Preliminary draft. Not for quotation or further circulation

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

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

June 2015

Abstract:
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

Preliminary draft. Not for quotation or further circulation

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 the social-psychological theory of "conviction narratives". This 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, 2015). It has been successfully used in other applications, for example as a measure of changing macroeconomic confidence or "animal spirits (Tuckett et al. 2015).

The theory 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 views 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 suppor 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, 2015). 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.

The three unstructured text-based data sources we analyse are: internal Bank of England daily commentary on market news and events; broker research reports; and Reuters' news articles in the United Kingdom. We make use of the conviction narrative methodology to 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.

At any given moment, there will be several narratives circulating among all financial agents. Some of these narratives, or pieces of them, are likely to be contained within the documents we analyse and reflect significant and meaningful shifts representing a dominance of one emotion over the other (Tuckett and Nikolic, 2015). The RSS measure we apply is aimed 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. The true power of this method lies in its top down approach, capturing aggregate shifts largely undetectable to the human eye. In particular, if the relative shifts in the emotional content correlate across several data sources, it would be reasonable to assume that at least some financial agents had adopted a subset of the narratives and held them as true. At the same time, it is critical 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. It is clear that this depends on the intention of the content creators.

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 become homogenous “to the detriment of the diversity that is indispensable for the smooth functioning of neoclassical markets” (Trichet, 2001). 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, or groupfeel, 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; Hansen et al, 2015) 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 on a limited range of sentiments departs from the above literature. 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 et al 2014). 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, in contrast to other studies, 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 such as from Twitter and/or Facebook. By contrast, we focus in this contribution on data sources more specifically connected to the financial system.

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. Two lists of previously applied and experimentally validated (Strauss, 2013) words, 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	Anxiety	Excitement	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.[2] 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'), nor do we control for their typical (expected) frequencies in ordinary usage. But this does not affect the nature of our key findings. This point is discussed in the Appendix.

The simplicity of our method is intentional for two main reasons. First, as we are not strictly interested in the exact relationship between the objects of a narrative and the emotions within it, we do not necessarily want to apply more sophisticated natural language processing techniques to establish these relationships. 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.[3] This technique gives

[2] 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.

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

us increased confidence that the meaning of individual words in our two lists does not change over time.

3. Results

3.1 The temporal 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.[4] 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.

It is important to note that the underlying theory of decision making which directs the construction of these series (conviction narrative theory) essentially refers to the *relative* level of sentiment, in other words (as defined above) excitement minus anxiety. However, we report the two component parts separately out of interest. 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.

[4] In particular, unlike event studies, we do not try to infer anything causal from the events that we depict on the charts.

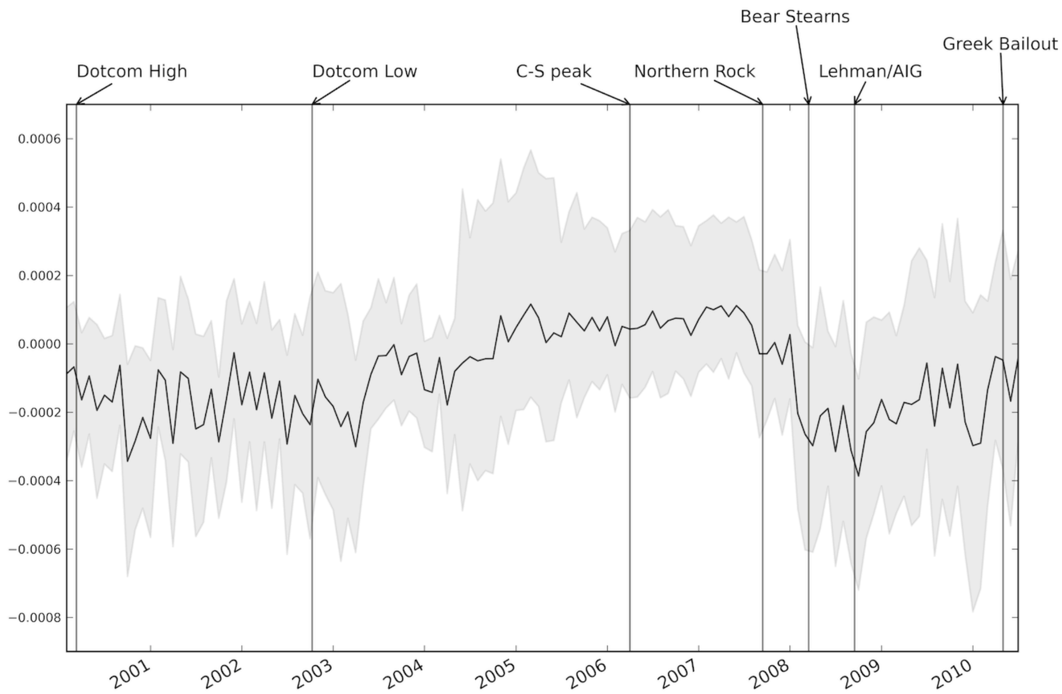


Figure 1: Relative sentiment of MCDAILY with 95% confidence bands generated by bootstrap

