

# **Computer-based Analysis of the Strategic Content of UK Annual Report Narratives**

Vasiliki Athanasakou (London School of Economics)

Mahmoud El-Haj (Lancaster University)

Paul Rayson (Lancaster University)

Martin Walker (Manchester University)<sup>†</sup>

Steven Young (Lancaster University)

This draft: April 2014

<sup>†</sup> Corresponding author: Room 621 Hankins Building, MBS, Booth Street West, Manchester, UK, M15 6PB, [Martin.Walker@mbs.ac.uk](mailto:Martin.Walker@mbs.ac.uk). We are grateful to the Economic and Social Research Council (Grant Ref: ES/J012394/1) and the Institute of Chartered Accountants of England and Wales for financial support.

## 1. Introduction

This project uses computer-based natural language processing (NLP) methods to analyse the content of annual report narrative disclosures by UK firms listed on the London Stock Exchange. Company law requires UK firms publish an annual report containing audited financial statements together with substantial narrative disclosures on a range of business and governance related topics. Published annual reports represent the most comprehensive systematic disclosure made by UK firms to shareholders and as such attract considerable attention from academic researchers, regulators, and the financial media. Other stakeholder groups including employees, creditors, customers, suppliers, etc. also rely in various degrees on these disclosures.

A substantial body of academic research examines aspects of annual report disclosures. However, two factors restrict the ability of researchers to analyse these important disclosures on a large sample basis. First, contrary to the structured reporting system under which US registrants operate for mandatory annual filings such as 10-Ks, UK firms have considerable discretion regarding both the content of their annual report narratives and the format in which these narratives are presented. For example, while UK company law and stock exchange listing rules mandate certain disclosures, and various codes of best practice provide general guidance on additional information that might be reported and the format should disclosures might take, management retain substantial discretion over disclosure categories, the level of detail provided for each category, and the format and ordering in which the information is presented. In practice, UK annual reports therefore display considerable variation both across firms and over time for a given firm. Second, electronic versions of UK annual reports are normally provided in PDF format, which severely restricts researchers ability to directly access and extract specific categories of narrative disclosure (e.g., the UK equivalent to the US Management Discussion and Analysis section.)

The reporting situation outlined above for the UK is common in most other non-US jurisdictions. As a consequence, researchers wishing to study disclosure policy for non-US registrants or conduct international comparative studies on disclosure policy have been restricted to manual analysis based on small samples of reports (see for example Clatworthy and Jones 2003, Lakhali 2005, O'Sullivan *et al.* 2008), or to examining only a subset firms cross-listed in the US (Bozzolan *et al.* 2009). Moreover, these document structure and accessibility problems mean that recent NLP advances by US accounting and finance researchers cannot be applied in many non-US settings (see Beattie *et al.* 2004). We are currently undertaking a programme of research to develop software solutions that enable

researchers to extract and analyse the content of UK annual report disclosures on a large scale, and to document the variety of practices in UK annual report narratives.

A specific area of disclosure where considerable variation in company reporting practice exists is in relation to the provision of information about the firm's strategy and business model. Such disclosures have become increasingly important in the UK over the last decade. This session will explain the progress we have made in identifying and scoring the strategic content of UK annual report narratives. In addition to outlining the main steps involved analysing UK annual report narratives for a comprehensive sample of UK firms, we will present the specific methods developed to score strategic content and report initial results regarding temporal variation in the use of strategic language, and the determinants of both the level and changes in the level of strategic language.

## **2. Sample and Data**

We obtain and examine annual reports for the financial years 2002 to 2013 issued by the majority of firm listed on the London Stock Exchange. The majority of these reports are provided in digital PDF format. In total we were able to analyse around 9,500 annual reports. We have invested considerable time to gain control over the section headings of the annual reports. A detailed report on this work is available in El-Haj et al (2014). Having control over the section headings is important because this allows the researchers to choose which sections are to be content analysed in relation to their particular research topic.

The 9,500 annual reports contain approximately 190,000 section headings in aggregate. We find significant variation in the number of headings across firms and in the particular names that firms used to describe particular parts of their report. A typical UK annual report contains sections relating to the following generic categories:

1. Highlights
2. Chairman's Statement
3. CEO's Review
4. Finance Director's Review
5. Other Reviews
6. Strategy and Business Model
7. Directors Report
8. Risk Report
9. Corporate Social Responsibility Report
10. Governance and Remuneration

11. Auditors Report
12. Accounting Policies
13. Financial Statements
14. Notes to the Accounts
15. Other Statutory Information

We retain headings for categories 1 to 7 for the purpose of scoring the strategic content of annual report narratives. Applying this filter reduces our original sample of sections from 200,000 down to approximately 60,000. These sections were then analysed for strategic content using the method outlined below.

