Journal of Accounting and Economics 60 (2015) 110–135

Contents lists available at ScienceDirect

Journal of Accounting and Economics

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

Textual analysis and international financial reporting: Large sample evidence ☆

Mark Lang *, Lorien Stice-Lawrence

The University of North Carolina at Chapel Hill, USA

ARTICLE INFO

Available online 22 October 2015

JEL classification:
G15
M40
M41
M48

Keywords:
International accounting
Textual analysis
Disclosure

ABSTRACT

We examine annual report text for over 15,000 non-US companies from 42 countries over the period 1998–2011, focusing on the length of disclosure, presence of boilerplate, comparability with US and non-US firms, and complexity. We find that textual attributes are predictably associated with regulation and incentives for more transparent disclosure and are correlated with economic outcomes such as liquidity, institutional ownership, and analyst following. Using mandatory IFRS adoption as an exogenous shock, annual report disclosure improved in the sense that quantity of disclosure increased, boilerplate was reduced, and comparability increased relative to both US and non-US firms. Firms with the greatest improvements in financial reporting experienced the greatest improvements in economic outcomes around IFRS adoption.

1. Introduction

A substantial and growing literature examines and compares the characteristics of accounting data, particularly earnings, for firms worldwide. However, much less is known about the associated textual disclosure, owing in part to the difficulty in obtaining full text annual reports and in part to the difficulty in quantifying textual data.

Our goal in this paper is to provide what is, to our knowledge, the first large-sample empirical analysis of annual report textual disclosure for over 15,000 non-U.S. firms from 42 countries over the period 1998–2011. We apply textual analysis to characterize disclosure across a number of dimensions: quantity of disclosure as measured by annual report length, use of boilerplate disclosure as measured by the use of standardized discussion, comparability as measured by the similarity of disclosure across firms within an industry relative to both US and non-US firms, and complexity of disclosure as measured by the Gunning Fog index.

We conduct analyses to understand cross-sectional determinants of disclosure attributes as well as trends over time. We also focus on one specific event, IFRS adoption, using a difference-in-differences design to assess the effects of an exogenous shock on disclosure. We correlate disclosure attributes with economic "outcomes" both in time series and cross-section, as well as around IFRS adoption, to assess how aspects of disclosure are associated with factors such as liquidity, analyst

☆ We thank S.P. Kothari (the editor), Feng Li (the discussant), Mary Barth, Mac Gaulin, K. Ramesh, an anonymous referee, and workshop participants at London Business School, the University of North Carolina, the 2014 Journal of Accounting and Economics Conference, and the Accounting Research Conference in Honor of Nick Dopuch at Washington University for their helpful comments.

* Correspondence to: Kenan-Flagler Business School, The University of North Carolina, Chapel Hill, NC 27516-3490, USA.
E-mail address: Mark_Lang@UNC.edu (M. Lang).

http://dx.doi.org/10.1016/j.jacceco.2015.09.002
0165-4101/

following, and institutional ownership. We focus on a broad set of firms, countries, and years to provide as complete a picture as possible of the characteristics, trends, and economic outcomes associated with textual disclosure.

From a descriptive perspective, we provide evidence of substantial cross-country variation in textual reporting. Perhaps most striking, though, are trends over time. The quantity of textual reporting has increased substantially. In addition, comparability has improved, both among non-US reports as well as between non-US and US reports. Boilerplate increased during the first half of the sample but has decreased in more recent years. Fog has generally increased over the sample period, perhaps reflecting the increase in disclosure of more complex topics (e.g., financial instruments, currencies, taxes, and share-based compensation).

