Accounting Forum xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Accounting Forum

journal homepage: www.elsevier.com/locate/accfor

# Stock market returns and the content of annual report narratives

Tomasz Piotr Wisniewski[a,*], Liafisu Sina Yekini[b,1]

[a] University of Leicester, School of Management, Ken Edwards Building, University Road, Leicester LE1 7RH, UK
[b] Coventry University, Coventry Business School, William Morris Building, Gosford Street, Coventry CV1 5DL, UK

ARTICLE INFO

*Article history:*
Received 28 November 2014
Received in revised form 24 August 2015
Accepted 19 September 2015
Available online xxx

*JEL classification:*
M41
G12
G14

*Keywords:*
Content analysis
Annual reports
Stock market returns

ABSTRACT

This paper uses the tools of computational linguistics to analyze the qualitative part of annual reports of UK listed companies. More specifically, the frequency of words associated with different language indicators is used to forecast future stock returns. We find that two of these indicators, capturing 'activity' and 'realism', predict subsequent price increases, even after controlling for a wide range of factors. Elevated values of these two linguistic variables, however, are not symptomatic of exacerbated risk. Consequently, investors are advised to peruse annual report narratives, as they contain valuable information that may not yet have been discounted in the prices.



## 1. Introduction

Annual reports are published to fulfill the reporting requirements imposed on listed companies and can be used by management to communicate with a variety of audiences (Stanton & Stanton, 2002). Each report invariably comprises two components, namely the unaudited narrative and the financial statements. The role that disclosed accounting numbers play in the context of capital markets has already been widely discussed in the extant literature (see for instance Ball & Brown, 1968; Bhandari, 1988; Lev, 1989; Livnat & Zarowin, 1990; Kothari, 2001; Chen & Zhang, 2007). The body of knowledge regarding the influence of qualitative information incorporated in annual reports is smaller, due to initial problems related to the objective quantification of language. Early attempts at using automated systems to analyze narrative accounting disclosures were made in the 1980s (Frazier, Ingram, & Tennyson, 1984) and the technologies as well as the computer software have been progressively developing ever since. The content analysis we employ in our study relies on computing frequencies of words relevant to a specific linguistic construct. Each construct is defined by its own dictionaries including words characterized by similar qualities or referring to the same theme. We subsequently try to evaluate whether these lexical variables forecast future stock market returns.

More specifically, we focus our attention on five master variables constructed from word frequencies by a linguistic software package called Diction. Two of these indicators, capturing 'activity' and 'realism', appear to be both statistically

* Corresponding author. Tel.: +44 116 252 3958.
*E-mail addresses:* t.wisniewski@le.ac.uk (T.P. Wisniewski), sina.yekini@coventry.ac.uk (L.S. Yekini).
[1] Tel: +44 024 7688 8440.

http://dx.doi.org/10.1016/j.accfor.2015.09.001

and economically significant in our models. We argue that companies with realistic plans and strategies and those able to compete actively in the marketplace will be well-regarded by stock market investors. This paper demonstrates that these two linguistic gauges can predict future one-year returns, even after we control for a range of company characteristics and performance indicators. Interestingly, the increased values of these linguistic measures should not be interpreted as risk contributors, as they are negatively related to the overall return standard deviation and the level of idiosyncratic risk. The most plausible interpretation of our results is that markets are to some extent informationally inefficient and that the price reaction to complex news is substantially delayed. To put it differently, we are observing a stock market anomaly that could potentially be exploited by traders.

Arguably, the motivations for drafting the narrative may extend beyond providing new material information. Annual reports can also be used to manage public impression (Neu, Warsame, & Pedwell, 1998) or be deployed as a marketing tool (Stanton & Stanton, 2002). Furthermore, a number of authors allude to the positivity bias frequently inherent in the reports (Hildebrandt & Snyder, 1981; Rutherford, 2005; Henry, 2008). Indeed, listed companies may be tempted to present their state of affairs in a favorable light in order to attract new investors, consumers, and business partners. Regulatory compliance may also be signaled to appease the regulators. Even if such practices are present, our paper shows that careful evaluation of the published narrative is a worthwhile and possibly lucrative exercise. This conclusion lends credence to the incremental information hypothesis outlined in Merkl-Davies and Brennan (2007), which states that managers may use discretionary narrative disclosures to reduce existing information asymmetries.

