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

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

February 2014

Preliminary draft. Not for quotation or further circulation

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 the emotional content in market narratives, as it develops over time, is highly correlated across data sources and clearly shows the formation (and subsequent collapse) of very high levels of sentiment – high excitement relative to anxiety – leading up to the global financial crisis. We also show that a new methodology that attempts to capture the emergence of consensus gives an intuitive representation of the period of stability and beliefs in a new paradigm prior to the crisis. With increasing consensus in narratives high in excitement and lacking anxiety likely to be an important warning sign of impending distress, the quantitative metrics we develop may complement other indicators and analysis in helping to gauge systemic risk and forecast developments in the financial system.

[1] Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Bank of England or to state Bank of England policy. This paper should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee or 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 such shifts in sentiment along with the extent of consensus in the market. We approach the data through the lens of the psychological theory of conviction narratives, which emphasises the role of narratives and emotional sentiment in driving decision-making under uncertainty (Chong and Tuckett, 2013). This motivates how the metrics we develop could potentially be used for systemic risk assessment and forecasting developments in the financial system.

We analyse three data sources: internal Bank of England daily commentary on market news and events; broker research reports; and Reuters news articles in the United Kingdom. The relative sentiment indices which 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. From mid 2007, a surge in anxiety drove rapid falls in sentiment which continued until soon after the collapse of Lehman Brothers. And there were further falls in sentiment prior to the height of the

Euro area sovereign crisis in 2011-2. At a higher frequency, we find that the sentiment series extracted from broker reports is highly correlated with the Michigan Consumer Sentiment index with approximately a 3-week lead, and that the sentiment series extracted from the Bank of England market commentary is somewhat negatively correlated with the VIX, an uncertainty measure based not on a survey-based assessment of market mood but on actual trading.

We also find, using a newly developed measure of narrative dispersal, 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. The theory predicts that such a situation is unsustainable, as it transpired to be, and as such the quantitative metrics we have developed could be useful alongside more traditional indicators and analysis (see, for example, Drehmann et al, 2011, Bank of England, 2014 or Giese et al, 2014) for gauging systemic risk.

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 the assessing trends in the economy (e.g., Varian). Our approach is related to studies which have attempted to quantify sentiment (or "animal spirits") via the use of market proxies or events, and investigate the effect of such sentiment on financial markets (e.g., Baker and Stein 2004; Baker and Wurgler 2006, 2007; Baker et al., 2012; Brown and Cliff,

2005; Edmans et al. 2007; Greenwood and Nagel 2009). More recently, studies have attempted to quantify sentiment using text data sources such as corporate reports and news media to try to predict asset prices (e.g., Tetlock 2007; Tetlock et al. 2008; Tetlock 2011; Soo 2013). There is also a growing literature on models of opinion dynamics, and how such models can add power in predictive settings (e.g., Lux 2008). Finally, another branch of relevant research has explored the effects of central bank communication on market variables such as private inflation forecasts (see for example Blinder et al., 2008, De Haan, 2008, Hubert 2012).

Our approach departs from most of this literature in two important ways. First, our primary focus is on gauging systemic risk which could precipitate financial crises, rather than on movements in particular asset prices. Second, we develop our measures of sentiment from the point of view of an established theory of psychology. We stress the importance of applying a theoretical filter when processing 'big data' in this way to avoid mining for seemingly significant correlations.

The specific theory we use is that of conviction narratives (for more details see Chong and Tuckett, 2013). In the context of radical, or Knightian, uncertainty, agents gain the conviction to act by creating a convincing enough story or argument that creates excitement about potential gains from that action while suppressing doubt and anxiety regarding potential losses. It is clear that in order to take any action, the potential gains must outweigh the potential losses. Thus, in the simplest case, the key variable of interest to us is the aggregate relative difference between *excitement* and *anxiety* and *shifts* in this difference over time.

This theory has been successfully used in other applications, for example as a measure of macroeconomic confidence (Tuckett, Ormerod, Smith and Nyman 2014) and to measure of shifts in emotion within particular conviction narratives (Tuckett, Smith and Nyman 2014). For a thorough exposition of the methods and their grounding in psychology, we refer the reader to these papers.

