ARTICLE IN PRESS
JID: KNOSYS [m5G;March 14, 2018;19:12]
Knowledge-Based Systems 000 (2018) 1–11
Contents lists available at ScienceDirect
Knowledge-Based Systems
journal homepage: www.elsevier.com/locate/knosys

Estimating the impact of domain-specific news sentiment on financial assets

Stephen Kelly*, Khurshid Ahmad

School of Computer Science and Statistics, Trinity College Dublin, Ireland

ARTICLE INFO

Article history:
Received 28 September 2017
Revised 28 February 2018
Accepted 2 March 2018
Available online xxx

Keywords:
News sentiment
Sentiment analysis
Domain terminology
Equity market
Oil commodities
Trading strategy

ABSTRACT

The influence of news on financial markets has been studied by extracting opinions and sentiment from text using content analysis and natural language processing techniques, and then using this measure to estimate the impact of sentiment on price changes. We present a method and implementation that analyses the content of news using multiple dictionaries that accounts for the specific use of terminology in a given domain. To evaluate our approach we build different collections of domain related news for two financial markets and examine the impact that topical news has on two financial benchmarks in the equity and oil markets. We examine how the level of news sentiment from different news sources influences financial returns over time. We create a trading signal based on the news impact that predicts next day returns in the Dow Jones Industrial Average and West Texas Intermediate crude oil. We find that incorporating news sentiment into a trading strategy increases annual returns over a simple buy and hold strategy for both markets.

1. Introduction

The number of variables that influence financial markets are innumerable and the time at which they influence markets varies. Contradictory opinions and beliefs about the movement of financial assets are so varied that accurately predicting future price changes is effectively not possible. However, variation in prices can be modelled and the likelihood of a price move can be at least estimated. We observe many external factors that cause theoretical models of price to deviate from their assumptions. The problem becomes how to explain these unexpected variations using means that traditional quantitative variables and measures have not succeeded in doing.

News stories and the accompanying sentiment contained in them represent a wealth of information that may not be fully reflected in financial market data. Increased interest has been seen in the finance community to employ methods of text and content analysis to automatically analyse this information in the hope of adding explanatory information to traditional financial models. Combining methods from two disciplines, that of text and econometric analysis, has in recent times garnered much attention [1–3]; with systems being developed to automate this process and fully integrate the methods and data processing tasks [4,5]. Creating a system to handle these vast amounts of news text and compute a model that can utilise this information poses an interesting challenge and is the subject of the work presented here.

In this paper we propose a method and implementation that can perform text analysis and statistical modelling, with an emphasis on creating a framework that can be generalised to allow the analysis of different text sources, news types, financial and time series data. We use a bag of words model with multiple dictionaries of categorised terms. The modelling component of the system evaluates the sentiment variable produced by estimating the inter-relationship between the news sentiment and financial returns using a rolling window vector autoregression model and a hypothesis test.

While many approaches exist in the literature for generating sentiment, and the impact of news has been investigated in several markets using various approaches, few methods have attempted to apply the same approach for estimating the impact of news in different markets in a general way. Classification based methods for instance need to be retrained for accurate application to different domains. The novelty of our approach lies in incorporating multiple dictionaries to deal with different categories and topics, in particular domain terminology, without the need for a large volume of training data. Combining dictionaries allows users to deal with domain subject matter, categorise documents according to a number of topics determined by the dictionary, and to disambiguate between sentiment terms and domain terms or phrases. We estimate the impact of this sentiment measure using robust statistical

* Corresponding author.
E-mail addresses: Stephen.Kelly@scss.tcd.ie (S. Kelly), Khurshid.Ahmad@scss.tcd.ie (K. Ahmad).

https://doi.org/10.1016/j.knosys.2018.03.004
0950-7051/

Please cite this article as: S. Kelly, K. Ahmad, Estimating the impact of domain-specific news sentiment on financial assets, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.03.004

methods and rely on the statistical significance of the result to aid decision making for potential trading opportunities in a parallel to traditional risk measures such as value at risk.

The rest of the paper is organised as follows. Section 2, we review the background work and literature surrounding content and sentiment analysis for financial data and similar systems developed. In Section 3 we describe our approach to content analysis accounting for domain terminology, and the rolling regression method to assess the impact of sentiment. Section 4 shows the results of our evaluation and result of the backtest of our system incorporating news sentiment into a simple trading model for financial returns. We conclude the paper with a short discussion and summary.

2. Background

2.1. Sentiment and financial markets

The price of assets traded on different markets are dynamic and constantly trying to find an equilibrium based on a presumed underlying value. Many interpretations and methods exist to try and explain why certain price changes have occurred. Explanations include reactions to macroeconomic news, market speculation, and changes in investor sentiment. These interpretations, although beneficial in explaining price changes, do not explain the entire process, often accounting for a small degree of change. Many theories have looked to investor and market sentiment to explain the more irrational trends in the market [6] with the stock market being the most widely studied market for monitoring and assessing the impact of sentiment [2,3,7].

Many of the studies have focused on stocks and forex markets [8] while less attention has been given to commodities markets. An investigation by Narayan and Narayan [9] finds that oil price news can impact certain sectors in the stock market, identifying a profitable approach to accounting for news. This study however only looks at the headlines as opposed to the full content of the text. Another study by Narayan and Bannigidadmath [10] extends the text data set of Garcia [2] and find the New York Times columns impacts different classes of stocks, demonstrating the reach and impact of news sentiment. These data are used again in [11] and it is shown that news can contribute higher returns to different portfolio trading strategies that are statistically robust. The impact of news sentiment on financial returns is evident in each of these studies, however their approaches are restricted in news type and do not account for the semantics or changing use of terms across domains.

Studies that have attempted to incorporate sentiment into models that describe market fluctuation have looked at the time varying nature of sentiment and how to account for volatility. Smales [12] examines the time varying nature of sentiment using rolling regressions, incorporating sentiment into the Fama–French pricing model. Smales finds a close relationship between the market risk premium (beta) and news sentiment, a linear relationship that varies in time.

The impact of sentiment varying through time has been studied in the context of volatility and business cycles also. Garcia [2] relates sentiment to recessionary and expansionary periods. The paper also examines issues of sentiment being related to volatility in a typical manner to other literature by including a volatility measure in the specified regression equations and also runs the same models but normalises the returns series with the conditional volatility series produced from a GARCH(1,1) model (degarching the series). Although sentiment is seen to vary with business cycles, the results of degarching returns shows the effect of sentiment isn't negated by accounting for volatility. Following from [1,2], the authors proposed a tentative analysis on the impact of sentiment on volatility and business cycles [13]. Using a volatility based regression it was found that sentiment and news volume had a small but statistically significant effect when considering these variables as extensions to the GARCH model. The authors in another study found links between the distribution of returns and sentiment by categorising them according to the different quantiles using locally weight regression and found a similar non-linear relationship between sentiment and different asset classes [14]. In Shen et al. [15] Baidu News is used as a proxy for the arrival and volume of information. One notable observation of their study is the lead-lag relationship between information flow and return volatility, supporting the financial theory of information arrival and its impact on volatility. This demonstrates the usefulness of information published online and news sentiment for financial analysis.