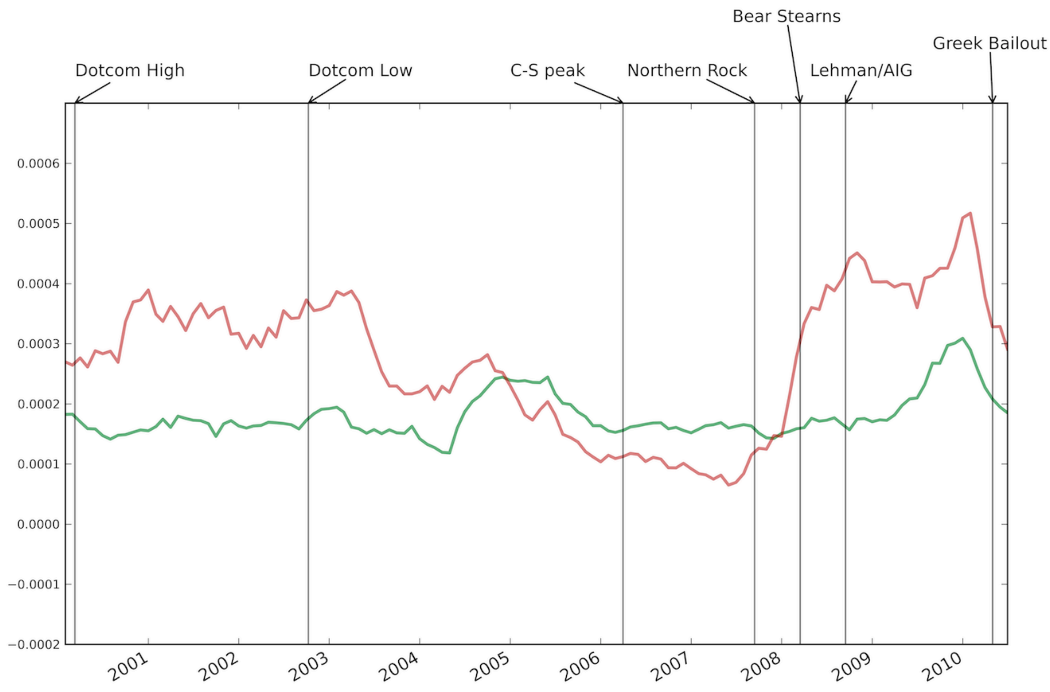


Figure 2: Emotional factors of MCDAILY; anxiety (red) and excitement (green)

