News and narratives in financial systems: exploiting big data for systemic risk assessment[1]

Rickard Nyman,[*] David Gregory[**], Sujit Kapadia[**], Paul Ormerod [*], David Tuckett[*] and Robert Smith[*]

July 2017

Abstract:

This paper applies algorithmic analysis to large amounts of financial market text-based data to assess how narratives and sentiment play a role in driving developments in the financial system. We find that changes in the emotional content in market narratives are highly correlated across data sources. They show clearly the formation (and subsequent collapse) of very high levels of sentiment – high excitement relative to anxiety – prior to the global financial crisis. Our metrics also have predictive power for other commonly used measures of sentiment and volatility. And we develop a new methodology that attempts to capture the emergence of narrative topic consensus. This gives an intuitive representation of the increasing homogeneity of beliefs around a new paradigm prior to the crisis. With increasing consensus around narratives high in excitement and lacking anxiety likely to be an important warning sign of impending financial system distress, the quantitative metrics we develop may complement other indicators and analysis in helping to gauge systemic risk.

---

[1] 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 its Policy Committees.

[*]University College London Centre for the Study of Decision-Making Uncertainty, R.Nyman@ucl.ac.uk; P.Ormerod@ucl.ac.uk; R.Smith@ucl.ac.uk; D.Tuckett@ucl.ac.uk

[**]Bank of England, sujit.kapadia@bankofengland.co.uk; dbjg21@gmail.com (David Gregory worked on this paper while at the Bank of England)

Corresponding Author: D.Tuckett@ucl.ac.uk

"The history of speculative bubbles begins roughly with the advent of newspapers…Although the news media…present themselves as detached observers of market events, they are themselves an integral part of these events. 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"
(Shiller, 2000)

1. Introduction

The years preceding the global financial crisis were characterised by widespread exuberance in the financial sector. As has often occurred throughout history (Reinhart and Rogoff, 2009), consensus emerged over a new paradigm, under which the greater efficiency of markets and distribution of risk around the system was thought to justify the strong positive sentiment. When the crash came during 2007 and 2008, sentiment reversed rapidly with fear and anxiety pervading the financial system.

Discussion of the key role played by sentiment in driving the economy and the financial system dates at least as far back as Keynes (1936). In his 2017 AEA Presidential Address, Shiller (2017) draws on psychological and epidemiological literature to discuss how the human brain is highly attuned to narratives which justify actions and can "go viral" in a way which drives economic and financial fluctuations. He concludes by discussing how textual and semantic analysis could be deployed to help understand the impact of narratives.

This paper takes a 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 financial market text-

based data to identify quantitative metrics that try to capture shifts in sentiment along with the extent of consensus in the market. We then attempt to evaluate the usefulness of these metrics by comparing them against notable events and developments in the financial system and against other commonly used measures, supporting our analysis with structural break and Granger causality tests, and we assess relationships with the wider economy in a simple vector autoregression (VAR) model.

In broad terms, we find that our metrics capture key developments in the financial system relatively well prior to, during and after the global financial crisis, with key shifts identified by structural break tests. At a higher frequency, they also have some predictive power for other commonly used measures of sentiment, confidence and volatility. And they appear to influence economic and financial variables in our VAR model. As such, we contend that our metrics have potential value for gauging sentiment and systemic risk in financial systems. And they may be useful in helping to signal the prospect of future distress as a complement to more traditional indicators and analysis (see, for example, Drehmann et al, 2011; Bank of England, 2014; or Giese et al, 2014).

The measures of sentiment we use are derived from pre-defined common English language word lists representing two specific emotional groups capturing approach or avoidance. The words were selected through the lens of a social-psychological theory of "conviction narratives" (CNT) (Tuckett and Nikolic, 2017), which draws on the idea that narratives are used by individuals to make sense of information, to simulate the outcome of proposed actions and to feel confident or otherwise about them. The approach can be seen as one way of empirically formulating the idea of "animal spirits" (Keynes, 1936) and has been used successfully used in other applications, for

example as a measure of changing macroeconomic confidence (Tuckett and Nyman 2017) .

As discussed by Bruner (1991) and Shiller (2017), the use of narratives to drive behaviour is arguably a fundamental form of human organisation.[2] CNT is a theory which emphasises the role of such narratives and particular groups of action-enabling or disabling emotions in driving decision-making under radical, or Knightian (Knight, 1921), uncertainty. It suggests that in this context agents do (and have to) construct narratives supporting their expectations and that these create a feeling of accuracy that readies them to act. Such *conviction narratives* combine cognition and emotion to interpret data, envision the future and support action.[3] A particular focus is on narratives that evoke attraction or *approach* to an object of investment (broadly conceived), versus emotions that evoke repulsion or *avoidance* of that object. This emphasis on approach and avoidance focuses the idea of sentiment on its implications for *action* in uncertain decision-making, thus focusing the often-vague topic of positive/negative sentiment. In more ordinary language we focus on *excitement* about the potential gains from an action relative to anxiety about the potential losses. If excitement comes to dominate relative to anxiety, investment will be undertaken (Tuckett and Nikolic, 2017). Thus, in the simplest case, the key variables of interest

---

[2] It has been argued that narrative allows experience to be ordered into "chunks" (Miller, 1956) with implicit relevance to plans (Pribriam et al, 1960) and causal models (Rottman and Hastie, 2013; Sloman and Lagnado, 2015) and so explanations and predictions about outcome.

[3] More precisely, conviction narratives enable actors to draw on the beliefs, causal models and rules of thumb situated in their social context to identify opportunities worth acting on, to simulate the future outcome of those actions and to feel sufficiently convinced about the anticipated outcomes to act. They are founded on biologically and socially evolved coping capacities that allow individuals to prepare to execute particular actions even though they cannot accurately know what the outcomes will be. Conviction narratives also provide an easy means for actors to communicate and gain support from others for their selected actions as well as to justify themselves. Ideas about the role of simulation and embodied cognition that are central to the supportive role of narratives in decision-making build on existing work in affective and cognitive neuroscience (e.g. Suddendorf and Corballis, 2007; Barsalou, 2008; Baumeister and Masicampo, 2010; Damasio and Carvalho, 2013). CNT is part of a growing body of work inside and outside economics developing broader models of human cognition and decision-making behaviour (for example, Bruner, 1991; Damasio, 1999; Lane and Maxfield, 2005; Mar and Oatley, 2008; Akerlof and Shiller, 2009; Beckert, 2011 and Tuckett, 2011)

are the aggregate relative difference between *excitement* and *anxiety* and *shifts* in this difference over time.