The relationship between volatility, macroeconomic fundamentals, and investor sentiment is examined using a structural VAR in [16]. They decompose volatility into long-run and short-run components finding the impulse response shock from investor sentiment (the U.S Crash Confidence Index is used a proxy for sentiment) to be more closely related to the transitory component. It is suggested that portfolio performance can be improved by considering the volatility induced by sentiment and including it as an additional risk factor. The idea of changing volatility and the dynamic influence of sentiment has also been highlighted in [17]. They follow the assumption that investors believe in an underlying dynamic market regime, such as a trending or mean-reverting market. They model these two regimes using a Markov process and view the reaction to news and investor sentiment through these regimes. Chung et al. [18] use the NBER defined business cycles with a Markov-switching model to examine the impact of sentiment in different economic regimes. They also find the predictive power of sentiment to change according to the regime they are currently in. Similarly, in [19] the impulse response and variance decomposition of the sentiment proxies are examined to see the effect that non-fundamentals have on economic fluctuations. Their findings suggest the relationship between sentiment and stock returns are related both to justified expectations and not just speculation based on improper information. These studies highlight the power of using structural VAR and its extensions to examine the contribution of a shock from the sentiment variable to the modelled economic variables and financial returns, and how the error variance is distributed amongst these variables. Our work presented here focuses primarily on how choice of news type, source, and domain can influence the creation of a sentiment variable, relying on a reduced form VAR to estimate the impact of the sentiment variable. Once a reliable sentiment proxy is found, principled econometric models and approaches such as those outlined can be employed to improve forecasting and estimate correlations with financial returns.

2.2. Domain news and sources of news sentiment

When analysing a collection of documents for sentiment, the documents are typically considered to be unified by topic [20]. When building a corpus (systematically organised collection of text) of text, consideration must be given to developing a balanced, representative collection of information regarding a specialist area or topic. Sentiment analysis applications in the area of finance have shown that by choosing authoritative sources, and topic relevant articles, the results will be more in tune with intuitive beliefs and also theoretical models of behaviour [2,3]. The construction of language resources such as the British National Corpus (BNC) have also shown that corpus linguists prefer reputable sources [21].

The use of informal sources of online messages and social media has also been prevalent in recent studies. The content and volume of messages boards have been shown to proxy for market behaviour [1,22]. Messages published in online columns have been linked to hypotheses in financial literature such as the price pressure hypothesis [23]. More recently with the advent of social media, great interest and emphasis has been put on looking at trends, volume, and analysing the content of social media and blogs [24–26]. Social media and many forms of online media tend to be unedited and uncensored. Due to this, the text data tends to be noisier, requires more effort in evaluating the messages, and is harder to extract useful information from.

Many sources of sentiment may exist, even for text based sentiment. The investigation in Dzielinski [27] uses Thomson Reuters News Analytics data, which classifies news into negative, neutral, and positive. This sentiment measure relies less on the level of negative sentiment, or relative negative sentiment as used in this study, but instead relates the occurrence of these news articles with the distribution of returns that day. The work presented in Smales [12,28] also relies on Reuters sentiment data and incorporates the relevance metric for each text. The studies presented by Smales highlight the importance of correct news categorisation for deriving sentiment that has a significant impact on market prices. The approach taken by Smales follows from Dzielinski where a weighted average of the prevailing sentiment is computed. The work in Groß-Klußmann and Hautsch [29] presents a system that incorporates Reuters news sentiment as well and finds the impact of this measure in an number of areas of the market microstructure. In their approach they also recognise the importance of filtering for relevant news articles and again use the Reuters scoring metric for news relevance. Although these studies recognise that article relevance is important they do not discriminate on the relevance of the articles based on the content, only incorporating relevance based on Reuters metrics. This approach although justified, may induce more noise as it doesn't disambiguate terms or sentiments from the Reuters metrics. Our method proposes looking at the relative frequency of terms, accounting for variation in news volume, and also attempts to disambiguate domain terms from sentiment terms. The financial trading system presented in Groß-Klußmann and Hautsch also assess the cumulative performance using abnormal returns. Our approach attempts to highlight potential trading opportunities as a next day risk estimation based on the level of negative news.

2.3. Sentiment and forecasting

Many of those in the domain of finance have used commercial off-the-shelf systems and sentiment data [12,27,28]. Although useful, these tools are often limited in their classification, limited in their analysis of affect and sentiment, and do not allow a user access to the linguistic rules and text processing [30]. So far, many of these approaches rely on the bag-of-words model. Single words are considered to be the simplest linguistic structure and provide a good baseline to start with when consider sentiment analysis applications. Recent advances in word vector representations have allowed richer semantics to be extracted from text [31] and their use with newer machine learning architectures has seen an increase in applications across a number of disciplines from semantic analysis, sentence classification, to image annotation [32–34]. A number of studies in financial literature have relied on previous systems such as the General Inquirer, Thomson Reuters News Analytics, and open source systems to generate a sentiment variable while other studies have developed their own approaches and frameworks. A review of some of these approaches and their applications is given in [35].

In more recent times literature from machine learning, and the deep learning trend, have used neural network based models for sentiment analysis and forecasting. Methods include convolutional neural networks [36], reinforcement learning [37], LSTM [38], and autoencoders [39] among the most popular approaches for financial analysis. Ensemble methods are still in use with traditional time series models like ARIMA being used with deep learning approaches [40]. A large proportion of these studies that use text sentiment in a prediction model for finance often focus on stock market and foreign exchange rate data, while fewer studies have focused on commodities and crude oil specifically [41]. Many of the more prominent studies attempting to incorporate natural language processing into financial forecasting are summarised in [8].

3. Methods

3.1. Data set

We carry out our evaluation using financial data from two different markets; the equity market using the Dow Jones Industrial Average (DJIA) and crude oil market using West Texas Intermediate (WTI) crude oil. We chose both of these financial assets due to their influence as benchmark assets for each of their respective markets and as a result each has a large number of news items released about them regularly.

For the equities case study we use opinion columns *Abreast of the Market* (AOTM) published in the *Wall Street Journal* which discusses news about DJIA companies and the *Lex column* from *Financial Times* (FT) which is considered to have a long term view of business and finance events.

In our study of the oil markets we examine changing the text type and source by using different sources including a corpus of *crude oil* related news sampled from the FT and message postings on the *Oildrum* blog. The text collections used in the commodity market evaluation are presented in Table 1. Using these sample data our system evaluates the differences that formal and informal news can have when constructing the sentiment proxy and its impact on returns. The text collections used for the equity and oil market evaluations are presented in Table 1.

In our work presented here we collect a volume of text ensuring the topic and source are consistent. Text were collected from specific sources, i.e The Financial Times and Lexis Nexis database. Within these databases the texts were tagged and classified according to topics. Each text corpus is considered consistent in that each text is from a single well regarded source and tagged or classified by the vendor as being relevant for that topic.

A dictionary of terms is central to the method of content analysis implemented by our system. The GI dictionary is used by default in the system to summarise the tone of the news articles. We evaluate the impact that negative sentiment has in each of our chosen markets (Table 2). Negative sentiment is predominantly used in many studies due to the over reaction to negative news typically observed, while this effect is less pronounced with positive sentiment and often the latter is only considered to have predictive power when aggregated with the negative sentiment measure [13,42,43].

When analysing general finance news, such as the *Abreast of the Market* column from the Wall Street Journal, the GI dictionary is used to estimate negative sentiment from the text. While examining the oil related text corpora, both the GI dictionary and a glossary of oil words, the Platts and Oil and Gas UK glossaries (Table 2), are used to summarise the negative sentiment in text. Using two dictionaries allows us to take into account potential miscounts and misinterpretations of domain terms and phrases. Using our approach to account for domain terms means potential miscounts of sentiment can be excluded. The assumption we draw on is that domain terms and phrases may be mistaken as being sentiment laden terms [21]. Although domain terminology may not be completely devoid of sentiment, by using domain knowledge

Table 1
Full sample of text collected for each corpora used in the equity market and sentiment evaluation.

