

News and narratives in financial systems: Exploiting big data for systemic risk assessment¹

Rickard Nyman^{*}, David Gregory^{**}, Sujit Kapadia^{**}, Paul Ormerod^{*},
David Tuckett^{*} & Robert Smith^{*}

September 2016

Abstract:

A number of recent contributions have tried to add to the understanding and forecasting of the macro economy by analysing news and narratives. This paper applies algorithmic analysis to large amounts of financial market text-based data to assess how narratives and emotions 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 – leading up to the global financial crisis. And we find that the shifts have predictive power for other commonly used measures of sentiment and volatility. We also show that a new methodology that attempts to capture the emergence of narrative topic consensus 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 author(s) and should not be taken to represent those of the Bank of England, the Monetary Policy Committee or the Financial Policy Committee.

^{*} University College London, Centre for the Study of Decision-Making Uncertainty

^{**} Bank of England

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.

This paper applies algorithmic analysis to large amounts of unstructured text-based data to identify quantitative metrics that try to capture shifts in sentiment along with the extent of consensus in the market. We find that these metrics capture key developments in the financial system relatively well prior to and during the global financial crisis, as well as having predictive power for other commonly used measures of sentiment and volatility. As such, these metrics could potentially be used for gauging systemic risk in financial systems and 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).

With rapid advances in ways to store and analyse large amounts of unstructured data, there is increasing awareness that these data may provide a rich source of useful information for assessing economic trends. For example, a growing literature exploits individual user-generated search engine data, such as *Google Trends*, to try to predict the current value of ('nowcast') economic variables such as GDP (see for example Choi and Varian, 2012). However, some recent studies suggest search engine data should be treated with care, either

because of a lack of transparency about how the data has been created (Lazer, et al. 2014) or uncertainty about the motivation for searching – independently or because of social influence (Ormerod et al., 2014).

By contrast, the measures of sentiment we use are pre-defined word lists representing two specific emotional groups. The words have been developed through the lens of a social-psychological theory of “conviction narratives” (CNT) (Tuckett and Nikolic, 2016), which is one way of empirically formulating the idea of “animal spirits” as described by Keynes (1936). CNT is a new theory which emphasises the role of narratives and particular groups of action-enabling or disabling emotions in driving decision-making under uncertainty (Chong and Tuckett, 2014; Tuckett and Nikolic, 2016). It has been successfully used in other applications, for example as a measure of changing macroeconomic confidence (Tuckett et al. 2015) and is part of a growing body of work inside and outside economics developing broader models of human cognition and decision-making behaviour (for example, Akerlof and Shiller, 2009; Bruner, 1990; Damasio, 1999; Lane and Maxfield, 2005; Mar and Oatley, 2008; Beckert, 2011 and Tuckett, 2011)².

² Narrative can be considered a fundamental form of mental organization (Bruner, 1991) that 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. 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 the actions by means of which they plan to achieve those opportunities and to feel sufficiently convinced about the anticipated outcomes to act (Tuckett and Nikolic, 2016). 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. Baumeister and Masicampo, 2010; Barsalou, 2008; Damasio and Carvalho, 2013; Suddendorf and Corballis, 2007.)

CNT suggests that within the context of deep, or Knightian (Knight, 1921), uncertainty (i.e., uncertainty characterised by a context in which the space of potential outcomes of some event cannot be articulated) 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. *Conviction narratives* contain a few fundamental components, notably a focus on the specific emotional elements of 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 in conviction narrative theory 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, 2016). Thus, in the simplest case, the key variables of interest are the aggregate relative difference between *excitement* and *anxiety* and *shifts* in this difference over time.

At any given moment, there will be several narratives and associated emotions circulating among all financial agents. Some of these narratives, or pieces of them, are likely to be contained within relevant text-based data sources. And if the relative shifts in the emotional content correlate across different sources, it is plausible that at least some financial agents had adopted a subset of the narratives and held them as true, though it is 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; broker research reports; and Reuters' news articles in the United Kingdom. Motivated by the conviction narrative methodology, we capture an emotional summary statistic (Relative Sentiment Shift or RSS) based on these sources, and explore changes in the statistic over time to assess 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, for example using different word lists, a key strength of this method lies in its top down approach, capturing aggregate shifts largely undetectable to the human eye.

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 or entities, 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.

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 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 – indeed, claiming to do so arguably became necessary for financial institutions to attract new investment (Aikman et. al., 2011). But such consensus in an environment of high sentiment could be suggestive of over-confidence or irrational exuberance and the theory predicts that such situations are likely to 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. In other words, our top-down text analysis methodology suggests evidence that consensual, conviction narratives emerged prior to the crisis in which anxiety and doubt substantially diminished, indicative of possible 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 MacDonald, 2011; Tetlock 2007; Tetlock et al. 2008; Tetlock 2011; Soo 2013) or to capture economic policy uncertainty (Baker et al., 2013). 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; Baker et al., 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, by focusing on a restricted dictionary of words we develop our measures of sentiment from the point of view of a social-psychological theory of action under

uncertainty (Tuckett and Nikolic, 2016). In this way we apply a theoretical filter which should more accurately detect features we hypothesise to be important and avoid some of the difficulties associated with data mining. When processing 'big data' there is a risk of obtaining seemingly significant correlations that do not generalise. Data-mining techniques may also generalise poorly to new data sources and to be highly context specific. Second our primary focus is specifically on gauging the systemic risk, rather than on movements in particular asset prices or broader macroeconomic developments. Conviction Narrative Theory postulates that systemic risk can be generated when market agents get captured by group narratives that open up a gap in the usual balance between excitement and anxiety in narratives, suggesting that some kind of "this time is different" process (Reinhart and Rogoff, 2009) is going on (Tuckett, 2011). If so, a severe correction can be expected to follow. Third, much current research that applies some form of text-based sentiment analysis to study the economy or financial markets tends to exploit social media generated data. By contrast, we focus on data sources more specifically connected to the financial system, including one source written within a central bank.

