Table 1
Emotion dictionary samples.

Anxiety		Excitement	
Jitter	Terrors	Excited	Excels
Threatening	Worries	Incredible	Impressively
Distrusted	Panics	Ideal	Encouraging
Jeopardized	Eroding	Attract	Impress

We analyse on average 26 documents per month from January 2000 until July 2010. The documents are typically relatively short, around 2–3 pages of email text. For the rest of the paper, we refer to these documents as ‘Market Commentary Daily (MCDAILY)’.

2.2. Broker reports

Broker research reports provide a large source of documents with relevance to financial markets and the macroeconomy. We analyse an archive of 14 brokers from June 2010 until June 2013, consisting of approximately 100 documents per month.

The documents are very long (up to 50 pages in some cases), and so we pick up on a large number of words. Visual inspection of a sample of these documents reveals that they primarily focus on macroeconomic developments in the major economies. We therefore expect the sentiment within these data to correlate most strongly with macroeconomic variables. Throughout the rest of the paper, we refer to this database as ‘Broker report (BROKER)’.

2.3. Reuters news archive

Finally, we use the Thomson-Reuters News archive, as also studied by Tuckett (2017) to assess macroeconomic trends. At the time of our analysis, the archive consisted of over 17 million English news wire articles.⁴ For most of this paper, we restrict our attention to news wire stories by Reuters in London during the period between January 1996 and September 2014, in which 6123 articles were published on average each month (after excluding all articles tagged by Reuters as ‘Sport’, ‘Weather’ and/or ‘Human Interest’). For the rest of the paper, we refer to this database as ‘Reuters (RTRS)’.

2.4. Relative sentiment shifts

A summary statistic of two emotional traits is extracted from our text data sources by a word count methodology described in more detail elsewhere (Tuckett et al., 2014). Two lists of previously applied and experimentally validated ordinary English words (Strauss, 2013), each of approximately size 150, are used, one representing *excitement* and one representing *anxiety*. Random samples of these words can be found in Table 1; a bigger random sample is included in the Appendix (section 1.1) and the full list is available on request from the authors.

For the summary statistic of a collection of texts T , we count the number of occurrences of excitement words and anxiety words and then scale these numbers by the total text size as measured by the number of characters (see Section 1.2 of the Appendix for the precise procedure).⁵ To arrive at a single Relative Sentiment Shift (RSS) statistic, we subtract the anxiety statistic from the excitement statistic, so that an increase in this relative emotion score is due to an increase in excitement and/or a decrease in anxiety. We compute this on a monthly basis.

$$\text{Sentiment}[T] = \frac{|\text{Excitement}| - |\text{Anxiety}|}{\text{Size}[T]} \quad (1)$$

As such, our RSS metrics should be viewed as first moment measures – of mean relative sentiment of a given period – rather than as second moment measures such as the VIX volatility index. But the relationship of RSS to such second moment measures is not clear a priori so subsequent sections compare RSS to other measures which are both first and second moment in their nature. The absolute magnitude of the RSS measure is unbounded and is influenced by writing style and, possibly, cultural norms. Since we do not adjust for the expected frequency of the words in the two emotion lists, we also do not know a-priori whether or not the measure will be negatively or positively centred. In practice, when comparing RSS series, we therefore scale them by subtracting the mean and dividing by the standard deviation. Empirically, we typically find RSS to be integrated of order one, so that the change in RSS from one period to the next tends to be stationary.

The simplicity of our method is intentional for three main reasons beyond the psychologically-motivated microfoundations. First, it is natural to consider whether simple text-based analysis which can be clearly understood and interpreted

⁴ Archive of Reuters news wire stories (<http://financial.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/ultra-low-latency-news.pdf>).

⁵ In some cases, it could be more suitable to scale by the number of documents. However, in this particular case, some documents contained tables and others did not, so the number of characters is a more appropriate choice.

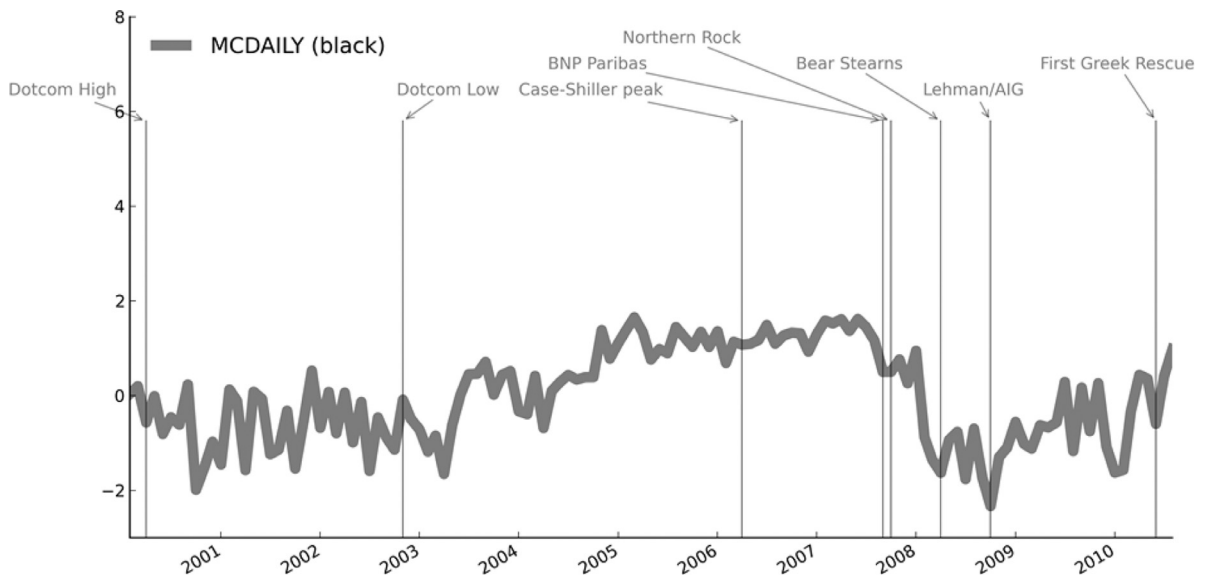


Fig. 1. Relative sentiment of MCDAILY. The y-axis displays the normalized values with 0 mean and standard deviation 1.

can be informative before moving to more complex methods. Second, simpler approaches might be more robust to including different data sources than more complex ones, as well as being more transparent. Finally, it also allows for an easier assessment of the robustness of the methodology. In particular, we can apply a bootstrap technique to compute 95% confidence intervals around the summary statistic. We sample new weights for each word in each dictionary (so that the sum of weights equals the size of the dictionary) and re-compute the statistic. Repeating the procedure gives a distribution from which to extract the confidence intervals. This technique gives us increased confidence that the meaning of individual words in our two lists does not change over time. Obviously, one can also imagine other methods of extracting confidence levels, e.g., to sample with replacement from the collection of texts.

As evidenced by the definition of the measure, we do not control for possible negations of these words (e.g. ‘not anxious’). We did, however, follow the procedure outlined by Loughran and McDonald (2011) to test (on the RTRS database) whether the presence of negation words would affect the sentiment series. Specifically, we excluded all words counted to produce the sentiment score if they were preceded within a window of three words by any of the words: “no”, “not”, “none”, “neither”, “never” or “nobody”. The resulting ‘negation aware’ series remained correlated with the original series as highly as 0.99, both in level and difference form (see Section 1.3 of the Appendix for further details). As discussed further in Section 1.4 of the Appendix, our results are also robust to the exclusion of words which could potentially have an economic meaning independent of emotional connotation.

Finally, we tested the robustness of the methodology to an alternative selection of sentiment wordlists. Specifically, we applied the sentiment methodology to the ‘positive’ and ‘negative’ wordlists produced by Loughran and McDonald (2011).⁶ A discussion comparing the two alternative RTRS series can be found in Section 3.1.

3. Results

3.1. The evolution of measures of sentiment

We explore the relative emotion series extracted from MCDAILY in Fig. 1 – a plot of the bootstrap 95% confidence bands around the MCDAILY series developed using the methodology described above can be found in section 1.5 of the Appendix. We annotate the chart with key events relating to financial stability for illustrative purposes – in particular, unlike event studies, we do not try to infer anything causal from the events that we depict on the charts.

The graph moves broadly as might be expected. In particular, it shows a stable increase during the mid-2000s. This is followed by a large and rapid decline from mid 2007, much of which occurs before the failure of Bear Stearns in March 2008 – strikingly, although this was already a period of turmoil in the financial system, the series hits very low levels before the worst parts of the crisis at around the time of the Lehman Brothers failure. As discussed below, these broad developments are also identified in formal structural break tests.

Although conviction narrative theory essentially refers to the *relative* level of sentiment – excitement minus anxiety – it is also interesting to consider the two component parts separately. Fig. 2 shows that the variation in anxiety levels is higher

⁶ The lists were downloaded from the website as they were available in 2011, www3.nd.edu/~McDonald/word_lists.html.

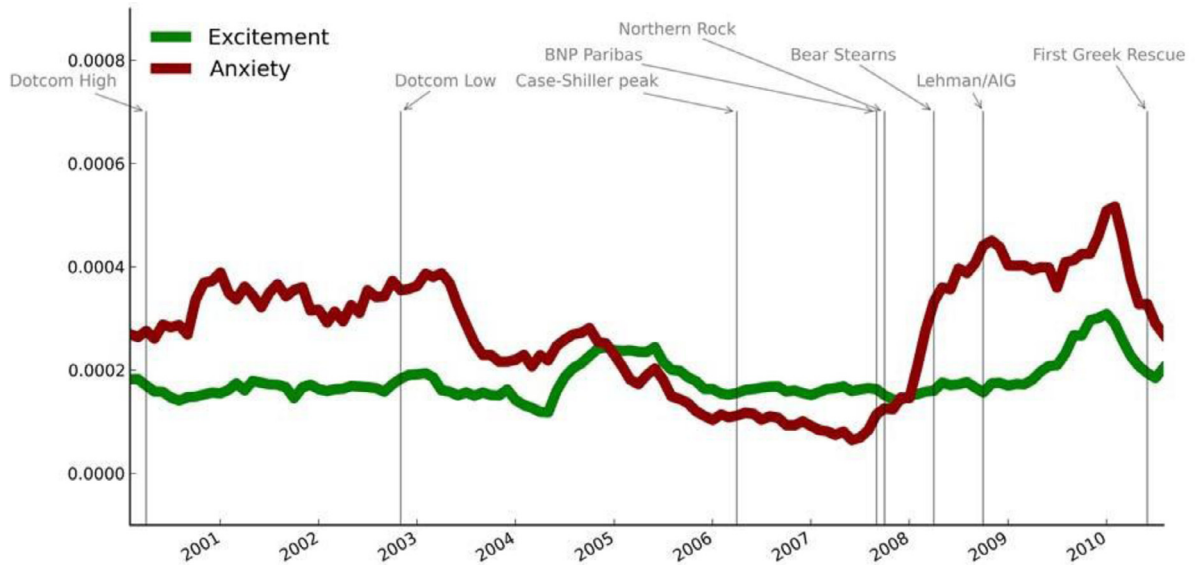


Fig. 2. Emotional factors of MCDAILY; anxiety (red) and excitement (green). The y-axis displays the individual aggregate word frequencies scaled by volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

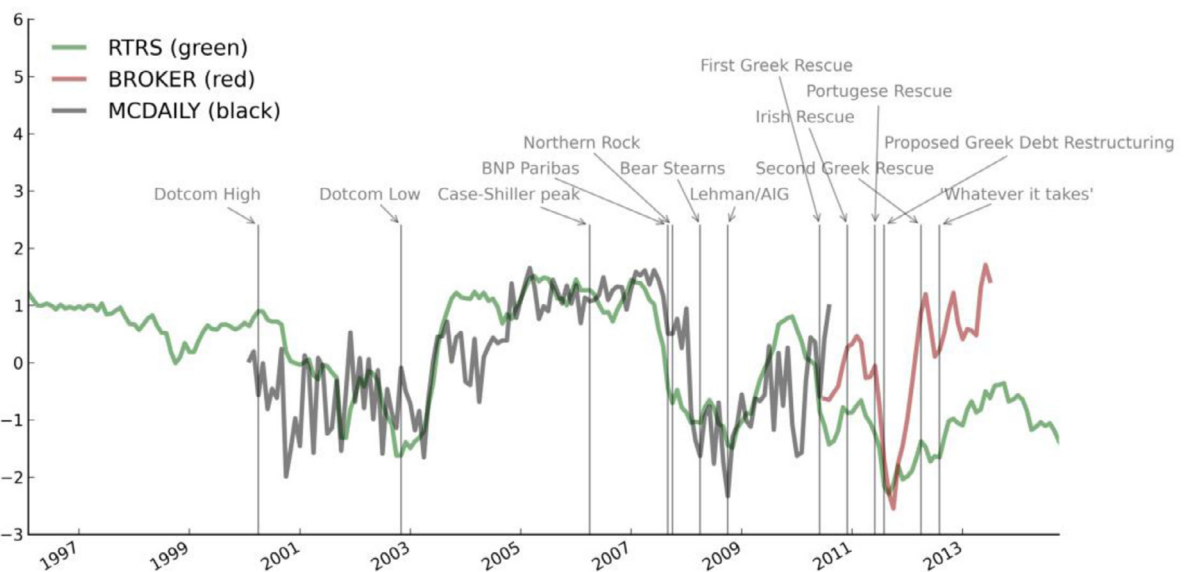


Fig. 3. Relative sentiment of MCDAILY (black), RTRS (green) and BROKER (red). The y-axis displays the normalized values with 0 mean and standard deviation 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

than that in excitement levels. This may reflect the fact that fear (or a lack of it) tends to drive movements in the financial system, consistent with heuristic-based approaches to Knightian uncertainty (Mousavi and Gigerenzer, 2014).

