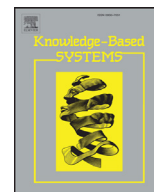




ELSEVIER

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.

© 2018 Elsevier B.V. All rights reserved.

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/© 2018 Elsevier B.V. All rights reserved.

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 Bannigidadmth [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 US 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

Download English Version:

<https://daneshyari.com/en/article/6861464>

Download Persian Version:

<https://daneshyari.com/article/6861464>

[Daneshyari.com](https://daneshyari.com)