Harnessing Machine Learning with News Data for Economic Insights

Research projects using TDM Studio at University of Toronto Rotman's Financial Innovative Hub

Winston (Yuntao) Wu is a MASc student at the department of Electrical and Computer Engineering, University of Toronto. He is also a Research Assistant at Rotman's Financial Innovative Hub (FinHub).

Rotman's Financial Innovative Hub (FinHub)
Rotman's FinHub provides a common forum for students, faculty, startup entrepreneurs and industry to develop expertise to drive ongoing innovation in the financial sector. Our mission: to expose and instruct students in cutting edge technological and finance innovations, to develop new research insights, and to develop and exchange ideas in financial innovation. (Source: Rotman School of Management, University of Toronto)

Introduction

This case study highlights Winston (Yuntao) Wu's collaborative research focused on the use of machine learning to explore the connection between news and market trends. Winston and colleagues at Rotman's Financial Innovative Hub analyzed text data to uncover sentiment and uncertainty, especially around central bank announcements. By understanding these soft indicators, they aimed to predict future market behavior and changes in macroeconomic factors.

Project

The mission of the research collaboration was to analyze how market trends and central bank announcements influence news coverage, and reciprocally, how news impacts future market behavior.

To accomplish the research goal, a substantial corpus of textual data, particularly news articles, was needed. This corpus needed to meet the following criteria:

• Encompass a significant historical timeframe, spanning back to at least the 1980s.

• Grant precise control of publication sources, including renowned outlets such as *The Wall Street Journal, The New York Times, The National Post, The Globe and Mail*, among others.

• Offer granularity in selecting the topics covered in news articles, ranging from monetary policy and central banks to inflation, commodities, and beyond.

• Enable the application of various state-of-the-art natural language processing techniques, facilitating the extraction of soft qualitative information like sentiment and uncertainty.

Research Tools

The team reviewed various research tools for accessing raw news data and discovered that many of them didn't meet the project's requirements. These tools were either constrained by time span, lacked detailed topic granularity, or necessitated purchasing the required data.

However, TDM Studio stood out as it provides daily raw news articles, enabling structured queries and filters such as date range and article types (news, magazine, journal, etc.). Additionally, it offers greater control over selecting relevant news articles, thereby reducing noise in the dataset and improving prediction accuracy. Consequently, the team opted for TDM Studio over other tools and online repositories because it aligned perfectly with their research needs.
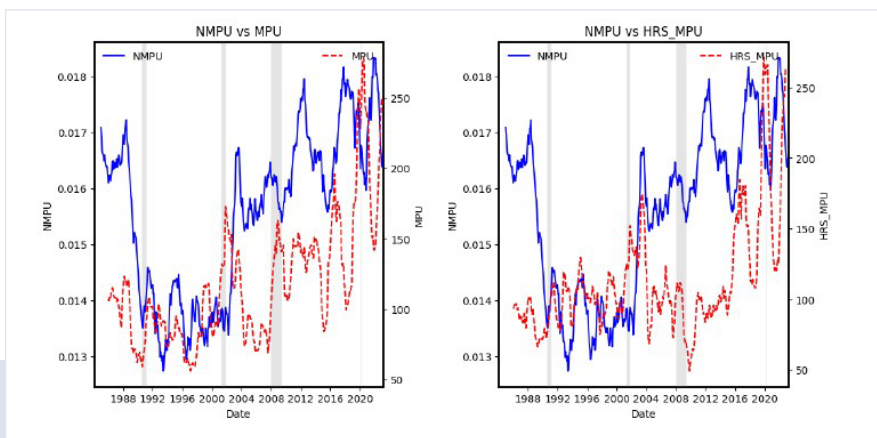
Graphs from final project, "Narrative Monetary Policy Uncertainty"

US NMPU, compared with MPU and HRS MPU: This figure shows the monthly US narrative monetary policy uncertainty indices (NMPU) with monetary policy uncertainty (MPU), Husted-Rogers-Sun MPU (HRS-MPU) from Jan., 1984. All the scores are smoothed by a 12-month moving average. The grey vertical bars are National Bureau of Economic Research (NBER) recessions.

Process

The research with TDM Studio focused on extracting uncertainty from news articles related to central bank announcements and general macroeconomic conditions. Thus, one of the searches used the query:

("monetary policy" or "monetary policies" or "interest rate" or "interest rates" or "Federal fund rate" or "Federal funds rate" or "Fed fund rate" or "Fed funds rate") and ("Federal Reserve" or "the Fed" or "Federal Open Market Committee" or "FOMC")

This resulted in 70k news articles from *The Wall Street Journal* and 21k news articles from *The New York Times* in the time range 1984-2023. The methodology didn't require the articles to contain explicitly the word "uncertain" or "uncertainty," which resulted in broader daily coverage.

Additional searches for news articles with specific topics on commodity, inflation and housing markets provided insights into how topic-based uncertainty interacts with the corresponding macroeconomic variables, such as the volatility of Goldman Sachs Commodity Index, changes in Consumer Price Index (CPI) level, and housing price returns.

The TDM Studio workbench offers analysis through an integrated Jupyter Notebook, a cloud-based environment that is preconfigured with standard data science libraries in R and Python.

 In the TDM Studio workbench and integrated Jupyter Notebook, the documents were cleaned up and the uncertainty score was computed using tools and techniques such as:

- Loughran-McDonald Master Dictionary - a lexicon used in natural language processing to analyze sentiment in documents such as financial news articles.
- BERT (Bidirectional Encoder Representations from Transformers) – a type of natural language processing model useful to understand context in language.

The uncertainty score for each article was calculated as the number of uncertain words divided by the total number of words in the article. The computed daily and monthly uncertainty scores were used to study the media response to future and past central bank announcements and predict market and commodity volatilities.

Results

Results are available in two publicly available papers that are based on the team's research with TDM Studio:

1. Martineau, Charles and Poulos, Zissis and Wu, Yuntao and Thompson, Cameron and Haghighi, Maryam and yuan, Jun and Hull, John C., Narrative Monetary Policy Uncertainty (September 1, 2023). Available at SSRN: https://ssrn.com/abstract=4573829 or http://dx.doi.org/10.2139/ssrn.4573829

Findings revealed heightened uncertainty before central bank announcements, followed by a notable drop post-announcement. However, negative news during announcements lead to increased uncertainty. At lower frequency, the uncertainty measure responded to changes in macroeconomic fundamentals such as unemployment, housing prices, and inflation.

2. Benson, Kevin and Hull, John C. and Nozawa, Yoshio and Poulos, Zissis and Strela, Vasily and Wu, Yuntao, Forecasting Using Text-Based Uncertainty Measures (November 9, 2023). Available at SSRN: https://ssrn.com/abstract=4628516

Findings revealed a strong relationship between narrative uncertainty and volatility in fixed income and commodity markets. Results also demonstrated that narrative uncertainty has predictive power when it comes to forecasting the MOVE index and inflation.

Conclusion

In their research, Winston (Yuntao) Wu and colleagues found that TDM Studio's access to newspapers with fine-grained search provided a large amount of useful text data for analysis with minimal noise. The computation power provided also greatly sped up the research process producing excellent research outcomes.

About TDM Studio

TDM Studio provides dataset creation and text analysis tools that make it possible to efficiently access and analyze over a billion documents across thousands of providers with consistent XML, opening the door to text mining and data science for more researchers.

Learn more at about.proquest.com/en/products-services/TDM-Studio/ or contact your ProQuest representative today.