At any given moment, there will be several narratives and associated emotions circulating among financial agents. Empirically, such narratives may manifest themselves in market movements. But, as they emerge, some of these narratives, or pieces of them, are likely to be contained within relevant text-based data sources with the common English language words used to describe them conveying a sense of sentiment. As such, text-based analysis offers a mechanism to test theories which assert a central role for narratives in driving economic and financial outcomes (see also Shiller, 2017). In particular, if the relative shifts in sentiment and emotional content correlate across different text sources, it is plausible that at least some financial agents had adopted a subset of the narratives and held them as true. It is, however, important to note that one cannot conclude, and it is in some cases highly unlikely (depending on the type of data), that the content creators *themselves* had adopted as true the narratives portrayed in their documents – for example, one can easily imagine a big difference between financial news documents and social media data, in the extent to which content creators *feel* what they write.

With this in mind, we analyse three unstructured text-based data sources of potential interest: internal Bank of England daily commentary on market news and events; Reuters' news wire articles in the United Kingdom; and broker research reports. Taking a narrative perspective, we capture an emotional summary statistic of sentiment (Relative Sentiment Shift or RSS) based on these sources, and explore how convincingly and robustly it measures shifts in confidence. This measure aims at capturing the *extent* to which the creators of the documents *portray* emotions within the narratives and, in particular, shifts in the balance between the proportions of excitement versus anxiety words. As with other text-based and big data approaches

which try to operationalise the concept of sentiment, a key strength of this method lies in its top down approach, capturing aggregate shifts largely undetectable to the human eye.

At a low frequency, the relative sentiment metrics that we extract appear, with the benefit of hindsight, to give early warning signs of significant financial events in recent years. In particular, overall sentiment was at very high and stable levels in the mid-2000s, arguably indicative of exuberance in the financial system and the risk of future distress. 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 our methods can be focussed on particular topics, such as 'property', thus potentially helping to shed light on specific sectors of the economy.

To gauge the robustness of our aggregate sentiment metrics, we compare them with both with standard aggregated measures of consumer confidence and market volatility and with some relevant but more atheoretic measures of uncertainty from the literature exploiting text-based information. Strikingly, we find that our sentiment metrics often act as a leading indicator of such other measures and can potentially help us to understand them. We also present tentative evidence from a simple VAR model that our measures might contain useful information in explaining economic developments. These results highlight the potential value of our metrics at a higher frequency as well.

Financial behaviour can also often be homogenous. Therefore, in the second, more exploratory, part of the paper, we ask whether we can measure *structural* changes in the distribution of narratives. Specifically, we develop a methodology to measure 'narrative consensus / disagreement' in the distribution of narratives as they develop

over time. This could be a relevant measure of the extent to which some narratives have spread through social networks (see, for example, Watts, 2002, Shiller, 2017 or Bailey et al, 2017), possibly amplified by herding ("keeping up with the Goldmans" – Aikman et al, 2015); via the media (Shiller, 2000); or been subject to social-psychological processes and adopted as true (*groupfeel* (Tuckett, 2011)). For example, prior to the global financial crisis, consensus appeared to develop across investors both about a new paradigm in the financial system and in the belief that it was possible to achieve higher returns than previously. But such consensus in an environment of high sentiment could be suggestive of over-confidence or irrational exuberance which may be unsustainable. The ability to measure the emergence of consensus or disagreement within text documents could therefore prove useful in identifying financial system risks.

Using our newly developed measure of narrative dispersal, we find that consensus in the Reuters news articles grew significantly over a period spanning several years prior to the global financial crisis. When viewed together with the sentiment series, this could be indicative of a growing, predominantly excited consensus about a new paradigm in the financial system prior to the crisis, in which anxiety and doubt substantially diminished – a possible signal of impending distress.

Other studies that attempt to quantify sentiment have used text-based data sources such as corporate reports and news media analysed with much more general word lists to capture emotion. They have attempted, for example, to predict various aspects of asset prices (e.g., Loughran and McDonald, 2011; Tetlock 2007; Tetlock et al. 2008; Tetlock 2011; Soo 2013) or to capture economic policy uncertainty (Baker et. al., 2016). A growing literature also exploits search engine data, such as *Google Trends*, to try to predict the current value of ('nowcast') economic variables such as GDP (e.g. Choi and Varian, 2012).

Research has also used text data to explore opinion formation in central banks (Hansen et al, 2014) or how the tone and language of statements by central banks may influence variables such as inflation forecasts and inflation expectations (see for example Blinder et al., 2008, Sturm & De Haan, 2011, Hubert 2012). More broadly, there is also 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, 2008; Greenwood and Nagel 2009).

Our emphasis departs from the above literature in several ways. First, we develop our dictionary and thus measures of sentiment from the point of view of a social-psychological theory of action under uncertainty (Tuckett and Nikolic, 2017). In this way we apply a theoretical filter drawn from psychological microfoundations which should more accurately detect features we hypothesise to be important. We also mitigate some of the difficulties associated with data mining, including the risk of obtaining seemingly significant correlations that do not generalise to new data or are highly context specific. For example, some recent studies suggest Internet search data (such as Google trends) should be treated with care, either because of a lack of transparency about how the data have been created (Lazer, et al. 2014) or uncertainty about the motivation for searching – independently or because of social influence (Ormerod et al., 2014). Second, and linked to the theoretical microfoundations, our word dictionaries comprise solely of 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, relative to survey-based measures of sentiment, our metrics are easy to implement across different countries and sources of data, and are available at high frequency and rapidly in real time.

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 with other economic and financial variables. Section 4 focuses on the measure of 'narrative consensus', explaining the methodology and results. Section 5 discusses how these measures might complement more traditional indicators and analysis used in systemic risk assessment, and Section 6 concludes.

2. Data and Methodology

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.

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 of clear 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 extensively studied by Tuckett et al. (2015) 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 6,123 articles were

---

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

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) and the full list is available on request from the authors.

Table 1: Emotion dictionary samples

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

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.[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.

---

[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.

$$Sentiment[T] = \frac{|Excitement| - |Anxiety|}{size[T]}$$

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 simplicity of our method is intentional for three main reasons. First, it is natural to consider whether simple text-based analysis 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 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. We also 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) that can be downloaded from the web.[6] For the RTRS source, the series produced using the alternative word list has a correlation of 0.84 with the series from our wordlists.