Section 2 explains the data we have analysed and focuses on our measure of emotion, or sentiment, and explains the methodology. Section 3 sets out results, including Granger causality tests between the emotion indices and various financial/economic indicators. 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. A certain amount of technical material is made available in an Appendix. Full details of all the

statistical tests involved in our analysis, along with further technical material, is in a Supplementary Material document, available on request from the authors.

2. Data and Methodology

We make use of a variety of data sources with a macroeconomic and 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 articles. For most of this paper, we restrict our attention to news published by Reuters in London during the period between January 1996 and September 2014, in which 6,123 articles were published on average each month (after excluding all articles tagged by Reuters as ‘Sport’, ‘Weather’ and/or ‘Human Interest’). For the rest of the paper, we refer to this database as ‘Reuters (RTRS)’.

2.4 Relative Sentiment Shifts

A summary statistic of two emotional traits is extracted from our text data sources by a word count methodology described in more detail elsewhere (Tuckett, Smith and Nyman, 2014). Two lists of previously applied and experimentally validated 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.

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.³ To arrive at a single statistic, highly relevant to the theory of conviction narratives, 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.

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

We compute this on a monthly basis. 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 carry out a test (on the RTRS database) if the presence of negation words would affect the sentiment series (following the procedure outlined in Loughran and McDonald, 2011) by excluding all words counted to produce the sentiment score if they were preceded (within a window of three words) by either of the words: “no”, “not”, “none”, “neither”, “never” or “nobody”. The ‘negation aware’ series thus produced remained correlated with

³ 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.

the original series as highly as 0.99, both in level and difference form. We also carried out a test of the robustness of the methodology to an alternative selection of sentiment wordlists. We applied the sentiment methodology to the ‘positive’ and ‘negative’ wordlists produced by Loughran and McDonald that can be downloaded from the web⁴. The series produced using the RTRS database is correlated with the series produced using our wordlists by 0.84.

The simplicity of our method is intentional for two main reasons. First, it is natural to consider whether simple text-based analysis can be informative before moving to more complex methods. Second, 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.⁵ This technique gives us increased confidence that the meaning of individual words in our two lists does not change over time.

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 purely 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

⁴ The lists were downloaded from the website as they were available in 2011, www3.nd.edu/~McDonald/word_lists.html

⁵ It is easy to imagine other methods of extracting confidence levels, e.g., to sample with replacement from the collection of texts.

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.

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 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.

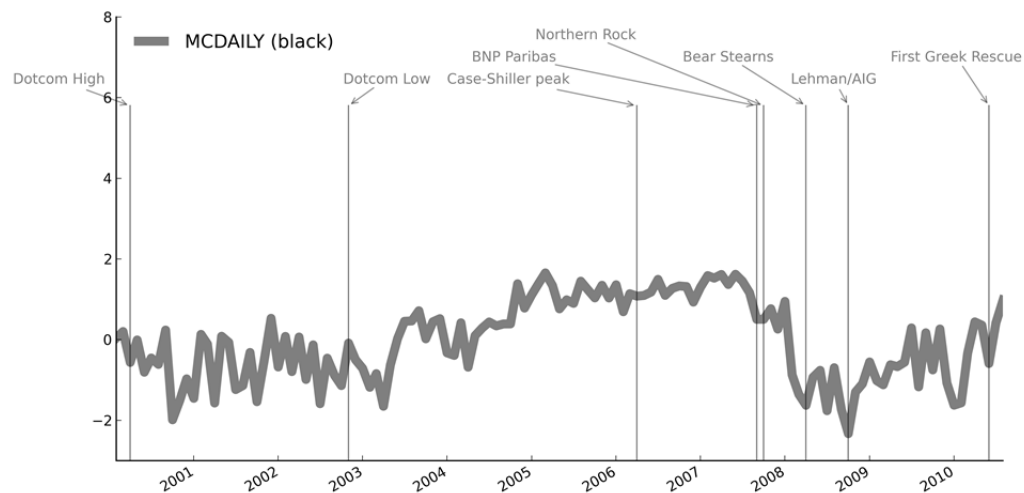


Figure 1: Relative sentiment of MCDAILY. The y-axis displays the normalized values with 0 mean and standard deviation 1