The MCDAILY series is compared with those extracted from the other two sources, namely RTRS and BROKER in Fig. 3. Each of the series is normalised with mean zero and standard deviation of 1 to facilitate comparison. The Fig. suggests that the series share a common trend. MCDAILY and BROKER are more volatile than RTRS (due to a much lower number of stories per month) and BROKER was available to us on a much shorter horizon than the other two archives. The exact correlations between the series are reported in Table 2 in Section 3.2.

Figs. 4 and 5 show the two component parts of the sentiment, excitement and anxiety, in RTRS and BROKER respectively. Again movements in anxiety appear to drive much of the fluctuation in overall sentiment.

Table 2
Correlations between relative sentiment series and common measures of sentiment, ignoring signs (-1 is t-1, +1 is t+1).

	MCD	RTRS	RTL	BRO	VIX(-1)	VIX	VIX(+1)	MCI(-1)	MCI	MCI(+1)	EPU(-1)	EPU	EPU(+1)	BoEU(-1)	BoEU	BoEU(+1)	CDS(-1)	CDS	CDS(+1)	PMI(-1)	PMI	PMI(+1)
MCD	1	0.59	0.60		0.54	0.62	0.66	0.24	0.26	0.27	0.30	0.43	0.41	0.43	0.54	0.61	0.63	0.67	0.63	0.38	0.38	0.43
RTRS		1	0.84	0.71	0.23	0.37	0.40	0.49	0.54	0.58	0.63	0.61	0.63	0.35	0.52	0.67	0.67	0.71	0.69	0.43	0.51	0.57
RTL			1	0.49	0.48	0.58	0.56	0.59	0.63	0.64	0.55	0.61	0.61	0.51	0.74	0.80	0.67	0.71	0.71	0.71	0.76	0.78
BRO				1	0.34	0.60	0.68	0.34	0.66	0.87	0.26	0.06	0.01	0.06	0.60	0.76	0.05	0.23	0.22	0.04	0.42	0.42

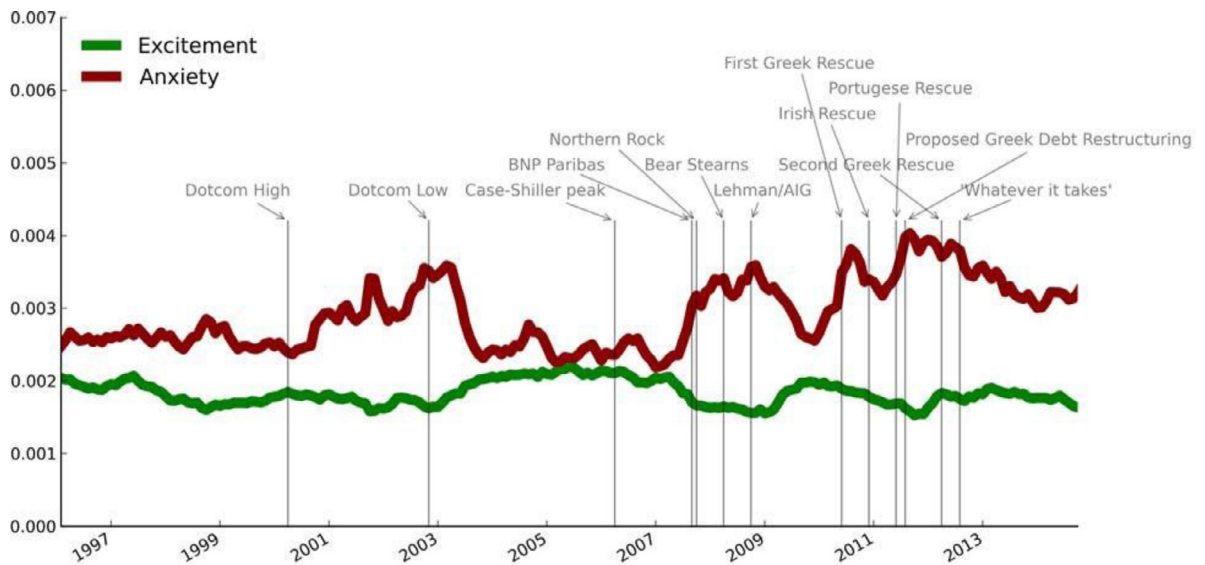


Fig. 4. Excitement (green) and Anxiety (red) in RTRS. The y-axis displays the individual aggregate word frequencies scaled by volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

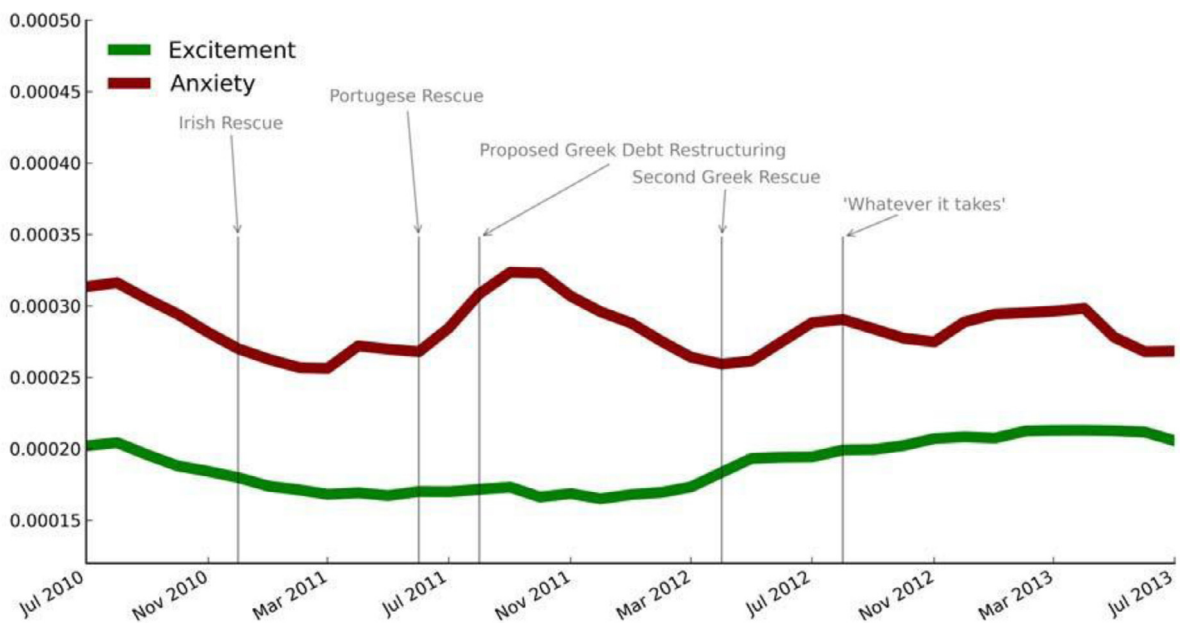


Fig. 5. Excitement (green) and Anxiety (red) in BROKER. The y-axis displays the individual aggregate word frequencies scaled by volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Sentiment Relating to Specific Topics

Thus far we have only discussed how the statistic can be extracted from a generic collection of texts. But it is also easy to filter for texts matching a given criteria, for example texts relating to a particular topic or entity. To illustrate this, we filtered for the mention of property-related terms in the Reuters’ news archive, using the terms “house prices”, “housing”, “property” and “real estate” to construct our filter. We chose to use a small set of related terms over a single term to increase the robustness of the series to temporal changes in language, though we also briefly compared our results with series constructed from using each term individually.

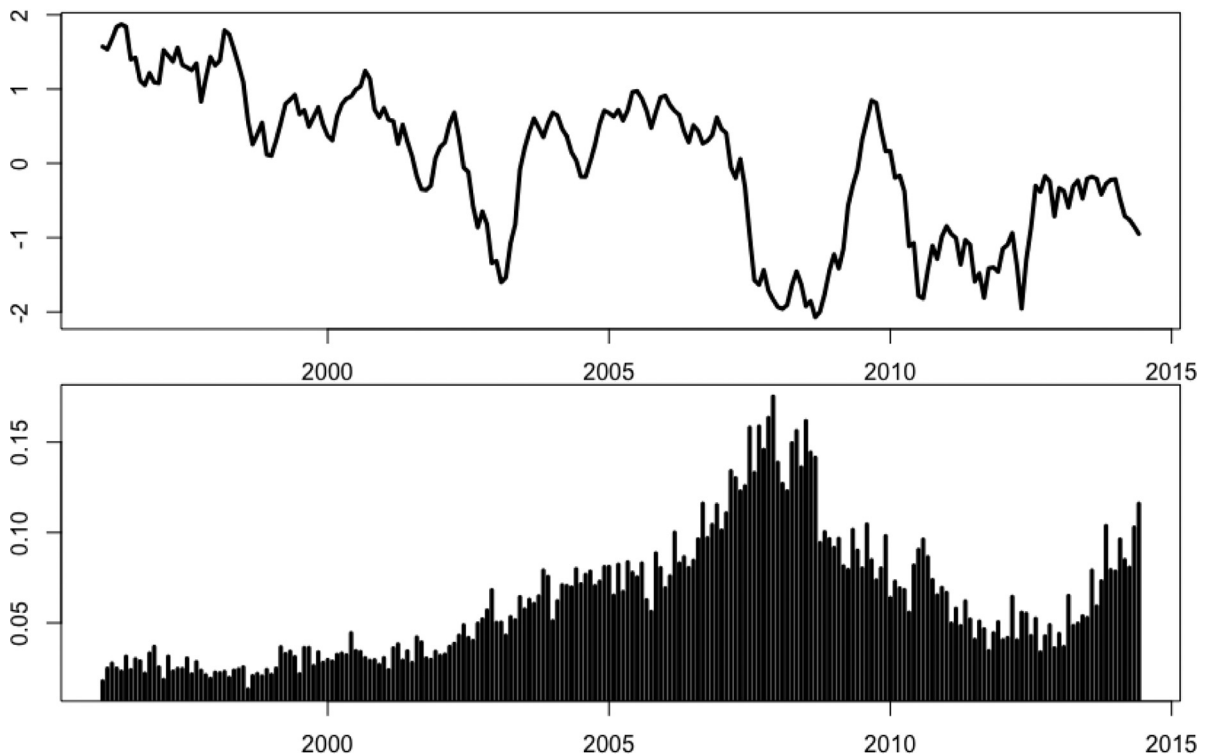


Fig. 6. Relative sentiment surrounding 'property' in RTRS (smoothed). The y-axis displays the normalized values with 0 mean and standard deviation 1.

Fig. 6 shows the results when we run the relative sentiment analysis only on articles containing at least one of the four property-related terms, with the number of articles shown in the bottom panel. The correlations between each individual series “house prices”, “housing”, “property” and “real estate” and the combined index are 0.59, 0.86, 0.86, 0.60 respectively. The average pairwise correlation between the four individual series is 0.47, indicating how the combined series may help to smooth out idiosyncrasies that may be associated with specific words.

It is particularly interesting to note the steady increase and later decline in the volume of property-related articles normalised by the total number of articles published in London, with the turning point occurring around the time of the bankruptcy of Lehman Brothers. The peak of the relative sentiment series appears to have occurred much before this, towards the end of 2006, after the series had undergone a steady increase for at least 4 years. The correlation between the property-based relative sentiment series and the more general index is 0.75. A test of Granger causality shows some evidence of causality going from the property series to overall RTRS (p -value 0.05) but not vice versa. This may be indicative of the fact that exuberance in property markets often precedes both wider financial sector exuberance and financial crises (Reinhart and Rogoff, 2009).