3.   Results

### 3.1   The evolution of measures of sentiment

We explore the relative emotion series extracted from MCDAILY in Figure 1, annotating 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. Figure 2 shows that the variation in anxiety levels is

---

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

higher 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).
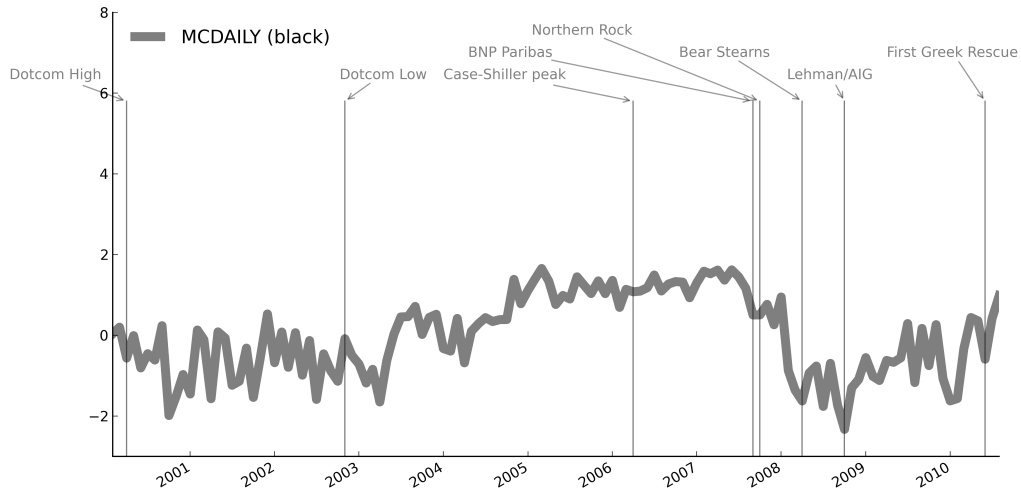
Figure 1: Relative sentiment of MCDAILY. The y-axis displays the normalized values with 0 mean and standard deviation 1
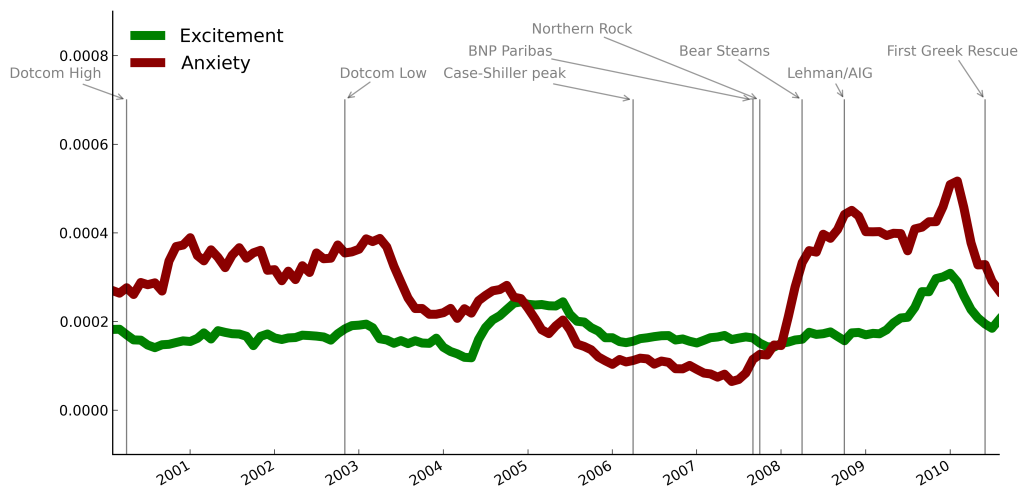
Figure 2: Emotional factors of MCDAILY; anxiety (red) and excitement (green). The y-axis displays the individual aggregate word frequencies scaled by volume

14
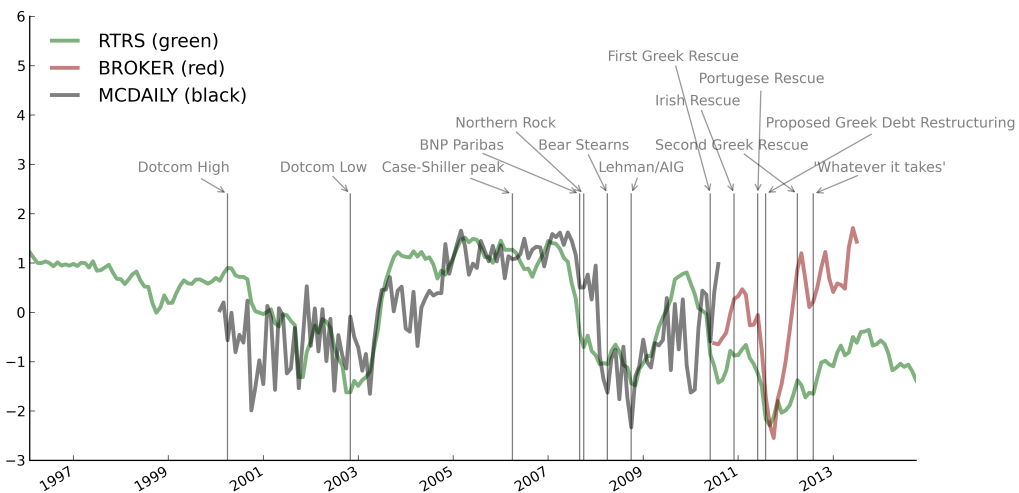
Figure 3**: Relative sentiment of MCDAILY (black), RTRS (green) and BROKER (red). The y-axis displays the normalized values with 0 mean and standard deviation 1**

The MCDAILY series is compared with those extracted from the other two sources, namely RTRS and BROKER in Figure 3. Each of the series is normalised with mean zero and standard deviation of 1 to facilitate comparison. The figure 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.

Figures 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.
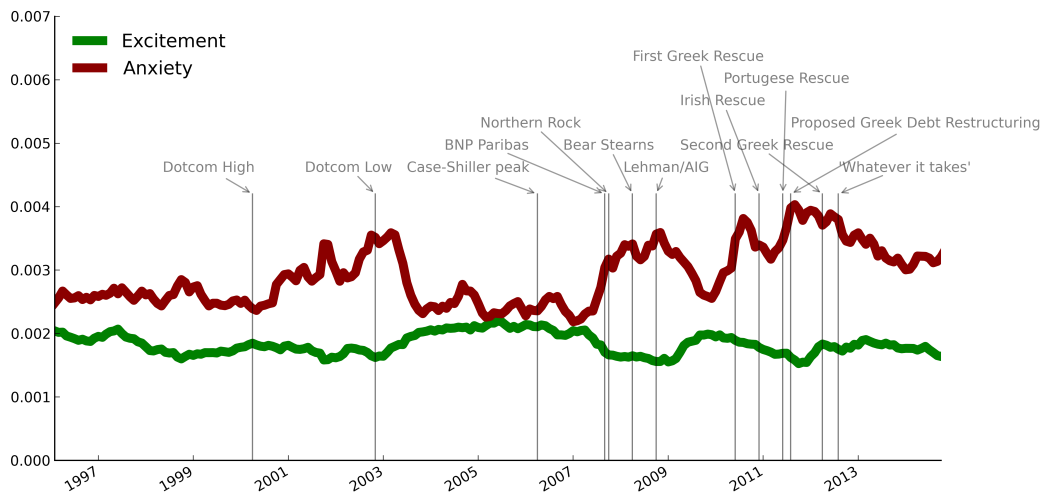
Figure 4: Excitement (green) and Anxiety (red) in RTRS. The y-axis displays the individual aggregate word frequencies scaled by volume
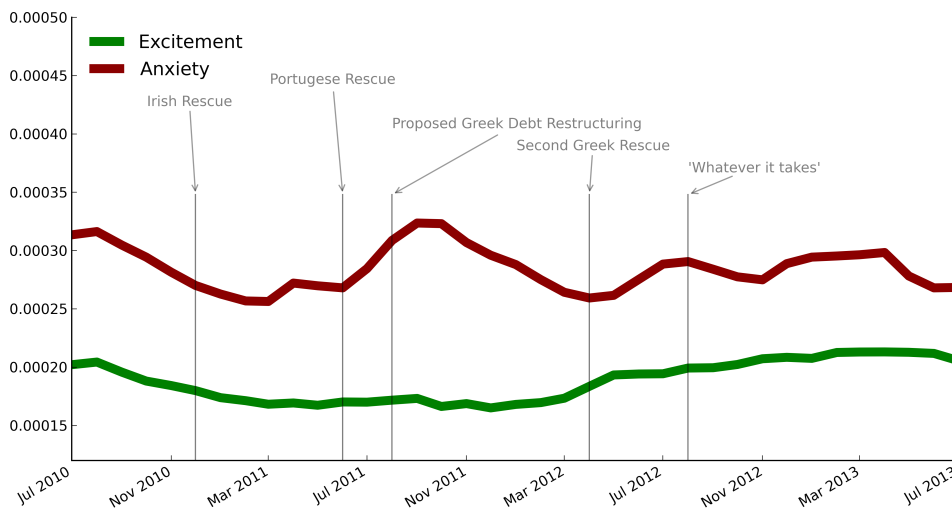

