



Contents lists available at ScienceDirect

International Review of Financial Analysis



Review

Textual sentiment in finance: A survey of methods and models[☆]Colm Kearney^{a,*}, Sha Liu^b^a Faculty of Business and Economics, Monash University, Sir John Monash Drive, Melbourne, Australia^b Business School, Trinity College Dublin, College Green, Dublin 2, Ireland

ARTICLE INFO

Article history:

Received 17 December 2013

Received in revised form 10 February 2014

Accepted 10 February 2014

Available online xxxx

JEL classification:

D80

D82

G02

G10

G12

G14

G30

G34

G38

M41

Keywords:

Behavioral finance

Textual sentiment

Internet messages

News

Market efficiency

ABSTRACT

We survey the textual sentiment literature, comparing and contrasting the various information sources, content analysis methods, and empirical models that have been used to date. We summarize the important and influential findings about how textual sentiment impacts on individual, firm-level and market-level behavior and performance, and vice versa. We point to what is agreed and what remains controversial. Promising directions for future research are emerging from the availability of more accurate and efficient sentiment measures resulting from increasingly sophisticated textual content analysis coupled with more extensive field-specific dictionaries. This is enabling more wide-ranging studies that use increasingly sophisticated models to help us better understand behavioral finance patterns across individuals, institutions and markets.

© 2014 Elsevier Inc. All rights reserved.

Contents

1.	Introduction	0
2.	Information sources	0
2.1.	Corporation-expressed sentiment	0
2.2.	Media-expressed sentiment	0
2.3.	Internet-expressed sentiment	0
2.4.	Comparing and contrasting the three main sentiment sources	0
3.	The methods of content analysis	0
3.1.	The dictionary-based approach	0
3.2.	Dictionaries and word lists	0
3.3.	Term weighting	0
3.4.	Machine learning	0
3.5.	Comparing and contrasting the textual analysis methods	0
4.	Measuring textual sentiment	0
4.1.	Measures using the dictionary-based approach	0
4.2.	Measures using machine learning	0
5.	Econometric models and hypothesis testing	0
5.1.	Linear regression models	0

[☆] The authors gratefully acknowledge financial support from *Enterprise Ireland* grant number CC-2011-2601-B.

* Corresponding author at: Faculty of Business and Economics, Monash University, Caulfield East, Melbourne, Australia. Tel.: +61 3 99031021.
E-mail addresses: colm.kearney@monash.edu (C. Kearney), lius2@tcd.ie (S. Liu).

5.2.	Vector autoregression models	0
5.3.	Logistic and probit regressions	0
5.4.	Volatility models	0
5.5.	Textual sentiment based trading strategies	0
6.	Main findings to date	0
6.1.	Sentiment and security prices and returns	0
6.2.	Sentiment and other market variables	0
6.3.	Sentiment, firm fundamentals and market efficiency	0
7.	Summary and conclusions	0
	References	0

1. Introduction

Almost half a century ago, Stone, Dunphy, Smith and Ogilvie (1966) described how words and sentences are quintessential human artifacts, the products of social constructs and experiences that provide essential evidence about individual and collective processes such as economic and financial activities and behaviors. They defined content analysis as any technique that enables inference by objectively and systematically identifying specified characteristics within text. By analyzing the raw data of words and sentences, behavioral scientists have become increasingly skilled at uncovering the evidence of sentiment or affect within text. Sentiment is now understood to be articulated in many forms of human discourse, public speeches, news reports, blogs and other forms of written, spoken and visual communication.

In behavioral finance, researchers have in the past decade intensified their efforts to understand how sentiment impacts on individual decision-makers, institutions and markets. Broadly speaking, two types of sentiment have been studied. The first is investor sentiment—beliefs about future cash flows and investment risks that are not justified by the facts at hand (Baker and Wurgler, 2007). A substantial body of this literature focuses on finding and quantifying the effects of investor sentiment on individual stocks and the overall market using various ways to measure investor sentiment. The second type of sentiment is text-based or textual sentiment—the degree of positivity or negativity in texts. In some studies, particularly those using corporate disclosures as the information source, the term ‘tone’ (positive or negative) is used to refer to sentiment. In broad terms, however, textual sentiment may also include affects other than positivity–negativity, such as strong–weak, and active–passive.