Our paper adds to the wealth of existing scholarship in several ways. First, we focus on UK data and, in doing so, we are able to make comparisons to the frequently examined US market. This fact may be crucial, as the US has adopted a rules-based approach to reporting, while the UK operates a principle-based system (Nobes & Parker, 2008). Relative to 10-K disclosures, the style of UK annual report narratives may allow managers more flexibility in terms of manipulating the tone and semantic characteristics of text either to influence investors' perception of the company's situation and its fundamental value or to engender an emotional response on their part. This could potentially be exacerbated by the fact that regulations against fraudulent and misleading disclosure in the UK (FRC/ASB/IASB) are less stringent than in the US (SEC, 1998, 2003; Sarbanes-Oxley Act 2002) and that the definition of fraud used by US courts is much broader (Davies, 2007: 44). As a result, disclosure-related litigation is much more common in the United States (Schleicher & Walker, 2010). The differing regulations may potentially affect the usefulness of corporate communications and provide a rationale for our study.

The second motivation for our paper relates to concentration of the previous literature on either the tone of earnings press releases (Henry, 2008; Demers & Vega, 2010; Davis, Piger, & Sedor, 2012), media reportage (Tetlock, 2007; Tetlock, Saar-Tsechansky, & Macskassy, 2008), Internet posts (Antweiler & Frank, 2004; Das & Chen, 2007) or the president's letter to shareholders (McConnell, Haslem, & Gibson, 1986; Swales, 1988; Abrahamson & Amir, 1996; Smith & Taffler, 2000). We analyze the text of the entire annual report, excluding the financial statements section and notes to accounts. These texts are substantially longer than the documents examined previously, which allows for more reliable identification of linguistic style (Grimmer & Steward, 2013: 6). Third, the results of our study may be of interest because, instead of focusing on a market reaction in a very short window surrounding the annual report disclosure date, we endeavor to make medium-term return predictions. Last but not least, we recognize that annual reports are multifaceted documents and therefore we attempt to measure several linguistic dimensions of the text. This approach, as the paper will demonstrate, proves to be insightful.

The remainder of the paper is organized as follows. The next section reviews the literature related to the design and purpose of annual report narratives, engages with studies applying methods of computational linguistics in the field of accounting and finance, and discusses prior applications of Diction software. Section 3 enumerates our data sources, elaborates on variable construction and presents summary statistics for our sample. Section 4 reports our empirical results both on return predictability and on whether the linguistic measures should be regarded as risk proxies. We then endeavor to frame our results within existing theoretical frameworks and end the paper with a set of concluding comments.

## 2. Literature review

### 2.1. Annual report narratives

While the process of drafting the annual report narrative is guided by a pre-existing set of conventions and regulations, companies still retain a large degree of flexibility in terms of content and linguistic characteristics. The design of the text is a complex, purposeful and well-considered process. Thomas (1997) points out that while the letters of the CEO and president may be written by the undersigned, a robust consultation process is typically in place involving lawyers and the chief financial officer. While large segments of a report may be prepared in-house, UK listed companies often resort to using external design agencies (Stanton & Stanton, 2002). Consequently, many departments and individuals may partake in the creation of the text, possibly with a common objective in mind.

The content of annual reports may vary substantially across different companies. However, due to regulatory constraints, these reports tend to share common characteristics. The narrative will typically open with non-statutory and non-audited Chairman's and CEO's statements. The Companies Act 2006 mandate large and medium quoted companies to include a business review section covering a description of company business, its performance, principal risks, position, trends and factors, as well as financial and non-financial key performance indicators (KPIs). It also obliges firms to report on environmental