Figure 5: Excitement (green) and Anxiety (red) in BROKER. The y-axis displays the individual aggregate word frequencies scaled by volume

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' in Reuters' news archive (Figure 6) and then ran the relative sentiment analysis only on the matching sentences within all articles, with the number of sentences reflected in the bottom panel. It is particularly interesting to note the

16

steady increase and later decline in volume of articles that matched the property criteria normed by the total number of articles published in London, 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 raw relative sentiment series correlates with RTRS at 0.57 with no statistical evidence of either a lead or lag.

This is indicative of how such focused analysis could potentially be 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. This particular example seems to indicate that the property sector became overly exuberant prior to the crisis.
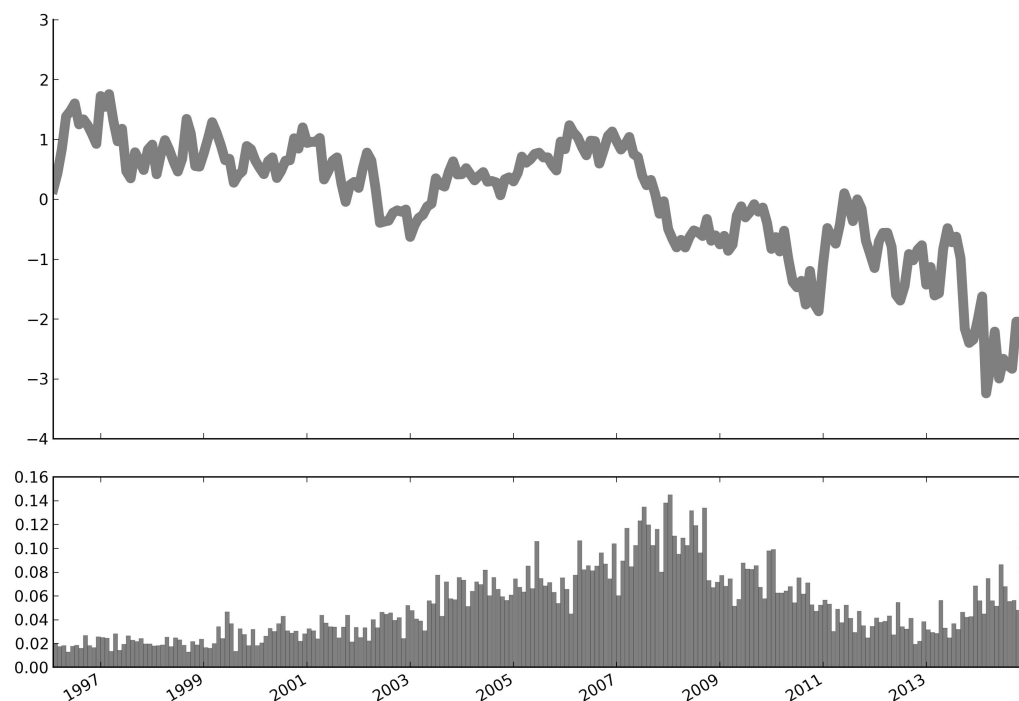
Figure 6: Relative sentiment surrounding 'property' in RTRS (smoothed). The y-axis displays the normalized values with 0 mean and standard deviation 1

3.2 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 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 – 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. The break in November 2004 in MCDAILY divided the period from April 2003 to the peak before the shift in December 2007

18

resulting in a more gradual increase than we can observe in the relative sentiment of RTRS. The breaks in both series 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 two structural breaks leading to three regimes: first a negative break in August 2007 at the onset of the financial crisis followed by a break in September 2011 marking the start of an even more negative regime.

3.3 Comparison with other measures

To illustrate how our measures compare with some other measures of uncertainty, Figure 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.
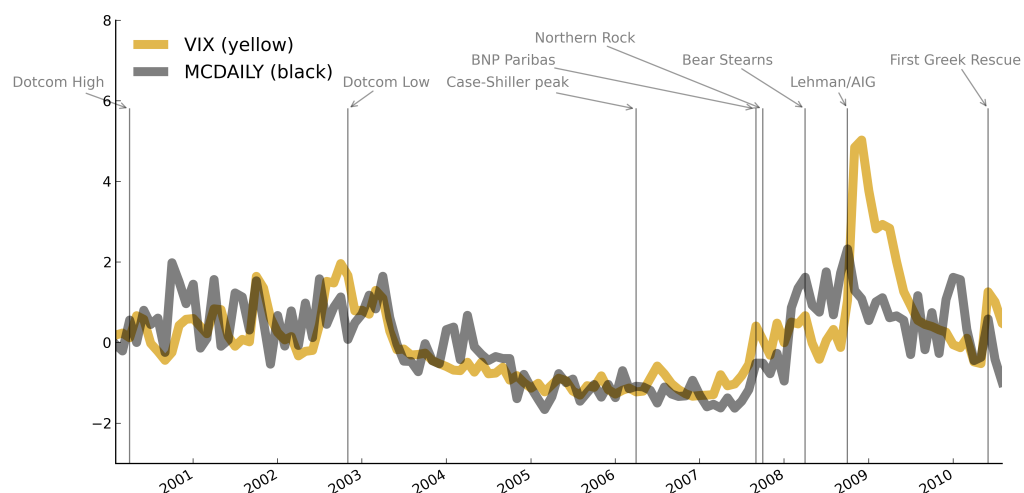
Figure 7: Relative sentiment of MCDAILY, inverted for convenience, (black) compared to the VIX (yellow). The y-axis displays the normalized values with 0 mean and standard deviation 1

More generally we can look at correlations between a wide range of variables of sentiment, confidence and uncertainty. To explore this, we computed simple pairwise correlations. The correlations with the Michigan Consumer Sentiment index[7] (MCI), the VIX[8], the economic policy uncertainty index of Baker et al. (2016)[9] (EPU), the Bank of England macroeconomic uncertainty index for the UK (BoEU – see Haddow et al. (2013))[10], senior CDS premia[11] 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 the absolute values of the correlations. 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. In some cases, the signs become less interpretable due to

---

[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, commonly 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 http://www.bankofengland.co.uk/financialstability/Pages/fpc/coreindicators.aspx. The series starts in January 2003.