The fundamental difference between investor sentiment and textual sentiment is that the former captures the subjective judgments and behavioral characteristics of investors, while the latter can include the former, but also includes the more objective reflection of conditions within firms, institutions and markets. The connection between textual sentiment and investor sentiment is complex, and the extent to which they are causally related has not yet been thoroughly examined or understood. It is also unclear how investors interpret textual sentiment. The existing studies tend not to make assumptions about investor rationality, or about the relationship between textual sentiment and investor behavior. In this sense, they transcend the boundaries between behavioral and traditional finance. The inclusion of qualitative information from textual sentiment into equity asset pricing models, for example, provides another perspective and potentially complementary information to quantitative informational measures in the price formation process. Qualitative information in publicly available documents or media articles may contain additional hard-to-quantify information. Li (2006) suggests that text-based information can potentially provide a more independent test of market efficiency than the number-based measures, because many of the latter are highly correlated so different anomalies may reflect the same empirical regularity. More generally, however, textual sentiment analysis provides an increasingly important approach to address many pivotal questions in behavioral finance.

In this paper, we review the burgeoning literature that uses textual analysis to extract sentiment from sources such as corporate

disclosures, media articles, and internet postings.¹ We describe the alternative content analysis methods including the dictionary-based approaches and the machine learning techniques that are commonly used to generate the sentiment series. Corporate disclosure studies usually aim to discover the fundamental relation between sentiment and future firm performance or other quantitative variables. Media articles and internet posting studies focus on the short-term effects of sentiment on market variables such as stock prices, returns, trading volumes and volatility. Each information source and linguistic method has its unique advantages and disadvantages which influence the research focus and limitations of each study. Because of the use of different types of sentiment and varied research focuses, a wide variety of models and methods has been used to test alternative hypotheses and to derive new findings.

Our paper is structured as follows. Section 2 describes and discusses the qualitative information sources used in the literature. Each of the most popular information sources – public corporate disclosures, news articles and internet messages – is described along with their unique features that are advantageous to others. Section 3 introduces and compares the most frequently used textual analysis methods: the dictionary-based approach and the machine learning approach. Sentiment measures derived from the different linguistic analysis methods are introduced in Section 4. Section 5 presents and reviews the empirical models that have been used in the literature to date, including contemporaneous linear regressions, univariate and multivariate time-series models, logistic regressions and volatility models. The general forms of models are summarized. This section also describes the expert sentiment-based trading strategies that are discussed in the literature. Section 6 summarizes the main findings of the literature to date. Section 7 summarizes the paper, draws together the most important conclusions and suggests future research directions.

2. Information sources

The qualitative information that has been analyzed by textual sentiment researchers in finance comes predominantly from three sources: public corporate disclosures/filings, media articles and internet messages. The sentiment expressed in these texts conveys market participants' and commentators' information and opinions about many aspects of developments in firms, institutions and markets. It also reflects how sentiment responds to these events. The most important work on sentiment analysis within corporate disclosures/filings includes Feldman, Govindaraj, Livnat and Segal (2008), Henry (2008), Henry and Leone (2009), Li (2006, 2010), Davis, Piger and Sedor (2011), Davis and Tama-Sweet (2011), Demers and Vega (2011), Doran, Peterson and Price (2010), Huang, Teoh and Zhang (2011), Loughran and McDonald (2011a, 2011b), Davis, Piger and Sedor (2011), Rogers, Van Buskirk and Zechman (2011), Ferris, Hao and Liao

¹ The studies we review are confined to those examining the role of positive and/or negative sentiment, or ‘uncertain’, ‘deceptive’ affects. Some other studies (e.g. Lehavy, Li and Merkle (2011), De Franco, Hope, Vyas and Zhou (2012) and Loughran and McDonald (2013b)) which investigate the complexity/readability of texts, are not included in this survey.

Download English Version:

<https://daneshyari.com/en/article/5084906>

Download Persian Version:

<https://daneshyari.com/article/5084906>

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