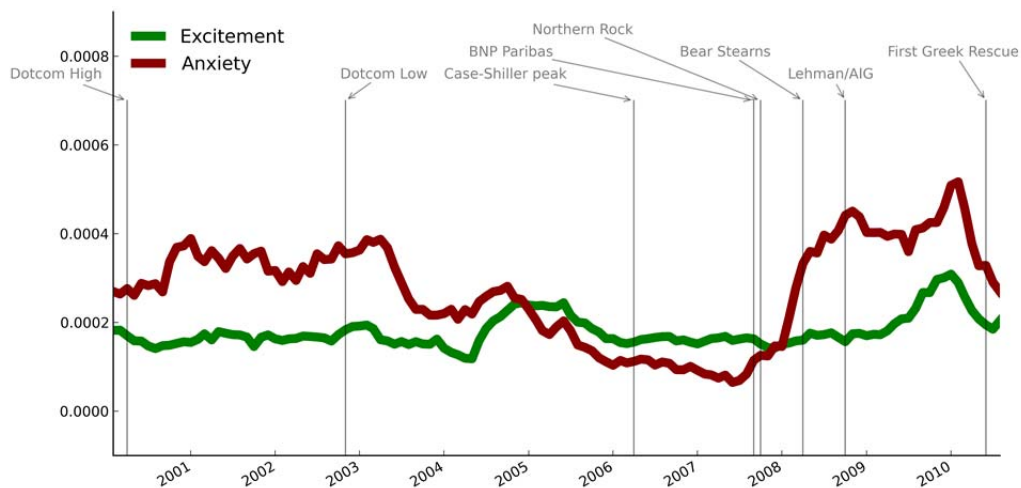


Figure 2: Emotional factors of MCDAILY; anxiety (red) and excitement (green). The y-axis displays the individual aggregate word frequencies scaled by volume

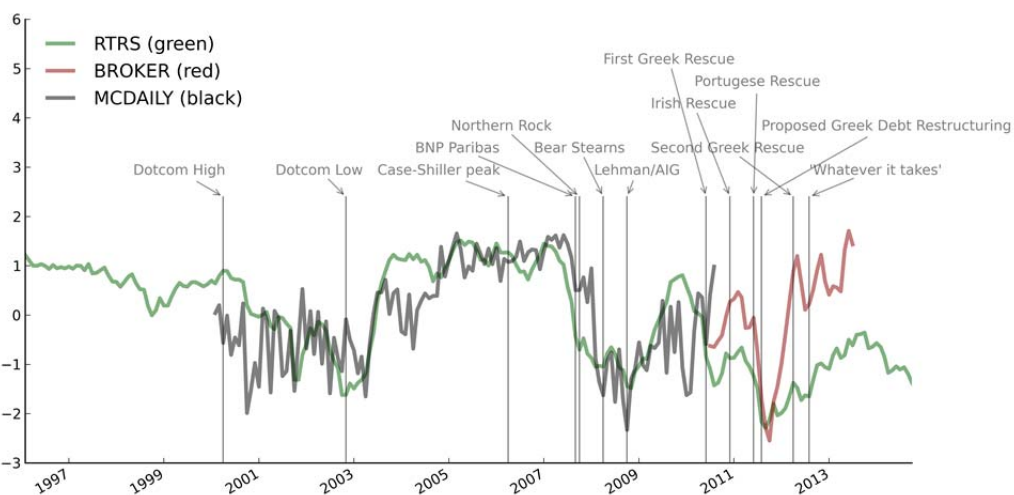


Figure 3: Relative sentiment of MCDAILY (black), RTRS (green) and BROKER (red). The y-axis displays the normalized values with 0 mean and standard deviation 1

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. We are primarily interested in the aggregate difference between the excitement and anxiety components, for theoretical reasons. However, the analysis is carried out using the two components separately and results are presented in the appendix.

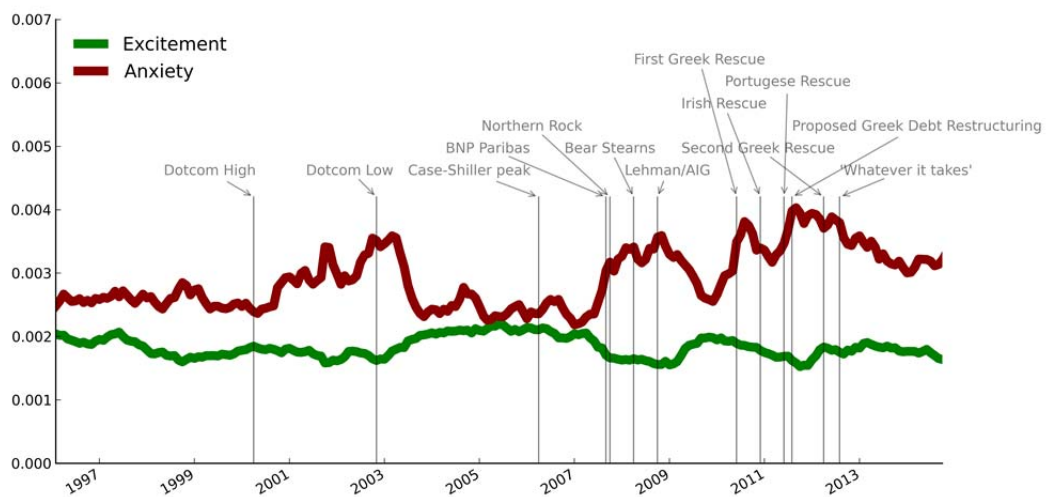


Figure 4: Excitement (green) and Anxiety (red) in RTRS. The y-axis displays the individual aggregate word frequencies scaled by volume