[12] Business expectations survey (Markit PMI). Based on answers to the question if 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.

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 sentiment series and the various other indicators considered above. We use the methodology described in Toda and Yamamoto (1995). 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 set of steps are described in the Appendix (section 2). 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 aggregate versions of each of the sentiment series (i.e., net balance between excitement and anxiety), as well as the component parts of each series, namely excitement and anxiety. All tests are carried out on unsmoothed relative sentiment series. We use monthly variables for all tests except for those involving BoEU. In those cases, as previously, 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,

Table 2

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

	MCD	RTRS	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	-	0.54	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	0.71	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
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

Table 3

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

	MCD	RTRS	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	-	0.20	-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	0.53	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
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

CDS and PMI to the various versions of the RTRS, BROKER and MCDAILY variables.[14]

Table 4

Wald test p-values of Granger-causality from the relative sentiment shift series RTRS,

BROKER and MCDAILY 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

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 sets out the corresponding p-values.

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	ΔMCI	ΔVIX	ΔBoEU	ΔEPU	ΔCDS	ΔPMI
ΔRTRS	0.0053**		0.001**		0.18	

RSS Series	ΔMCI	ΔVIX	ΔBoEU	ΔEPU	ΔCDS	ΔPMI
ΔMCDAILY	0.66	0.96		0.26		0.11
ΔBROKER		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	ΔMCI	ΔVIX	ΔBoEU	ΔEPU	ΔCDS	ΔPMI
ΔRTRS	0.21		1.4e-06**		0.34	
ΔMCDAILY	0.87	0.41		0.25		0.4
ΔBROKER		0.96		0.98	0.61	0.31

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

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 measure 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.

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. In particular, they may be useful to gauge future financial market volatility, consumer confidence and various measures of uncertainty.

To illustrate this, we follow Nyman et al. (2014) and show how BROKER can be used to predict, out-of-sample, the change in the Michigan Consumer Sentiment index. For each prediction, a simple regression model is estimated using data up to time t and is used to

predict the change in MCI from time t to t+1. This replicates as far as possible a genuine forecasting situation.

The adjusted R-squared when regressing the predictions on the actual changes in the MCI is 0.49. The predictions are unbiased as the constant term in the regression is not significantly different from 0 and the coefficient on the predictions is not significantly different from 1. This compares to an equivalent adjusted R-squared of 0.11 when consensus forecasts made by economists and published in Reuters are regressed on the actual changes in the MCI.

Figure 8 illustrates the difference in accuracy of the forecasts made using BROKER and the consensus forecasts. The solid grey line (MCI) shows the actual changes in the Michigan index. The solid blue line (RSS BROKER) shows the predictions made using BROKER and the dotted black line (CONSENSUS) shows the consensus predictions.
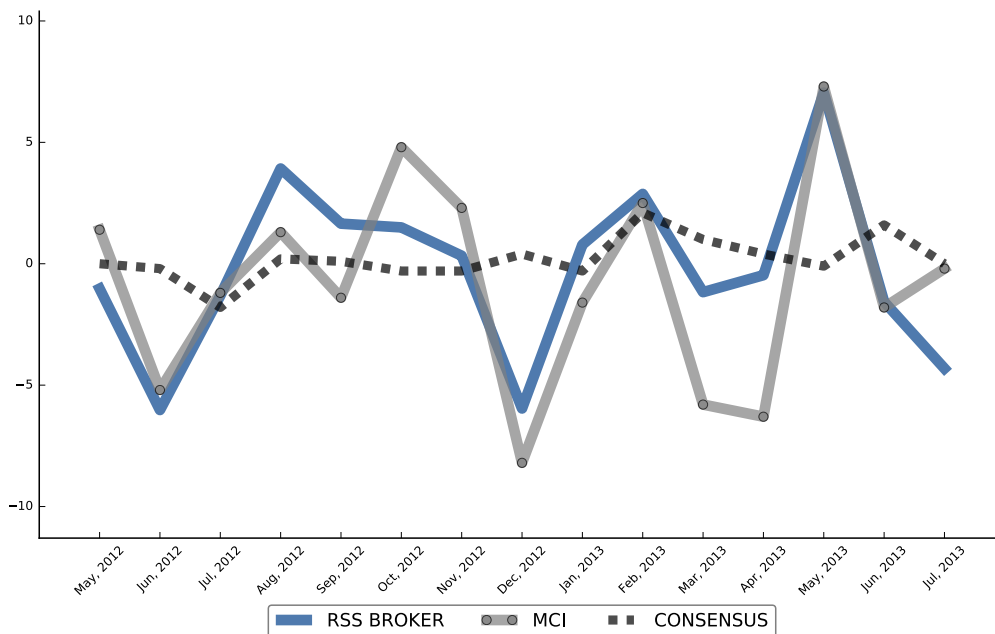
Figure 8: Change in MCI compared to forecasts of the change made using BROKER and consensus economist forecasts

3.4 Effect of relative sentiment on the UK economy

25

We now explore the relationship between relative sentiment and economic activity in the context of vector auto regression (VAR). In this exercise we use the RTRS series extracted from news for two reasons: 1) it is the longest relative sentiment series of the three and 2) emotions expressed in general economic, financial and business news are arguably more likely to be related to economic activity than, e.g., financial commentary.

VAR models have been used 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, the log of the relevant stock price index, official interest rate, log employment, and log industrial production.

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

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

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, 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 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 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.