In the first part of our paper, we make use of the conviction narrative methodology to capture an emotional summary statistic of sentiment in the financial system based on several different financial text data sources, and explore changes in this statistic over time to assess how it convincingly and robustly it measures shifts in confidence and how it correlates with financial market events. It has been argued that shifts in (as well as the volatility of) the emotional content of narratives can, at times, indicate changes in macroeconomic confidence, or "animal spirits", and be a driving force of financial and economic investment (Tuckett, Ormerod, Smith and Nyman 2014). Therefore, we have reason to believe that the summary statistic we develop could be used to gauge systemic risk in financial systems.

Financial behaviour can also often become homogenous "to the detriment of the diversity that is indispensable for the smooth functioning of neoclassical markets" (Trichet, 2003). Therefore, in the second part of the paper, which is more exploratory in nature, we ask whether we can measure *structural* changes in the distribution of narratives. Specifically, we develop a methodology to measure 'consensus' or (alternatively) '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 became necessary for financial institutions to attract new investment (Haldane, 2011). But such consensus in an environment of high sentiment could be suggestive of over-confidence or irrational exuberance – the ability to measure the emergence of consensus or disagreement within text documents could therefore prove useful in identifying financial system risks.

The rest of the paper is structured as follows. Section 2 explains the data we have analysed. Section 3 focuses on our measure of emotion, or sentiment, and explains the methods in 3.1 and the findings in 3.2. Section 4 focuses on the measure of 'consensus', and explains the methods in 4.1 and the results in 4.2. Finally, section 5 concludes.

2. Data

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

2.1 Broker reports

Broker research reports provide a large source of documents of clear relevance to financial markets. We analyse an archive of 14 brokers from January 2008 through June 2013 consisting of documents of a primarily global economic focus. The archive consists of approximately 104 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. Throughout the rest of the paper we refer to this database as BROKER.

2.2 Bank of England internal market commentary

The Market 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. 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. In fact, 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.

We analyse on average 26 documents per month from January 2000 through July 2010. For the rest of the paper, we refer to these documents as MCDAILY ('Market Commentary Daily').

2.3 Reuters News Archive

Finally, we use the Reuters News archive, as also extensively studied by Tuckett, Ormerod, Smith and Nyman (2013) to assess macroeconomic trends. At the time of our analysis, the archive consisted of over 14 million English news articles. For most of this paper, we restrict our attention to news published by Reuters in London during the period January 2003 through July 2013, in which 6123 articles were published on average each month (after we exclude all articles tagged by Reuters as 'Sport', 'Weather' and/or 'Human Interest'). For the rest of the paper, we refer to this database as RTRS.

3. Relative Sentiment Shifts

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 some of the documents we analyse. The emotional content of these narratives can at times show significant and meaningful shifts representing a dominance of one emotion over the other [add reference]. The measure of relative emotion 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.

If the relative shifts in the emotional content correlates 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. In other words, it would provide evidence for the existence of and the accuracy to which one can measure some latent emotional variables as they evolve dynamically in a market economy.

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 they selected to portray. For example, one can easily imagine a big difference, in the extent to which the content creators *feel* what they write, between financial news documents and social media. It is clear that this depends on the intention of the content creators.

3.1 Methods

A summary statistic of two emotional traits is extracted from text data by a word count methodology. The simplicity of this method is intentional. 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 need to apply more sophisticated natural language processing techniques to establish these relationships. Two lists of previously applied and experimentally validated (Strauss, 2013) words are used (each of approximately size 150), one representing *excitement* and one representing *anxiety,* examples of which can be found in Table 1.

