Thesis Week 1 Assignment

Group 7: Victor Xiao, Zhaochen Jiang, Changle Li

Feb 1st 2025

Potential Topics

Topic 1: Agent-Based Black-Litterman Framework for Multi-Asset Allocation

This thesis proposes an agent-based extension of the Black-Litterman model, where specialized forecasting agents for different asset classes (equities, fixed income, FX, commodities) generate heterogeneous views on expected returns. By aggregating these views and blending them with market equilibrium weights, the framework aims to produce a unified multi-asset allocation that adapts to different market conditions while maintaining robust risk management.

Methodology

- **Data Collection:** Compile historical returns, fundamentals, macro indicators, and sentiment data for different asset classers from WRDS, Bloomberg, and other sources.
- **Agent Construction:**
  - *Equity:* Sector-based fundamental, technical factor models.
  - *Fixed Income:* Yield curve dynamics, inflation signals, credit spreads.
  - *FX:* Interest rate differentials, monetary policy communication, macro sentiment.
  - *Commodity:* Supply-demand metrics, futures term structures, inventory data.
- **Black-Litterman Integration:** Derive market equilibrium weights, incorporate agent "views" on expected returns with confidence levels, and compute the posterior return vector and covariance matrix.
- **Portfolio Optimization & Backteting**

Topic 2: Constructing a Text-Based Indicator to Track and Analyze Chinese Government Policy Signals

This thesis explores the application of natural language processing (NLP) techniques to analyze Chinese government policy documents, particularly Politburo meeting statements. The goal is to construct a quantitative indicator that tracks shifts in policy tone and focus over time, offering a systematic way to assess policy directions and their economic implications.

Research Objectives

The study will have two main objectives:

1. **Developing a Policy Signal Tracker**: Using NLP methods such as topic modeling, sentiment analysis, and word embeddings to extract key themes and sentiment shifts from policy documents.
2. **Analyzing Policy Impact**: Investigating how changes in policy tone correlate with macroeconomic indicators and market responses, using econometric models such as Structural Vector Autoregression (SVAR).

Methodology

The proposed framework includes:

- Collecting and preprocessing Politburo meeting statements and other government policy documents.
- Applying NLP techniques (e.g., Latent Dirichlet Allocation, sentiment scoring, and transformer-based embeddings) to quantify policy stance.
- Constructing an index that tracks policy shifts over time.
- Evaluating the relationship between policy language and economic variables using econometric models.

Topic 3: Constructing a High-Frequency Indicator to Nowcast and Assess Structural Shocks in China's GDP

This thesis aims to construct a comprehensive high-frequency indicator to nowcast and track China's official GDP, evaluating the accuracy and reliability of reported data. Building on the methodology proposed by Chen, Higgins, and Zha (2024), the indicator will incorporate a broader range of macroeconomic variables beyond the existing *China Cyclical Activity Tracker (C-CAT)* to enhance its representativeness.

Research Objectives

The study will have two main objectives:

1. **Nowcasting and Tracking GDP**: Using principal component analysis (PCA) or machine learning techniques, the constructed indicator will be benchmarked against China's official GDP data to assess potential inconsistencies or smoothing issues.
2. **Assessing Structural Shocks**: Applying a Structural Vector Autoregression (SVAR) model, the study will identify and analyze key macroeconomic shocks affecting China's GDP, such as *external demand shocks, policy stimulus, and consumption-constrained shocks* (e.g., COVID-19 lockdowns).

Methodology

The proposed indicator will be developed using a two-stage approach:

- Extracting the principal components from a set of high-frequency macroeconomic variables to construct a cyclical activity tracker.
- Employing an SVAR model to assess the effects of structural shocks on China's GDP dynamics.

Topic 4: Automated Equity Research via LLM-Based Prompt Engineering

This thesis develops an automated equity research pipeline using Large Language Model (LLM) APIs (e.g., ChatGPT or DeepSeek) to process newly released financial statements. By creating robust prompts, the system will generate timely insights on EPS, forward-looking recommendations, and company fundamentals. The focus lies on evaluating accuracy, cost-effectiveness, and feasibility within existing investment workflows.

Research Objectives

1. **Accuracy and Reliability**: Assess how closely LLM-driven analyses align with traditional equity research outputs, particularly regarding financial metrics and investment recommendations.
2. **Cost and Operational Feasibility**: Investigate the financial and technical viability of implementing LLM-based solutions, considering factors such as inference costs, resource allocation, and infrastructure.
3. **Scalability and Usability**: Evaluate how effectively prompt-engineered methods adapt to different market conditions and corporate reporting standards, ensuring broad applicability across sectors.

Methodology

- **Data Collection and Preprocessing**: Gather recent and historical financial statements from multiple publicly listed firms, cleaning and structuring data to ensure reliable input for LLM prompts.
- **Prompt Engineering and Model Integration**: Design targeted prompts to extract key financial indicators (e.g., EPS, revenue growth) and integrate LLM outputs into an investment-focused workflow. Emphasize best practices in prompt design for clarity and relevance.
- **Evaluation Metrics and Benchmarking**: Compare LLM-generated insights against analyst reports and market data, using both quantitative (e.g., accuracy of earnings estimates) and qualitative (e.g., clarity, completeness) metrics.
- **Cost-Benefit and Feasibility Analysis**: Examine API usage costs, resource overheads, and practical hurdles. Propose strategies to optimize prompt usage patterns for long-term scalability in equity research.

By converging prompt engineering, financial statement analysis, and cost-benefit considerations, this research outlines a robust, scalable framework for automated equity research—potentially reducing turnaround times and expanding coverage across a diverse range of firms.

Team Structure for Thesis Project

Shared Responsibilities

- **Literature Review & Research Question Development**: Analyzes existing studies and refines key research questions.
- **Writing & Editing**: Each member reviews others' work and revises drafts together to maintain clarity.
- **Meetings & Decision-Making**: The team meets regularly to discuss progress, address any issues, and make necessary adjustments to project goals.

Individual Focus Areas

Victor Xiao

- **Project Management & Coordination**: Manages the project timeline, arranges meetings, and ensures tasks are completed on schedule.
- **Communication & Reporting**: Serves as the main point of contact with professors and discussants, providing updates and incorporating feedback.
- **Methodology Design & Data Collection**: Plans the overall research approach, selects appropriate methods, and oversees data-gathering tasks.

Zhaochen Jiang

- **Data Acquisition & Preprocessing**: Finds relevant data sources, cleans and organizes datasets, and prepares them for later analysis.
- **Analysis & Interpretation**: Helps verify analytical outputs and interpret results in a financial or econometric setting.
- **Data Analysis & Visualization**: Handles coding duties, applies modeling tools or APIs, and creates charts or dashboards to present findings.

Changle Li

- **Output Review**: Checks the accuracy of generated outputs.
- **Results Interpretation & Documentation**: Summarizes findings for presentations and final reporting, ensuring technical details are clear and well-documented.

Discussant Team

Group 8: Haining Han, Angelina Bu, Tao Luo