Figure 9 shows the impulse response of a 2.5 standard deviation negative shock to RSS on IP, Employment and the FTSE 100 index. These provide indicative evidence that RSS has a significant impact on industrial production (IP), employment and the UK stock market. In each case, the effect lasts for almost 20 months.
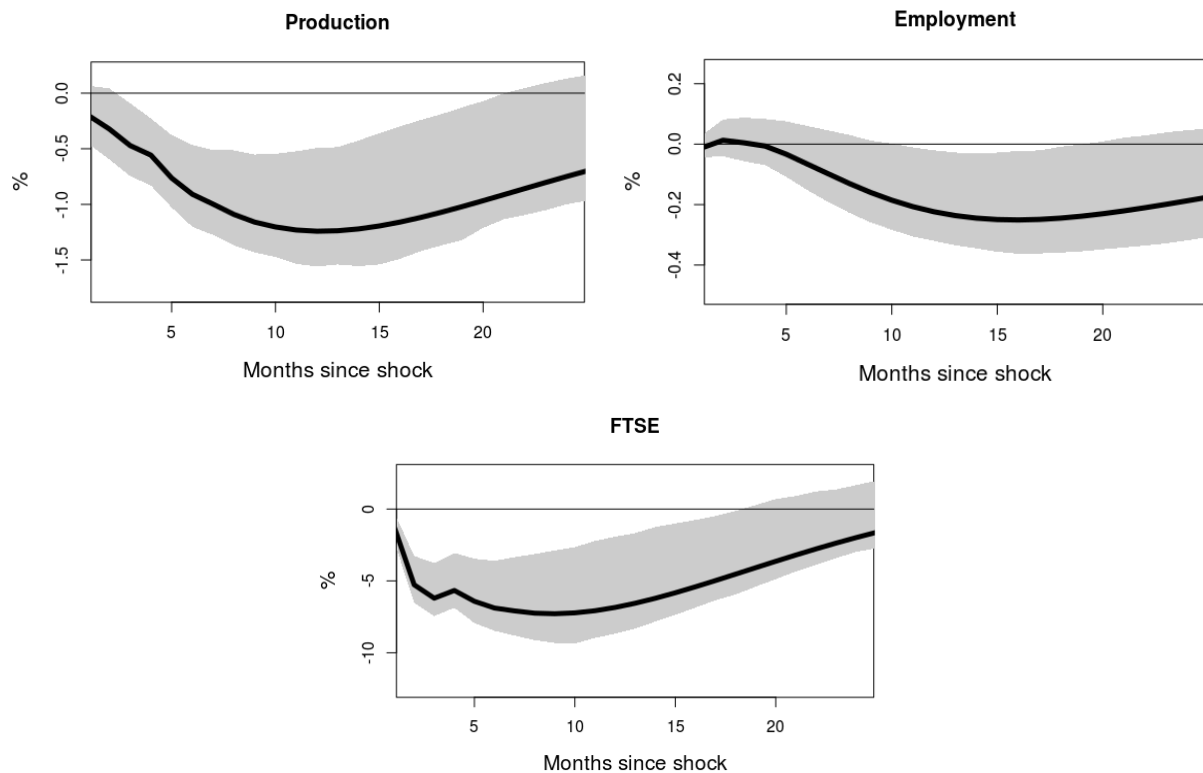
Figure 9: Impact of a negative shock to RSS on industrial production, employment and the FTSE 100

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 changing the order of the variables by placing RSS last instead of first does not change the impulse responses and neither does adding a trend.

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 financial markets 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 focused 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 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 Figure 11. 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 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 from the largest topic cluster at this time reveals a common theme about weak credit conditions and economic uncertainty.

An analysis of structural breaks, with the same parameters as used in the analysis of breaks for the RSS series, 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.

29

Table 8 shows the outcome of Wald tests of Granger causality between the entropy series and other measures which speak to uncertainty: the VIX, the Bank of England Uncertainty Index (BoEU) and the EPU. The p-values of Granger-causality in the converse direction can be seen in Table 9.[15] According to these tests, there is evidence of Granger causality from the entropy variable to both the BoEU index and the EPU. This is indicative of hour our measure of consensus / disagreement may give an early signal of changes in uncertainty.
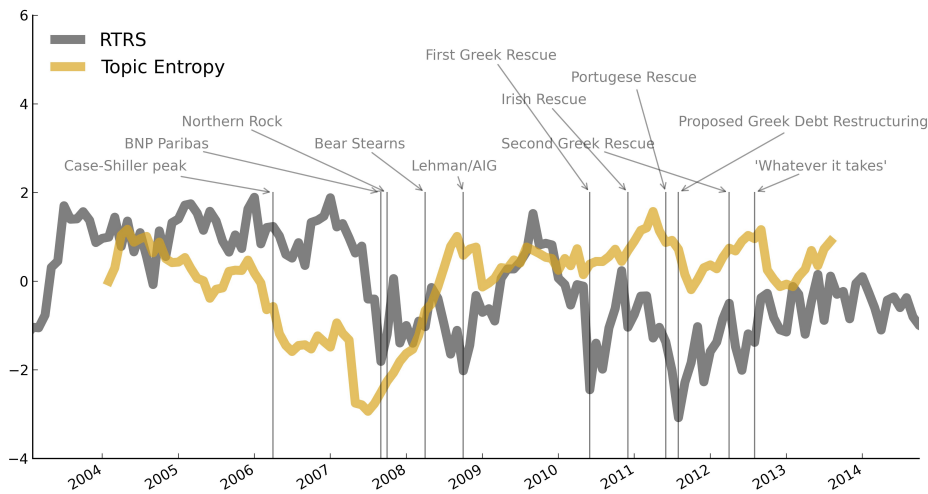

Figure 11: Relative sentiment (black) and entropy (yellow) in Reuters' London news. The y-axis displays the normalized values with 0 mean and standard deviation 1

Table 8

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		

Table 9

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

---

[15] We use monthly variables in each case except that involving BoEU for which we produce a quarterly entropy series by averaging the months within the quarter.

	VIX	BoEU	EPU
Entropy	0.27	0.98	0.52

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

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. At a lower frequency, and with the benefit of hindsight, the metrics also appear to signal rising concerns prior to the global financial crisis.   In this section, we focus on the potential uses of indicators for emerging financial system stress, but we note that the text sources linked more closely to macroeconomic variables could be useful in forecasting or 'now-casting' economic activity (Tuckett and Nyman, 2017).

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.[16] Many authorities use indicator dashboards or cobwebs, including the European Systemic Risk Board, the Office of Financial Research in the United States, the World Bank,

---

[16] See Aikman et al. (2009), Burrows et al. (2012) and Kapadia et al. (2013) for a discussion of the Bank of England's approach to stress testing and its "RAMSI" model. On early warning indicator models, see Kaminsky and Reinhart (1999), Drehmann et al. (2011), Borio and Lowe (2002, 2004), Barrell et al. (2010) 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).

the Reserve Bank of New Zealand and the Norges Bank.[17] 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).[18]

Recognising that no single set of indicators or models can ever provide a perfect guide to

systemic risk, due to the complexity of financial interlinkages and the tendency for the

financial system to evolve over time, and time lags before risks become apparent, judgement

also plays a crucial role in specifying any policies to tackle threats to the financial system.

And qualitative information, including from market and supervisory intelligence typically

also helps to support such judgements.

As we have shown in previous sections, 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

---

[17] 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). On the cobweb approach used in
New Zealand and Norway, see Bedford and Bloor (2009) and Dahl et al. (2011), respectively.
[18] See also Giese et al. (2014).

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 the outcome of such procedures can lead to some intuitive and useful representations of financial market sentiment. At a low frequency, the shifts correlate well with financial market events and, at a higher frequency, the appear to lead a number of financially oriented economic indicators.