In terms of determinants, and controlling for a variety of other factors, our textual attributes appear to be significantly influenced by regulation and the demand for information. Annual report disclosure is significantly longer for firms applying IFRS or US GAAP and firms which are audited by Big-5 auditors. In addition, disclosure for IFRS and US GAAP firms contains less boilerplate and tends to be more comparable with both US and non-US firms. The analysis for IFRS and US GAAP is robust to inclusion of firm and year fixed effects, suggesting that differences in firm-level economics or time periods do not drive the results.[1]

Next, we correlate our disclosure measures with liquidity, institutional ownership, and analyst following. There is intuitive appeal and theoretical support for the notion that longer annual reports with less boilerplate and greater comparability should be correlated with greater transparency leading to greater liquidity, analyst following, and institutional ownership. However, given that our constructs are, by their very nature, imprecise and there is little existing literature applying them, we want to establish that they do, indeed, correlate with informational variables. For parsimony, we first aggregate our textual variables using factor analysis and then assess whether those aggregated measures are correlated with economic outcomes.

We find that liquidity, institutional ownership, and analyst following tend to be higher for firms with greater lagged quantities of disclosure, less boilerplate, and greater comparability with US firms. While we are hesitant to draw strong conclusions on causality, results are robust to inclusion of firm fixed effects and lagged independent variables, as well as an instrumental variables analysis, suggesting that increases in our annual report variables tend to precede increases in liquidity, institutional ownership, and analyst following.

Next, we investigate one particular innovation in more detail: the mandatory adoption of IFRS. IFRS adoption provides an interesting context for at least two reasons. First, from the firm's perspective, it represents a significant exogenous shock to accounting standards and regulation. One of the difficulties with disclosure research in general is the fact that it suffers from potential endogeneity concerns. However, mandatory IFRS adoption was largely outside of the firm's control. Second, and more importantly, while there is a substantial body of research examining the characteristics of accounting information around IFRS adoption, we know much less about its effect on textual disclosure. Using a difference-in-differences design, we document striking changes in disclosure around IFRS adoption. Relative to our control sample, IFRS adoption increased the "quality" of annual reports in the sense that it increased the amount of disclosure, as well as decreasing repetitive disclosure and increasing similarity with both US and non-US peer firms. It is notable that, while empirical evidence on the impact of IFRS with respect to quantitative data is mixed (see, for example, Barth et al. (2008) and Ahmed et al. (2013)), the effects on textual characteristics are striking, suggesting that the impact of IFRS on textual disclosure may have been at least as pronounced as its effect on measurement.

Finally, we correlate changes in our measures of textual disclosure with economic outcomes around IFRS adoption. We first use mandatory IFRS adoption as well as peer disclosure variables as instruments for firm-level disclosure in an instrumental variables analysis to document that changes in disclosure are associated with economic outcomes. We also examine variation in the extent to which firms benefitted from mandatory IFRS adoption by showing that firms that experienced the largest increases in the quantity and comparability of disclosure also enjoyed greater increases in liquidity, institutional ownership, and analyst following. Firms that experienced increases in fog and boilerplate showed little or no benefit of IFRS adoption.

Our analysis is subject to several caveats. First, our approach and methods are somewhat exploratory because development of textual measures, especially in the international context, is still at an early stage which, in some cases, means that we develop our own metrics. We focus on measures that we believe are likely to be of importance to regulators and investors globally (e.g., quantity of disclosure, comparability, boilerplate, and complexity). While we take some comfort in the fact that our textual measures correlate with determinant and outcome variables in predictable ways, and change notably around IFRS adoption, we recognize that there is necessarily noise in our measurement. That being said, we believe that it is important to explore approaches for quantifying global textual data, appropriately caveated.

Second, we adopt a large sample approach. There are advantages and disadvantages to a broad-brush approach relative to one that focuses on specific samples or selections of words. While our approach inevitably ignores some of the subtleties

[1] We take several steps to ensure that our comparisons are "apples to apples." First, the fact that results are robust to firm and year fixed effects suggests that we are not capturing country-level, firm-level or time period idiosyncrasies. Second, the reports are checked by the data provider to ensure basic consistency. Third, the data analysis software we apply eliminates tabular and other information which appears inconsistent with English textual disclosure. Fourth, we also exclude reports that are US regulatory filings. Finally, we hand check a sample of cases which appear to be unusual in length or content to ensure that these procedures are effective.