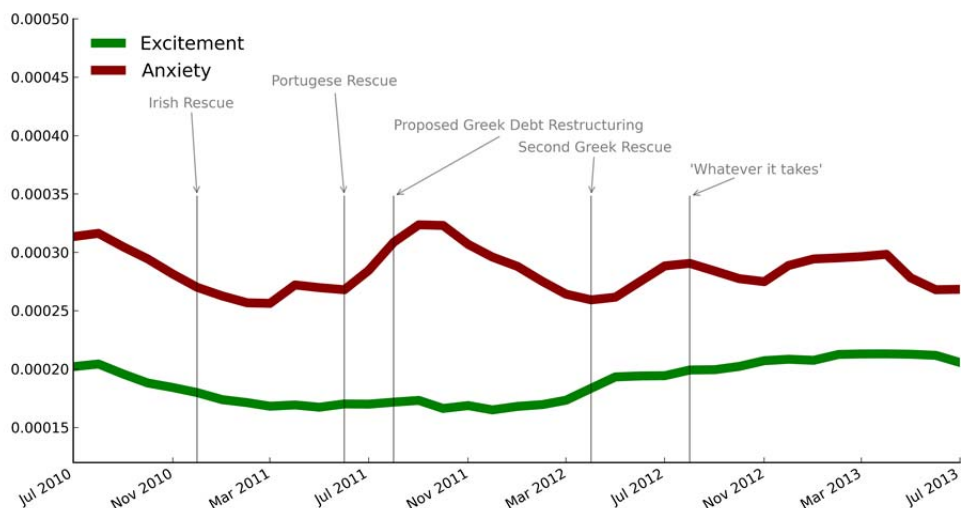


Figure 5: Excitement (green) and Anxiety (red) in BROKER. The y-axis displays the individual aggregate word frequencies scaled by volume

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 explore the potential of such an approach, 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 (recall these are articles published in London). It is particularly interesting to note the 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 (which has here been smoothed with parameter $\alpha = 0.3$) 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 (anxiety relative to excitement steadily dropped out of the discussion). The raw relative sentiment series correlates with RTRS at 0.57 with no statistical evidence of either a lead or lag.

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.

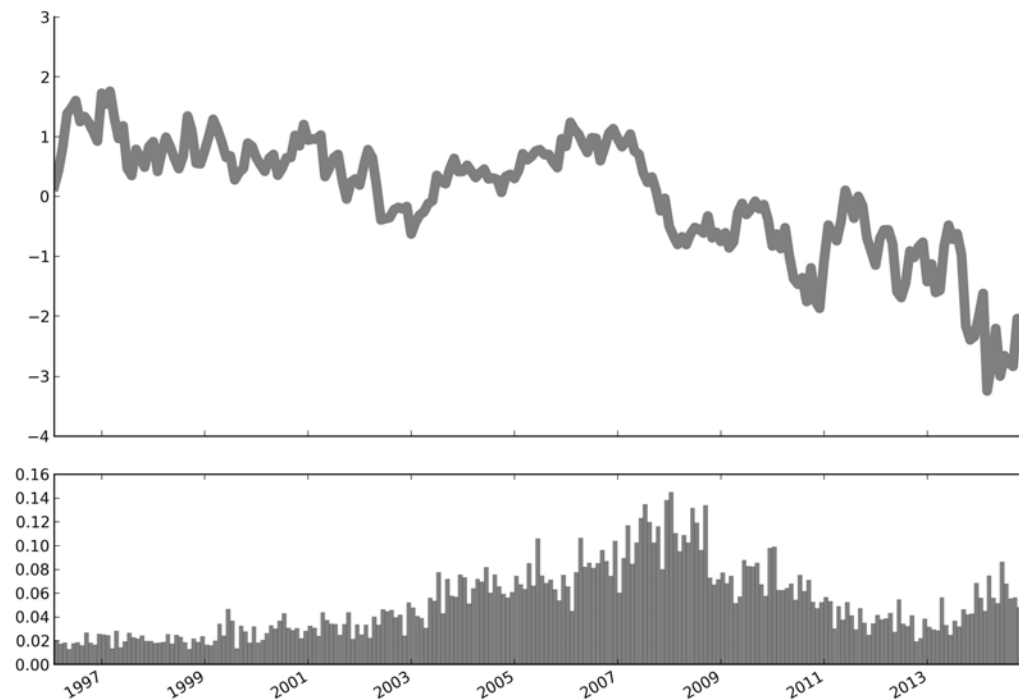


Figure 6: Relative sentiment surrounding 'property' in RTRS. The y-axis displays the normalized values with 0 mean and standard deviation 1

3.2 Structural breaks

Importantly, 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 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. This yields four structural breaks for the RTRS series on August 2000, May 2003, May 2007 and April 2010. We find three breaks for the MCDAILY series on April 2003, November 2004 and December 2007. We find two breaks for the BROKER series on May 2011 and October 2011.

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.