We have also developed a novel methodology to measure consensus in the distribution of narratives. This metric can potentially be used to measure homogenisation among market participants. Greater consensus, when viewed together with an increase in the relative sentiment series, 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 behaviour leading up to the financial crisis.

Overall, the relative sentiment and consensus summary statistics developed 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.

Acknowledgments

References

Aikman, D, Alessandri, P, Eklund, B, Gai, P, Kapadia, S, Martin, E, Mora, N, Sterne, G and Willison, M (2009), 'Funding liquidity risk in a quantitative model of systemic stability', *Bank of England Working Paper No. 372*.

Aikman, D, Nelson, B and Tanaka, M (2015), 'Reputation, risk-taking and macroprudential policy', *Journal of Banking and Finance*, 50, pages 428-439.

Akerlof, G A & Shiller, R. (2009). Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism. Princeton: Princeton University Press.

Bai J., Perron P. (2003), Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, **18**, 1-22.

Bailey, M., Cao, R., Kuchler, T. and Stroebel, J., 2016. *Social networks and housing markets* (No. w22258). National Bureau of Economic Research.

Baker, M., Stein, J. S. (2004). Market Liquidity as a Sentiment Indicator. Journal of Financial Markets, 7(3): pp. 271-299.

Baker, M., Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, 61(4): pp. 1645-1680.

Baker, M., Wurgler, J. (2007). Investor Sentiment in the Stock Market. The Journal of Economic Perspectives, 21(2): pp. 129-151.

Baker, M., Wurgler, J., Yuan, Y. (2012). Global, Local and Contagious Investor Sentiment. The Journal of Financial Economics, 104(2): pp. 272-287.

Baker, S. R., Bloom, N., Davis, S. J. (2016). 'Measuring Economic Policy Uncertainty', *Quarterly Journal of Economics*, 131(4), pages 1593-1636.

A. Banerjee, J. J. Dolado, J. W. Galbraith, and D. F. Hendry (1993): *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.

Bank of England (2014), 'The Financial Policy Committee's powers to supplement capital requirements: a Policy Statement', January.

Barrell, R., E. Davis, D. Karim, and I. Liadze (2010). "Bank regulation, property prices and early warning systems for banking crises in OECD countries", Journal of Banking and Finance, vol. 34, no. 9, pp. 2255-2264.

Barsky, R.B. and Sims, E.R., 2012. Information, animal spirits, and the meaning of innovations in consumer confidence. *The American Economic Review*, *102*(4), pp.1343-1377.

Baumeister, R.F., & Masicampo, E.J. (2010). Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal-culture interface. Psychological Review, 117, 945-971.

Berry, M. W. (2004). Survey of text mining: Clustering, classification, and retrieval. Springer.

Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., Jansen, D. J., (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence. European Central Bank, Working Paper Series No 898.

Bloom, N (2009), 'The impact of uncertainty shocks', *Econometrica*, Vol. 77, pages 623–85.

Borio, C., and P. Lowe (2002). "Assessing the risk of banking crises", BIS Quarterly Review, December, pp 43-54.

Borio, C., and P. Lowe (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. The Journal of Business, 78(2): 405-440.

Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry*, 1-21.

Burrows, O, Learmonth, D and McKeown, J (2012). RAMSI: a top-down stress-testing model. Bank of England Financial Stability Paper No. 17.

Carver, C. S., & Harmon-Jones, E. 2009. Anger is an approach-related affect: evidence and implications. Psychological bulletin, 135(2): 183.

Chong, K. and Tuckett, D. (2015) Constructing Conviction through Action and Narrative: How Money Managers Manage Uncertainty and the Consequences for Financial Market Functioning. Socioecon Rev (2015) 13 (2): 309-330. DOI: https://doi.org/10.1093/ser/mwu020

Deerwester, S. et al. (1988). Improving information retrieval with latent semantic indexing. In: Proceedings of the 51st Annual Meeting of the American Society for Information Science, Vol. 25, 36-40.

Drehmann, M, Borio, C and Tsatsaronis, K (2011), 'Anchoring countercyclical capital buffers: the role of credit aggregates', *International Journal of Central Banking*, Vol. 7, No. 4, pages 189–240.

Edmans, A., Garcia, D., Norli, A. (2007). Sports Sentiment and Stock Returns. The Journal of Finance, 62(4): 1967-1998.

Fernandez-Corugedo, E and Muellbauer, J (2006), 'Consumer credit conditions in the United Kingdom', *Bank of England Working Paper No. 314.*

Giese, J, Andersen, H, Bush O, Castro, C, Farag, M and Kapadia, S (2014), 'The credit-to-GDP gap and complementary indicators for macroprudential policy: evidence from the UK'. *International Journal of Finance and Economics*, 19(1), 25-47.

Gray, D. F., R. C. Merton, and Z. Bodie (2008). A new framework for measuring and managing macrofinancial risk and financial stability, Harvard Business School Working Paper no. 09-015.

Gray, D. F., and A. Jobst (2011). "Modeling systemic financial sector and sovereign risk", Sveriges Riksbank Economic Review, September.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37(3): pp. 424-438. DOI: 10.2307/1912791.

Greenwood, R., S. Nagel (2009). Inexperienced Investors and Bubbles. Journal of Financial Economics, 93(2): 239-258.

Haddow, A, Hare, C, Hooley, J and Shakir, T (2013), 'Macroeconomic uncertainty; what is it, how can we measure it and why does it matter?', Bank of England Quarterly Bulletin, Vol. 53, No. 2, pages 100-109.

Hansen, S, McMahon, M and Prat, A (2014), "Transparency and Deliberation within the FOMC: a Computational Linguistics Approach," CEPR Discussion Paper 9994.

Harmon-Jones, E., Harmon-Jones, C., & Price, T. F. 2013. What is Approach Motivation? Emotion Review, 5(3): 291-295.

Holló, D., M. Kremer, and M. Lo Duca (2012). CISS - a composite indicator of systemic stress in the financial system, European Central Bank Working Paper, no. 1426.

Hubert, P. (2012). Disentangling Qualitative and Quantitative Central Bank Influence. Preliminary Version.

Illing, M, and Liu, Y (2006). "Measuring financial stress in a developed country: an application to Canada", *Journal of Financial Stability*, vol. 2, no. 4, pp. 243-265.

Kaminsky, G.L and Reinhart, C.M (1999) The twin crises: the causes of banking and balance-of-payments problems. *American Economic Review*, vol. 89, no. 3, June, pp. 473-500.

Kapadia, S., M. Drehmann, J. Elliott, and G. Sterne (2013). "Liquidity risk, cash flow constraints, and systemic feedbacks", in J. G. Haubrich, and A. W. Lo (eds.), Quantifying Systemic Risk, University of Chicago Press.

Keynes, J.M. (1936). *The General Theory of Employment, Interest and Money*. Palgrave Macmillan.

