Title: SEC Risk Factors: a Structural Topic Model (STM) Approach

Authors: Brandon Trieu and Kelli Duncan

Faculty Sponsor: Dr. Martin Grace, Department of Risk Management & Insurance

Introduction: Since 2005 publically traded corporations in the United States are required by law to include a risk factor analysis in their annual 10-K, a report filed with the Securities and Exchange Commission (SEC). This is also the firm's annual statement and is supposed to provide investors with material information about the company's financial condition and risks.

Purpose: Our project aims to "data mine" the 10-Ks to find out the types of risks discussed in the annual report. We follow the basic idea of Campbell et al. (2014) but focus on having the data provide us with the risk topics in our data using Roberts et al. (2014) Structural Topic Modeling approach. Our results will provide a list of topics and their importance to each company in the sample.

Method: We chose the companies on the 2006 Fortune 100 list and copied their risk factor disclosures from their 10-K filings for every year between 2006 and 2014. We cleaned this data to remove extraneous characters. We have a two stage method. In the first stage, using the R language we analyze the 10-Ks and we obtain statistics regarding the words, length of the document, sentiment (positive and negative), and indices of readability. Secondly, we estimate a Structural Topic Model, which is like a multinomial logit model which predicts the prevalence of words in a document using a set of covariates. Our covariates are the year and firm identification as well as the various statistics we obtained in stage one.

Results: Our expected results are a list of important risk topics and indications of their changes over time. At this point we have finished stage 1 and stage 2 is nearly complete.

Conclusion: Our tentative conclusions are that there are approximately 30 distinct topics in the set of 10Ks such as credit risk, health risk, government regulation risk and the like.

Potential Applications: When completed our results will be able to be used in other models that examine the firm's risk factors and how these factors relate to import risks like corporate governance, cyber risk, and economic crises.

KEY WORDS
SEC 10K, Structural Topic Model (STM), Risk Factors, Sentiment Analysis, Text Mining

Bibliography

Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu, and Logan B. Steele. "The information content of mandatory risk factor disclosures in corporate filings." *Review of Accounting Studies* 19, no. 1 (2014): 396-455.
Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. "stm: R Package for Structural Topic Models." Available at http://cran.r-project.org/web/packages/stm/index.html