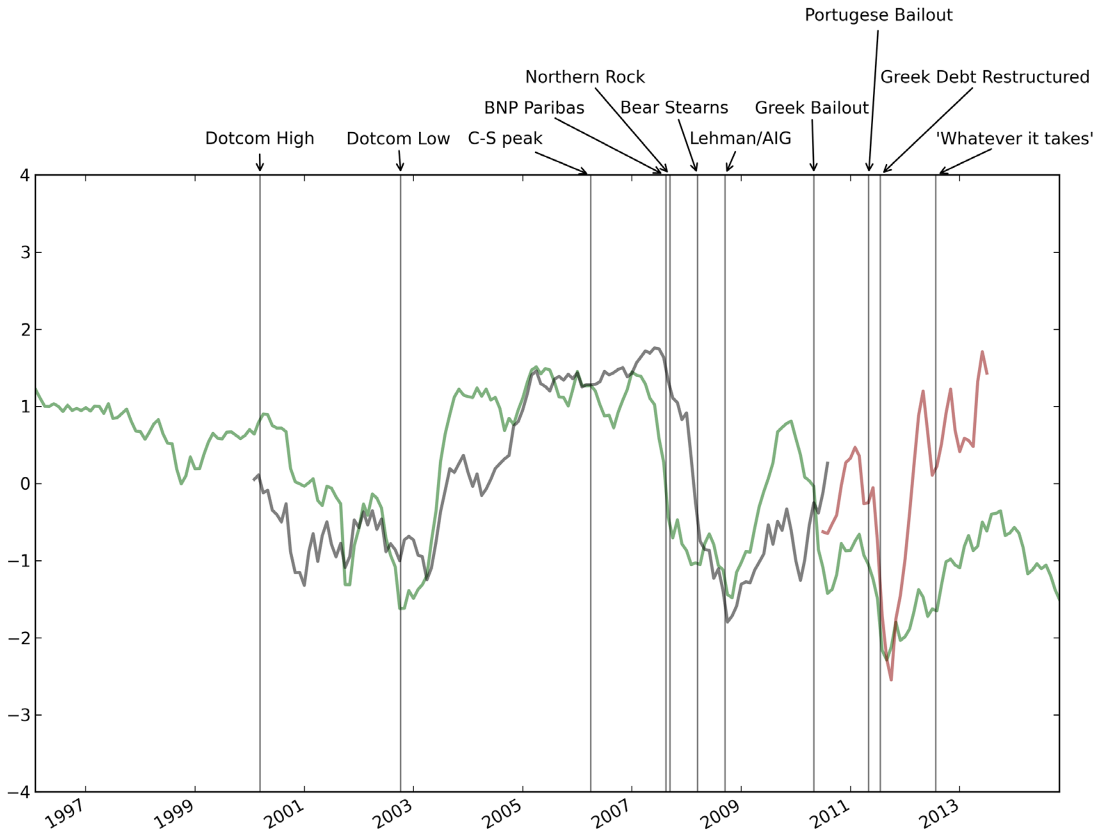


Figure 3: Relative sentiment of MCDAILY (black), RTRS (green) and BROKER (red)

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 clearly shows that the series tend to move together.

Importantly, both the MCDAILY and RTRS show sharp falls well in advance of the financial crisis (the third, BROKER, only became available during 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 – 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.

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.