Table 1: Emotion dictionary samples

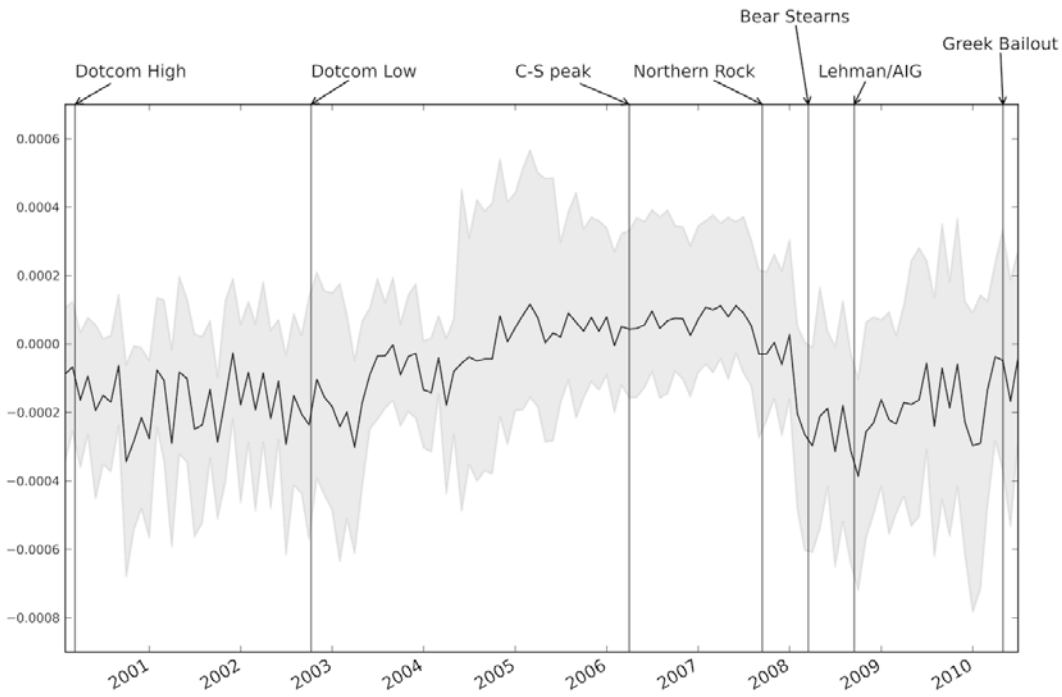
Excitement	*Anxiety*
amaze	anxiety
amazed	anxious
attract	avoid
attracted	avoids
beneficial	bother
boost	bothers
confident	attract
confidently	attracted

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 in 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.

3.2 Results

We explore the relative emotion series extracted from MCDAILY in Figure 1. The graph, which is annotated with events of relevance to financial stability moves broadly as might be expected. In particular, ut 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 actual crisis at around the time of the Lehman Brothers failure. Figure 2 shows that the biggest driver behind the relative emotion statistic appears to be changes in anxiety levels (series are here smoothed with parameter



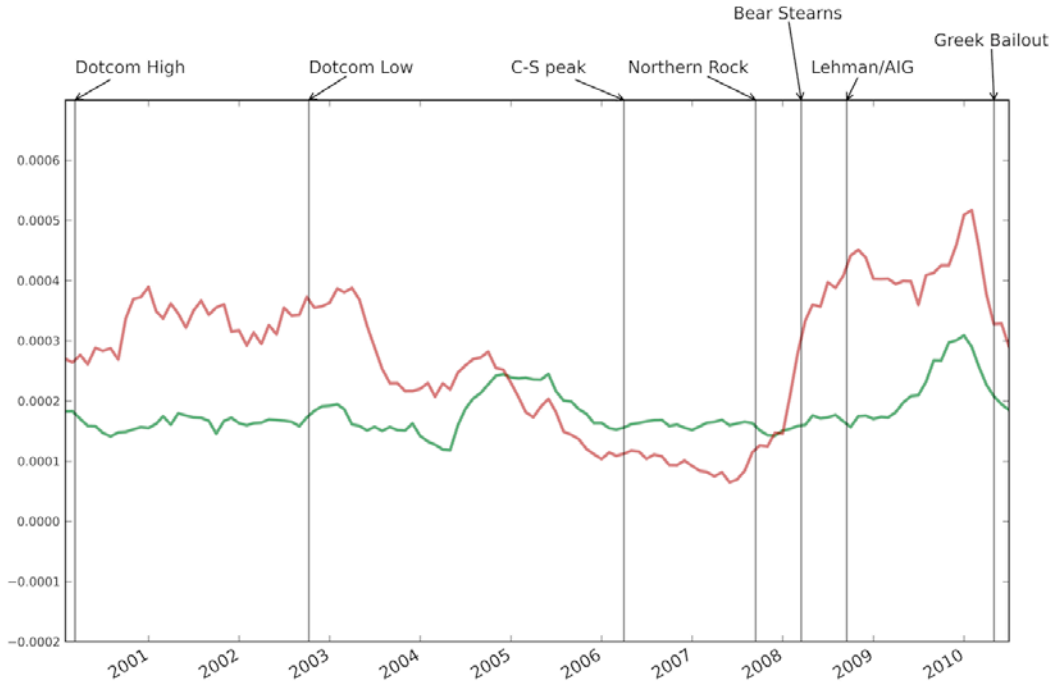


Figure 2: Emotional factors of MCDAILY; anxiety (red) and excitement (green)