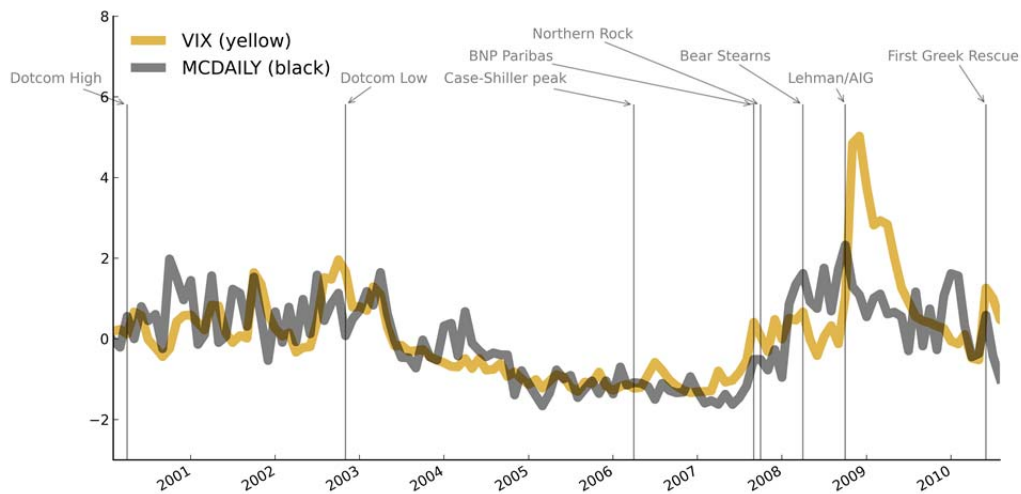


Figure 7: Relative sentiment of MCDAILY, inverted for convenience, (black) compared to the VIX (yellow). The y-axis displays the normalized values with 0 mean and standard deviation 1

More generally we can look at correlations between variables. To explore how our measures compare to a range of other financial and economic indicators, we conducted a simple correlation based pairwise comparison study. The correlations with the Michigan Consumer Sentiment index⁶ (MCI), the VIX⁷, the economic policy uncertainty index of Baker et al. (2013)⁸ (EPU), the Bank of England macroeconomic uncertainty index (BoEU – see Haddow et al. (2013)), senior CDS premia⁹ and PMI¹⁰ together with the correlations between the

⁶ 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.

⁷ 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'.

⁸ We use the UK version of the series available at http://www.policyuncertainty.com/europe_monthly.html. The series starts in January 1997.

⁹ 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.

individual relative sentiment series are presented in Table 2.¹¹ To facilitate comparisons and since the sign of all correlations is as expected, we give the absolute values of the correlations. It is clear that our three sentiment measures are fairly highly correlated with all of the other measures.

Table 2

Correlations between relative sentiment series and common measures of sentiment, ignoring signs

	MCD	RTRS	BRO	VIX	MCI	EPU	BoEU	CDS	PMI
MCD	1	0.59	-	0.65	0.26	0.43	0.54	0.67	0.38
RTRS	-	1	0.71	0.37	0.54	0.61	0.52	0.71	0.51
BRO	-	-	1	0.57	0.66	0.06	0.60	0.23	0.42

Lead-lag correlations between the sentiment series and the economic variables are presented in Table 3a and 3b.

Table 3a

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

	VIX(-1)	VIX(+1)	MCI(-1)	MCI(+1)	EPU(-1)	EPU(+1)	BoEU(-1)	BoEU(+1)
MCD(t)	0.56	0.65	0.24	0.27	0.30	0.41	0.43	0.61
RTRS(t)	0.26	0.37	0.49	0.58	0.63	0.63	0.35	0.67
BRO(t)	0.34	0.65	0.34	0.87	0.26	0.01	0.06	0.76

Table 3b

¹⁰ 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.

¹¹ 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. We do not do this for the VIX and the MCI as that is of less relevance here.

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

	CDS(-1)	CDS(+1)	PMI(-1)	PMI(+1)
MCD(t)	0.63	0.63	0.38	0.43
RTRS(t)	0.67	0.69	0.43	0.57
BRO(t)	0.05	0.22	0.04	0.42

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), as 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.

We carry out tests using the aggregate versions of each of the sentiment series (ie 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. Table 4 below shows results obtained testing Granger-causality from the various versions of the RTRS, BROKER and MCDAILY variables to MCI, VIX, BoEU, EPU, CDS and PMI. Table 5 below shows results obtained testing Granger-causality between the same variables in the reverse direction, from MCI, VIX, BoEU, EPU, CDS and PMI to the various versions of the RTRS, BROKER and MCDAILY variables.¹²

Table 4

¹² The missing entries in both tables could not be determined because of some form of VAR misspecification

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

RSS Series	MCI	VIX	BoEU	EPU	CDS	PMI
RTRS EXC-ANX	0.005**	0.28	4e-06**	0.3		0.0002**
RTRS EXC	0.032*	0.044*	0.0013**	0.03*	0.05*	0.05*
RTRS ANX	0.003**	0.56	7e-05**	0.1		0.0004**
MCDAILY EXC-ANX	0.5	0.09	6e-05**	0.05*	0.09	0.06
MCDAILY EXC	0.8	0.44	0.13	0.85	0.57	0.57
MCDAILY ANX	0.8	0.38	0.001**	0.06	0.12	0.33
BROKER EXC-ANX	2e-11**	0.18		0.92	0.6	0.1
BROKER EXC	0.022*	0.84		0.77	0.43	0.82
BROKER ANX	3e-05**	0.12		0.72	0.68	0.03*

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, and their
component parts

RSS Series	MCI	VIX	BoEU	EPU	CDS	PMI
RTRS EXC-ANX	0.29	0.093	0.022*		0.57	0.08
RTRS EXC	0.39	0.22	0.038*	0.62	0.24	0.24
RTRS ANX	0.03*	0.0013**	0.21		0.85	0.12
MCDAILY EXC-ANX	0.95	0.39	0.58	0.18	0.89	0.49
MCDAILY EXC	0.94	0.61	0.47	0.76	0.32	0.03*
MCDAILY ANX	0.92	0.52	0.69	0.08	0.98	0.95
BROKER EXC-ANX	0.94	0.16		0.73	0.97	0.41
BROKER EXC	0.22	0.084		0.72	0.001**	0.72
BROKER ANX	0.72	0.33		0.98	0.78	0.67

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, which is 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 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, as well as their use to gauge future financial market volatility, consumer confidence and various measures of uncertainty. The authors show in Nyman et al. (2014) how BROKER can be used to predict, out-of-sample, the change from the previous final estimate of the MCI to the current preliminary estimate. The adjusted R squared of the predictions when regressed on the actual changes is 0.486 compared to 0.114 for the consensus forecasts made by economists and published in Reuters. The forecasts are unbiased as the constant term is not significantly different from 0 and the coefficient on the predictions is not significantly different from 1. Figure 8 illustrates the difference in accuracy of the forecasts made using BROKER and the consensus forecasts.