One may contend that results derived from a manual selection of terms ex-post benefit from hindsight as they cannot fully avoid future information potentially creeping into the index. But policymakers regularly identify potential risks in real time via a range of other means, including other economic indicators and models, market intelligence and judgement, and such risks can be used to guide the choice of terms. Many risks, including those related property, have also played a recurrent role in financial crises throughout history (Reinhart and Rogoff, 2009) and so general terms related to such risks could plausibly be generated ex ante. Therefore, we believe that such focused analysis remains of value if trying to monitor the emergence of exuberance in property markets, or indeed changes in risk-taking sentiment in any specific sector of the economy, provided that it avoids a focus on obvious time-specific terminology such as ‘subprime mortgages’. At the same time, rather than selecting words manually one could clearly use word embedding models to derive a set of related terms or apply unsupervised machine learning methods, like generative topic models, to identify topics that require special attention in real time, but we leave this for future work.

3.3. Structural breaks

As noted above, viewing the series at a low frequency, both the MCDAILY and RTRS show sharp falls well in advance of the financial crisis (we only have data on the third, BROKER, from 2010). For example, the mean value of MCDAILY over the

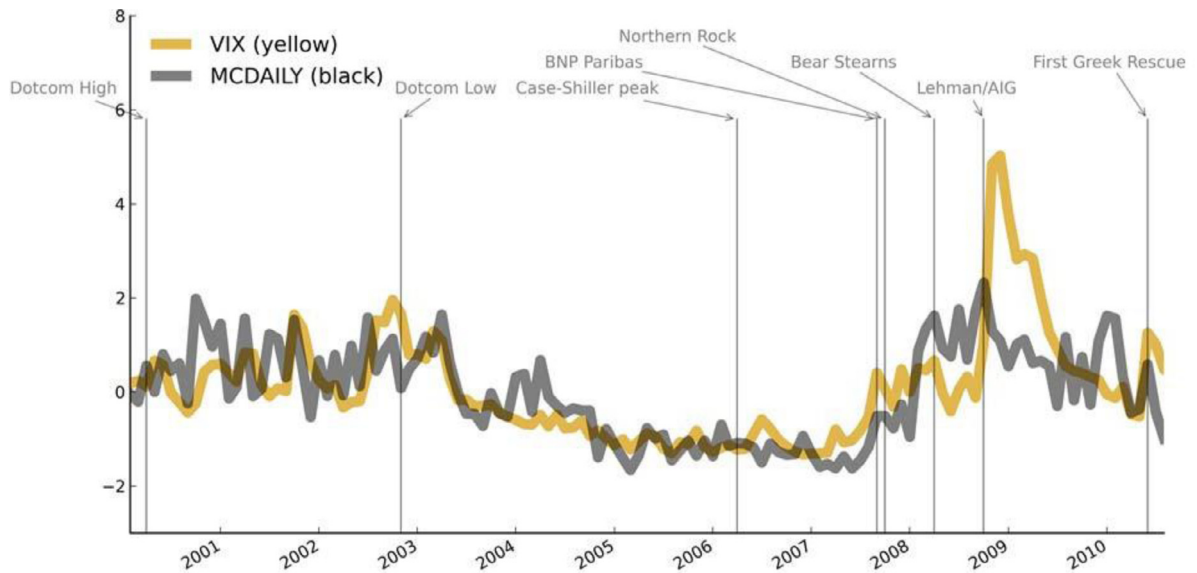


Fig. 7. Relative sentiment of MCDAILY, inverted for convenience, (black) compared to the VIX (yellow). The y-axis displays the normalized values with 0 mean and standard deviation 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

boom period July 2003 through June 2007 is 0.916, with a standard deviation of 0.567. The August 2007 value fell to 0.506, and in the second half of 2007, the mean value was 0.691. In January 2008, however, there was a sharp fall to -0.868 , 3.15 standard deviations below the mean of the July 2003 to June 2007 period, and the series continued to fall well in advance of the failure of Lehman Brothers.

The break in trends in the RTRS series was even earlier. Over the July 2003 to June 2007 period, this averaged 1.083 with a standard deviation of 0.472. As early as June 2007, the RTRS fell to -0.399 , 3.14 standard deviations below its 2003–2007 mean. By August 2007, it was 6.11 standard deviations below.

We conduct a simple formal statistical test for structural breaks in the sentiment series for which we have pre-crisis data (MCDAILY and RTRS) using the method of Bai and Perron (2003). The number of breakpoints m is estimated using Bayesian Information Criterion (BIC) and their positions are estimated by minimising the residual sum of squares of the $m+1$ resulting line segments. We set a maximum number of breakpoints to be found to 5.

We find four structural breaks for the RTRS series in August 2000, May 2003, May 2007 and April 2010 and three for the MCDAILY series in April 2003, November 2004 and December 2007. The break in August 2000 resulted in a dramatic negative shift in the RTRS series (it is only very soon after the MCDAILY series starts). This can most likely be attributed to the burst of the dotcom bubble. The breaks in April and May 2003 are likely to represent a dramatic positive shift in the series out of the post-dotcom slump, though it is also plausible that the swift conclusion of the Iraq war may have contributed to the rise in sentiment. The break in November 2004 in MCDAILY divided the period from April 2003 to the peak before the shift in December 2007 resulting in a more gradual increase than we can observe in the relative sentiment of RTRS. The breaks in both series in May and November 2007 mark the beginning of a dramatic negative shift in sentiment prior to the financial crisis. And the final break in the RTRS series in April 2010 likely relates to the European debt crisis and problems in Greece. This results in a further drop in the series to its absolute lowest levels.

The relative sentiment surrounding 'property' in RTRS exhibit three structural breaks leading to four regimes: first a negative break in October 2000, followed by a positive break in August 2003 and finally a negative break in February 2007, showing an early indication of the recession to follow.

3.4. Comparison with economic and financial metrics

To illustrate how our measures compare with commonly used economic and financial metrics, Fig. 7 shows MCDAILY plotted against the VIX, with the MCDAILY variable inverted for ease of comparison. It is clear that the measures track each other closely.

More generally we can look at correlations between a wide range of economic and financial variables of sentiment, confidence and uncertainty. To explore this, we computed simple pairwise correlations. The correlations with the Michigan

Consumer Sentiment index (MCI),⁷ the VIX,⁸ the economic policy uncertainty index (EPU) of Baker *et al.* (2016),⁹ the Bank of England macroeconomic uncertainty index for the UK (BoEU – see Haddow *et al.* (2013)),¹⁰ senior CDS premia¹¹ and PMI,¹² together with the correlations between the individual relative sentiment series at a monthly frequency are presented in Table 2.¹³ To facilitate comparisons and since the sign of all correlations is as expected, we give absolute values. We also report on lead and lag correlations between the sentiment series and the financial and economic indicators. It is clear that our three sentiment measures are fairly highly correlated with all of the other measures and that the correlations tend to be stronger when the sentiment variables lead the other variables than when they lag them. To illustrate the short-run dynamics of the series, Table 3 presents correlations between the first order differences of the variables. The main columns of interest in this table are those with the suffix '+1', which show to what extent sentiment can be used as a *leading* indicator. In other words, these columns indicate the extent to which a change in the sentiment variable precedes a change in the economic variable. In some cases, the signs become less interpretable due to negative autocorrelations of some of the first order differences. The correlation between BROKER at time t and MCI at time $t+1$ is particularly high.

To formally test potential lead-lag relationships we report the results of Granger causality tests between the three RSS sentiment series and the various other indicators considered above. We use the methodology described in Toda and Yamamoto (1995). Although this method does not distinguish between long-run and short-run dynamics, it has the advantage of requiring fewer statistical tests, which is helpful given the fairly large number of causality tests we undertake. The main step is to first add lags to the VAR specification equal to the largest order of integration of the two series and then to include further lags as necessary to make the residuals behave like white noise. The full procedure and supporting tables are presented in the Appendix (sections 2 and 4). Here, we simply report the final step in the process which provides the evidence on the existence or otherwise of Granger causality. In summary, this method tests for Granger causality in level form, i.e., without reducing the variables to stationary series, and thus tests for both long-run and short-run causality.

We carry out tests using the unsmoothed aggregate versions of each of the sentiment series (i.e., net balance between excitement and anxiety). We use monthly variables for all tests except for those involving BoEU. In those cases, we create quarterly versions of the monthly series by averaging the months in a quarter. Table 4 below shows results obtained testing Granger-causality from the various versions of the RTRS, BROKER and MCDAILY variables to MCI, VIX, BoEU, EPU, CDS and PMI. Table 5 below shows results obtained testing Granger-causality between the same variables in the reverse direction, from MCI, VIX, BoEU, EPU, CDS and PMI to the various versions of the RTRS, BROKER and MCDAILY variables.¹⁴

As well as testing for Granger causality between the series in level form, we also test for causality between the first differences of the series. Table 6 and 7 set out the corresponding p -values.

Overall, there is some evidence of Granger causality from our text-based sentiment measures to the metrics we consider but less causality in the opposite direction. In particular, the RTRS measure is significant in many of the tests. As we might expect, RTRS and BROKER, sources more reflective of broad macroeconomic commentary, appear to relate most closely to the MCI and PMI, which are the most macroeconomic measures of comparison. By contrast, MCDAILY, a source which reflects financial market commentary, exhibits much lower p -values in relation to the VIX, CDS premia and BoEU measures.¹⁵

As well as being suggestive of the robustness and usefulness of these measures, these results are indicative of the potential use of the relative sentiment measures as short-term forecasting devices in addition to their possible usefulness at a lower frequency as discussed above. Indeed, in a horse race comparing several different text-based measures, Kalamara *et al.* (2020) find some evidence that a metric constructed by applying the dictionary used in this paper to UK newspapers can yield near-term out-of-sample forecast improvements for some macroeconomic variables such as business investment and unemployment when comparing against certain standard benchmark forecasting models.

⁷ The MCI was created as a means to assess consumers' ability and willingness to buy. The survey is carried out with at least 500 phone interviews, during a period of around 2 weeks, in which approximately 50 questions are asked. Survey results are released twice each month at 10.00 a.m. Eastern Time: preliminary estimates are published usually (variations occur during the winter season) on the second Friday of each month, and final results on the fourth Friday.

⁸ The VIX, sometimes known as the 'fear' index, is a measure of implied volatility derived from the price of S&P500 options. We consider an average of VIX, computed using closing prices of all trading days for a given month, thus making the series comparable to the relative sentiment series, which are also monthly 'averages'.

⁹ We use the UK version of the series available at http://www.policyuncertainty.com/europe_monthly.html. The series starts in January 1997.

¹⁰ This is a composite measure derived from seven measures: FTSE option-implied volatility, sterling option-implied volatility, dispersion of company earnings forecasts, dispersion of annual GDP growth forecasts, GfK unemployment expectations balance, CBI 'demand uncertainty limiting investment' score, and number of press articles citing 'economic uncertainty'.

¹¹ Senior CDS premia accessible from the Bank of England's set of core indicators of financial stability <https://www.bankofengland.co.uk/-/media/boe/files/core-indicators/countercyclical-capital-buffer.xlsx>. The series starts in January 2003.

¹² Business expectations survey (Markit PMI). Based on answers to the question of whether business activity is expected to be higher, lower or stay the same in 12 months. The series starts in April 1997.

¹³ Correlations are computed on the full available range of overlapping data. Here MCD = MCDAILY and BRO=BROKER. Since the BoEU index is a quarterly series we create quarterly series of the three sentiment indicators by averaging the values within each quarter when computing correlations involving BoEU. All other correlations are based on monthly data.

¹⁴ The missing entries in both tables could not be determined because of some form of VAR misspecification.

¹⁵ As a further exercise, we also conducted tests against an estimate of the variance risk premium from Kaminska and Roberts-Sklar (2015) which builds on the methodology of Bollerslev *et al.* (2009). But we found no evidence of Granger causality in either direction. In principle, one could also test our metrics against other measures of tail risk, including those constructed from option prices (eg; Anderson *et al.*, 2015).

Table 3Correlations between first order differences of relative sentiment series and common measures of sentiment (-1 is $t-1$, $+1$ is $t+1$).

	MCD	RTRS	RTLTM	BRO	VIX(-1)	VIX	VIX(+1)	MCI(-1)	MCI	MCI(+1)	EPU(-1)	EPU	EPU(+1)	BoEU(-1)	BoEU	BoEU(+1)	CDS(-1)	CDS	CDS(+1)	PMI(-1)	PMI	PMI(+1)
MCD	1	0.48	0.39		0.20	-0.13	-0.17	-0.05	0.02	-0.01	0.13	-0.28	-0.01	0.19	-0.15	-0.33	0.08	-0.31	-0.11	-0.03	-0.08	0.16
RTRS		1	0.74	0.53	0.27	-0.37	-0.15	-0.16	0.12	0.14	0.13	-0.25	0.00	0.38	-0.15	-0.52	0.04	-0.31	0.08	-0.10	0.13	0.27
RTLTM			1	0.49	0.07	-0.44	-0.11	-0.09	0.21	0.09	0.02	-0.26	-0.03	0.24	-0.59	-0.44	-0.01	-0.24	0.01	0.02	0.21	0.28
BRO				1	-0.09	-0.37	-0.17	-0.29	0.17	0.66	0.04	-0.23	0.03	0.51	-0.27	-0.57	0.03	-0.38	0.22	-0.08	0.44	0.18

Table 4

Wald test p-values of Granger-causality from the relative sentiment shift series RTRS, BROKER and MCDAILY to MCI, VIX, BoEU, EPU, CDS and PMI.