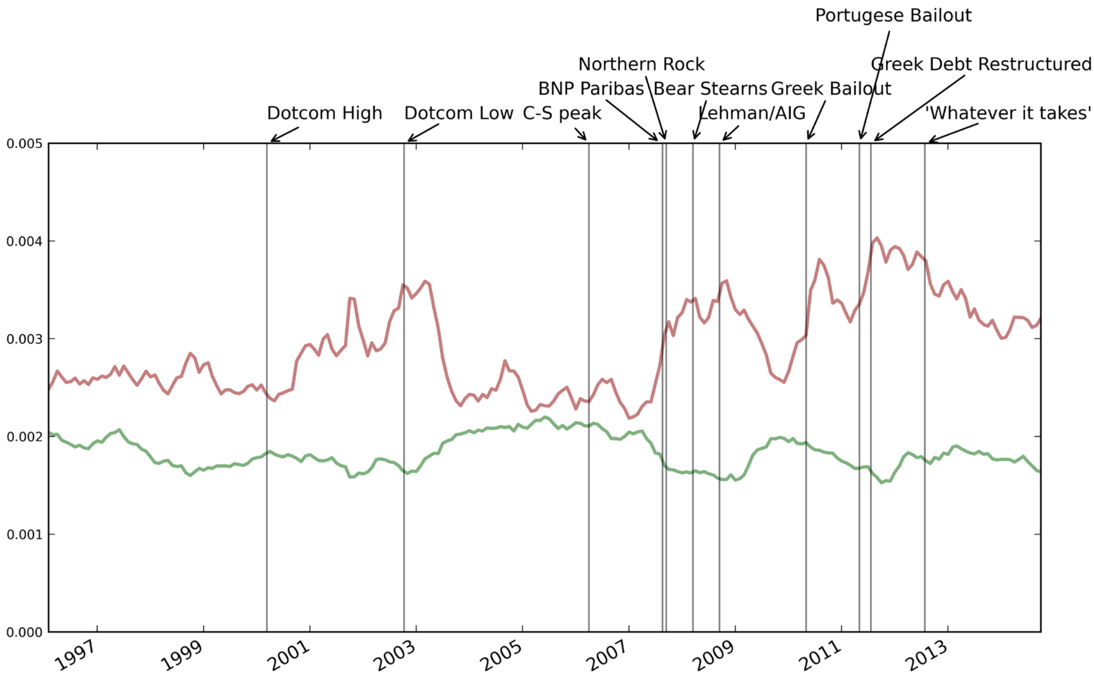


Figure 4: Excitement (green) and Anxiety (red) in RTRS

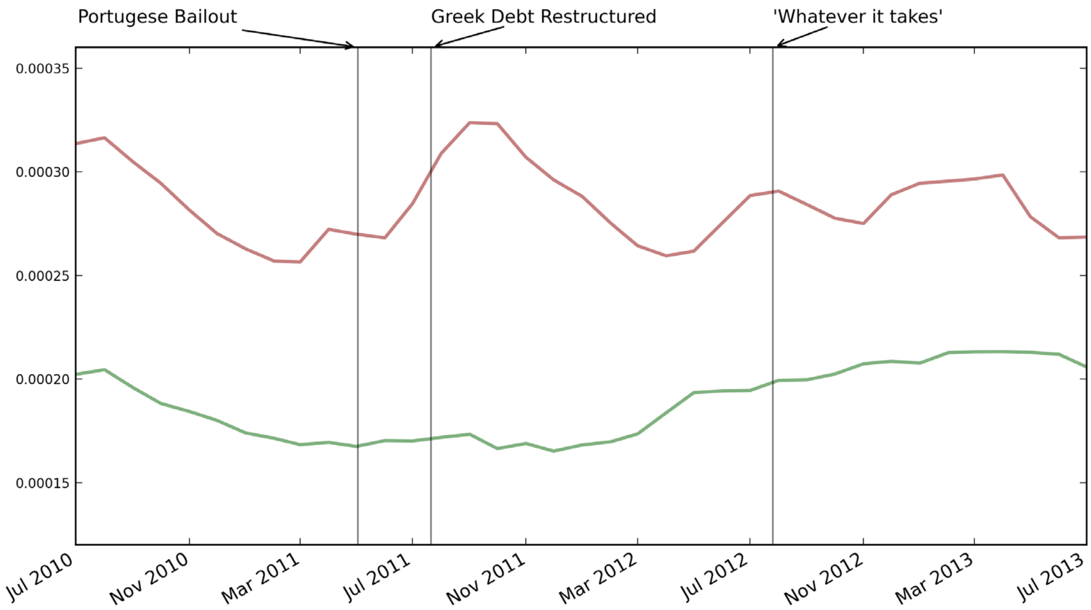


Figure 5: Excitement (green) and Anxiety (red) in BROKER

3.2 Comparison with other measures

To explore how the statistic compares to a range of other financial and economic indicators at a higher frequency, we conducted a simple correlation based pairwise comparison study. The correlations with the Michigan Consumer Sentiment index[5] (MCI), the VIX[6], the economic policy uncertainty index of Baker et al. (2013) (EPU) and the Bank of England macroeconomic uncertainty index (BoEU – see Bank of England (2013)), together with the correlations between the individual relative sentiment series are presented in Table 2[7].

[5] 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.

[6] 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'.

[7] 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.

Table 2: Correlations between relative sentiment series and common measures of sentiment

	MCD	RTRS	BRO	VIX	MCI	EPU	BoEU
MCD	1	0.59	-	-0.62	0.26	-0.59	-0.54
RTRS	-	1	0.71	-0.37	0.54	-0.68	-0.52
bRO	-	-	1	-0.60	0.66	-0.64	-0.60

Figure 6 plots RTRS, MCI, VIX, EPU and BoEU, and Figure 7 shows MCDAILY and the VIX.

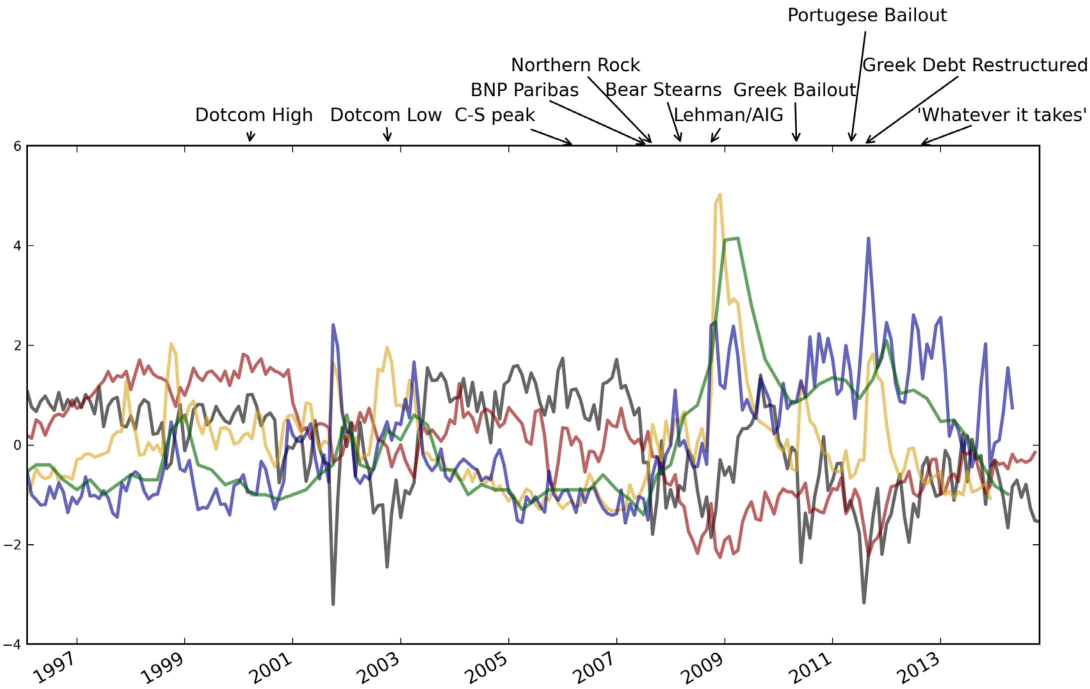


Figure 6: Relative sentiment of RTRS (black) compared to the MCI (red), the VIX (yellow), the EPU (blue) and the BoEU (green)