Text	Description	Article	Words	Period	Type
Abreast of the Market	Opinion column run in the WSJ (Daily until 2008)	5466	5,663,193	03/01/1989 31/12/2008	Editorial, Op-Ed
FT Lex Column	Only articles included in the Lex Column of FT	20,195	6,559,921	02/01/2005 07/11/2014	Editorial
FT crude oil news	Sampled across the FT using the key word "crude oil"	23,157	12,252,484	04/01/2000 08/12/2014	Reportage, Editorial
Oildrum Blogs	Posts and commentary on the Oildrum blog	80,766 excerpts	5,838,633	26/01/2007 01/09/2013	Blogs

Table 2
Lexical resources used in the market evaluation.

Word list	Description	Words
GI Negative	Category contained in the General Inquirer dictionary constructed from the Harvard IV-4 dictionary, the Lasswell value dictionary.	2005
Platts	Glossary from Platts who provide information and price and estimation on industry benchmarks in the physical energy markets	704
Oil and Gas UK	Glossary from the Oil and Gas UK a leading representative body for the UK offshore oil and gas industry	126

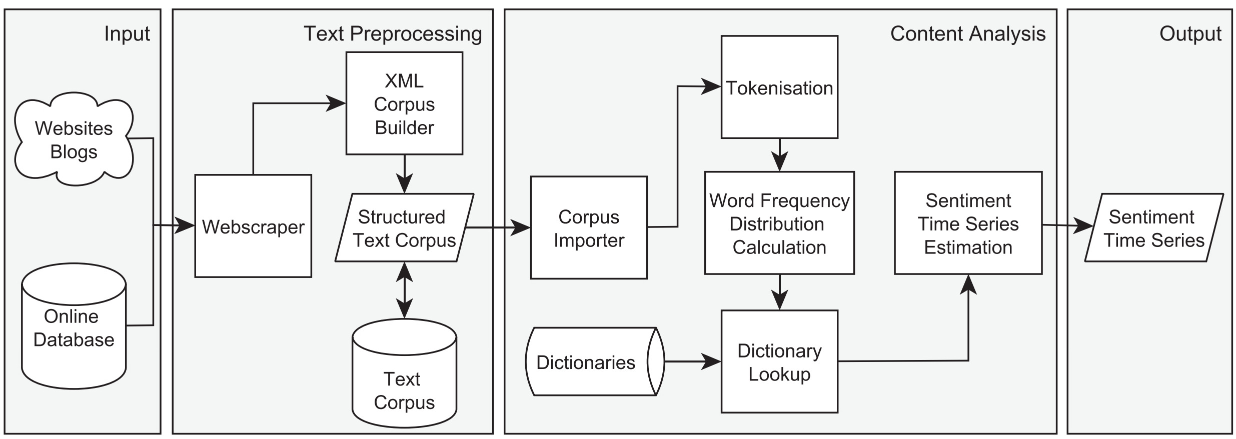

Fig. 1. A system diagram of the text analysis component developed for the system that computes a time series of sentiment.

sentiment may be better summarised for an article than if it was considered in general language use. We give a brief illustration of this in our results section (Section 4).

3.2. Content analysis for domain text

The overall architecture of the text analysis component is shown in Fig. 1. The main functionality is divided into two main components, the first component deals with collecting, structuring, and aggregation of text documents from online news sources to build a text corpus. The second component performs content analysis using the bag-of-words model and a given dictionary(s) of categorised terms. The output of this component is a time series of sentiment extracted from the corpus of text. This time series is then passed to the second part of the system which aggregates the time series output with financial data and performs the statistical modelling, described in the subsequent section.

3.3. Text processing

The text analysis component analyses the content of the text by looking at the frequency at which pre-categorised terms (negative category of terms from the GI dictionary), defined in the supplied dictionary, occur in a text.

When constructing the sentiment measure we compute the relative frequency of negative terminology as it appears in all documents for a day. This score is less sensitive to changes in news volume day to day. The relative frequency can be defined by saying that for k number of w words in a category of a dictionary, the total frequency of their occurrence in a corpus of text D can be denoted as the summation of $f(w_D^k)$. If we denote time as t, then the frequency of all words in a category of a dictionary that occur in all articles published during some time period (during one day for example) can be written as the sum of $f(w_t^k)$. If the frequency of a word or token n that occurs in all articles published during a single time period is the sum of $f(w_t^n)$, then the relative frequency of a category of terms from a dictionary is calculated as:

$$Relative\ frequency = \frac{\sum_{i=1}^{k} f(w_t^k)}{\sum_{i=1}^{n} f(w_t^n)} \quad (1)$$

where:

n = total number of words in all documents published during a time period

k = number of terms in the dictionary category

$f()$ = frequency of a token

t = time period

In our evaluation we use an oil benchmark and oil related news. In this evaluation we examine two domain glossaries for the oil industry these include the Platts[1] glossary, and the Oil and Gas UK glossary[2] (Table 2). These lists cover common terms and abbreviations from the oil and energy industries. Both sources are seen as being reputable; Platts is a leading source of information on energy markets and provider of benchmark prices, while Oil and Gas UK is the leading trade association for the UK's oil and gas industry. The Platts glossary contains 704 words with expanded abbreviations,

[1] wwww.platts.com/glossary.
[2] http://www.oilandgasuk.co.uk/glossary.cfm.

Please cite this article as: S. Kelly, K. Ahmad, Estimating the impact of domain-specific news sentiment on financial assets, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.03.004

while the Oil and Gas UK glossary contains 133 words including expanded abbreviations. The count of affect terms will not be altered if GI negative and positive terms occur independently of the domain glossaries. If the term “crude” occurs separately from the compound term “crude oil”, the latter will not increase the negative word count but the former will still be accounted for.

This method of disambiguation for specialist text has been incorporated into the dictionary lookup function of our content analysis component. This procedure is summarised by Algorithm 1 and is implemented in our system.

Algorithm 1 Generate a time series of sentiment for a domain specific corpus.

Input: Base dictionary C_m consisting of m number of C categories
Domain Dictionary C_s consisting of s number of C categories
Corpus of text documents D organised into n time periods
Output: Time series $F_C = \{f_1, \ldots, f_n\}$ where $f(n)$ is the sum of the word frequencies during time period n for category C

1: **for** $i = 1$ to n **do**
2: $w_i \leftarrow \text{Tokenise}(D_i)$
3: **for all** C **in** C_m **and** C_s **do**
4: **if** w_i **in** C_m **and** C_s **then**
5: $f(w_n^s) \leftarrow \text{Count}(w_i)$
6: $f_n^s \leftarrow \sum_{j=1}^{i} f(w_i^s)$
7: **end if**
8: **if** w_i **in** C_m **and not** C_s **then**
9: $f(w_i^m) \leftarrow \text{Count}(w_j)$
10: $f_i^m \leftarrow \sum_{j=1}^{i} f(w_i^m)$
11: **end if**
12: **if** w_i **in** C_s **and not** C_m **then**
13: $f(w_i^s) \leftarrow \text{Count}(w_i)$
14: $f_i^s \leftarrow \sum_{j=1}^{i} f(w_i^s)$
15: **end if**
16: $f_i \leftarrow \{f_i^s, f_i^m\}$
17: **end for**
18: $F_C \leftarrow \{f_1, \ldots, f_n\}$
19: **end for**

In Algorithm 1 content analysis is performed on all documents that occur during a given time period. For our evaluation daily frequency is chosen. Daily news articles can be more easily aggregated to this level to give a more consistent time series with less likelihood of missing observations. The input to Algorithm 1 consists of a dictionary and corpus of text. A general language dictionary is used, in our evaluation we use the GI dictionary by default in our system, and becomes the *base dictionary* consisting of several categories of terms. A second dictionary or glossary of terms can be input also and used in conjunction with the base dictionary. This domain dictionary can account for potential misinterpretations of the base dictionary terminology for domain text. The corpus of text is organised in time and is done by aggregating news and text published for a particular day (for daily time frequency). The dictionaries and text corpus are passed to the algorithm and for each day, or time period, all news articles with a time stamp for that day are aggregated together and tokenised into n-gram length words (Algorithm 1 lines 1 and 2). For each category in the supplied dictionaries the frequency of occurrence of terms in each category in the corpus is calculated (lines 3–16). If a word from the corpus appears in both the base and specialist dictionary, then the specialist dictionary takes precedence and the count for the category in the specialist dictionary is increased (lines 4–6). The frequency of occurrence for all the specialist dictionary words that occur in all articles that day is computed and added to the overall frequency for the corresponding category in the specialist dictionary (line 6). This function is used again if the word from a text appears exclusively in the specialist dictionary (lines 12–14). If the token from the text only appears in the base dictionary then the frequency count is increased for this corresponding base dictionary category (lines 8–10). The time series of term frequency for the category is computed and is the final total frequency score for that day (line 16). The daily frequency of a category is then aggregated into a time series and output by the algorithm (line 18).