RSS Series	MCI	VIX	BoEU	EPU	CDS	PMI
RTRS	0.005**	0.28	4e-06**		0.3	0.0002**
MCDAILY	0.5	0.09	6e-05**	0.05*	0.09	0.06
BROKER	2e-11**	0.18		0.92	0.6	0.1

Note: *p<0.05; **p<0.01.

Table 5

Wald test p-values of Granger-causality from MCI, VIX, BoEU, EPU, CDS and PMI to the relative sentiment shift series RTRS, BROKER and MCDAILY.

RSS Series	MCI	VIX	BoEU	EPU	CDS	PMI
RTRS	0.29	0.093	0.022*		0.57	0.08
MCDAILY	0.95	0.39	0.58	0.18	0.89	0.49
BROKER	0.94	0.16		0.73	0.97	0.41

Note: *p<0.05; **p<0.01.

Table 6

Wald test p-values of Granger-causality from the first difference of relative sentiment shift series RTRS, BROKER and MCDAILY to MCI, VIX, BoEU, EPU, CDS and PMI.

RSS Series	MCI	VIX	BoEU	EPU	CDS	PMI
RTRS	0.0053**		0.001**		0.18	
MCDAILY	0.66	0.96		0.26		0.11
BROKER		0.42		0.0068**	0.51	0.3

Note: *p<0.05; **p<0.01.

Table 7

Wald test p-values of Granger-causality from the first difference of MCI, VIX, BoEU, EPU, CDS and PMI to the relative sentiment shift series RTRS, BROKER and MCDAILY.

RSS Series	MCI	VIX	BoEU	EPU	CDS	PMI
RTRS	0.21		1.4e-06**		0.34	
MCDAILY	0.87	0.41		0.25		0.4
BROKER		0.96		0.98	0.61	0.31

Note: *p<0.05; **p<0.01.

3.5. Comparison with Loughran and McDonald sentiment index

As well comparing our novel sentiment index against economic and financial metrics, it is also natural to consider how it compares against other commonly used text-based sentiment measures. This helps both to provide further external validity and serve as a robustness check. While a full forecasting horse race is beyond the scope of this paper (see [Kalamara et al., 2020](#), for an example of this), we conduct some simple exercises to compare the sentiment indicators generated using the original Loughran and McDonald positive and negative word lists (downloaded from the University of Notre Dame website in 2014) to those generated by our word lists. An important difference between our RSS lists and the lists proposed by Loughran and McDonald is that the RSS lists have remained the same since they were initially constructed, making it easier to make comparisons across various studies and databases.

We undertake a number of comparisons and checks: (i) we look at the word lists to qualitatively identify any key differences, e.g., in terms of 'topics' represented; (ii) we compare the timings of large movements in the two series; (iii) we compute lead-lag correlations between the two series; and (iv) we carry out a Granger causality test between the two series.

A key feature of the RSS lists is that we have tried as far as possible to exclude words that could be directly linked with economic or financial activity. This is in contrast with Loughran and McDonald who explicitly included terms such as 'recession', 'crisis', 'collapse', 'illiquid', 'illiquidity', 'liquidate', etc. We hypothesise that including words such as these risks making the resulting sentiment index contemporaneous with certain economic events and therefore less likely to offer predictive signals.

[Table 8](#) compares the movements of the two series around key financial turning points in the first half of 2003 and the second half of 2007, with the sentiment data scaled for each series by subtracting its mean and dividing by its standard

Table 8
Movements of scaled sentiment around key financial turning points in the first half of 2003 and the second half of 2007. RTLM generated using Loughran & McDonald and RTRS generated using RSS.

2003	RTLM	RTRS	2007	RTLM	RTRS
Feb 2003	-0.85	-1.06	Apr 2007	1.33	0.62
Mar 2003	-1.27	-0.77	May 2007	1.28	0.78
Apr 2003	0.07	0.30	Jun 2007	0.83	-0.41
May 2003	0.46	0.44	Jul 2007	0.36	-0.41
Jun 2003	0.81	1.68	Aug 2007	-1.40	-1.81
Jul 2003	1.09	1.38	Sep 2007	-1.13	-1.20

deviation, this making the scales comparable and implying that positive numbers are above the mean of the respective series and vice versa.

We have highlighted in bold the months for which the RSS series appears to shift in advance of the Loughran and McDonald series. Perhaps most notably, RSS drops sharply below its mean in June 2007, whereas it takes another two months for the LM series to do the same. There is also some evidence that RSS shifts upwards a month earlier in 2003.

Tables 2 and 3 show estimated lead-lag correlations between the various RSS measures as well as the LM version of RTRS (RTLM). The correlation between RTRS (RSS) and RTLM is 0.84 in level form (Table 2) and 0.74 in first difference form (Table 3). In broad terms, the series therefore appear to follow a similar pattern, which implies the approaches are both capturing some common latent factor, thus providing external validation of our approach. However, slight but significant differences still remain as also discussed above. In addition, Table 3 showing short-term lead lag relationships between changes in sentiment and various economic variables suggests a slight advantage in using RSS to foresee changes in the economy. In particular, the columns with suffix '+1' show correlations of changes in sentiment at time t and the economic variables at time t+1. In most cases, the absolute value of that correlation is higher for RSS than it is for LM.

Finally, we carried out a Granger causality test between the two series, following the methodology laid out in section 3.4. The p-value of causality from RSS to LM is 0.24 and in the reverse direction it is 0.64. We can therefore conclude that over the full period there is no evidence of a systematic lead or lag between the two metrics. This result was largely expected since most of the time, the various economic terms present in the LM dictionaries should have little to no effect on the sentiment series, apart from the times when significant economic events occurred.

Overall, the comparison with the Loughran and McDonald index provides a degree of external validity to our novel metric. At the same time, we find tentative evidence that the novel metric we develop may have slightly stronger leading properties, particularly around key financial turning points.

3.6. Effect of relative sentiment on the UK economy

We now explore the relationship between relative sentiment and economic activity in the context of a VAR model. A range of previous work uses VAR models to estimate the effect of uncertainty on the economy (e.g. Bloom (2009), Haddow (2013) and Baker et al (2016)). It is commonly found that shocks to such (proxy) measures of uncertainty have a significant and negative impact on economic activity.

To estimate the empirical effect of relative sentiment on the UK economy we estimate a monthly VAR(p) over the period Jan 1996 through September 2014. We adopt the same model specification as Baker et al (2016), using a Cholesky decomposition with the following ordering to recover orthogonal shocks: RSS, as measured by RTRS, the log of the relevant stock price index, the official interest rate, log employment, and log industrial production.¹⁶ We use the RTRS series in this exercise because it is the longest relative sentiment series of the three and because emotions expressed in general economic, financial and business news are arguably more likely to be related to economic activity than, for example, financial market commentary.

The model is specified as the following VAR(3), with lag parameter $p = 3$ chosen using AIC:

$$\begin{bmatrix} RSS_t \\ LFTSE_t \\ R_t \\ LEMP_t \\ LPROD_t \end{bmatrix} = A_1 \begin{bmatrix} RSS_{t-1} \\ LFTSE_{t-1} \\ R_{t-1} \\ LEMP_{t-1} \\ LPROD_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} RSS_{t-2} \\ LFTSE_{t-2} \\ R_{t-2} \\ LEMP_{t-2} \\ LPROD_{t-2} \end{bmatrix} + A_3 \begin{bmatrix} RSS_{t-3} \\ LFTSE_{t-3} \\ R_{t-3} \\ LEMP_{t-3} \\ LPROD_{t-3} \end{bmatrix} + \varepsilon_t \tag{2}$$

where RSS_t is the monthly relative sentiment shift series for the UK, $LEMP_t$ is log employment, $FTSE_t$ is the FTSE 100 index, $LPROD_t$ is log production, and R_t is the level of Bank Rate. We test the stability of the impulse responses by ordering RSS last in the VAR and including a deterministic trend.

We also follow Baker et al. (2016) in considering shocks to RSS which are equivalent to the difference between the mean value in 2005–2006 and the mean value in 2011–2012 – periods either side of the crisis dominated by relatively stable and

¹⁶ Note that RTRS is integrated of order one so the VAR specification is sensible.

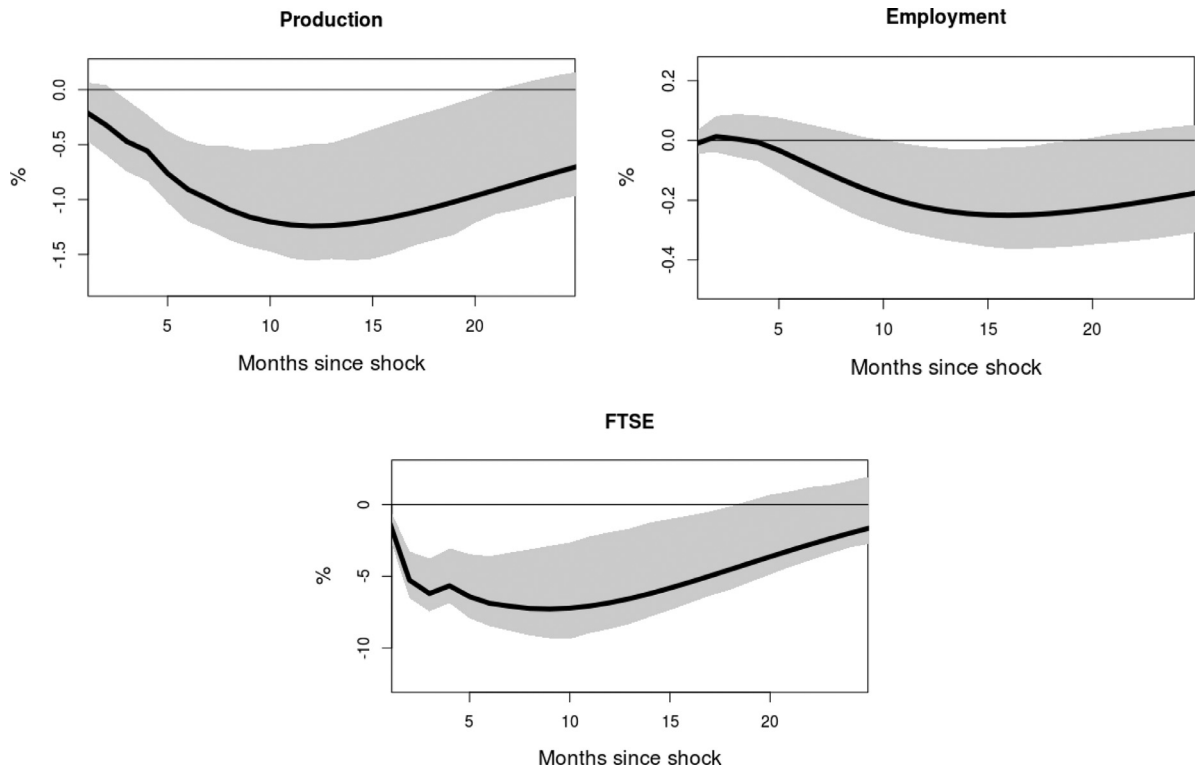


Fig. 8. Impact of a negative shock to RSS on industrial production, employment and the FTSE 100 with 95% confidence intervals.

high levels of RSS and by volatile and low levels of RSS respectively. The difference between the two periods represents 2.5 standard deviations of RSS.

Fig. 8 shows the impulse response of a 2.5 standard deviation negative shock to RSS on industrial production, Employment and the FTSE 100 index. These provide indicative evidence that RSS has a significant impact on industrial production, employment and the UK stock market. In each case, the effect lasts for about 20 months.

The maximum increase in the FTSE resulting from a 2.5 standard deviation shock in RSS is 7.28%. For employment and industrial production the corresponding figures are 1.24% and 0.25% respectively. Over the same two periods, the EPU index increased by approximately 2 standard deviations (standard deviation measured over the January 1996 through September 2014 period). Baker et al. (2016) show how such a shock results in a maximum of 1.1% drop in industrial production and a maximum of 0.35% drop in employment. Robustness tests show how adding a trend or changing the order of the variables by placing RSS last instead of first does not materially change the impulse responses.