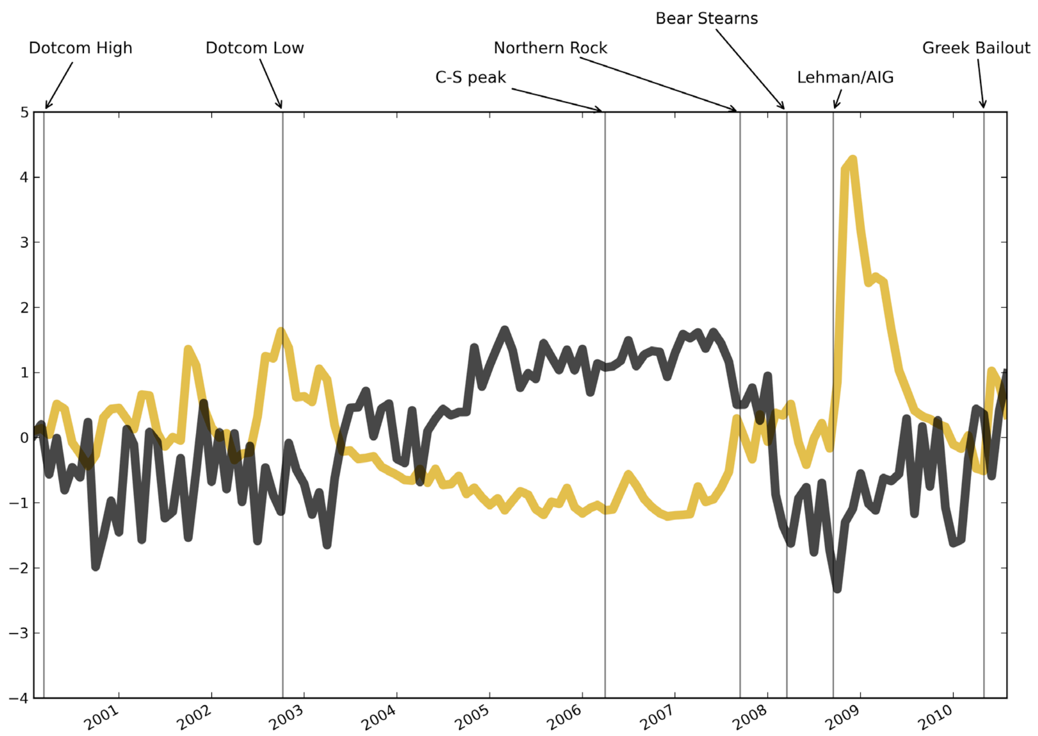


Figure 7: Relative sentiment of MCDAILY (black) compared to the VIX (yellow)

In this section, we report the results of Granger causality tests between the three sentiment series and the various economic indicators plotted above.

We use the methodology described in Toda and Yamamoto (1996). This process involves six distinct steps, carried out in sequence. We describe these in the Appendix. In the main text, we simply report the final step in the process, which is the step which provides the evidence on the existence or otherwise of Granger causality. However, the five previous steps all require various statistical tests to be performed prior to the final test step being taken.

Initially, we carry out tests using the aggregate versions of each of the sentiment series, where by the phrase 'aggregate version' we mean the net balance between excitement and anxiety. We go on to report tests for the component parts of each series, namely excitement and anxiety separately. Table 3 below shows results obtained using the aggregate version of the RTRS variable.

Table 3

Wald test statistics of Granger-causality between the relative sentiment shift series RTRS and the economic indicators MCI, VIX, and BoEU. Aggregate version of the sentiment series i.e. excitement minus anxiety

Direction	Chi-Sq	d.f.	p-value
RTRS -> MCI	12.7	3	0.005***
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	4e-06***
BoEU -> RTRS	7.7	2	0.022**

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

For the aggregate series, it was not possible to obtain meaningful results between RTRS and EPU using the procedure described above[8]. Tables 4 and 5 show the results for the component parts of RTRS, excitement and anxiety respectively.

Table 4

Wald test statistics of Granger-causality between the excitement component of the relative sentiment shift series RTRS and the economic indicators MCI, VIX, EPU and BoEU.

Direction	Chi-Sq	d.f.	p-value
RTRS -> MCI	8.8	3	0.032**
MCI -> RTRS	3.0	3	0.39
RTRS -> VIX	9.8	4	0.044**

[8] Technically, The VAR equations show high levels of serial correlation of the residuals up to and including 16 lags.

VIX -> RTRS	5.7	4	0.22
RTRS -> EPU	17.5	8	0.026**
EPU -> RTRS	22.3	8	0.004***
RTRS -> BoEU	13.3	2	0.0013***
BoEU -> RTRS	6.5	2	0.038**

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

Table 5

Wald test statistics of Granger-causality between the anxiety component of the relative sentiment shift series RTRS and the economic indicators MCI, VIX, EPU and BoEU.

Direction	Chi-Sq	d.f.	p-value
RTRS -> MCI	15.8	4	0.003***
MCI -> RTRS	10.7	4	0.03**
RTRS -> VIX	1.2	2	0.56
VIX -> RTRS	8.7	2	0.0013***
RTRS -> BoEU	19.0	2	7e-05***
BoEU -> RTRS	3.1	2	0.21

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

There is significant causality from the sentiment series to the MCI in both the relative sentiment variable RTRS, and in its two component parts, namely excitement and anxiety. There is evidence of two way causality between the RTRS and the VIX, although perhaps surprisingly it is significant in the direction from RTRS to the VIX in the case of the excitement series but not the anxiety one. For the anxiety component of BROKER, it was not possible to obtain meaningful results between BROKER and EPU using the procedure described above. The VAR equations show high levels of serial correlation of the residuals up to and including 16 lags.