Kohonen, Teuvo (1982). "Self-Organized Formation of Topologically Correct Feature Maps". Biological Cybernetics 43 (1): pp. 59–69.

Knight, F. H. (1921). Risk, Uncertainty and Profit. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics* **54**, 159–178.

Loughran, Tim and Bill McDonald, 2011, "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks," Journal of Finance, 66:1, 35-65.

Lucas, Robert (1976). "Econometric Policy Evaluation: A Critique". In Brunner, K.; Meltzer, A. The Phillips Curve and Labor Markets. Carnegie-Rochester Conference Series on Public Policy 1. New York: American Elsevier. pp. 19–46

Lux, T. (2008). Sentiment Dynamics and Stock Returns: The Case of the German Stock Market. Kiel Working Papers No 1470.

Mousavi, Shabnam and Gigerenzer, Gerd, Risk, Uncertainty, and Heuristics (March 19, 2014). Journal of Business Research, Vol. 67, No. 8, March 2014. Available at SSRN: https://ssrn.com/abstract=2657782

Miller, G.A. (1956), The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review, 63, 81-97.

Nyman, R., Ormerod, P., Smith, R., Tuckett, D. (2014). Big Data and Economic Forecasting: a Top-Down Approach using Directed Algorithmic Text Analysis. Workshop on using big data for forecasting and statistics. European Central Bank. 7-8 April, 2014. Available at: https://www.ecb.europa.eu/events/pdf/conferences/140407/TuckettOrmerod_BigDataAndEconomicForecastingATop-DownApproachUsingDirectedAlgorithmicTextAnalysis.pdf

OFR annual report (2012) https://financialresearch.gov/annual-reports/files/office-of-financial-research-annual-report-2012.pdf

Panksepp, J. 2013. Cross-species neuroaffective parsing of primal emotional desires and aversions in mammals. Emotion Review, 5(3): 235-240.

Pelleg, D., Moore, A. (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. Proceedings of the Seventeenth International Conference on Machine Learning. Pp. 727-734.

Phelps, E. A. 2006. Emotion and cognition: insights from studies of the human amygdala. Annu. Rev. Psychol., 57: 27-53.

Porter, M. F. (1980). An Algorithm for Suffix Stripping. Program, 14(3): pp. 130-137.

Pribram, Karl H.; Miller, George A.; Galanter, Eugene (1960). Plans and the structure of behavior. New York: Holt, Rinehart and Winston

Reinhart, C, and Rogoff, K (2009). *This time is different: Eight Centuries of Financial Folly*, Princeton University Press.

Rolls, E. T. 2013. What are emotional states, and why do we have them? Emotion Review, 5(3): 241-247.

Rottman, M.B., and Hastie, R. (2013) Reasoning About Causal Relationships: Inferences on Causal Networks. Psychological Bulletin. Psychol Bull. 2014 Jan; 140(1): 109–139. doi: 10.1037/a0031903

Said, S. E. and D. A. Dickey (1984): Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* **71**, 599–607.

Shannon, Claude E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal **27** (3): 379–423.

Shiller, R (2000). *Irrational Exuberance*. Princeton University Press.

Shiller, R (2017). 'Narrative economics', *Cowles Foundation Discussion Paper*, Number 2069.

Sloman, S.A and Lagnado, D (2015) Causality in Thought. Annu. Rev. Psychol. 2015. 66:3.1–3.25

Soo, C. K. (2013). Quantifying Animal Spirits: News Media and Sentiment in the Housing Market. The Stephen M. Ross School of Business at the University of Michigan Research Paper Series, Forthcoming. Available at SSRN: http://ssrn.com/abstract=2330392

Sturm, J. E., De Haan, J., (2011). Does central bank communication really lead to better forecasts of policy decisions? New evidence based on a Taylor rule model for the ECB. Rev World Econ. DOI 10.1007/s10290-010-0076-4

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." The Journal of Finance, 62(3): pp. 1139–1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. 2008."More than Words: Quantifying Language to Measure Firms' Fundamentals." The Journal of Finance, 63(3): pp. 1437–1467.

Tetlock, P. C. (2011). All the News That's Fit to Reprint: Do Investors React to Stale Information? Review of Financial Studies 24, pp. 1481-1512.

Toda, H.Y. and T. Yamamoto (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes, Journal of Econometrics, Vol. 66, pp. 225-250

Tuckett, D. (2011). Minding the Markets: An Emotional Finance View of Financial Instability. Palgrave Macmillan.

Tuckett, D.A and Nikolic, M., (2017) The Role of Conviction in Decision-Making Under Radical Uncertainty. Theory and Psychology. Online early at Sage Open. https://doi.org/10.1177/0959354317713158

Tuckett, D., Smith, R. E., & Nyman, R. (2014). Tracking phantastic objects: A computer algorithmic investigation of narrative evolution in unstructured data sources. Social Networks, 38, 121–133. doi:10.1016/j.socnet.2014.03.001

Tuckett, D and Nyman, R., (2017) the relative sentiment shift series for tracking the economy. (under review). https://www.researchgate.net/publication/316923347_THE_RELATIVE_SENTIMENT_SHIFT_SERIES_FOR_TRACKING_THE_ECONOMY

Watts, D (2002), 'A simple model of global cascades on random networks', *Proceedings of the National Academy of Sciences*, Vol. 99, pages 5,766-71

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.

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

1.2 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 quarter. We can then compute the relative sentiment score as described in the main text.

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 either 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 remain correlated with the original series as highly as 0.999 in level form and 0.999 in difference form. Although this might initially seem counterintuitive, on second thought it is in fact trivial to understand that negation will only effect the movement of the series if there is a systematic bias in its use for a given period. In other words, 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 level form and 0.99 in difference form. We can conclude that the RSS measure is likely not effected by the presence of economic terms.

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)

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 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=trunc($(N$-1$)^{\wedge}(1/3))$, 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

45

```
➤ 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 (Porter, 1980)

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 rarely help us distinguish between topics and

    the least frequent words typically introduce too much noise and fail to show

consistent patterns. Another commonly used technique is to remove all words in a predefined list, so called 'stopwords'.

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,

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

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),

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

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

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

For $t > 0$:

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

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

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.

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

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

Note: *p<0.1; **p<0.05; ***p<0.01

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 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

Note: *p<0.1; **p<0.05; ***p<0.01

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 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

Note: *p<0.1; **p<0.05; ***p<0.01

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

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

Note: *p<0.1; **p<0.05; ***p<0.01

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

Note: *p<0.1; **p<0.05; ***p<0.01

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

Note: *p<0.1; **p<0.05; ***p<0.01

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*

Note: *p<0.1; **p<0.05; ***p<0.01

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

Note: *p<0.1; **p<0.05; ***p<0.01

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

Note: *p<0.1; **p<0.05; ***p<0.01

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

Note: *p<0.1; **p<0.05; ***p<0.01

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

Note: *p<0.1; **p<0.05; ***p<0.01