4. Measuring consensus

We turn now to our second, more exploratory, line of investigation: can we measure structural changes in the variability of narratives – in particular, at a given point in time, is there consensus over particular narratives or a wide dispersion of narratives (disagreement)? The objective is to investigate if we can detect when some narratives grow to become dominant, arguably to the detriment of the smooth functioning of the financial system and potentially hinting at impending distress if also associated with strongly positive aggregate sentiment. Here a narrative is defined as a collection of articles and consensus is defined as a ‘lack of a natural division’ of articles into clearly defined separate groups. We are therefore not referring to consensus in terms of ‘views’ or ‘opinions’, but rather as ‘concentration of articles’. We introduce a novel methodology to explore this.

4.1. Methodology

For this investigation, we focus on RTRS as it generally seems to perform well and has a larger sample than the other sources, which is helpful for the techniques we apply. To measure consensus, we make use of modern information retrieval methods. The main challenge is to find a good methodology for automatic topic detection. Many such approaches exist in the literature (Berry, 2004), but we rely on the straightforward approach of clustering the articles in word-frequency space (after removal of commonplace words) to form topic groups, whereby each article belongs to a single distinct topic. We then

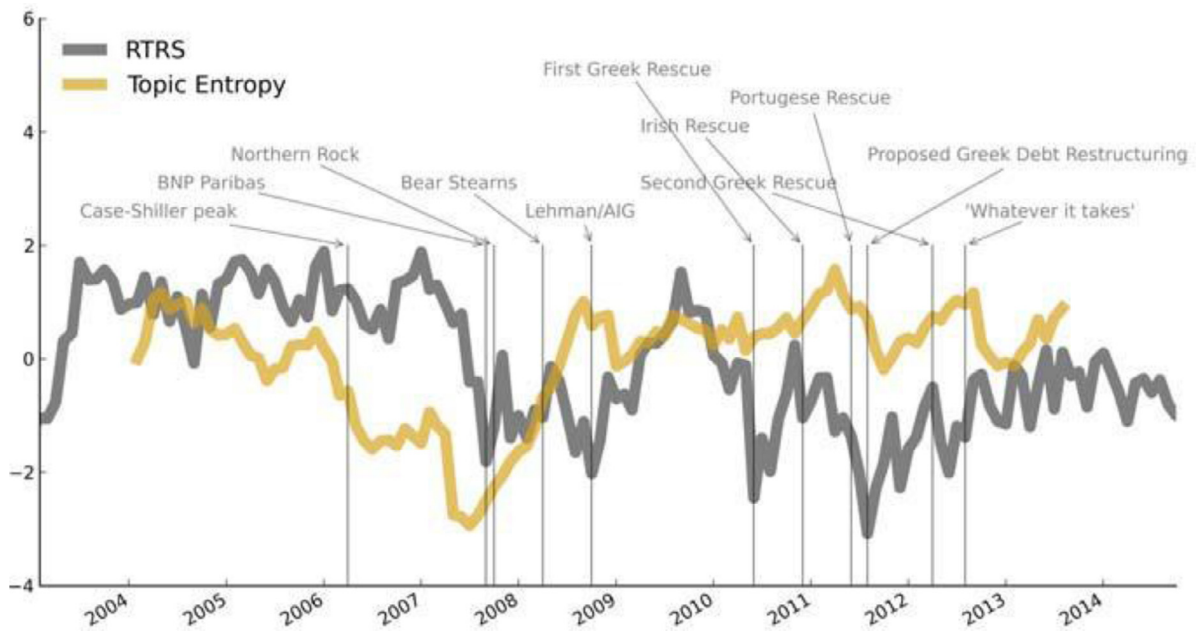


Fig. 9. Relative sentiment (black) and entropy (yellow) in Reuters' London news. The y-axis displays the normalized values with 0 mean and standard deviation 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 9

Wald test p-values of Granger-causality from Entropy to VIX, BoEU and EPU.

	VIX	BoEU	EPU
Entropy	0.12	9.1e-06**	0.03*

Note: * $p < 0.05$; ** $p < 0.01$.

measure the uncertainty (entropy) in the distribution of the articles across the topic groups. We consider an increase in the uncertainty (entropy) of the topic distribution as a decrease in consensus and vice versa. The details and justification of the construction can be found in the Appendix (Section 3).

4.2. Results

We plot the narrative consensus found in RTRS in Fig. 9.¹⁷ The graph shows a clear increase in consensus (decrease in entropy) preceding the crisis period and much more disagreement subsequently. Having decomposed the narrative discourse into one index measuring shifts in emotion (the previous section) and another measuring structural changes in consensus (entropy), it appears from these results that a predominantly excited consensus emerged prior to the crisis, driven by low levels of anxiety. This seems consistent with the convergence of beliefs on the idea that a new paradigm could deliver permanently higher returns in the financial system than previously without threatening stability. With the onset of the crisis, this eventually shifted into predominantly anxious disagreement, as might be expected in an environment of fear and uncertainty. Interestingly, however, the narrative consensus series peaks in mid-2007, just as anxiety starts to dominate. Exploring sample articles and frequent words from the largest topic cluster at this time reveals a common theme about weak credit conditions and economic uncertainty.

An analysis of structural breaks, using the same methodology as previously, reveals two breaks resulting in three main regimes: one downwards break March 2006 and one upwards break in March 2008. In other words, the series moved into a regime of relatively high consensus as early as March 2006 and remained in this regime for approximately 2 years.

Table 9 shows the outcome of Granger causality tests between the entropy series and other measures which speak to uncertainty: the VIX, the BoEU and the EPU. The p-values of Granger-causality in the converse direction can be seen in Table 10.¹⁸ According to these tests, there is evidence of Granger causality from the entropy variable to both the BoEU index

¹⁷ The series spans a shorter period than the full RTRS series, focussing only on the period around the crisis, for two main reasons: (i) the analytics are computationally very intense with run times taking days; and (ii) the method suffers when there are not much data available (i.e., in the pre 2000 period).

¹⁸ We use monthly variables in each case except those involving BoEU for which we produce a quarterly entropy series by averaging the months within the quarter. We also tested against the variance risk premium but did not find any evidence of causality in either direction.

Table 10
Wald test p-values of Granger-causality from VIX, BoEU and EPU to Entropy.

	VIX	BoEU	EPU
Entropy	0.27	0.98	0.52

Note: * $p < 0.05$; ** $p < 0.01$.

and the EPU. This is indicative of how our measure of consensus / disagreement may give an early signal of changes in uncertainty.

Overall, the consensus series captures both the presence of predominantly excited consensus and predominantly anxious consensus. This highlights how the two measures, of emotion and narrative consensus, might therefore beneficially be interpreted side by side.

5. Discussion

Our results highlight how our measures of sentiment and narrative consensus correlate well with, and in some cases even appear to ‘cause’, certain economic and financial variables. Depending on the text source, some perform better with financial variables, some with macroeconomic variables. Although left to future work, this suggests that the text sources linked more closely to macroeconomic variables could also be useful in forecasting or ‘now-casting’ economic activity (Tuckett and Nyman, 2017; Kalamara et al., 2020). At a lower frequency, and with the benefit of hindsight, the metrics appear to signal rising concerns prior to the global financial crisis. As we now discuss, this highlights the possible value of these indicators for signalling the potential for financial system stress.

There are many different approaches for identifying and modelling threats to the financial system, including the use of stress tests, early warning models, composite indicators of systemic risk, and Merton-based models of systemic risk that use contingent claims analysis.¹⁹ Many authorities use indicator dashboards, including the European Systemic Risk Board, the Office of Financial Research in the United States and the World Bank.²⁰ In the United Kingdom, the Financial Policy Committee routinely reviews a set of core indicators which have been helpful in identifying emerging risks to financial stability in the past, and which therefore might be useful in detecting emerging risks (Bank of England, 2014).²¹

Recognising that no single set of indicators or models can ever provide a perfect guide to systemic risk, judgement also plays a crucial role in specifying any macroprudential policies to tackle threats to the financial system. And qualitative information, including from market and supervisory intelligence, typically helps to support such judgements.

As we have shown, our measures of sentiment and consensus, extracted from text-based information, appear to be informative of episodes of emerging systemic risk and high market volatility. As such, they offer a potential mechanism for extracting quantitative metrics from qualitative, text-based information that is used to inform policy making and might therefore be one component of indicator dashboards, complementing other approaches used to detect systemic risk. These measures could also be calculated on a real-time basis, offering them an important advantage over some more conventional indicators. Arguably, they are also likely to be more robust to the Lucas (1976) critique because the writers of individual documents are very unlikely to respond collectively by adapting their writing tones or styles because an indicator based on vast numbers of documents is used as one guide for helping to set policy.

At the same time, it is clearly important to test these indicators further. For example, which particular text-based sources should be the focus of attention, how good are the metrics in distinguishing signal from noise, and how do they compare with more conventional indicators in this respect? We leave these questions for further work.

6. Conclusion

In this paper, we have explored the potential of using algorithmic text analysis, applied through the lens of conviction narrative theory, to extract quantitative summary statistics from novel data sources which have largely only been used qualitatively thus far. We have demonstrated that our approach can lead to some intuitive and useful representations of financial market sentiment. At a low frequency, the shifts correlate well with financial market events, with key shifts identified by structural break tests. And at a higher frequency, they appear both to lead other commonly used measures of sentiment, confidence and volatility, and to be informative for developments in the economy.

¹⁹ See Aikman et al. (2009), Kapadia et al. (2013) and Bank of England (2015) on stress testing models and approaches developed at the Bank of England. On early warning indicator models, see Kaminsky and Reinhart (1999), Drehmann et al. (2011), Borio and Lowe (2002, 2004), Barrell et al. (2010), Schularick and Taylor (2012) and Giese et al. (2014). On composite indicator models, see Illing and Liu (2006) and Holló et al. (2012). On contingent claims models, see Gray et al. (2008) and Gray and Jobst (2011).

²⁰ For the US, see Section 3 of the OFR Annual Report (2012). The ESRB’s Risk Dashboard is published on the web (see <http://www.esrb.europa.eu/pub/rd/html/index.en.html>).

²¹ See also Giese et al. (2014).

We have also developed a novel methodology to measure consensus in the distribution of narratives. This metric can potentially be used to measure homogenisation in the financial system. Greater consensus, when viewed together with an increase in sentiment, may also be interpreted as an increase of predominantly excited consensus of narratives prior to the global financial crisis. Thus, we appear to have found novel empirical evidence of groupfeel and the build-up of systemic risk-taking behaviour leading up to the financial crisis.

Overall, our metrics may be useful in gauging risks to financial stability arising from the collective behaviour discussed. While further work is needed to refine these metrics, including in relation to both the methods and the data inputs used, they have the potential to provide a useful quantitative, analytical perspective on text-based market information which could help to complement more traditional indicators of systemic risk.

Acknowledgements

The views expressed in this paper are solely those of the authors and should not be taken to represent those of the Bank of England or the European Central Bank.

The paper draws very extensively on the Bank of England Staff Working Paper (Nyman *et al.*, 2018) that we wrote together with David Gregory, Paul Ormerod and Robert Smith. We thank them for their tremendous contribution to this work.

We would also like to thank Jon Rand and Chris Hare at the Bank of England for providing access to data and Eric Fishkin, Chrystia Freeland, Richard Brown and Maciej Pomalecki of Thomson Reuters for arranging access to the Reuters News archive. Thanks are also due to David Aikman, Saleem Bahaj, Johan Bollen, Oliver Burrows, Jennie Castle, Ambrogio Cesa-Bianchi, Kimberly Chong, Mathias Drehmann, Peter Eckley, Laurent Ferrara, Julia Giese, Andy Haldane, Mikael Juselius, Iryna Kaminska, George Kapetanios, Andrew Lo, Tuomas Peltonen, Giorgio Primiceri, Leif Anders Thorsrud, Philip Treleven and Pawel Zabczyk for helpful comments and advice. We would also like to thank seminar participants at the Bank of England, the Bundesbank, the NBER Summer Institute Workshop on 'Forecasting and Empirical Methods' (Cambridge, MA, July 2017), the MIT Golub Center for Finance and Policy 3rd Annual Conference (Cambridge, MA, September 2016), the Norges Bank-ESCB-CEBRA workshop on 'Financial Stability and Macroprudential Policy' (Oslo, August 2016), the 9th ECB Workshop on 'Forecasting Techniques: Forecast Uncertainty and Macroeconomic Indicators' (Frankfurt, June 2016), the 2015 RiskLab / Bank of Finland / ESRB Conference on 'Systemic Risk Analytics' (Helsinki, September 2015), the 2014 FRB Cleveland-OFR Financial Stability Conference (Washington D.C, December 2014), the 'Concluding conference of the macro-prudential research network of the ESCB' (Frankfurt, June 2014), and the ECB Workshop 'Using big data for forecasting and statistics' (Frankfurt, April 2014). David Tuckett wishes to acknowledge support from the Institute of New Economic Thinking (grants no. IN01100025 and IN1300051), the Eric Simenhauer Foundation of the Institute of Psychoanalysis (London) and to the UK Research Councils (EPSRC grant reference EP/P016847/1 and a grant from the ESRC-NIESR Rebuilding Macroeconomics network). Rickard Nyman has been supported by a grant to the Centre for the Understanding of Decision-Making Uncertainty by the Institute of New Economic Thinking (Grant no INO16-00011).