For both the EPU and the BoEU, again there is some evidence of two-way causality, but this is mainly, and is more pronounced, from sentiment to the uncertainty indices rather than in the reverse direction.

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 and consumer confidence. The analysis of RTRS, the EPU and BoEU indices suggests the methodology applied in this paper can help us understand recent measures of uncertainty.

Tables 6, 7 and 8 show results of the Granger causality test procedure with the BROKER series. Given the short length of BROKER, it was not possible to analyse in any meaningful way the BoEU series in conjunction with it, because the latter is on a quarterly basis.

Table 6

Wald test statistics of Granger-causality between the BROKER series and the economic indicators MCI, VIX, and EPU. Aggregate version of the sentiment series i.e. excitement minus anxiety

Direction	Chi-Sq	d.f.	p-value
BROKER -> MCI	44.9	1	2e-11***
MCI -> BROKER	0.005	1	0.94
BROKER -> VIX	3.4	2	0.18
VIX -> BROKER	3.7	2	0.16
BROKER -> EPU	18.9	7	0.008***
EPU -> BROKER	12.2	7	0.093*

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

Table 7

Wald test statistics of Granger-causality between the excitement component of the BROKER series and the economic indicators MCI and VIX .

Direction	Chi-Sq	d.f.	p-value
BROKER -> MCI	5.2	1	0.022**
MCI -> BROKER	1.5	1	0.22
BROKER -> VIX	0.85	3	0.84
VIX -> BROKER	6.6	3	0.084*

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

Again, there were persistent problems of residual autocorrelation in the VAR models using the EPU, so we do not report results for this variable.

Table 8

Wald test statistics of Granger-causality between the anxiety component of the BROKER series and the economic indicators MCI, VIX and EPU.

Direction	Chi-Sq	d.f.	p-value
BROKER -> MCI	20.6	2	3e-05***
MCI -> BROKER	0.7	2	0.72
BROKER -> VIX	2.4	1	0.12
VIX -> BROKER	0.95	1	0.33
BROKER -> EPU	4.8	2	0.090*
EPU -> BROKER	2.6	2	0.27

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

The results obtained using the BROKER series are qualitatively similar, but perhaps not as well determined, as those obtained with RTRS.

Somewhat surprisingly, carrying out the analysis using the MCDAILY series failed in general to find evidence of significant Granger causality. The exceptions were, first, the aggregate series and the VIX, where the Wald test of causality from MCDAILY to the VIX is significant at a p-value of 0.092. Second, MCDAILY and BoEU, where there is strong evidence of causality from both the anxiety component of MCDAILY and the aggregate MCDAILY to the BoEU. These results are reported in the Supplementary Material.

3.3 A further illustrative result

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 8) 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, 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 the emergence of a bubble in any specific sector of the economy. This particular example seems to indicate that the property sector became overly exuberant prior to the crisis.