3.4. Statistical modelling

The modelling component of the system computes a VAR model to estimate the inter-relationships between sentiment, returns, and a series of control variables that might also act as a proxy for the same effects as the sentiment variable. A rolling regression model is also employed to examine the changing influence of sentiment in time. These are combined with a hypothesis test to estimate the statistical confidence in using the sentiment variable in order to explain changes in financial assets. The specification for the VAR model is shown in Eq. (2).

$$r_t = \phi_0 + \sum_{i=1}^{5} \phi_{i,1} r_{t-i} + \sum_{i=1}^{5} \phi_{i,2} s_{t-i} + \sum_{i=1}^{5} \phi_{i,3} v_{t-i} + \sum_{i=1}^{5} \phi_{i,4} o_{t-i} + \sum_{i=1}^{5} \phi_{i,5} Exog_t + \varepsilon_t \tag{2}$$

where:
r_t = Financial returns
s_t = News Sentiment
v_t = Log detrended trading volume
o_t = VIX returns
$Exog_t$ = Matrix of exogenous variables
ε_t = Uncorrelated white-noise disturbances

A number of transformations were also performed on our data to create stationary or weak stationary series. We calculate returns using the log difference of prices. We standardised the sentiment time series so as to have zero mean and unit variance. We find no significant variation or outliers in the mean or variance of our sentiment time series over the sample periods chosen. The mean and variance of the sentiment series remain consistent over the sample series and results are not impacted by the standardisation. Standardising the data in this way allows the coefficient values of the regression to be interpreted as unit deviation changes in the distribution of negative terms occurring in the corpus of news. That is, a one standard deviation (one unit) change in the frequency of negative sentiment will impact the returns by the coefficient amount estimated.

For the log detrended trading volume (*Vlm*), the log of the volume series is de-trended using a moving average of the log volume series for a period of 60 observations (days). The market level trading volume is used in the form of the New York Stock Exchange (NYSE) group trading volume when examining the Dow Jones Industrial Average and futures contract trading volume for the oil market case study. For volatility the VIX series (Chicago Board Options Exchange Market Volatility Index) is used. Dummy variables to account for the day-of-the week effect and January effect are also included. The control variables and these transformations described have been drawn from models defined in the literature [2,3]. Finally, we include several lags in the equation ($\sum_{i=1}^{5}$)

All regression results presented use Newey–West standard errors to give a heteroskedasticity corrected co-variance matrix in the model estimation. This is particularly relevant for financial returns where changing variance can be an issue.

The attributes of financial series are known to change in time. To examine this behaviour we use a rolling window with the VAR model to assess how market changes can influence the results

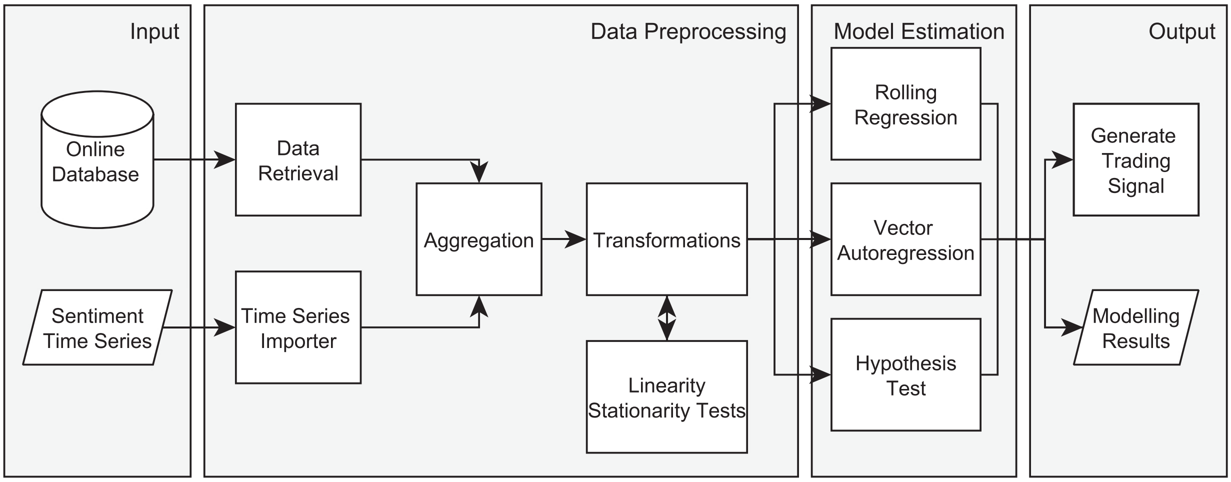

Fig. 2. A system diagram of the Statistical processing component developed which performs data processing and modelling.

found. We perform a hypothesis test to examine the statistical and explanatory power of the sentiment variable for financial returns. This becomes the basis of our basic trading scheme also. We compute the z-score for the sentiment variable in each rolling window period. If for a particular time period, the hypothesis test is significant for negative sentiment predicting a decrease in next day returns we create a sell signal. Otherwise we create a buy and hold signal. The data aggregation and modelling procedure is summarised in Fig. 2.

All time series data are aligned to the financial time series (DJIA, WTI), weekends and public holidays are dropped as no trading happens on these days. Dates where data does not exist for any of the time series variables in the aggregated data are dropped to ensure the best possible alignment of available data. We omit news from non-trading days such as weekends. Volume of financial news if often seen to decrease substantially on weekends with some columns only rarely publishing articles (FT Lex column for instance). Our preliminary tests showed little change in impact from news on returns when we offset weekend news to Monday. We found this marginal impact to already be accounted for using a Monday dummy variable. The Monday dummy has been used in financial literature to account for trading effects and the market micro-structure more so than weekend news. We make the assumption that news has the greatest impact the day it is released and during trading hours when traders can process and act on the information quickly.

All time series data were obtained through the Tradestation platform and the associated exchanges.[3]

4. Results and evaluation

We present two evaluations in the equity and commodity markets. In each case the impact and explanatory power of the sentiment variable produced by the system from different text collections is assessed. First a long-term VAR model is estimated with sentiment, then a rolling regression model is then estimated to assess the changing impact of sentiment over time and the resulting hypothesis test is used to create a trading signal for returns based on the level of negative sentiment.

The VAR equation used has been specified by Eq. (2). Using this model the average impact of sentiment is computed by the system.

[3] DJIA and WTI futures contracts were obtained from the Intercontinental Exchange. The WTI futures contracts were compared to contracts quoted by Stephens Analytics premium data on Quandl with no significant difference being quoted in prices or contact roll over being noted.

