

European Accounting Review >

Volume 27, 2018 - [Issue 3](#)

4,208 Views

89 CrossRef citations to date

0 Altmetric

Research Note

# Corporate Risk Disclosure and Audit Fee: A Text Mining Approach

Rong Yang , Yang Yu, Manlu Liu & Kean Wu

Pages 583-594 | Received 01 Feb 2016, Accepted 01 Apr 2017, Published online: 23 May 2017

 Cite this article

 <https://doi.org/10.1080/09638180.2017.1329660>



Sample our  
Economics, Finance,  
Business & Industry Journals  
>> **Sign in here** to start your access  
to the latest two volumes for 14 days

 Full Article

 Figures & data

 References

 Supplemental

 Citations

 Metrics

 Reprints & Permissions

Read this article

Share

## Abstract

The aim of this study is to introduce an innovative text mining approach to assess firms' risks using unstructured textual disclosure from annual reports. Specifically, we use Natural Language Processing techniques to extract firms' self-identified risks including financial, strategic, operational, and hazard risks based on an enterprise risk management framework. We examine the association between these four risk measures derived from the risk factor section in 10-K filings and audit fees. The results show that audit fees are significantly and positively related to firm-specific financial, strategic, and operational risks, indicating the informativeness of corporate textual risk disclosures. This study provides direct support for the recent US reporting regulatory requirement of adding a new section on risk factors in corporate annual reports.

## Acknowledgements

We very much appreciate the insightful comments and suggestions of Laurence van Lent (editor) and two anonymous referees of the *European Accounting Review* that greatly improved this paper. We also thank for Yi Liang, Botao Chen, Alfred Zhu Liu and seminar participants at the 2016 AAA Annual Conference, Mid-Atlantic Region Meeting and CSWIM Meeting. All authors gratefully acknowledge financial support from Saunders College of Business. All remaining errors or omissions are the authors' responsibility.

---

## Notes

1 Unstructured data refer to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. It is typically text-heavy and often includes multimedia content. While some files may have an internal structure, they are still considered unstructured because the data do not fit in a database (Feldman & Sanger, [2007](#)). In contrast, structured data like spreadsheets are easily searchable by basic algorithms. Therefore, it is more difficult to analyze unstructured textual documents than structured ones.

2 The number of observations drops by 10,266 due to the calculation of RESTPROB and ACCR. In untabulated analysis, we re-run our tests excluding these two variables and the results still hold.

3 In the main analysis, we use industry-adjusted risks, measured as the difference between the number of risk sentences from the risk factor section and the industry-median of the sentence count number based on the three-digit SIC industry code in a given year. In the untabulated robustness check, we also use raw risks, measured as the count number of risk sentences from the risk factor section. Our results remain qualitatively unaltered.

4 In addition to the total number of sentences in 10-Ks as a control variable, we recalculate the level and change of raw risk measures deflated by the total number of sentences in 10-Ks and re-estimate the model (1). Our main results still hold.

5 Following Doyle et al. ([2007](#)), for our pre-404 period, we start with firms that had a 404 audit in the initial year of the 404 audit, that is, the fiscal year ending on or after

15 November 2004. Then, we trace back one year to obtain ICMW in year 2003. In untabulated tests, our main results presented in [Table 5](#) remain almost identical no matter we code ICMW using IC weakness under SOX 302 in 2003 or for the entire sample period.

6 For example, one standard deviation of increase in financial risk for a median firm is related to increase in audit fee by \$18,256 ( $\$1,005,504 \times 0.001 \times 18.156$ ).

7 The insignificant coefficient on the hazard risk could be the results of imprecise estimation due to the low variance of independent variables.

---

## Additional information

### Funding

All authors gratefully acknowledge financial support from Saunders College of Business.

---

### Related research

People also read

Recommended articles

Cited by  
89

## Information for

[Authors](#)

[R&D professionals](#)

[Editors](#)

[Librarians](#)

[Societies](#)

## Opportunities

[Reprints and e-prints](#)

[Advertising solutions](#)

[Accelerated publication](#)

[Corporate access solutions](#)

## Open access

[Overview](#)

[Open journals](#)

[Open Select](#)

[Dove Medical Press](#)

[F1000Research](#)

## Help and information

[Help and contact](#)

[Newsroom](#)

[All journals](#)

[Books](#)

## Keep up to date

Register to receive personalised research and resources by email



Sign me up



Copyright © 2026 Informa UK Limited [Privacy policy](#)

[Cookies](#) [Terms & conditions](#) [Accessibility](#)

Registered in England & Wales No. 01072954  
5 Howick Place | London | SW1P 1WG



**Taylor & Francis**  
by informa