Appendix

1. Extracting the relative sentiment series

1.1. Wordlists

Table A1 contains a random sample of 40 anxiety words and 40 excitement words. Note that when the same word is spelled differently in American and British English we have included both variants in the list.

1.2. Article tokenization

In order to count the frequency of the words in our emotion dictionaries we carry out a simple tokenization strategy. We split each article into a 'bag-of-words' (i.e., an unordered set of words) using the following procedure:

- 1 Convert the full article into lowercase letters only (to match our lists of lowercase emotion words)
- 2 Remove each occurrence of the quotation marks ' and '
- 3 Replace each non alphabetic character by a single space character
- 4 Split the text into words whenever we encounter a sequence of at least one whitespace character (including newlines, tabs and spaces)
- 5 Remove any remaining whitespace before or after the resulting words

Technically, we achieve steps 2–5 by replacing each match of the regular expression '['"']' by the empty string "", and then replacing each match of the regular expression '[^a-zA-Z]+' by the space character ' ', and finally splitting the text at each match of the regular expression '\s+'. From the set of remaining words we then count how many matches there are with the two emotion word dictionaries. The relevant anxiety and excitement word counts are aggregated over the articles in the given period, which in this case is a given month. We can then compute the relative sentiment score as described in the main text.

Table A1

Randomly drawn selection of words indicating excitement (about gain) and anxiety (about loss).

Anxiety	Anxiety	Excitement	Excitement
Jitter	Terrors	Excited	Excels
Threatening	Worries	Incredible	Impressively
Distrusted	Panics	Ideal	Encouraging
Jeopardized	Eroding	Attract	Impress
Jitters	Terrifying	Tremendous	Favoured
Hurdles	Doubt	Satisfactorily	Enjoy
Fears	Traumatised	Brilliant	Pleasures
Feared	Panic	Meritorious	Positive
Traumatic	Imperils	Superbly	Unique
Fail	Mistrusts	Satisfied	Impressed
Erodes	Failings	Perfect	Enhances
Uneasy	Nervousness	Win	Delighted
Distressed	Conflicted	Amazes	Energise
Unease	Reject	Energizing	Spectacular
Disquieted	Doubting	Gush	Enjoyed
Perils	Fearing	Wonderful	Enthusiastic
Traumas	Dreads	Attracts	Inspiration
Alarm	Distrust	Enthusiastically	Galvanized
Distrusting	Disquiet	Exceptionally	Amaze
Doubtable	Questioned	Encouraged	Excelling

1.3. Emotion word negation

We test for the potential impact of negation on the movements in the RTRS RSS series. We apply the simple method of negation detection reported in [Loughran and McDonald \(2011\)](#). We proceed by excluding any emotion word found in the text that is preceded, within three words, by any of the words ‘no’, ‘not’, ‘none’, ‘neither’, ‘never’ or ‘nobody’. In other words we do not consider the word at all if this is the case, as opposed to changing the meaning of the word and treating it as belonging to the ‘opposite’ category.

The negation-modified RTRS series remains correlated with the original series as highly as 0.999 in both level form and difference form. Although this might initially seem counterintuitive, it is in fact trivial to understand that negation will only affect the movement of the series if there is a systematic bias in its use for a given period, i.e., if, for a period of time, a given word is more likely to be negated than at other times. Negation could have an effect on the overall level of the series, if, for example, excitement words are more likely negated than anxiety words. However, since we are concerned with the movements of the series over time, as opposed to the actual levels, this does not concern us.

1.4. Orthogonality of wordlist

We test the hypothesis that the RTRS RSS series is orthogonal to the economy, and as such to fundamental news. We exclude all words in our excitement and anxiety lists that could potentially have economic meaning independently of emotional connotation. The words ‘uncertain’ and ‘uncertainty’ from the anxiety dictionary and the words ‘boost’, ‘boosted’, ‘boosts’, ‘exuberance’ and ‘exuberant’ from the excitement dictionary were excluded on this basis and a new RSS series produced using the remaining words. The new series remains correlated with the original RSS series at 0.99 in both level and difference form. We can conclude that the RSS measure is unlikely to be affected by the presence of economic terms.

1.5. Bootstrap confidence intervals of MCDAILY

[Fig. A1](#) shows the bootstrap confidence interval generated by random sampling of word weights from the two emotion dictionaries using the methodology described in the main text.

2. Granger causality procedure

2.1. Methodology

We use the methodology described in [Toda and Yamamoto \(1996\)](#). In outline, in investigating Granger causality between any two series, this is as follows:

- 1 Check the order of integration of the two series using Augmented Dickey-Fuller ([Said and Dickey 1984](#); p-values are interpolated from [Table 4.2, p. 103 of Banerjee et al. 1993](#)) and the [Kwiatowski-Phillips-Schmidt-Shin \(1992\)](#) tests. Let m be the maximum order of integration found.
- 2 Specify the VAR model using the data in levelled form, regardless of what was found in step 1, to determine the number of lags to use with standard method. We use the Akaike Information Criteria.
- 3 Check the stability of the VAR (we use OLS-CUSUM plots).

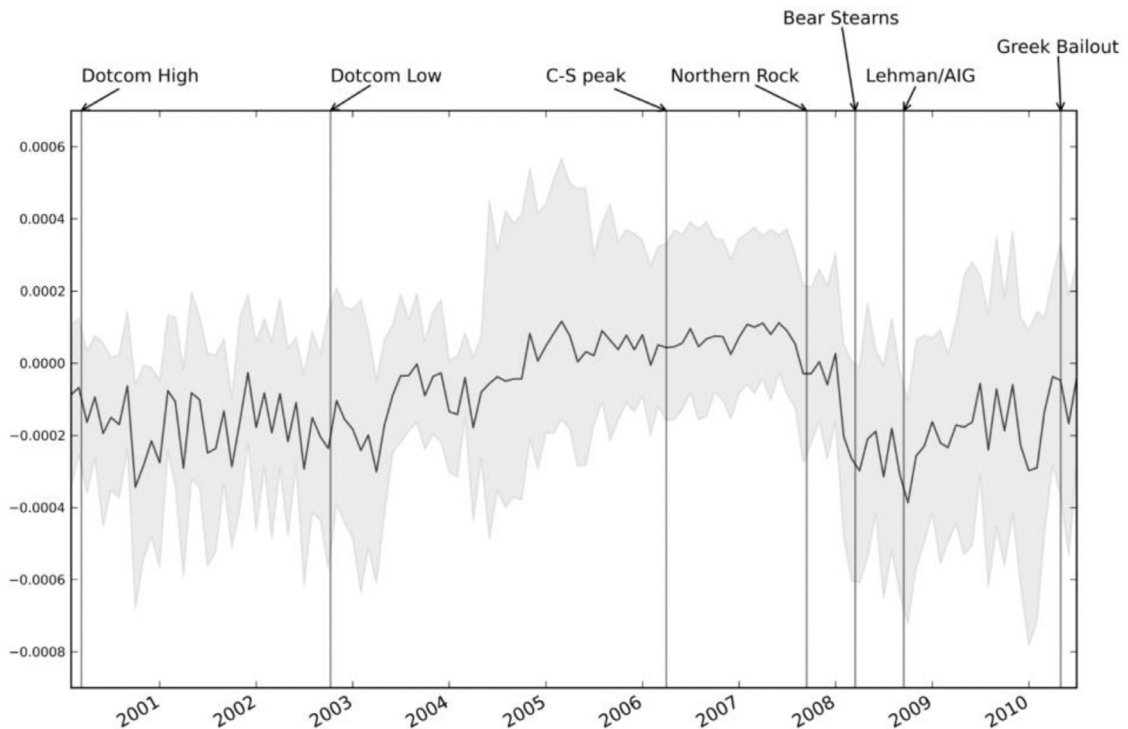


Fig. A1. MCDAILY with confidence intervals.

- 4 Test for autocorrelation of residuals. If autocorrelation is found, increase the number of lags until it goes away. We use the multivariate Portmanteau- and Breusch-Godfrey tests for serially correlated errors. Let p be the number of lags then used.
- 5 Add m extra lags of each variable to the VAR.
- 6 Perform Wald tests with the null being that the first p lags of the independent variable have coefficients equal to 0. If this is rejected, we have evidence of Granger-causality from the independent to dependent variable.

We used the statistical program R to carry out the analysis. The various packages used to carry out the above Toda-Yamamoto procedure, the details of the specific results obtained and a description of the various R packages using the test procedure are available on request from the authors.

2.2. Granger causality R packages

The packages in R used in the Toda-Yamamoto procedure to investigate Granger causality are as follows:

- *tseries* – we use the two functions *adf.test* and *kpss.test* (the Augmented Dickey-Fuller test and Kwiatkowski-Phillips-Schmidt-Shin test respectively) to check if series are stationary or contain unit roots. The *adf.test* function allows you to define the alternative hypothesis by the “alternative” argument. We use the default option of “stationary”. The function also allows you to manually specify the lag order k to calculate the test statistic. We use the default option $k = \text{trunc}((N-1)^{1/3})$, where N is the length of the series and *trunc* is a function built into R truncating the value towards zero.
- *vars* – we use the function *VARselect* to compute the Akaike Information Criteria for VAR(p) processes with p from 1 through 20. The function takes a number of arguments. We make use of the “lag.max” argument, which we set to 15 and the “type” argument which we set to “const”, indicating that information criteria should be computed for lags from 1 through 15 and that a constant term should be included in the VAR, respectively. We use the *VAR* function for estimating a VAR(p) process. Similarly to *VARselect* we use the “p” argument specifying the number of lags to include and the “type” argument specifying whether to include a constant term, or a trend or both. In all cases we set this argument to “const”, indicating that only a constant term should be included. We use the function *serial.test* to compute the multivariate Portmanteau- and Breusch-Godfrey tests for serially correlated errors in a VAR(p) process. We use the default number of lags for each test. In the case of the Portmanteau test we keep the default value of the “lags.pt” argument at 16 and in the case of the Breusch-Godfrey test we keep the default value of the “lags.bg” argument at 5. We set the “type” parameter to either “PT.asymptotic” or “BG” to compute the Portmanteau- or Breusch-Godfrey test respectively. We use the function *stability* to compute empirical fluctuation processes according to the OLS-CUSUM method. We use the default values of each argument, in particular the “type” argument which defaults to “OLS-CUSUM” for the OLS-CUSUM method. The

figures for the empirical fluctuation processes are generated by the use of the built in *plot* function on the returned object from the call to *stability*.

- *aod* – we use the function *wald.test* to perform the Wald tests for granger causality. We use three of the function's available arguments. The argument "Terms" specifying which terms of the model to include in the null hypothesis of the Wald test, given as a vector of term indices. The argument "b" specifying a vector of the coefficients of the model. The argument "Sigma" specifying the variance-covariance matrix of the model. To specify the values of the latter two arguments, we use the *coef* method on the relevant equation from the VAR to extract the relevant coefficients and the *vcov* method on the relevant equation from the VAR to extract the relevant variance-covariance matrix.

Here is an illustration of the full procedure when testing for Granger causality between RTRS and Michigan Consumer Sentiment Index (MCI).

```
# create time series variables from vectors
∅ rss <- ts(RTRS, start=c(1996,1), freq=12)
∅ mci <- ts(MCI, start=c(1996,1), freq=12)
∅ data <- cbind(mci, rss)
∅ colnames(data) <- c("MCI", "RTRS")

# test variables for order of integration
∅ adf.test(rss)
∅ kpss.test(rss)
∅ adf.test(diff(rss))
∅ kpss.test(diff(rss))

# similarly for mci. Let m be the maximum order of integration found
# compute information criteria for VAR models of different lags
∅ VARselect(data, lag=15, type="const")

# test residuals of chosen lag p
∅ serial.test(VAR(data, p=p, type="const"))
∅ serial.test(VAR(data, p=p, type="const"), type="BG")
∅ plot(stability(VAR(data, p=p, type="const")))

# if OK proceed and create a new VAR model adding m lags.
∅ var.model <- VAR(data, p=(p+m), type="const")

# perform a Wald test on the first p lags of the independent variable for both equations.
# E.g., if p = 1 and we test Granger causality from RTRS to MCI:
∅ wald.test(b=coef(var.model$varresult[[1]]), Sigma=cvov(var.model$varresult[[1]]), Terms=c(2))

# and to test the second equation, Granger causality from MCI to RTRS:
∅ wald.test(b=coef(var.model$varresult[[2]]), Sigma=cvov(var.model$varresult[[2]]), Terms=c(1))

# note that the independent variables are ordered by the number of lags
```

3. Narrative consensus

3.1. Constructing the narrative consensus series

We proceed as follows, following well-established methods:

- 1 Pre-process all documents by representing them as 'bags-of-words' in which word order is ignored and word-endings are removed using a standard English word stemmer, known as the Porter stemmer.
- 2 Compute a word by document frequency matrix, with words as rows and documents as columns (each entry ij is the frequency of word i in document j).
- 3 Remove uninformative rows (words at the extremes of the total word frequency distribution). We remove words at the top of the cumulative distribution (the smallest number of words accounting for a fixed percentage of the total word count) and at the bottom (the largest number of words accounting for a fixed percentage of the total word count) as the most frequent words (e.g., words such as 'and', 'or' and 'the') rarely help us distinguish between topics and the least frequent words typically introduce too much noise and fail to show consistent patterns. This can also be achieved using a frequency statistic called TFIDF (term frequency-inverse document frequency), that estimates a word's importance taking into account how often they appear in the corpus (term frequency) and how unique they are in terms of the documents they appear in (inverse document frequency). Another commonly used step is to remove all words in a predefined list, so called 'stopwords'. We experimented with both but found the simpler method using only the frequency distribution gave very similar results but executed much faster.