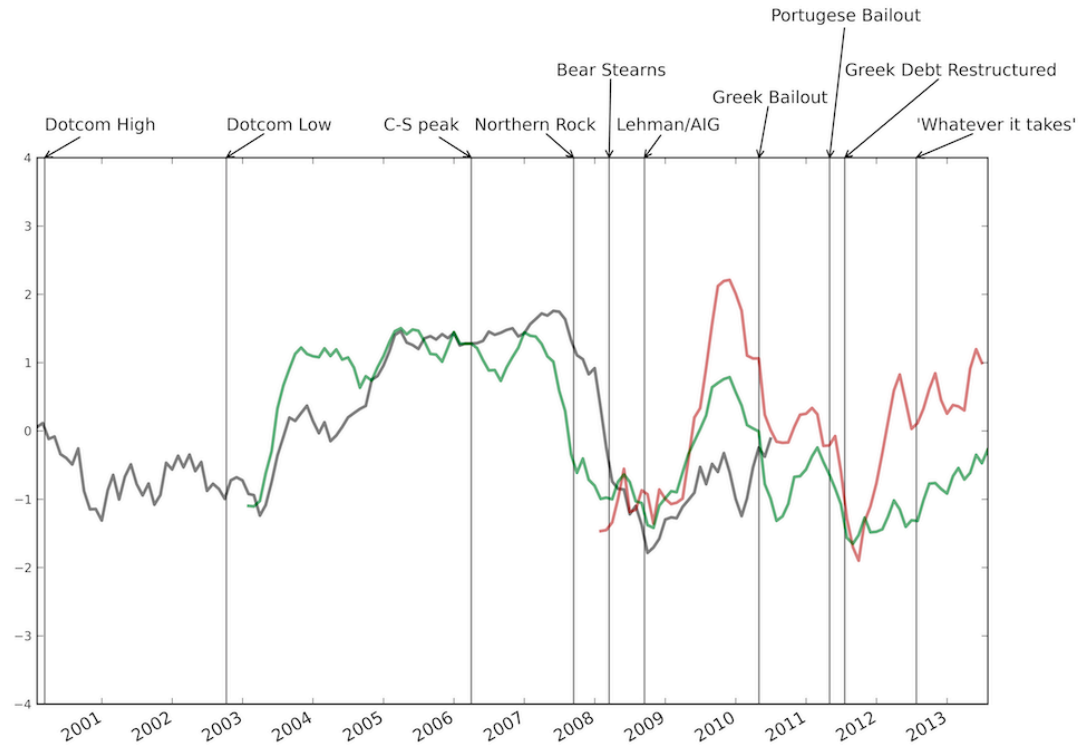


Figure 3: Relative sentiment of MCDAILY (black), RTRS (green) and BROKER (red)

This series is compared with that extracted from other sources, such as RTRS and BROKER in Figure 3 (series are here smoothed with parameter