Table 3
Coefficients for the sentiment proxy from Eq. (2). Negative sentiment was computed from the Abreast of the Market column in the Wall Street Journal (AOTM), and *Lex column* uses negative sentiment extracted from the *Lex* column in the Financial times. The model was built using data from 1989-01-03 to 1999-11-16 ($n = 2717$). Sentiment extracted from the FT *Lex* column was for the period of 2005-01-03 to 2014-11-07 ($n = 2409$). Coefficients are presented in basis points, one basis point equals a percentage point of 0.01%. The significance for each coefficient is given at 1% (Bold Italic), 5% (Bold) and 10% (Italic) levels.

	AOTM	*Lex* column
$NegSent_{t-1}$	**−4.70**	***−8.5***
$NegSent_{t-2}$	−1.10	**5.1**
$NegSent_{t-3}$	1.40	0.0
$NegSent_{t-4}$	**5.30**	1.1
$NegSent_{t-5}$	**4.60**	−1.3
$\chi^2[NegSent]$	***16.37***	**14.17**

For each model the statistical significance of the coefficients computed are reported. The hypothesis test is also computed and gives a better indication if sentiment is a statistically valid predictor of returns ($\chi^2[NegSent]$). For brevity, we present just the coefficients of the sentiment variable as opposed to the full VAR. For further details of full VAR model computed in our evaluations see [13].

4.1. The impact of sentiment on financial returns

4.1.1. Impact of sentiment on equity markets

The results presented in Table 3 show the coefficient values of the lagged negative sentiment variable predicting DJIA returns for two different periods and using opinion news from two financial news sources. This result summarises the average effect that the negative sentiment variable has on the daily returns of the Dow Jones Industrial Average. The negative sentiment series (*NegSent*) is the standardised relative count of negative affect terms occurring daily in the column *Abreast of the market* from the Wall Street Journal and in a second model the *Lex* column of the Financial Times, categorised according to the GI affect dictionary [13,42].

The results in Table 3 show a similar impact of negative sentiment with DJIA returns. A one standard deviation increase in the frequency of negative terms predicts a 4.7 basis points decrease in the Dow Jones returns on average for the period using AOTM sentiment and −8.5 basis points for the Lex column negative sentiment. A reversal of this impact is also seen across the lagged variables. That is, the initial negative shock of the downward pressure on returns is temporary and is almost fully reversed. The fourth

Please cite this article as: S. Kelly, K. Ahmad, Estimating the impact of domain-specific news sentiment on financial assets, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.03.004

Table 4
Coefficients for the sentiment proxy from Eq. (2). Negative sentiment was computed from the FT crude oil corpus, and *Oildrum* uses negative sentiment extracted from the *Oildrum* blogs archived online. The model was built using data from 2000-01-04 to 2014-12-08 ($n = 3553$) for the FT crude oil data. Sentiment extracted from the Oildrum blogs was for the period of 2007-01-26 to 2013-08-30 ($n = 1182$). Coefficients are presented in basis points, one basis point equals a percentage point of 0.01%. The significance for each coefficient is given at 1% (Bold Italic), 5% (Bold) and 10% (Italic) levels.

	FT crude oil	*Oildrum* blog
$NegSent_{t-1}$	−2.9	4.48
$NegSent_{t-2}$	**−8.5**	−8.08
$NegSent_{t-3}$	**−7.9**	*−20.26*
$NegSent_{t-4}$	−5.8	−1.74
$NegSent_{t-5}$	0.7	**−21.51**
$\chi^2[NegSent]$	***14.322***	*9.3*

and fifth lags are also statistically significant although positive in their influence on returns for AOTM sentiment, supporting the idea of a reversal or return to mean for returns as stated in the literature [3]. The inclusion of the sentiment variable also gives a statistically significant chi-squared test $\chi^2[NegSent] = 16.368$ for AOTM and $\chi^2[NegSent] = 14.17$ supporting the contribution of sentiment to the explanatory power of the model. A proxy for positive sentiment was also computed by the system from the news columns but no statistically significant impact on DJIA returns was found. All series are found to be stationary according to the augmented DickeyFuller (ADF) test with the variance bias and heteroskedasticity accounted for using Newey West adjusted standard errors.

4.1.2. Impact of sentiment on crude oil

We present the results showing the impact of sentiment on the price of crude oil using WTI crude oil futures in Table 4.

The result shows that negative sentiment has a negative relationship with WTI futures returns accounting for an 8 basis point impact on returns on average for the period of the sample set. Much of the impact of sentiment is dispersed during the week, across the lag values. As previous studies have shown, the timing of news plays a role in the predictability and usefulness of sentiment derived from news content. Agenda setting columns such as the *Abreast of the Market* and the FT's *Lex column* typically have timely and succinct information and are often self-contained. Articles in the oil related text corpus used in our evaluation contain news and commentary about events that are on-going and the stories follow these events accordingly. The inclusion of the sentiment variable gives a statistically significant chi-squared test ($\chi^2[NegSent] = 14.322$) supporting its contribution to the explanatory power of the model. Extracting sentiment from the oil and energy blog *Oildrum*, also sees a statistically significant impact and hypothesis test on WTI futures.

Incorporating the oil glossary and computing the relative frequency of negative sentiment using Algorithm 1, we see a decrease in the frequency of negative terms is seen. For the FT crude oil corpus the frequency of negative terms decreases by 24% (from 113,556 to 85,795) and the *Oil drum* blogs sees a smaller decrease in negative sentiment at 16% (169,904–143,357).

News published in authoritative newspapers is written so as to be understood by a wide audience as the readership is large and varied. Our experiments with the oil commodities sees a very small decrease in the impact of sentiment when the GI negative category is combined with the domain glossary slightly. The statistical significance and overall explanatory power of the sentiment variable computed using the negative category with domain word lists decreases. For the Oildrum blogs, the coefficient for the negative sentiment variable increases as does its statistical significance after filtering for domain terms and phrases. This may be due to the higher number of domain specific phrases and words being present in the Oildrum posts. Typically the blog posts on the Oildrum tend to be more specialist discussing aspects of the industry in greater depth, using more domain specific words and language. Overall, accounting for domain language when analysing domain related text can help improve sentiment extraction and identification.

4.2. Robustness of results

Our sample periods were chosen due to the availability of text data. To illustrate our approach and system we gave attention to the domain terminology present in the text, and the type of text being collected. In this way our study focuses on the key terms used in the text and each corpora selected was constructed to contain a specific type of news such as an opinion column (WSJ Abreast of the market, FT Lex column), internet blogs (Oildrum), and industry related news (FT Oil and Gas news). In each case we collected as much news as was available to us. The Oildrum has ceased publication but was at one stage considered a useful and authoritative source of internet based oil news. The availability of the WSJ Abreast of the Market column decreases to weekly publications by the end of 2008, the study is extended to as far as there are daily observations. This allows us to create a daily time series and still have comparable results to the original study proposed by Tetlock [3]. We initially tested the results of the system with those of the literature, and our extended dataset and found the results to be consistent [13]. In the case of the FT Lex column dataset we collected all articles that were available to us from the FT archive. A slightly larger corpus was collected for FT oil related news as it sampled more types of articles from the FT. In each case we were able to compare our results with at least ten years of text data, except for the Oildrum blogs which have six years. However, this still represents a good sample of consistent and authoritative news.

For each model estimation we ensured the results were not spurious and to address concerns of model uncertainty we included control variables based on previous justifications from the literature as described in Eq. (2). We also ensured the time series variables were consistent and did not contain large gaps or many missing dates, that each series was stationary, large outliers were not skewing results, and that the coefficients are individually or jointly significant.

4.3. News trading strategy

We examine the time varying influence of sentiment on returns by employing a moving window function with the VAR model. We compute a rolling method with a windowed time horizon of 30 days accounting for a month of trading data. We update the model everyday as new news is processed and incorporated into the model creating an overlapping window model. We compute the standardised score of the news count for each windowed period so as to account for the possibility of the mean and standard deviation changing over time.