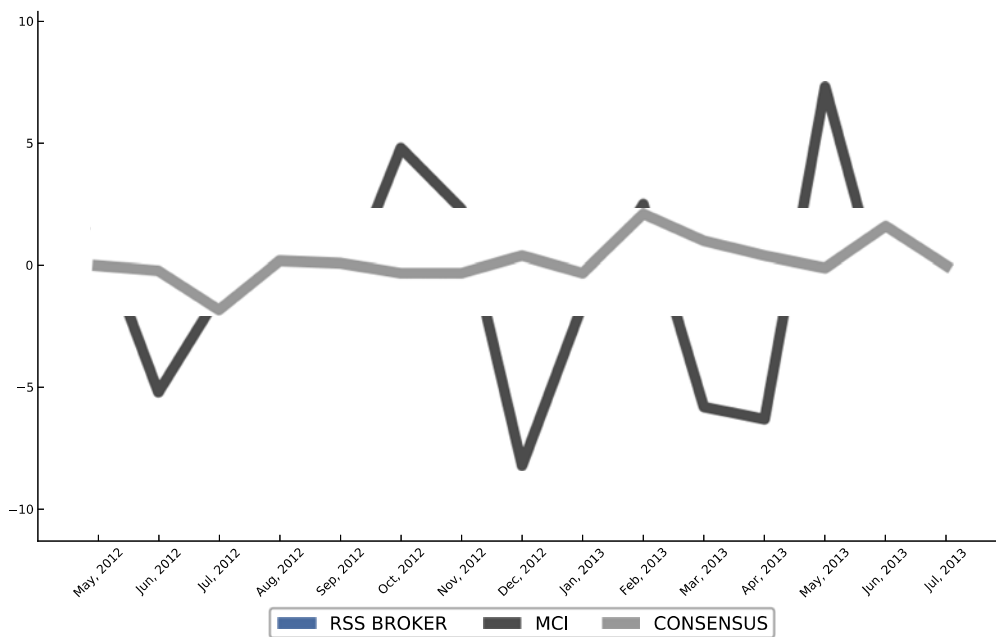


Figure 8: Change in MCI compared to forecasts of the change made using BROKER and consensus economist forecasts

3.4 Effect of relative sentiment on the UK economy

We further 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 is 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) and Haddow (2013). 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 use the model in Haddow et al. (2013) and replace their measure of uncertainty

by the UK RSS RTRS series (aggregated quarterly by averaging the months in a quarter). The model is specified as the following VAR(2):

$$\begin{bmatrix} rss_t \\ GDP_t \\ L_t \\ CPI_t \\ r_t \\ credit_t \end{bmatrix} = A_1 \begin{bmatrix} rss_{t-1} \\ GDP_{t-1} \\ L_{t-1} \\ CPI_{t-1} \\ r_{t-1} \\ credit_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} rss_{t-2} \\ GDP_{t-2} \\ L_{t-2} \\ CPI_{t-2} \\ r_{t-2} \\ credit_{t-2} \end{bmatrix} + \varepsilon_t,$$

where rss_t is the quarterly relative sentiment shift series for the UK, GDP_t is the quarterly level of GDP (in log deviations from a statistical trend), L_t is the quarterly level of employment in hours worked (in log deviations from a statistical trend), CPI_t is the seasonally adjusted level of the consumer price index (in log deviations from a statistical trend), r_t is the level of Bank Rate and $credit_t$ is an indicator of credit conditions. In each case, the statistical trend is estimated using a Hodrick-Prescott filter with smoothing parameter set to 1600. We use the same data as used to estimate the model in Haddow et al.¹³

We first replicate the model using the Haddow et al. uncertainty variable over the period 1989Q2 – 2012Q2 and confirm that a one standard deviation shock to the uncertainty index does have a significant and negative impact on activity. This remains true (although, only just) when we shorten the period to start in 1996Q1 (the quarter from which we have relative sentiment data for the UK).

When we replace the uncertainty series with the relative sentiment shift series, we find a qualitatively similar response of activity to a one standard deviation shock, but it is not statistically significant. Figure 9 shows the respective impulse response functions of GDP to shocks to uncertainty and RSS.