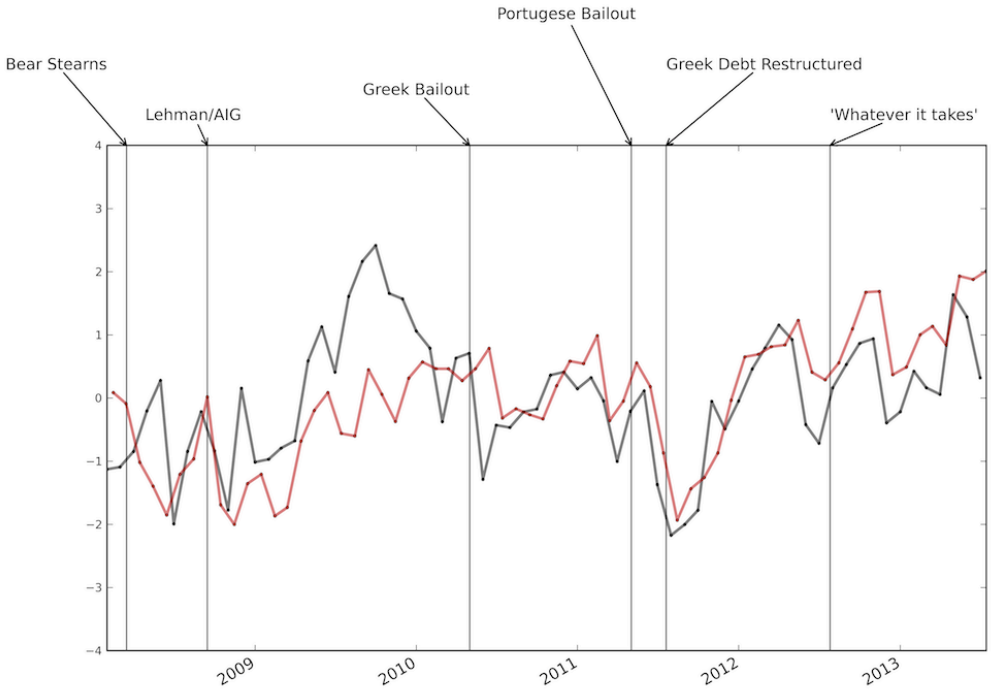


Figure 4: Relative sentiment of BROKER (black) compared to the Michigan Consumer sentiment index (red)

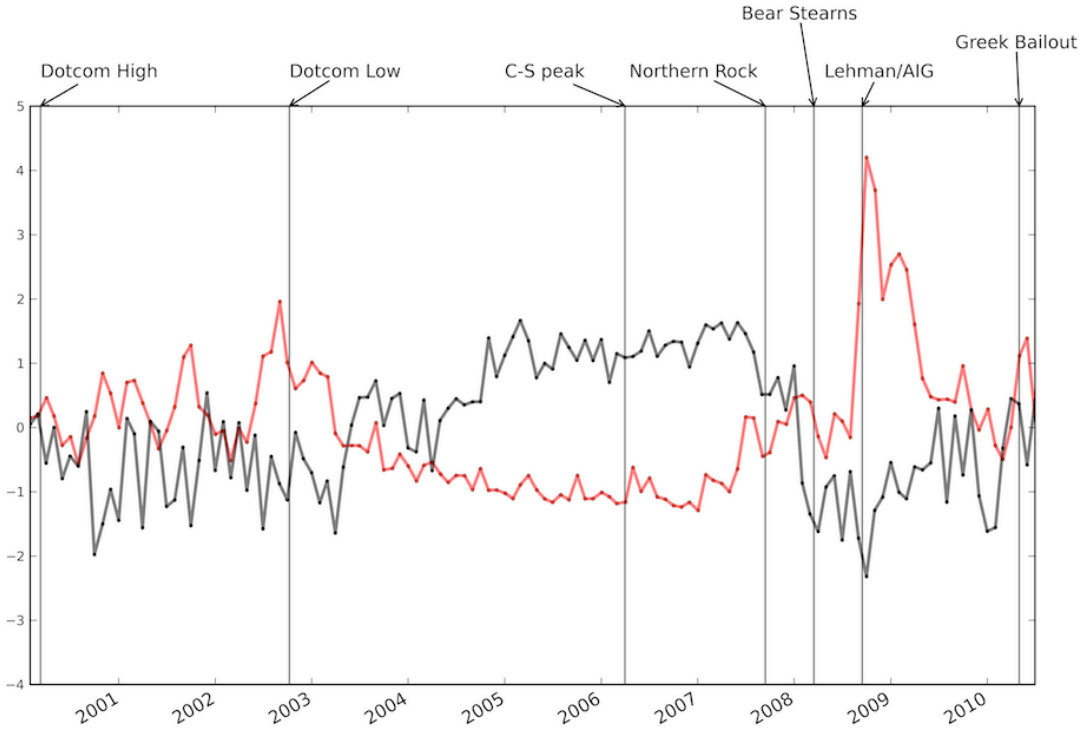


Figure 5: Relative sentiment of MCDAILY (black) compared to the VIX (red)

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, such as matching a pattern, as we have done with, e.g., 'property' or 'Property' in Reuters news archive (Figure 6). In other words, we target the analysis on a particular topic or entity. The bottom chart measures the total number of characters found in articles matching the above patterns in the entire Reuters news archive. It is particularly interesting to note the steady increase and later decline in volume, the turning point occurring a couple of months after the bankruptcy of Lehman Brothers and the corresponding surge in anxiety. Such focused analysis could potentially be of value in monitoring the emergence of a property bubble.