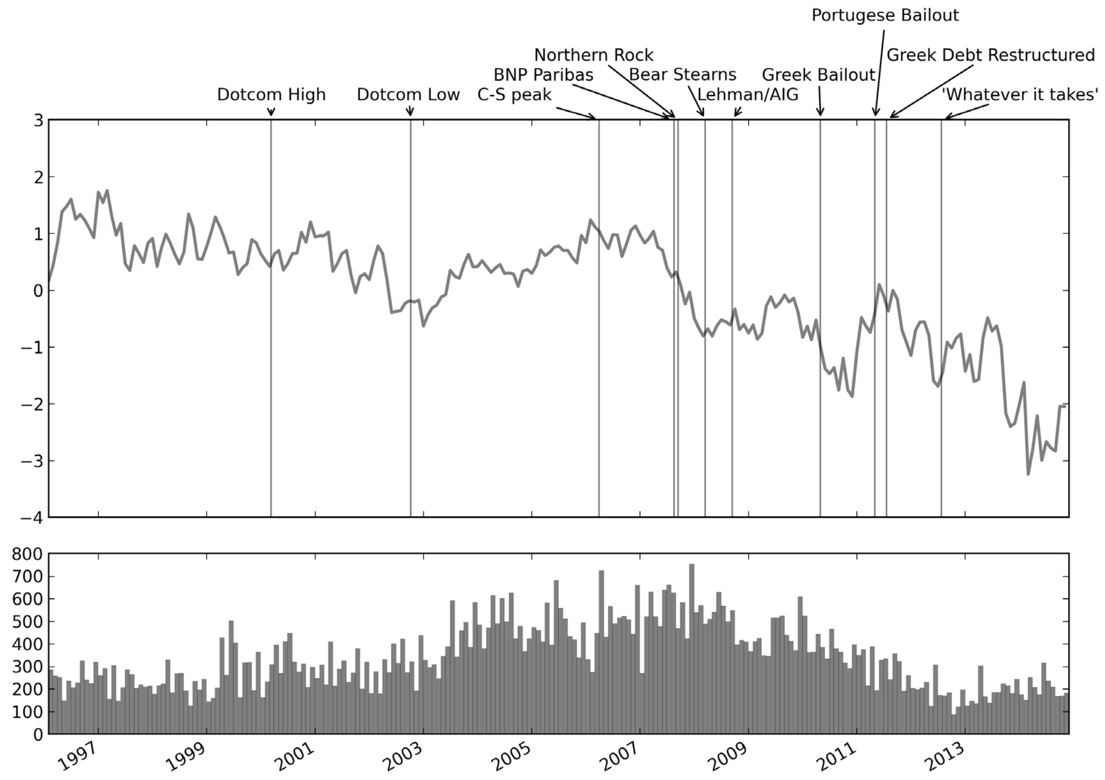
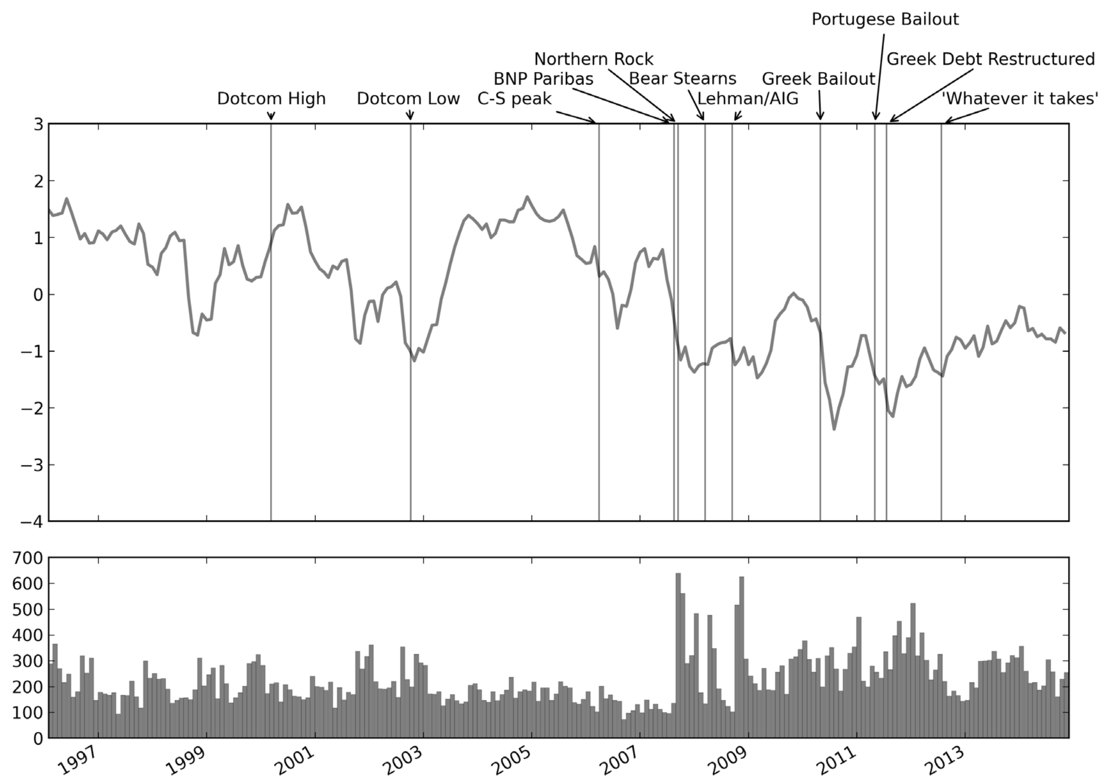


Figure 8: Relative sentiment surrounding 'property' in RTRS



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 is of larger volume 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.

4.2 Results

We plot the narrative consensus found in RTRS in Figure 9. The graph shows a clear increase in consensus (decrease in entropy) preceding the crisis period and much more disagreement subsequently.[9] Having decomposed the narrative discourse into one index measuring shifts in emotion (the previous section) and another measuring structural changes in consensus (entropy) and, 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 9 shows the outcome of Wald tests of Granger causality between the entropy series and the Bank of England Uncertainty Index (BoEU). Again, we use the procedure described in section 3 above, and report step 6 here. Full details of steps 1-5 are in the Supplementary Material.

[9] 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.