¹³ The $credit_t$ index pre-1995 is taken from Fernandez-Corugedo and Muellbauer (2006) and from 1995 onwards it is a weighted average of interest rates facing households for credit card loans, personal loans and mortgages

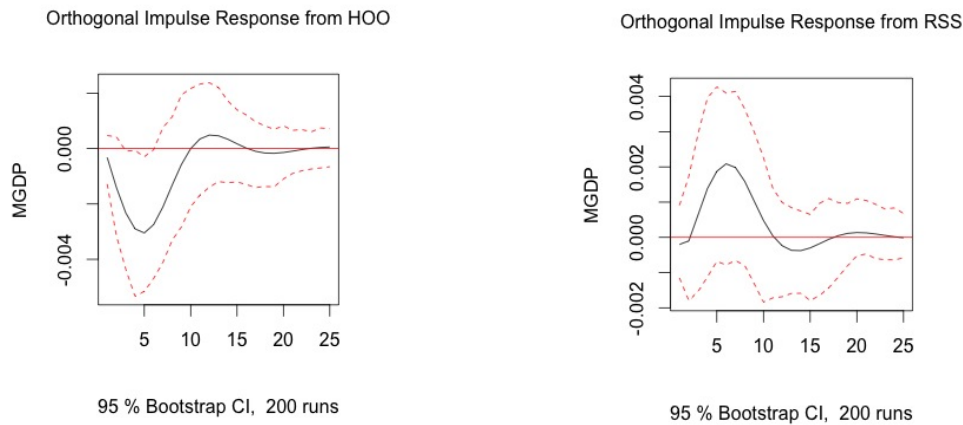


Figure 9: Impact of one standard deviation shocks on GDP from BoEU (left) and RTRS RSS (right)

Note: MGDG represents quarterly level of GDP, as log deviations from a statistical trend.

However, the Granger causality results described above suggest that causality flows from RSS to the Haddow et al. uncertainty variable. We confirm this in the context of the VAR model set out above by extending it including both RSS and the Haddow et al. uncertainty variable in the model. We observe the impact of each variable on the other in the VAR model. We find a significant impact of RSS on uncertainty that persists for about 5-6 quarters but no significant impact in the reverse direction. This result remains true regardless of the order of the two variables during model estimation. The respective impulse response functions can be seen in Figure 10.

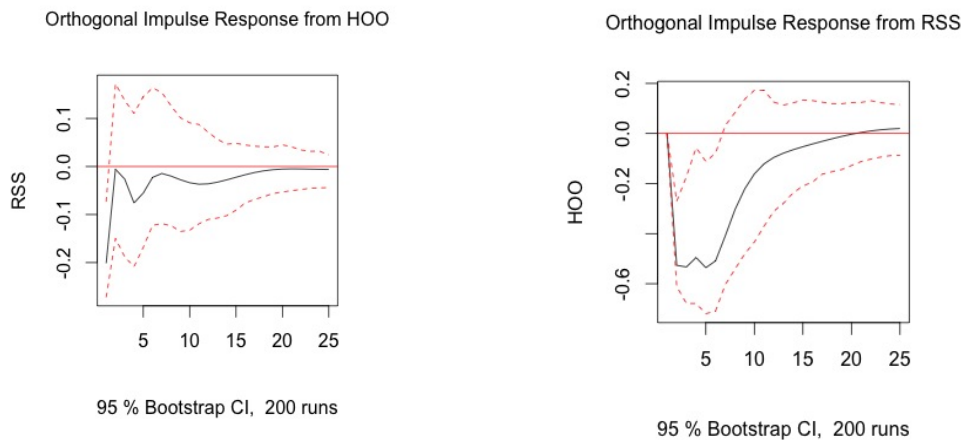


Figure 10: Impact of one standard deviation shocks of uncertainty on RSS (left) and vice versa (right)

Note: HOO represents the uncertainty variable of Haddow et al and RSS the relative sentiment series.

This suggests that it is in fact relative sentiment that drives the *perception* of uncertainty rather than any true ‘probabilistic’ uncertainty, ultimately slowing down the economy. This motivates further research into the effects of narratives on the economy and financial stability.

We check the robustness of the above statements to the ordering of the variables in the VAR (placing the variables of interest both first and last in the equation) and to trending (checking the results both with and without de-trending the relevant variables using the Hodrick-Prescott filter). The respective impulse response functions can be found in the supplementary material.

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. 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 4).

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.¹⁴ Having decomposed the narrative discourse into one index measuring shifts in emotion (the previous section) and

¹⁴ We also investigated the robustness of the narrative consensus metric, a process which involves some technical analysis essentially based upon the degree to which documents at a point in time are similar. We describe this in the Appendix.

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.

Table 6 shows the outcome of Wald tests of Granger causality between the entropy series and the VIX, the Bank of England Uncertainty Index (BoEU), the EPU, CDS and PMI. We again use the procedure described in section 3. The p-values of Granger-causality in the converse direction, in the direction of the entropy series from the other three variables, can be seen in Table 7. The only significant directions of Granger-causality are from the entropy variable to the BoEU index and from the entropy variable to EPU.

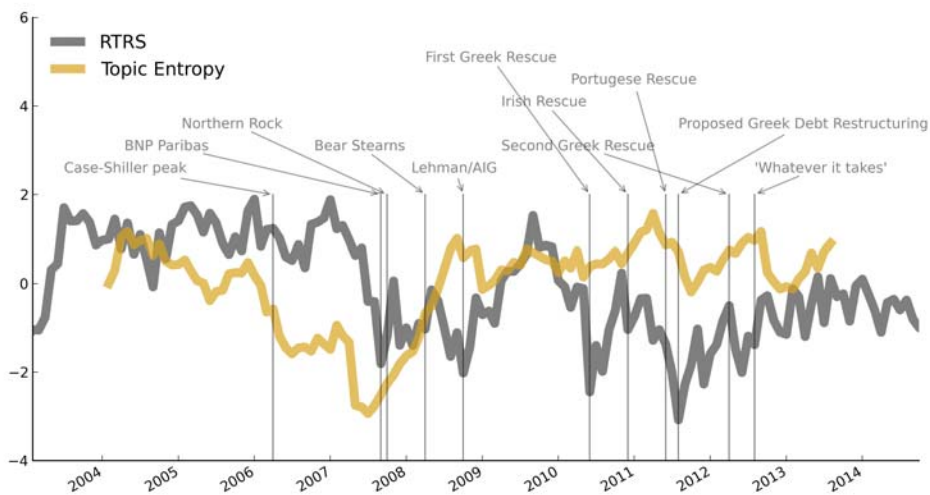


Figure 11: Relative sentiment (black) and entropy (yellow) in Reuters' London news. The y-axis displays the normalized values with 0 mean and standard deviation 1

