LLM Financial Constraints Index

Overview

This document provides a brief overview of the publicly available **LLM Financial Constraints Index**, an annual firm-level measure of financial constraints constructed from the Management Discussion and Analysis (MD&A, Item 7) section of firms' SEC Form 10-K filings. The dataset covers fiscal years 1997–2022 and is intended for academic research. A complete description of the construction, validation, and implementation of the index is available in the Internet Appendix accompanying our paper, "Debt and Assets."

The index is generated using a large language model (LLM) that classifies firms based solely on the information contained in their MD&A disclosures. Higher values indicate greater financial constraints.

Dataset Contents

The dataset contains one observation per firm-year and includes the following variables:

Variable	Description
gvkey	Compustat firm identifier
year	Calendar year of 10-K filing
index	LLM Financial Constraints Index

The index takes the following values:

Value	Interpretation
1	Financially unconstrained
2	Neutral
3	Financially constrained (including firms classified as either *potentially financially constrained* or *financially constrained* by the LLM)

Construction of the Index

The index is constructed from firms' Item 7 (Management Discussion and Analysis) disclosures in annual 10-K filings obtained from the SEC EDGAR database. Each firm-year disclosure is evaluated using ChatGPT, which classifies the firm's financial constraint status based exclusively on the narrative discussion contained in the filing.

The original four-category classification (financially unconstrained, neutral, potentially financially constrained, and financially constrained) is aggregated into the three-level index released in this dataset.

Recommended Use

The LLM Financial Constraints Index is an **ordinal** measure of financial constraints. Researchers may use the index as:

- an ordinal variable taking values from 1 to 3;
- a binary indicator for financially constrained firms (llm_fc = 3)

Higher values indicate greater financial constraints.

Coverage and Missing Observations

The dataset covers fiscal years 1997–2022. Not all Compustat firm-years are included because an Item 7 (MD&A) disclosure could not be reliably extracted for every filing.

Citation

If you use this dataset, please cite our paper "Debt and Assets" forthcoming at the *Journal of Financial Economics.* The paper describes the construction and validation of the LLM Financial Constraints Index.

Contact

Questions regarding the dataset may be directed to the authors.