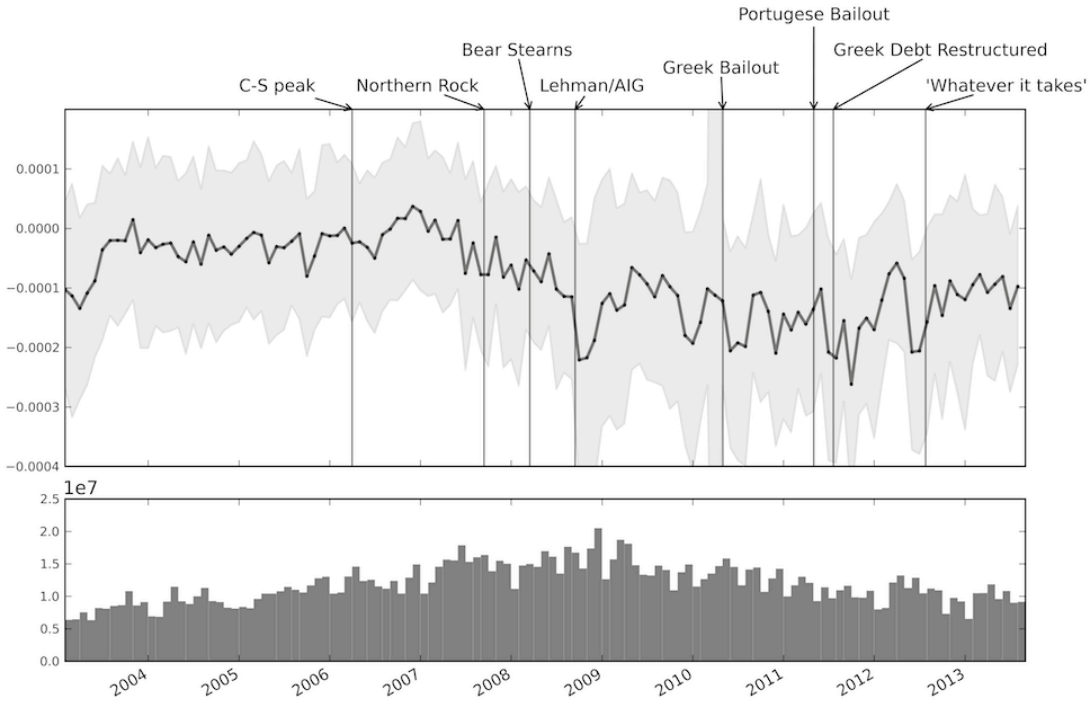


Figure 6: Relative sentiment surrounding 'property' in RTRS

4. Measuring consensus

We turn now to our second question: can we measure structural changes in the distribution of narratives? The objective is to investigate if we can detect when some narratives grow to become dominant, potentially to the detriment of the smooth functioning of financial markets and hinting at impending distress. We introduce a novel methodology to explore this.

4.1 Methods

For this investigation we focused on RTRS as it is of larger volume than the other sources. To measure consensus, we make use of current state of the art information retrieval methods. The main challenge is choosing a good methodology for automatic topic extraction. Our approach relies on clustering the articles into topic groups after first mapping each to a numerical vector. 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 (cite).
2. Compute a word by document frequency matrix, with words as rows and documents as columns (each entry

accounting for

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 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 the *information entropy* (cite Shannon entropy). For a discrete distribution the entropy is simply a logarithmically weighted sum of probabilities,

It is clear from this that entropy is increasing logarithmically as *k* increases. In other words, the entropy is like an inverse consensus measure. For example, 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.

4.2 Results

We run the algorithm across several choices of parameters (the full list of model parameters, and combinations used (e.g., '40, 5, 100' and '50, 2, 100'), can be found in Table 2) and smooth (using double exponential smoothing) and average the results to arrive at the red line in Figure 7.

The graph shows a clear increase in consensus (decrease in entropy) preceding the crisis period and much more disagreement subsequently. Having decomposed the narrative discourse into one index measuring structural changes in consensus (entropy) and one index measuring shifts in emotion (the previous section), it appears from a result that a predominantly excited consensus emerged leading up to the crisis, driven by low levels of anxiety. This seems consistent with belief 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 shifted into predominantly anxious disagreement, as might be expected in an environment of fear and uncertainty.