Table 6

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

RSS Series	VIX	BoEU	EPU	CDS	PMI
Entropy	0.12	9.1e-06**	0.03*	1.0	0.65
Note:	*p<0.05; **p<0.01				

Table 7

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

RSS Series	VIX	BoEU	EPU	CDS	PMI
Entropy	0.27	0.98	0.52	0.43	0.81
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 ‘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 et al, 2015).

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.¹⁵ 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, the Reserve Bank of New Zealand and the Norges Bank.¹⁶ In the United Kingdom, the Financial Policy Committee

¹⁵ 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).

¹⁶ 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.

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).¹⁷

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.

¹⁷ See also Giese et al. (2014).

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

6. Conclusion

In this paper, we have explored the potential of using algorithmic text analysis, applied through the lens of conviction narrative theory, to extract quantitative summary statistics from novel data sources, which have largely only been used qualitatively thus far. We have demonstrated that the outcome of such procedures can lead to some intuitive and useful representations of financial market sentiment. The shifts correlate well with financial market events and 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

The authors would like to thank Jon Rand at the Bank of England [and others] for supporting the project and providing access to data. The authors would also like to thank Chrystia Freeland, Richard Brown and Maciej Pomalecki of Thomson Reuters for arranging access to the Reuters News archive.

David Tuckett's involvement in this work was made possible through Institute of New Economic Thinking Grant no IN01100025 and a grant from the Eric Simenhauer Trust of The Institute of Psychoanalysis.

Thanks are also due to Oliver Burrows, Andy Haldane, Philip Treleaven and Kimberly Chong for advice and support as well as to colleagues in the PhD Financial Computing Centre, the Psychoanalysis Unit at UCL and to the Anna Freud Centre. We have also benefitted greatly from discussion at the conferences and other meetings at which the work has been presented.

Appendix

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

2. Granger causality procedure

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, and the various packages used to carry out the above Toda-Yamamoto procedure are documented in the Supplement. The details of the specific results obtained, and a description of the various R packages, using the test procedure are available on request in the form of a Supplementary Material document.

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 penalised 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).

Definition: 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.

Definition: 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 A1) and smooth (using double exponential smoothing, with $\alpha = \beta = 0.3$) and average the results across parameter runs to arrive at the yellow line in Figure 9.

Table A2: Consensus parameters

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.

3.2 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.

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.
- Bai J., Perron P. (2003), Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, **18**, 1-22.
- 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. (2013). Measuring Economic Policy Uncertainty. Chicago Booth Research Paper No 13-02. Available at SSRN: <http://ssrn.com/abstract=2198490>
- 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.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, 59, 617-645.
- 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.

- 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.
- Choi, H., and Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record Special Issue: Selected Papers from the 40th Australian Conference of Economists. Volume 88, Issue Supplement s1 pp. 2-9*
DOI: 10.1111/j.1475-4932.2012.00809.x
- Chong, K. and Tuckett, D. (2014) Constructing Conviction through Action and Narrative: How Money Managers Manage Uncertainty and the Consequences for Financial Market Functioning. *Socio-Economic Review*. 1-26.
doi:10.1093/ser/mwu020
- Damasio A, Carvalho G.B. (2013) The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews. Neuroscience* [2013, 14(2):143-152]
- 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.
- 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.
- 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.
- 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.
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science* 343, no. 14 March: 1203-1205. Copy at <http://j.mp/1ii4ETo>
- 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.
- 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.
- OFR annual report (2012) <https://financialresearch.gov/annual-reports/files/office-of-financial-research-annual-report-2012.pdf>
- Ormerod, P., Nyman, R., Bentley, R.A, (2014). *Nowcasting economic and social data: when and why search engine data fails, an illustration using Google Flu*

Trends. arXiv:1408.0699 [physics.soc-ph]

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.

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.

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

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

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.

Suddendorf, T. & Corballis, M. C. 2007 The evolution of foresight: what is mental time travel and is it unique to humans? Behav. Brain Sci. 30, 299–313. (doi:10.1017/S0140525X07001975)

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., (2016) The Role of Conviction in Decision-Making Under Radical Uncertainty. Presented to the International Conference on Thinking, Brown University Rhode Island, August 4, 2016. Available at: https://www.researchgate.net/publication/307905267_The_Role_of_Conviction_and_Narrative_in_Decision_Making_under_Radical_Uncertainty

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., Ormerod, P., Nyman, R., and Smith R.E. (2015) Information And Economics: A New Way To Think About Expectations And To Improve Economic Prediction. <http://206.191.128.204/ideas-papers/research-papers/information-and-economics-a-new-way-to-think-about-expectations-and-to-improve-economic-prediction> Paper presented at the OECD/INET conference, Paris, April.