- 4 Reduce the dimensionality of the document vectors (columns), to d dimensions, by the use of Singular Value Decomposition (SVD). In the information retrieval literature, the method we use is referred to as Latent Semantic Analysis (Deerwester, 1988) and has proved highly successful in a wide range of applications. LSA is naturally able to model important language structures, such as the similarity between synonyms.
- 5 Cluster the document vectors. The clustering algorithm must automatically determine the number of clusters used to model the data. There are several such algorithms. We pick an extension of the popular K-means algorithm known as X-means (Pelleg and Moore, 2000), which iteratively decides whether or not to split one cluster into two using the Bayesian Information Criteria (BIC) as a measure of model fit. BIC measures how well a model fits data by the level of observed noise given the model while penalising linearly for the number of model parameters (i.e., penalising over-fitting).

This procedure gives us a distribution of the number of documents in each cluster (e.g., 1000 articles on sovereign debt, 100 articles on crude oil, etc.) and the total number of clusters found.

Using this distribution, we want our measure of consensus to have two intuitive properties:

- If the number of topics (clusters) is reduced while the size of each cluster is held fixed and equal – consensus should increase.
- If given a fixed number of topics, any particular topic grows in proportion to the others – consensus should increase.

A measure of the topic distribution, which would give us these properties, is *information entropy* (Shannon, 1948).

3.2. Discrete entropy

For a discrete distribution, such as in our particular case, the entropy is simply a logarithmically weighted sum of probabilities,

$$-\sum_{i=1}^k p_i \log(p_i) = -\sum_{i=1}^k \frac{n_i}{N} \log\left(\frac{n_i}{N}\right) = \log(N) - \frac{1}{N} \sum_{i=1}^k n_i \log(n_i), \tag{3}$$

where n_i is the number of articles in cluster i , N is the total number of articles and k is the number of clusters.

The entropy is maximised (for a fixed number of clusters k) when documents are uniformly distributed over the clusters. As the distribution moves away from uniformity, the entropy will decrease. To better understand how entropy changes with k (the number of found clusters), we can simplify the equation as follows (if we assume a uniform distribution of documents across the clusters),

$$\log(N) - \frac{1}{N} k \frac{N}{k} \log\left(\frac{N}{k}\right) = \log(N) - \log\left(\frac{N}{k}\right) = \log(k). \tag{4}$$

It is clear from this that entropy is increasing logarithmically as k increases. In other words, the entropy is like an inverse consensus measure. Thus, if a narrative grows to dominate the news, for example narratives such as sovereign debt, structured finance or housing, the narrative entropy will decrease showing an increase in consensus. Similarly, if the total number of narratives decrease, all else fixed, consensus will increase (again signified by a decrease in narrative entropy).

We smooth the result using a method known as double exponential smoothing. Double exponential smoothing is often chosen as an alternative to the simple single exponential smoothing when it is believed that the underlying data contains a trend component.

3.3. Double exponential smoothing

Given series $x_t = \{x_0 \dots x_n\}$ we decompose it into a smoothed series s_t and a trend component b_t by the procedure

$$s_0 = x_0, \quad b_0 = \frac{x_1 - x_0}{2} \tag{5}$$

For $t > 0$:

$$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} - b_{t-1}) \tag{6}$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \tag{7}$$

for some $\alpha, \beta \in [0, 1]$. In our case we discard b_t after using it to estimate the smoothed series s_t .

We run the algorithm across several choices of parameters (the list of model parameters, and combinations used (e.g., ‘40, 5, 100’ and ‘50, 2, 100’), can be found in Table A2) and smooth (using double exponential smoothing, with $\alpha = \beta = 0.3$) and average the results across parameter runs.

Table A2

Consensus parameter combinations used to construct the Narrative ENTROPY index.

Parameter	Values Considered
Upper word bound	40,50,50,60,40,50,50,60
Lower word bound	5,2,10,10,5,2,10,10
Vector Dimensionality	100,100,100,100,200,200,200,200

Note: the considered values were combined in the ordered they are listed, i.e. (40, 5, 100), (50, 2, 100), etc.

3.4. Constructing narrative consensus proxy measures

To investigate the robustness of the narrative consensus metric we devise two further methodologically distinct approaches to capture proxies for narrative consensus.

- 1 Average document 'overlap'
- 2 Average document 'similarity'

We compute (1) from the word-by-document frequency matrix (after removing the 'uninformative' words) by simply dividing the number of non-zero entries in the matrix by the total number of entries, giving us a comparable time series (in which higher document overlap is a proxy for higher narrative consensus). We compute (2) by repeatedly sampling pairs of document vectors and computing the angle between them (a standard similarity metric for document vectors). We repeat the procedure 1000 times and compute the mean angle. In this case, a mean angle closer to zero is a proxy for higher narrative consensus.

4. Granger causality tables

This section presents results of relevant statistical tests performed. In particular:

- Tests to determine the order of integration of each variable
- Residual tests of each VAR model used to test for Granger causality
- Wald test statistics for each test of Granger causality

The Augmented Dickey-Fuller lag denotes lag order and the Kwiatkowski-Phillips-Schmidt-Shin lag denotes the truncation lag parameter. The analysis implies that all variables should be treated as integrated of order 1, except for the excitement component of BROKER which should be treated as integrated of order 2.

Table A3

Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests for stationarity of monthly sentiment series; the periods considered are those for which the respective text data was available.

Variable	ADF (lag)	p-value	KPSS (lag)	p-value
RTRS Level	-3.33 (6)	0.07	1.82 (3)	< 0.01
RTRS Diff	-6.21 (6)	< 0.01	0.02 (3)	> 0.1
MCDAILY Level	-1.68 (5)	0.71	0.70 (2)	0.01
MCDAILY Diff	-6.66 (4)	< 0.01	0.05 (2)	> 0.1
BROKER Level	-2.68 (3)	0.31	0.47 (1)	0.05
BROKER Diff	-3.97 (3)	0.02	0.03 (1)	> 0.1

The Augmented Dickey-Fuller lag denotes lag order and the Kwiatkowski-Phillips-Schmidt-Shin lag denotes the truncation lag parameter. The analysis implies that all variables should be treated as integrated of order 1, except for the aggregate version of MCDAILY, the anxiety component of MCDAILY, and the Narrative ENTROPY series which should all be treated as integrated of order 2.

Table A4

Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests for stationarity of quarterly sentiment series and the Narrative ENTROPY series; the periods considered are those for which the respective text data was available.

Variable	ADF (lag)	p-value	KPSS (lag)	p-value
RTRS Level	-2.71 (4)	0.29	1.32 (1)	< 0.01
RTRS Diff	-4.37 (4)	< 0.01	0.03 (1)	> 0.1
MCDAILY Level	-1.99 (3)	0.58	0.39 (1)	0.08
MCDAILY Diff	-2.86 (3)	0.24	0.08 (1)	> 0.1
MCDAILY 2Diff	-3.81 (3)	0.03	0.03 (1)	> 0.1
ENTROPY Level	-2.12 (3)	0.52	0.47 (1)	0.05
ENTROPY Diff	-2.76 (3)	0.28	0.10 (1)	> 0.1
ENTROPY 2Diff	-5.07 (3)	< 0.01	0.04 (1)	> 0.1

The Augmented Dickey-Fuller lag denotes lag order and the Kwiatkowski-Phillips-Schmidt-Shin lag denotes the truncation lag parameter.

Table A5

Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests for stationarity of selected comparative variables; MCI (January 1996 through September 2014), VIX (January 1996 through September 2014), EPU (January 1997 through September 2014), BoEU (1996Q1 through 2014Q1), PMI (April 1997 through September 2014), CDS (January 2003 through September 2014).

Variable	ADF (lag)	p-value	KPSS (lag)	p-value
MCI Level	-2.30 (6)	0.45	3.95 (3)	< 0.01
MCI Diff	-6.58 (6)	< 0.01	0.06 (3)	> 0.1
VIX Level	-3.04 (6)	0.14	0.28 (3)	> 0.1
VIX Diff	-7.35 (6)	< 0.01	0.04 (3)	> 0.1
EPU Level	-3.42 (5)	0.05	3.24 (3)	< 0.01
EPU Diff	-7.40 (5)	< 0.01	0.02 (3)	> 0.1
BoEU Level	-2.32 (4)	0.45	1.17 (1)	< 0.01
BoEU Diff	-3.81 (4)	0.02	0.08 (1)	> 0.1
PMI Level	-4.23 (5)	< 0.01	1.29 (3)	< 0.01
PMI Diff	-5.56 (5)	< 0.01	0.05 (3)	> 0.1
CDS Level	-0.30 (5)	0.99	3.24 (2)	< 0.01
CDS Diff	-5.22 (5)	< 0.01	0.16 (2)	> 0.1

Table A3-A13

Table A6

Multivariate Portmanteau- and Breusch-Godfrey tests for serially correlated errors of VAR models involving RTRS and MCI, VIX, BoEU, CDS and PMI.

VAR model	Portmanteau	d.f.	Breusch-Godfrey	d.f.
RTRS/MCI	60.83	52	19.36	20
RTRS/VIX	64.63	52	15.30	20
RTRS/BoEU	42.57	56	19.02	20
RTRS/CDS	73.506	56	28.13	20
RTRS/PMI	48.458	44	19.888	20

Table A7

Multivariate Portmanteau- and Breusch–Godfrey tests for serially correlated errors of VAR models involving BROKER and MCI, VIX, EPU, CDS and PMI.

VAR model	Portmanteau	d.f.	Breusch–Godfrey	d.f.
BROKER/MCI	41.68	60	15.60	20
BROKER/VIX	40.52	56	25.61	20
BROKER/EPU	46.683	60	19.799	20
BROKER/CDS	56.996	60	18.738	20
BROKER/PMI	54.433	60	25.587	20

Table A8

Multivariate Portmanteau- and Breusch–Godfrey tests for serially correlated errors of VAR models involving MCDAILY and MCI, VIX, BoEU, EPU, CDS and PMI.

VAR model	Portmanteau	d.f.	Breusch–Godfrey	d.f.
MCDAILY/MCI	60.799	56	23.738	20
MCDAILY/VIX	52.54	56	22.58	20
MCDAILY/BoEU	29.946	40	30.802	20
MCDAILY/EPU	55.391	56	19.918	20
MCDAILY/CDS	69.989	56	31.381	20
MCDAILY/PMI	59.838	56	21.621	20

Table A9

Wald test statistics of Granger-causality between the RTRS series and MCI, VIX, BoEU, CDS and PMI.

Direction	Chi-Sq	d.f.	p-value
RTRS -> MCI	12.7	3	0.0053
MCI -> RTRS	3.8	3	0.29
RTRS -> VIX	3.8	3	0.28
VIX -> RTRS	6.4	3	0.093
RTRS -> BoEU	24.5	2	4.7e-06
BoEU -> RTRS	7.7	2	0.022
RTRS -> CDS	2.4	2	0.3
CDS -> RTRS	1.1	2	0.57
RTRS -> PMI	24.9	5	0.00015
PMI -> RTRS	9.9	5	0.077

Table A10

Wald test statistics of Granger-causality between the BROKER series and MCI, VIX, EPU, CDS and PMI.

Direction	Chi-Sq	d.f.	p-value
BROKER -> MCI	44.9	1	2.1e-11
MCI -> BROKER	0.0052	1	0.94
BROKER -> VIX	3.4	2	0.18
VIX -> BROKER	3.7	2	0.16
BROKER -> EPU	0.01	1	0.92
EPU -> BROKER	0.12	1	0.73
BROKER -> CDS	0.28	1	0.6
CDS -> BROKER	0.0013	1	0.97
BROKER -> PMI	2.7	1	0.1
PMI -> BROKER	0.68	1	0.41

Table A11

Wald test statistics of Granger-causality between the MCDAILY series and MCI, VIX, BoEU, EPU, CDS and PMI.