We run a backtest using the trading signal generated on negative news predicting a decrease in next day returns for our two chosen markets. We report the cumulative performance of a simple buy and hold strategy versus trading on news and also the annualised returns and Sharpe ratio for this strategy in both markets. Our approach relies on the output of the hypothesis test and whether negative sentiment will impact returns. In our approach we detect short term impacts from news and this suggests a potential trade opportunity as opposed to making a point estimate for the next day returns.

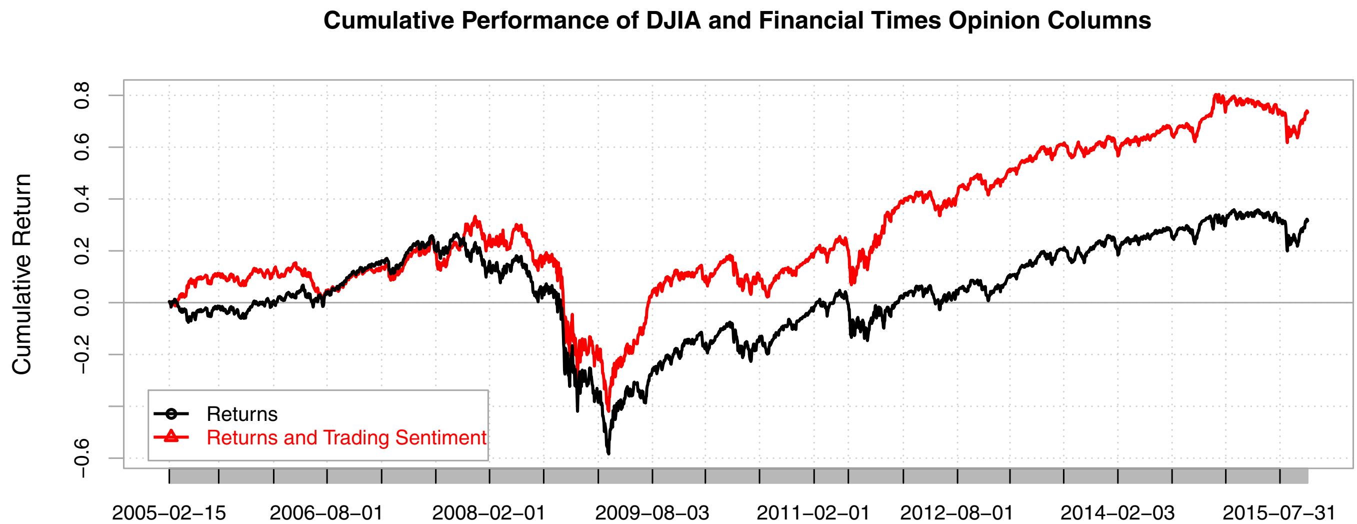

Fig. 3. Rolling VAR (30 day window) model with DJIA returns and negative sentiment for the period from 2005-01-02 to 2015-11-02. The negative sentiment variable was extracted from the Financial Times Lex Column.

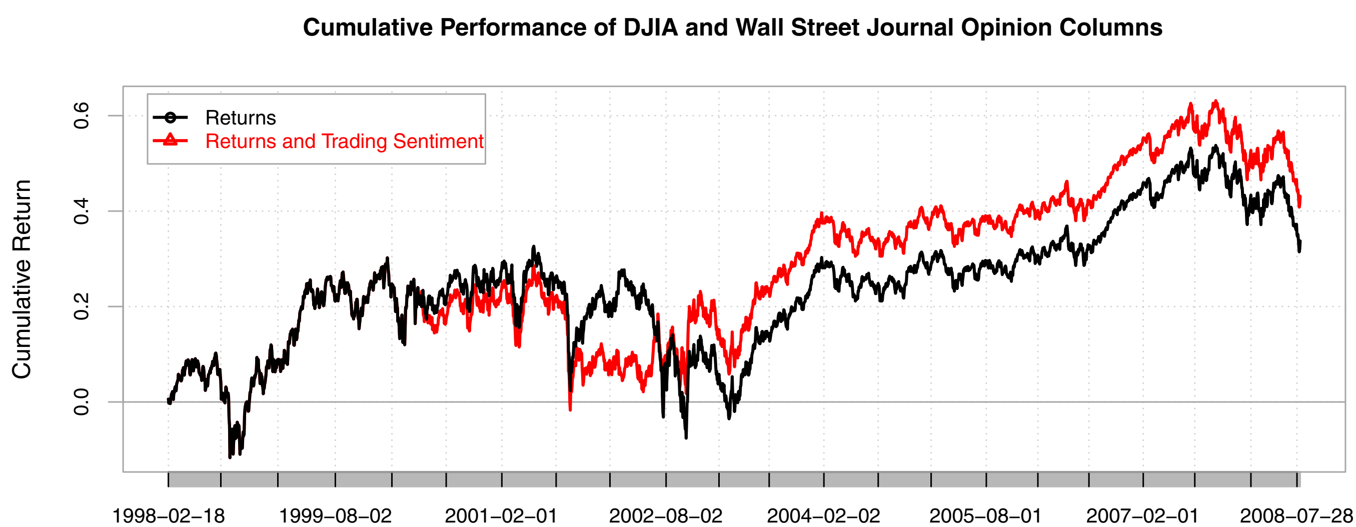

Fig. 4. Rolling VAR (250 day window) model with DJIA returns and negative sentiment for the period from 1998-01/03 to 2008-08-25. The negative sentiment proxy was derived from the Wall Street Journal *Abreast of the Market* opinion column.

Table 5
Performance summary of news trading with the Dow Jones Industrial Average (DJIA) and news from the Financial Times (FT) and Wall Street Journal (WSJ).

	DJIA	DJIA with FT Lex	DJIA	DJIA with WSJ AOTM
Annualised return	3.1%	7.3%	3.5%	4.5%
Sharpe ratio (10^{-2})	16.4	39.0	19.9	25.5
Maximum drawdown	57.3%	52.9%	33.1%	27.3%

Table 6
Performance summary of news trading with West Texas Intermediate (WTI) crude oil futures and oil news from the Financial Times (FT) and Oildrum blog.

	WTI	WTI with FT oil news	WTI	WTI with Oildrum Blogs
Annualised return	−0.7%	8.2%	−7.9%	17.7%
Sharpe ratio (10^{-2})	−2.0	22.7	−18.8	42.2
Maximum drawdown	82.8%	54.8%	82.8%	54.0%

We show the results of a backtest for ten year periods where our text data was available (Figs. 3–6). In each case we see having a basic news trading strategy can make higher returns than a simple buy hold of financial returns. In some examples we include the 2008 financial market crash where high returns are seen by accounting for negative news and speculation. We find that omitting this period and incorporating news into a trading strategy still yields higher returns.

