

## ***Modelling Loss Given Default with Text Data from 10-Ks***

The analysis of text-as-data is an area of increasing interest in financial research. Some scholars have employed various probabilistic models and deep learning techniques to analyse the vast amount of textual data created from a wide range of sources. The aim of text analysis is to describe the content, structure and function of the information embedded in the text. Early textual studies in the field of finance mainly use simple probabilistic language models to analyse sentiment in the text that are collected from news, social media, SEC filings or earnings calls, and attempt to examine its impact on stock returns, corporate financial performance and risks. Previous research indicates that text information is useful in many tasks in finance, even after controlling for traditional quantitative variables from financial statements. Among various text data sources, the company's annual report (10-K) has proved especially helpful in complementing the understanding of the risk level of the company, since it contains subjective judgments and risk assessments of the management. However, in contrast to other common text data sources such as news and media coverage, annual reports usually consist of sections with very long text, that increase the difficulty in information extraction and the data processing time. Thus, the information contained in annual reports has not been thoroughly explored. In this study, we use the text information in companies' annual reports to estimate their corporate bonds' loss given default (LGD) from multiple perspectives including sentiments, readability, semantics and extracted keywords/topics. Through these analyses, we aim to find signals and reveal potential mechanisms between 10-K textual information and the LGD. In particular, we focus on how these text characteristics change immediately prior to the default and the recovery of the corporate bonds.

Keywords: Loss given default; Natural language processing; Annual report