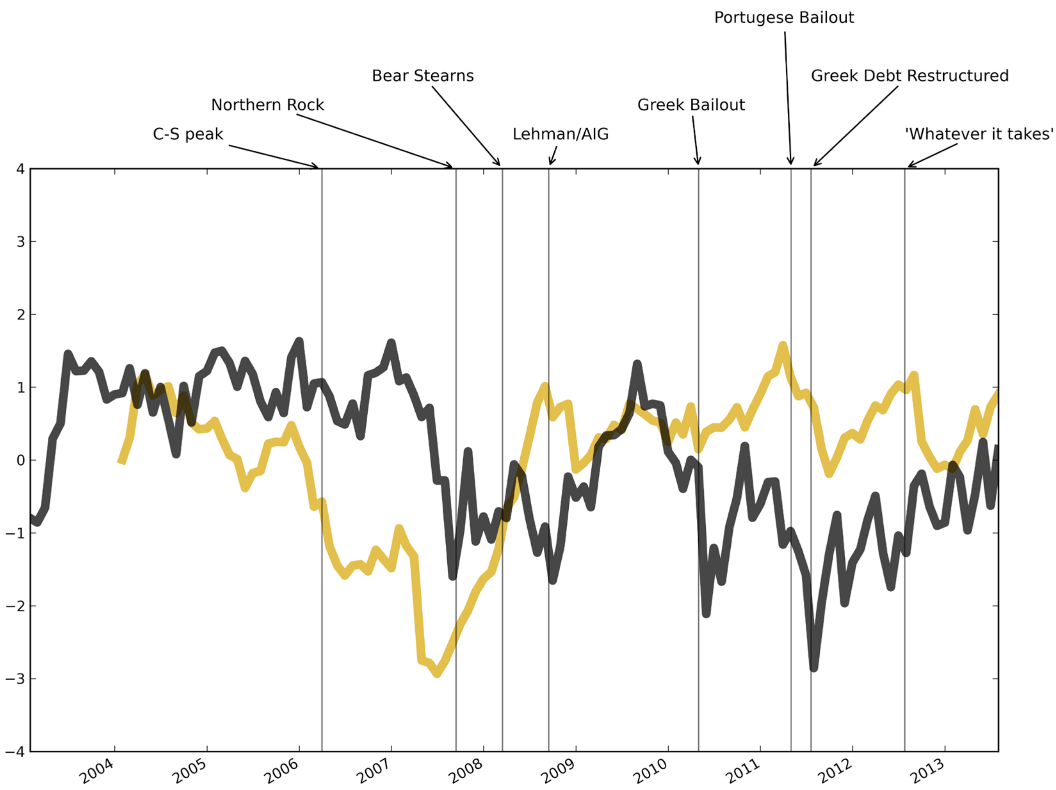


Figure 9: Relative sentiment (black) and entropy (yellow) in Reuters' London news

Table 9

Wald test statistics of Granger-causality between the entropy series shown in Figure 9 and the BoEU series.

Direction	Chi-Sq	d.f.	p-value
ENTROPY -> BoEU	33.3	6	9e-06***
BoEU-> ENTROPY	1.0	6	0.98

Note: *p<0.1; **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.[10] 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.[11] In the United Kingdom, the Financial Policy Committee routinely reviews a set of core indicators which have been helpful in identifying

[10] 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).

[11] 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.

emerging risks to financial stability in the past, and which therefore might be useful in detecting emerging risks (Bank of England, 2014).[12]

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 correlated, indeed to be leading indicators 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. In particular, the basis of these measures in social-psychological theory rather than realised economic or financial variables can provide a different perspective on systemic risk. These measures could also be calculated on a real-time basis, offering them an important advantage over 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.

---

[12] 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 emotional shifts correlate well with financial market events and appear to drive 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.

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.

References

Aikman, D., Haldane, A. G., Nelson, B. D. (2011). Curbing the credit cycle. DOI: 10.1111/ecoj.12113

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' available at http://www.bankofengland.co.uk/financialstability/Documents/fpc/policystatement140113.pdf

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.

Brown, G. W., Cliff, M. T., (2005). Investor Sentiment and Asset Valuation. The Journal of Business, 78(2): 405-440.

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

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, C Borio , K Tsatsaronis (2011). Anchoring countercyclical capital buffers: The role of credit aggregates. International Journal of Central Banking, 7.

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, Kapadia S (2014), 'The credit-to-GDP-gap and complementary indicators for macroprudenital policy: evidence from the UK', International Journal of Finance and Economics, Vol 19, pp. 25-47

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.

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

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

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

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 Leaning. Pp. 727-734.

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

Reinhart, C. M., and Kenneth S. Rogoff. 2009. "The Aftermath of Financial Crises." American Economic Review, 99(2): 466-72. DOI**:** 10.1257/aer.99.2.466

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

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.

Trichet, J. C. (2001). Preserving financial stability in an increasingly globalised world. Keynote speech by Mr Jean-Claude Trichet, Governor of the Banque de France, at the European Financial Markets Convention (EFMC 2001), Paris, 15 June 2001.

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

Tuckett, D.A and Nikolic, M., (2015) The Role of Conviction in Decision-Making Under Deep Uncertainty. DOI: 10.13140/RG.2.1.3901.3281

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.