The results presented in Tables 5 and 6 show a summary of news trading strategy versus the buy and hold strategy. The annualised return consists of the yearly average of compounded daily returns for our sample periods. We present the Sharpe Ratio which acts as a risk-adjusted measure of return incorporating variance (standard deviation) as a measure of volatility and risk. A higher Sharpe ratio is considered to have a better risk to return performance. Maximum drawdown is also used as an indication of risk and measures the largest peak-to-trough decline (downward trend) in a series or portfolio. An interactive version of the system has been implemented and hosted as a web application built using the

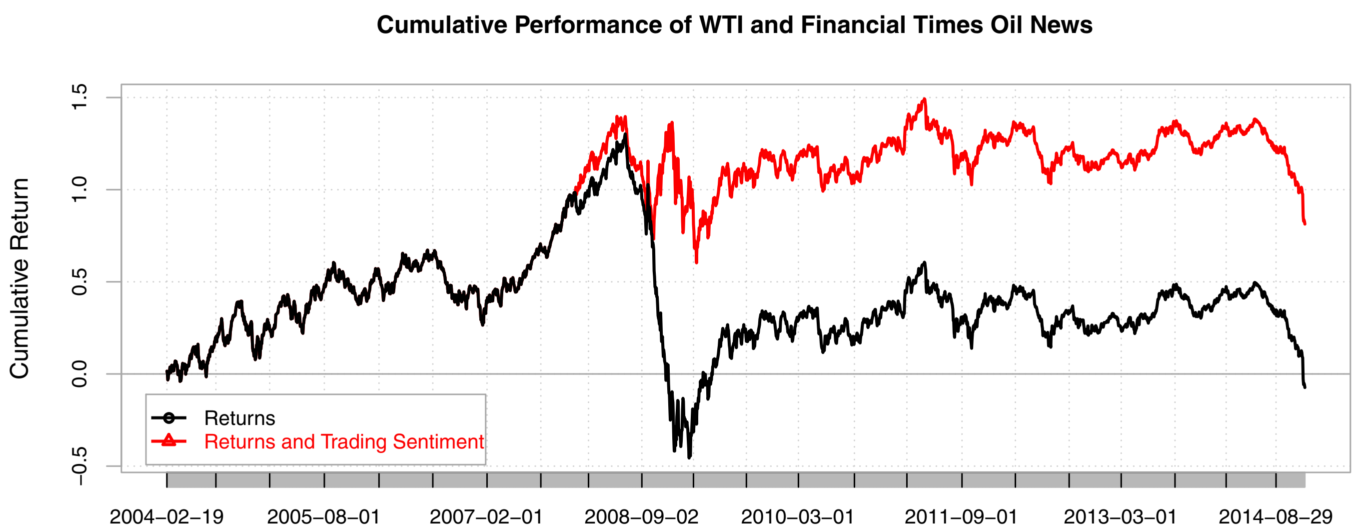

Fig. 5. Rolling VAR (250 day window) model with WTI futures returns and negative sentiment for the period from 2004-01-01 to 2014-12-31. The negative sentiment proxy was derived from crude oil related news published in the Financial Times.

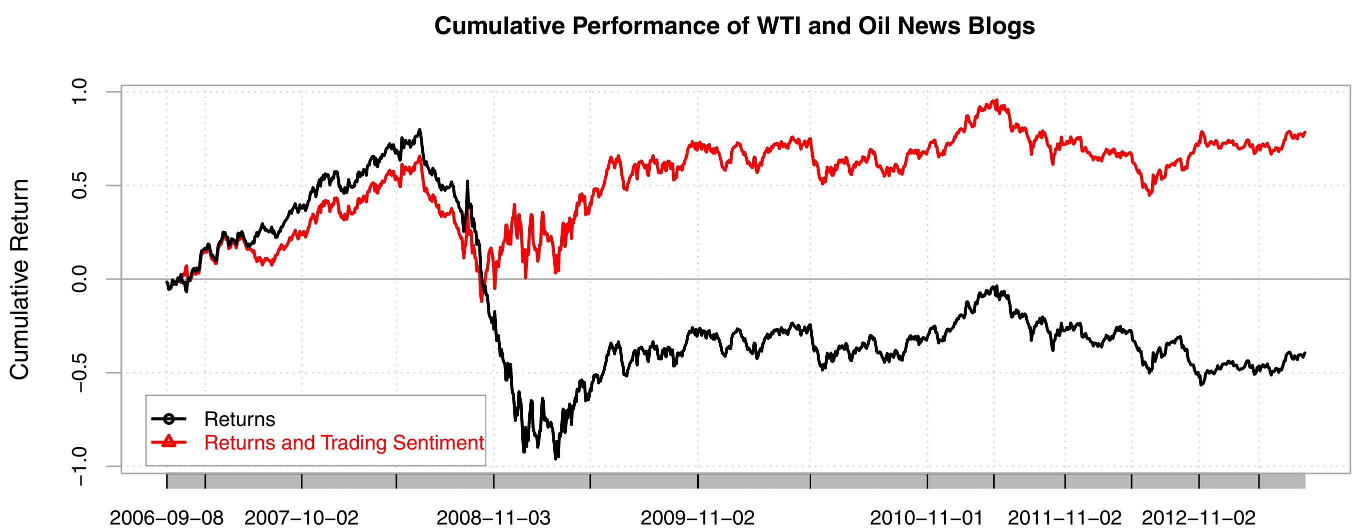

Fig. 6. Rolling VAR (250 day window) model with WTI futures returns and negative sentiment for the period from 2006-06-14 to 2013-09-01. The negative sentiment proxy was derived from news published on the *Oildrum* blog.

R package *Shiny*, a prototype demo is hosted online with limited functionality for demonstration purposes.[4]

For each of our evaluations, and by incorporating sentiment from each of our sources, we see that average annual returns are higher. The performance measures also show an improvement incorporating news versus a simple buy and hold strategy. The Sharpe ratio is higher particularly for the crude oil market and sees a smaller drawdown by accounting for negative sentiment. Overall we demonstrate the benefit of incorporating qualitative information in a quantitative model and use of this strategy as a simple indicator in a larger, more thorough trading system may increase overall returns.

5. Discussion

Generating a reliable variable to represent news sentiment requires the construction of a text corpus with a reasonable justification for the choice of source, article type (such as an opinion column), and topic. The impact that different text types can have on the results and explanatory power of the news media proxies was examined previously by Kelly [13] and has motivated the choice of text here. The consistency of news, its topic of discussion, type, and source can play a role in the performance of the sentiment proxy extracted. News that is on topic and relevant or specific to a domain, such as the opinion columns *Abreast of the Market* from the WSJ and the *Lex* column from FT, which both discuss market level news, or the *Oil & Gas* section of FT, which discusses oil industry news, have a more consistent and statistically significant impact on returns than general widely sampled news. The high readership of these sources also has an influence on whether sentiment extracted will have a significant impact on returns.

Accounting for domain terminology, although gives a more reliable summary of sentiment contained in domain text, has less statistical and economic impact on returns. Using a reliable base dictionary and single words are sufficient to estimate the influence of news sentiment on financial returns. Furthermore, the use of domain terminology will improve results only marginally further lin-

[4] https://stevjk.shinyapps.io/nia_13_aib.

Please cite this article as: S. Kelly, K. Ahmad, Estimating the impact of domain-specific news sentiment on financial assets, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.03.004

guistically motivated investigation would need to be considered as to why this may be the case.

Although several approaches for automatically extracting keywords and terms have been investigated, using a formally constructed terminology has shown to give good results for creating a text feature vector for use in applications such as document categorisation and news impact forecasting [44]. Semantically rich feature vectors such as those produced by the word2vec model have dominated recent literature while it and variations of it will likely become more prominent in future work. Previous investigations by the authors have shown support vector machines and neural network based models to give good results as compared to regression based analysis [44]. The future direction of such forecasting methods will continue to favour deep learning based methods [8].

6. Conclusion and future work

In this paper we have presented an approach and implementation to generate and evaluate news sentiment as a variable for explaining changes in financial returns. We demonstrated that sentiment has a statistically significant impact on returns and follows from similar results first presented in the literature [2,3,42]. Through our two evaluations in the equity and commodity markets, we find that a simple trading system can benefit from incorporating news sentiment into next day trading decisions. We show that a proxy for sentiment extracted from relevant news and text has explanatory power for more than one type of asset.

Our suggestions for future work would include a more thorough investigation into different extensions of the VAR model, in particular looking at the historical variance attributed to the sentiment relative to the control variables using impulse response analysis and variance decomposition. The changing effects of sentiment according to market conditions, in particular volatility, and how these differences in regimes influences the forecasting ability of news sentiment posses an interesting avenue of research to incorporate into the overall method. Lastly, using unsupervised methods of extracting sentiment that can also take into account the changing use of language or sentiment terms would be very beneficial to our proposed approach. A time varying language model such as the one proposed in [45] combined with a statistical estimation accounting for volatility would be a very useful dynamic method of forecasting and estimating the impact of news sentiment over time.