Direction	Chi-Sq	d.f.	p-value
MCDAILY -> MCI	1.4	2	0.5
MCI -> MCDAILY	0.11	2	0.95
MCDAILY -> VIX	4.8	2	0.09
VIX -> MCDAILY	1.9	2	0.39
MCDAILY -> BoEU	29.0	6	6e-05
BoEU -> MCDAILY	4.7	6	0.58
MCDAILY -> EPU	5.9	2	0.052
EPU -> MCDAILY	3.5	2	0.18
MCDAILY -> CDS	4.9	2	0.085
CDS -> MCDAILY	0.23	2	0.89
MCDAILY -> PMI	5.7	2	0.059
PMI -> MCDAILY	1.4	2	0.49

Table A12

Multivariate Portmanteau- and Breusch–Godfrey tests for serially correlated errors of VAR models involving the ENTROPY series and VIX, BoEU and EPU.

VAR model	Portmanteau	d.f.	Breusch–Godfrey	d.f.
ENTROPY/VIX	38.767	52	8.2147	20
ENTROPY/BoEU	30.58	40	28.38	20
ENTROPY/EPU	59.622	52	10.888	20

Table A13

Wald test statistics of Granger-causality between the ENTROPY series and BoEU.

Direction	Chi-Sq	d.f.	p-value
ENTROPY -> VIX	5.9	3	0.12
VIX -> ENTROPY	3.9	3	0.27
ENTROPY -> BoEU	33.3	6	9.1e-06
BoEU -> ENTROPY	1.0	6	0.98
ENTROPY -> EPU	9.1	3	0.027
EPU -> ENTROPY	2.2	3	0.52

References

- Aikman, D., Alessandri, P., Eklund, B., Gai, P., Kapadia, S., Martin, E., Mora, N., Sterne, G., Willison, M., 2009. Funding liquidity risk in a quantitative model of systemic stability. *Bank of England Working Paper*, Number 372.
- Aikman, D., Nelson, B., Tanaka, M., 2015. Reputation, risk-taking and macroprudential policy. *J. Bank. Financ.* 50, 428–439.
- Akerlof, G. A., Shiller, R., 2009. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press, Princeton.
- Ardia, D., Bluteau, K., Boudt, K., 2019. Questioning the news about economic growth: sparse forecasting using thousands of news-based sentiment values. *Int. J. Forecast.* 35 (4), 1370–1386.
- Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. *J. Appl. Econometr.* 18, 1–22.
- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., 2018. The economic effects of social networks: evidence from the housing market. *J. Polit. Econ.* 126 (6).
- Baker, M., Stein, J.S., 2004. Market liquidity as a sentiment indicator. *J. Financ. Mark.* 7 (3), 271–299.
- Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *J. Financ.* 61 (4), 1645–1680.
- Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. *J. Econ. Perspect.* 21 (2), 129–151.
- Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. *Q. J. Econ.* 131 (4), 1593–1636.
- Bank of England (2014). 'The Financial Policy Committee's powers to supplement capital requirements: a Policy Statement', January.
- Bank of England (2015). 'The Bank of England's approach to stress testing the UK banking system', October.
- Barrell, R., Davis, E., Karim, D., Liadze, I., 2010. Bank regulation, property prices and early warning systems for banking crises in OECD countries. *J. Bank. Financ.* 34 (9), 2255–2264.
- Barsky, R.B., Sims, E.R., 2012. Information, animal spirits, and the meaning of innovations in consumer confidence. *Am. Econ. Rev.* 102 (4), 1343–1377.
- Berry, M.W., 2004. *Survey of Text Mining: Clustering, Classification, and Retrieval*. Springer, Hershey, N.Y.
- Bloom, N., 2009. The impact of uncertainty shocks. *Econometrica* 77, 623–685.
- Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. *Rev. Financ. Stud.* 22 (11), 4463–4492.
- Borio, C., Lowe, P., 2002. Assessing the risk of banking crises. *BIS Quarterly Review* (December) 43–54. https://www.bis.org/publ/qtrpdf/r_qt0212e.pdf.
- Borio, C., Lowe, P., 2004. Securing sustainable price stability. Should credit come back from the wilderness? *BIS Working Papers*, no. 157.
- Brown, G.W., Cliff, M.T., 2005. Investor sentiment and asset valuation. *J. Bus.* 78 (2), 405–440.
- Bruner, J., 1990. *Acts of Meaning*. Harvard University Press, Cambridge, Mass and London, England.
- Carver, C.S., Harmon-Jones, E., 2009. Anger is an approach-related affect: evidence and implications. *Psychol. Bull.* 135 (2), 183.
- Collier, P., Tuckett, D., 2021. Narratives as a coordinating device for reversing regional disequilibrium. *Oxford Rev. Econ. Policy* 37 (1), 97–112. doi:10.1093/oxrep/gra0060.
- Damasio, A., Damasio, H., 2006. Minding the Body. *Daedalus* 135 (3), 15–22. On Body in Mind (Summer, 2006).

- Deerwester, S., et al., 1988. Improving information retrieval with latent semantic indexing. *Proc. 51st Annual Meet. Am. Soc. Inf. Sci.* 25, 36–40.
- Drehmann, M., Borio, C., Tsatsaronis, K., 2011. Anchoring countercyclical capital buffers: the role of credit aggregates. *Int. J. Central Bank.* 7 (4), 189–240.
- Edmans, A., Garcia, D., Norli, A., 2007. Sports sentiment and stock returns. *The J. Financ.* 62 (4), 1967–1998.
- Gennaioli, N., Shleifer, A., 2018. *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press, Princeton and Woodstock.
- Giese, J., Andersen, H., Bush, O., Castro, C., Farag, M., Kapadia, S., 2014. The credit-to- GDP gap and complementary indicators for macroprudential policy: evidence from the UK. *Int. J. Financ. Econ.* 19 (1), 25–47.
- Gray, D.F., Merton, R.C., Bodie, Z., 2008. A new framework for measuring and managing macrofinancial risk and financial stability. *Harvard Business School Working Paper* no. 09-015.
- Gray, D.F., Jobst, A., 2011. Modeling systemic financial sector and sovereign risk. *Sveriges Riksbank Econ. Rev.* http://archive.riksbank.se/Upload/Rapporter/2011/POV_2/pov_2011_2_Gray_Jobst.pdf.
- Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross- spectral methods. *Econometrica* 37 (3), 424–438. doi:10.2307/1912791.
- Greenwood, R., Nagel, S., 2009. Inexperienced Investors and Bubbles. *J. Financ. Econ.* 93 (2), 239–258.
- Haddow, A., Hare, C., Hooley, J., Shakir, T., 2013. Macroeconomic uncertainty; what is it, how can we measure it and why does it matter? *Bank Eng. Q. Bull.* 53 (2), 100–109.
- Harmon-Jones, E., Harmon-Jones, C., Price, T.F., 2013. What is approach motivation? *Emotion Rev.* 5 (3), 291–295.
- Holló, D., Kremer, M., Lo Duca, M., 2012. CISS - a composite indicator of systemic stress in the financial system. *European Central Bank Working Paper*, no. 1426.
- Illing, M., Liu, Y., 2006. Measuring financial stress in a developed country: an application to Canada. *J. Financ. Stability* 2 (4), 243–265.
- Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., Kapadia, S., 2020. Making text count: economic forecasting using newspaper text. *Bank of England Staff Working Paper*, Number 865.
- Kaminska, I., Roberts-Sklar, M., 2015. A global factor in variance risk premia and local bond pricing. *Bank of England Staff Working Paper*, Number 576.
- Kaminsky, G.L., Reinhart, C.M., 1999. The twin crises: the causes of banking and balance-of-payments problems. *Am. Econ. Rev.* 89 (3), 473–500.
- Kapadia, S., Drehmann, M., Elliott, J., Sterne, G., 2013. Liquidity risk, cash flow constraints, and systemic feedbacks. In: Haubrich, J.G., Lo, A.W. (Eds.), *Quantifying Systemic Risk*. University of Chicago Press.
- Keynes, J.M., 1936. *The General Theory of Employment, Interest and Money*. Palgrave Macmillan.
- Kindleberger, C. P., Aliber, R. Z., 2005. *Manias, Panics, and Crashes: A History of Financial Crises*. Wiley, Hoboken, NJ.
- Knight, F.H., 1921. *Risk, Uncertainty and Profit*. Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics* 54, 159–178.
- Larsen, V.H., 2020. Components of uncertainty. *International Economic Review* In press.
- Larsen, V.H., Thorsrud, L.A., 2019. The value of news for economic developments. *J. Econometrics* 210 (1), 203–218.
- Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014. The parable of Google Flu: traps in big data analysis. *Science* 343 (6176), 1203–1205.
- Loughran, T., McDonald, W., 2011. When is a Liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J. Finance* 66 (1), 35–65.
- Lucas, R., 1976. Econometric policy evaluation: a critique. In: *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy 1. American Elsevier, New York, pp. 19–46.
- Lux, T., 2011. Sentiment dynamics and stock returns: the case of the German stock market, 3rd ed, 41. *Empirical Economics*, pp. 663–679.
- Mousavi, Shabnam, Gigerenzer, Gerd, 2014. 'Risk, uncertainty, and heuristics' (March 19, 2014). *J. Bus. Res.* 67 (8).
- Nimark, K., Pitschner, S., 2019. News media and delegated information choice. *J. Econom. Theory* 181, 160–196.
- Nyman, R., Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P., Smith, R., 2018. News and narratives in financial systems: exploiting big data for systemic risk assessment. *Bank of England Staff Working Paper*, Number 704.
- OFR annual report (2012) <https://financialresearch.gov/annual-reports/files/office-of-financial-research-annual-report-2012.pdf>.
- Pelleg, D., Moore, A., 2000. X-means: extending K-means with efficient estimation of the number of clusters. In: *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 727–734.
- Rambaccussing, D., Kwiatkowski, A., 2020. Forecasting with news sentiment: evidence with UK newspapers. *Int. J. Forecast.* 36 (4), 1501–1516.
- Reinhart, C., Rogoff, K., 2009. *This time is different: Eight Centuries of Financial Folly*. Princeton University Press.
- Rolls, E., 2014. *Emotion and Decision-Making Explained*. Oxford, OUP.
- Schularick, M., Taylor, A., 2012. Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870–2008. *Am. Econ. Rev.* 102 (2), 1029–1061.
- Shannon, Claude E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (3), 379–423.
- Shiller, R., 2000. *Irrational Exuberance*. Princeton University Press.
- Shiller, R., 2019. *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.
- Soo, C.K., 2013. Quantifying animal spirits: news media and sentiment in the housing market. *Ross School of Business Working Paper*, Number 1200.
- Strauss, Viktor, M., 2013. *Emotional Values of Words in Finance: Anxiety about Losses and Excitement about Gains*. M.Sc. thesis in Social Cognition. University College London.
- Tetlock, Paul C., 2007. Giving content to investor sentiment: the role of media in the stock market. *J. Financ.* 62 (3), 1139–1168.
- Tetlock, Paul C., Saar-Tsechansky, Maytal, Macskassy, Sofus, 2008. More than words: quantifying language to measure firms' fundamentals. *J. Financ.* 63 (3), 1437–1467.
- Tetlock, P.C., 2011. All the News that's fit to reprint: do investors react to stale information?. *Rev. Financial Studies* 24, 1481–1512.
- Toda, H.Y., Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes. *J. Econometrics* 66, 225–250.
- Tuckett, D., 2011. *Minding the Markets: An Emotional Finance View of Financial Instability*. Palgrave Macmillan.
- Tuckett, D.A., Nikolic, M., 2017. The role of conviction in decision-making under radical uncertainty. *Theory Psychol.* 27 (4), 501–523.
- Tuckett, D., Smith, R.E., Nyman, R., 2014. Tracking phantastic objects: A computer algorithmic investigation of narrative evolution in unstructured data sources'. *Soc. Netw.* 38, 121–133. doi:10.1016/j.socnet.2014.03.001.
- Tuckett, D., Holmes, D., Pearson, A., Chaplin, G., 2020. Monetary policy and the management of uncertainty: a narrative approach. *Bank of England Staff Working Paper*, Number 870. Bank of England.
- Tuckett, D., 2017. The future of macroeconomics: why observation of the behaviour of human actors and how they combine within the economy, is an important next step. Conference paper, Institute of New Economic Thinking (INET) October, 2017. <https://www.ineteconomics.org/uploads/papers/TUCKETT-The-future-of-macroeconomics-Final.pdf>.
- Watts, D., 2002. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci.* 99, 766–771.