Machine Learning in Finance: Project Proposal

Joseph Bridges

Spring 2024

1 Introduction

Every year, companies are required by law to provide an in depth look into their internal finances, decision making, and outlook in their 10K statements. EDGARS, the U.S. Security Exchange Commission's Electronic Data Gathering, Analysis, and Retrieval system, provides public electronic access to these annual 10Ks.

In company 10Ks, one can expect to find descriptions of outstanding debt, changes in the debt, and risk factors associated with current debt. Part of the risk assessment of outstanding debt is an assessment of the impact of covenants on potential future performance. Covenants are obligations incurred by the issuing party that restrict potential future actions and can be complicated. It may be hard for the market to price debt that contains substantial covenants. Covenants, generally, help debt holders reduce risk associated with the debt, and arguably allow debt holders to purchase debt with lower yield to maturity. This opens the door for equity investors to systematically overvalue firms with debt that includes substantial covenants, as they may overlook hidden debt costs associated with covenants.

Using modern natural language processing tools such as Hugging Face's transformer API, company 10Ks can be parsed quickly to extract pertinent debt information and KPIs. The 10Ks can be provided en masse via scrapping tools such as Selenium, which bypass the need to query the EDGAR API directly. This project would scrape 10Ks using Selenium, use Hugging Face's python module to extract soft debt covenant information, and then quantitatively assess the extracted information so that a sorting variable can be formed on covenant proportion.

2 Background

The effect of debt covenants on future firm performance has been studied empirically in the past. Oee such paper, "High-Yield Debt Covenants and Their Real Effects" by Bräuning, Ivashina, and Ozdagli (2022), looked at this exact problem. The paper differentiates between traditional loans, which have maintenance covenants requiring borrowers to continuously comply with certain thresholds (e.g., maintaining a leverage ratio below a certain value), and high-yield debt characterized by incurrence covenants. Incurrence covenants restrict some actions of the borrowers if the covenant thresholds are crossed but do not lead to violations of the contracts and associated shifts in control rights. The study found that in the leveraged-loan market, incurrence covenant restrictions have significant real effects, including a drop in the investment rate and a reduction in the debt-to-assets ratio, even before firms default or declare bankruptcy. This suggests a new shock amplification mechanism through contractual restrictions in a highly levered corporate sector [1].

With this in mind, a natural extension of these findings would be to examine if a 'substantial

incurrence covenants' effect persists in market returns. It is the goal of this project to develop an effective measure for examining this effect.

3 Hypothesis

Hypothesis: Firms with a high proportion of high yield debt covenants to market cap will underperform in months when they have already had poor performance in the prior quarter.

The investment story of the hypothesis is this: investors systematically overvalue companies with high debt covenants as they underestimate the hidden cost of debt imposed by those covenants. When a company begins to struggle, equity holders start to worry about cost of debt, as debt costs tend to manifest in times of volatility or when bankruptcy risks are elevated. After investors understand the debt better, they price in debt costs. If we short stocks before investors price in the cost of debt associated with high yield covenants, then we can profit from the resulting dip in share price.

4 Proposed Methodology

3.1 Web Scraping with Selenium

Selenium will be used to navigate the EDGAR database and automate the download of 10-K filings. The process will involve:

- Identifying the target companies and their respective CIK (Central Index Key) numbers.
- Constructing URLs to access the filings archive for each company.
- Parsing the HTML structure to locate and download the 10-K filings.

3.2 Natural Language Processing with Hugging Face

The extracted 10-K filings will be processed using NLP techniques to identify and extract information about high yield debt covenants. The process will involve:

- Preprocessing the text data to remove irrelevant sections and clean the text.
- Utilizing Hugging Face's transformer models with prompts such as "What substantial high yield debt covenants is this company beholden to?" to extract relevant information.
- Analyzing the extracted data to identify key terms and conditions related to high yield debt covenants.

3.3 Quantitative Metric Development

A quantitative metric will be developed to measure the ratio of high debt covenants to market capitalization. The process will involve:

- Calculating the total value of high yield debt covenants extracted from the 10-K filings.
- Obtaining market capitalization data for the corresponding companies.

- Developing a formula to express the ratio of high debt covenants to market capitalization.

3.4 Backtest

Using the metric developed from 3.4, a sorting variable will be developed that sorts firms based on proportion of high yield debt covenants. This can be combined with the previously explored Markov Chain approach of incorporating market timing information to test the hypothesis using portfolios and regressions.

Citations

[1] Falk **Bräuning**, Victoria Ivashina, and Ali Ozdagli "High-Yield Debt Covenants and Their Real Effects" *Federal Reserve Bank of Boston - Research Department Working Papers*, 2022. 22-5.

$$CovD/E = \frac{\text{(Debt (with incurrence covenants) Total)} \cdot \prod Confidence}{\text{Market Cap}}$$