Despite the interpretation of what behaviour or market anomaly is being detected, by using content analysis to generate a proxy for news sentiment, we find that the sentiment variable produced from reliable news sources can help predict changes in financial returns and is economically significant in a basic trading strategy.

Acknowledgements

The authors would like to thank the hospitality of Trinity College Dublin, the School of Computer Science and Statistics, for scholarship support that led to the production of this work.

References

[1] W. Antweiler, M.Z. Frank, Is all that talk just noise? The information content of internet stock message boards, J. Financ. 59 (3) (2004) 1259–1294.
[2] D. Garcia, Sentiment during recessions, J. Financ. 68 (3) (2013) 1267–1300.
[3] P.C. Tetlock, Giving content to investor sentiment: the role of media in the stock market, J. Financ. 62 (3) (2007) 1139–1168.
[4] R.C. Cavalcante, R.C. Brasileiro, V.L. Souza, J.P. Nobrega, A.L. Oliveira, Computational intelligence and financial markets: a survey and future directions, Expert Syst. Appl. 55 (2016) 194–211.
[5] X. Li, H. Xie, L. Chen, J. Wang, X. Deng, News impact on stock price return via sentiment analysis, Knowl. Based Syst. 69 (2014) 14–23.
[6] M. Baker, J. Wurgler, Investor sentiment and the cross-section of stock returns, J. Financ. 61 (4) (2006) 1645–1680.
[7] Z. Da, J. Engelberg, P. Gao, The sum of all fears investor sentiment and asset prices, Rev. Financ. Stud. 28 (1) (2015) 1–32.
[8] F.Z. Xing, E. Cambria, R.E. Welsch, Natural language based financial forecasting: a survey, Artif. Intell. Rev. 50 (2018) 1–25.
[9] S. Narayan, P.K. Narayan, Are oil price news headlines statistically and economically significant for investors? J. Behav. Financ. 18 (3) (2017) 258–270.
[10] P.K. Narayan, D. Bannigidadmath, Does financial news predict stock returns? New evidence from islamic and non-islamic stocks, Pac.-Basin Financ. J. 42 (2017) 24–45.
[11] P.K. Narayan, D.H.B. Phan, S. Narayan, D. Bannigidadmath, Is there a financial news risk premium in Islamic stocks? Pac.-Basin Financ. J. 42 (2017) 158–170.
[12] L.A. Smales, Time-variation in the impact of news sentiment, Int. Rev. Financ. Anal. 37 (2015) 40–50.
[13] S. Kelly, News, Sentiment, and Financial Markets: A Computational System to Evaluate the Influence of Text Sentiment on Financial Assets (2016).
[14] Z. Zhao, S. Kelly, K. Ahmad, Finding sentiment in noise: non-linear relationships between sentiment and financial markets, in: Proceedings of International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2017, pp. 580–591.
[15] D. Shen, X. Li, W. Zhang, Baidu news information flow and return volatility: evidence for the sequential information arrival hypothesis, Econ. Model. 69 (2018) 127–133.
[16] C.-W. J. Chiu, R.D. Harris, E. Stoja, M. Chin, Financial Market Volatility, Macroeconomic Fundamentals and Investor Sentiment (2016).
[17] N. Barberis, A. Shleifer, R. Vishny, A model of investor sentiment, J. Financ. Econ. 49 (3) (1998) 307–343.
[18] S.-L. Chung, C.-H. Hung, C.-Y. Yeh, When does investor sentiment predict stock returns? J. Empir. Financ. 19 (2) (2012) 217–240.
[19] R. Verma, P. Verma, Are survey forecasts of individual and institutional investor sentiments rational? Int. Rev. Financ. Anal. 17 (5) (2008) 1139–1155.
[20] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.
[21] K. Ahmad, S. Kelly, K. Ahmad, Being in text and text in being: notes on representative texts, in: Incorporating Corpora: Multilingual Matters, Clevedon, 2008, pp. 60–91.
[22] S.R. Das, M.Y. Chen, Yahoo! for amazon: sentiment extraction from small talk on the web, Manag. Sci. 53 (9) (2007) 1375–1388.
[23] Y. Zhang, W. Song, D. Shen, W. Zhang, Market reaction to internet news: information diffusion and price pressure, Econ. Model. 56 (2016) 43–49.
[24] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (1) (2011) 1–8.
[25] B. O'Connor, R. Balasubramanyan, B.R. Routledge, N.A. Smith, From tweets to polls: linking text sentiment to public opinion time series, Proceedings of International AAAI Conference on Weblogs and Social Media, ICWSM 11 (2010) 122–129.
[26] Y. Yu, W. Duan, Q. Cao, The impact of social and conventional media on firm equity value: a sentiment analysis approach, Decis. Support Syst. 55 (4) (2013) 919–926.
[27] M. Dzielinski, News Sensitivity and the Cross-section of Stock Returns, Available at SSRN (2011).
[28] L.A. Smales, News sentiment in the gold futures market, J. Bank. Financ. 49 (2014) 275–286.
[29] A. Groß-Klußmann, N. Hautsch, When machines read the news: using automated text analytics to quantify high frequency news-implied market reactions, J. Empir. Financ. 18 (2) (2011) 321–340.
[30] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.
[31] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.
[32] A. Karpathy, L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3128–3137.
[33] Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).
[34] I. Chaturvedi, Y.-S. Ong, I.W. Tsang, R.E. Welsch, E. Cambria, Learning word dependencies in text by means of a deep recurrent belief network, Knowl. Based Syst. 108 (2016) 144–154.
[35] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowl. Based Syst. 89 (2015) 14–46.
[36] X. Ding, Y. Zhang, T. Liu, J. Duan, Deep learning for event-driven stock prediction, in: Proceedings of International Joint Conference on Artificial Intelligence, IJCAI, 2015, pp. 2327–2333.
[37] Y. Deng, F. Bao, Y. Kong, Z. Ren, Q. Dai, Deep direct reinforcement learning for financial signal representation and trading, IEEE Trans. Neural Netw. Learn. Syst. 28 (3) (2017) 653–664.
[38] D.M. Nelson, A.C. Pereira, R.A. de Oliveira, Stock market's price movement prediction with lstm neural networks, in: Proceedings of International Joint Conference on Neural Networks, IJCNN, IEEE, 2017, pp. 1419–1426.
[39] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PLoS One 12 (7) (2017) e0180944.

Please cite this article as: S. Kelly, K. Ahmad, Estimating the impact of domain-specific news sentiment on financial assets, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.03.004

[40] C. Liu, S.C. Hoi, P. Zhao, J. Sun, Online Arima Algorithms for Time Series Prediction (2016).
[41] S. Feuerriegel, D. Neumann, News or noise? How news drives commodity prices, in: Proceedings of the International Conference on Information Systems, ICIS 2013, Milano, Italy, December 15–18, 2013, 2013.
[42] S. Kelly, K. Ahmad, The impact of news media and affect in financial markets, in: Proceedings of International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2015, pp. 535–540.
[43] T. Murphy, S. Kelly, K. Ahmad, Innovations in the Crude Oil Market: Sentiment, Exploration and Production Methods, (2015).
[44] P. Manomaisupat, Automatic Term Extraction and Text Categorisation, University of Surrey (United Kingdom)., 2006 Ph.D. thesis.
[45] R. Bamler, S. Mandt, Dynamic word embeddings, in: Proceedings of International Conference on Machine Learning, 2017, pp. 380–389.