### 3. Scoring Strategic Content

The starting point for our analysis is a dictionary of 1,156 strategic key words and phrases constructed using the index pages of the following leading textbooks on business strategy: Porter (1985), Barney and Clark (2007), Rumelt (2011), Magretta (2012), and Grant (2013). In addition we used the key words and phrases in a recent survey of the development of the concept of business strategy (Ronda-Pupol, G.A., and Guerras-Martin L.A.N., 2012).

We then use a Java script to count the number of keyword occurrences by sections, which yields keyword scores for  $1,156 \times 60,000$  ( $1,156 \times 9698$ ) sections (annual reports). We pay particular attention to category 6 (Strategy and Business Model) sections, which we refer to hereinafter as strategic sections (SS). We use the contents of these sections to weight the strategic content in the whole of the annual report. Specifically, we calculate the conditional probability of a section being a strategic section given that a particular keyword appears in the section ( $\Pr(SS|keyword)$ ) using Bayes rule:

$$\Pr(SS | keyword) = \frac{\Pr(keyword | SS) \times \Pr(keyword)}{\Pr(SS)}, \quad (1)$$

where

$\Pr(keyword)$  = number of times a keyword appears/Total count of all keywords

$\Pr(SS)$  = total count of all key words in all strategy sections / total count of all keywords in all sections

$\Pr(keyword|SS)$  = number of times a keyword appears in a strategy section / total count of all keywords in all strategy sections

To score an individual annual report we calculate the weighted sum of all the keywords appearing in the annual report. We call this the *Bayesscore* of the strategic content:

$$Bayesscore = \sum_{keyword} (keyword\ count) * (Prob(SS/keyword)) \quad (2)$$

#### 4. Results

The session will provide a review of the empirical findings we have from the econometric analysis of the *Bayesscore*. Initial findings indicate:

1. Significant average growth over time in the *Bayesscore*, roughly doubling since 2002.
2. A high level of persistence in the *Bayesscore*. A panel regression of current *Bayesscore* on prior year *Bayesscore* has an adjusted R square of around 66%
3. Firm size and industry membership both appear to be highly significant determinants of the cross sectional variation in the *Bayesscore*.
4. Firm level changes in the *Bayesscore* over time are significantly related to prior changes reported in accounting performance and stock returns.
5. Firm level changes in the *Bayesscore* over time are significantly related to how the firm's prior year *Bayesscore* compares with that year's industry practice (i.e. the industry *Bayesscore*).

#### 5. Conclusions

This session will report on the success we have had in quantifying the strategic content of UK annual report narratives and modelling how this varies across firms and over time. We will also discuss opportunities for research on other key sections of the annual report.

## References

- Barney J.B. and D.N. Clark. Resource-based theory (2007). Creating and sustaining competitive advantage. Oxford University Press.
- Beattie, V., McInnes, B., Fearnley, S. 2004. A methodology for analyzing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes. *Accounting Forum* 28, 205–236.
- Bozzolan, S., Trombetta, M. and Beretta, S., 2009. Forward looking disclosures, financial verifiability and analysts' forecasts: A study of cross-listed European firms. *European Accounting Review*, 18 (3), 435–473.
- Clatworthy, M. and Jones, M.J., 2003. Financial reporting of good news and bad news: Evidence from accounting narratives. *Accounting and Business Research*, 33 (3), 171–185.
- Grant R.M. (2013). *Contemporary Strategy Analysis*. Wiley (8<sup>th</sup> Edition).
- Lakhal, F., 2005. Voluntary earnings disclosures and corporate governance: evidence from France. *Review of Accounting and Finance*, 4 (3), 64–85.
- Magretta J (2012). *Understanding Michael Porter*. Harvard Business Press Books.
- Mahmoud El-Haj, Paul Rayson, Steven Young, Martin Walker (2014) Detecting Document Structure in a Very Large Corpus of UK Financial Reports. *Language Resources and Evaluation Conference*, 26-31 May 2014
- Porter M. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press.
- Ronda-Pupol, G.A., and Guerras-Martin L.A.N., 2012 Dynamics of the Evolution of the Strategy Concept 1962–2008: A Co-Word Analysis, *Strategic Management Journal*, 33: 162–188.
- Rumelt,R (2011). *Good Strategy Bad Strategy: The Difference and Why it Matters*. Crown Business.