Table 2: Consensus parameters

Parameter	*Notation*	*Values considered*
Upper word bound		40,50,50,60,40,50,50,60
Lower word bound		5,2,10,10,5,2,10,10
Document vector dimensionality		100,100,100,100,200,200,200,200

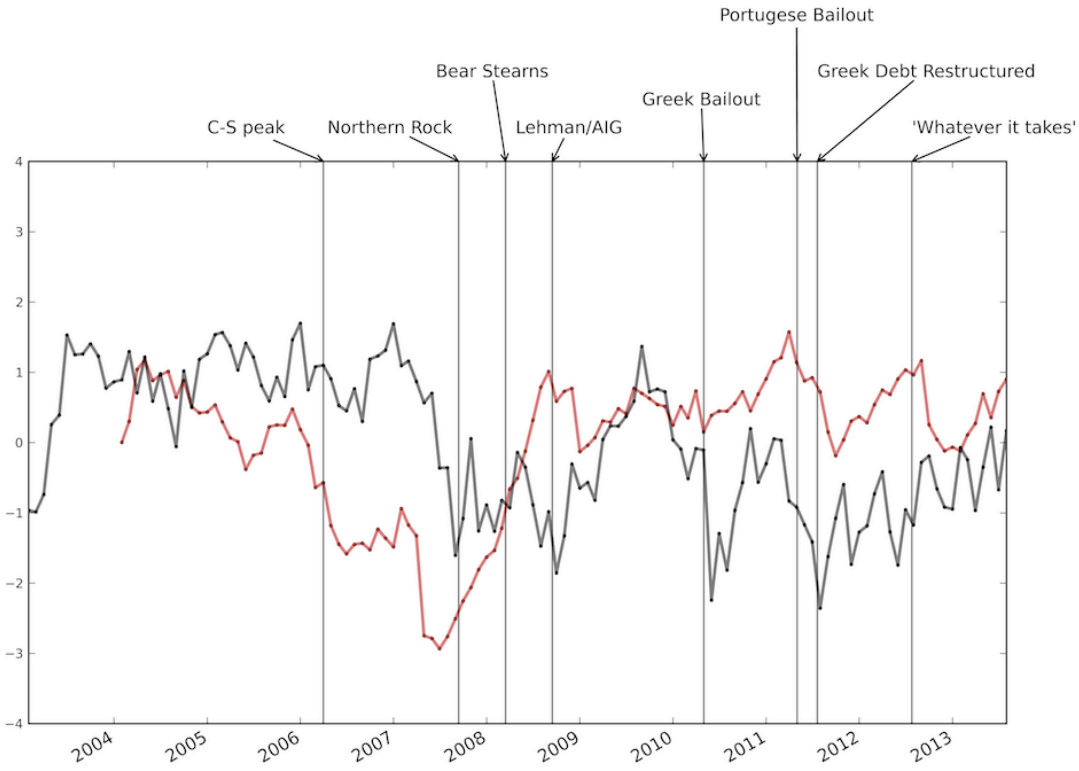


Figure 7: Relative sentiment (black) and consensus (red) in Reuters' London news

5. Discussion

In this paper, we have explored the potential of using algorithmic text analysis, applied through the lens of conviction narrative theory, to extract summary statistics from data archives largely unused by policy makers 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 are robust to the choice of words, as well as across all data sources analysed. A comparative analysis shows that the emotion statistic in BROKER is highly correlated with the Michigan Consumer Sentiment Index and that of MCDAILY is somewhat negatively correlated with the VIX. We can conclude that our metric gives additional information. When we focus on particular topics or entities, such as 'property', we find that the methods could be useful as tools to monitor potential risks to economic and financial stability.

Furthermore, we have created a novel methodology to measure consensus in the distribution of narratives. This metric can potentially be used to measure the build-up of financial instability resulting from homogenisation among market participants. The increased consensus metric, when viewed together with the relative sentiment shift series, can 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 recent 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 Paul Ormerod, 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.

Bibliography

Tuckett, Smith and Nyman, 2013

Tuckett, Ormerod, Smith and Nyman 2013

Chong and Tuckett, 2013

Tuckett, 2011

Haldane

De Haan

Blinder et al.

Paul Hubert

Soo, Cindy

Strauss, 2013

FSR

LSA

X Means

Shannon Entropy

Etc.

Appendix

Latent Semantic Analysis

Double exponential smoothing