MATH 279 Project Proposal: Geopolitical Alpha via Sparse Bipartite Networks – A HAR-Lasso Approach to Commodity-Equity Lead-Lag

Ivan Sit

University of California, Los Angeles

Supervisor: Mihai Cucuringu
January 27, 2026

up to 2-3 collaborators

1 Introduction

Asset prices in the energy sector are theoretically driven by the underlying commodities (oil, natural gas). However, during geopolitical shocks, the transmission mechanism from commodity markets to equity markets is often delayed or non-linear. Standard correlation analyses fail to capture the multi-scale nature of this relationship (e.g., does Exxon react to daily oil moves or monthly trends?).

This project proposes a **Sparse Bipartite Network Model** to map the lead-lag structure between Commodities and Energy Stocks. By integrating a **Heterogeneous Autoregressive (HAR)** feature set with a **Lasso Regression** framework, we aim to construct a dynamic graph where edges represent statistically significant predictive links. We utilize the lead-lag detection frameworks discussed in Cartea et al. (2023) to identify when these links generate actionable "Geopolitical Alpha."

2 Methodology

Data Source (CRSP & Commodities). We will utilize daily closing price and volume data from **CRSP** for S&P 500 Energy constituents (Set S) and major commodity futures (Set C: Crude Oil, Brent, Natural Gas, Heating Oil) over the period 2010–2025.

Feature Construction: The HAR Framework. Following the Heterogeneous Autoregressive (HAR) model structure (Corsi, 2009), we decompose commodity price information into three distinct time horizons to capture short-term shocks and long-term trends. For each commodity $c \in C$, we construct the vector X_c:

$$X_{c,t} = [r^{(d)}_{c,t-1}, r^{(w)}_{c,t-1}, r^{(m)}_{c,t-1}]$$

where $r^{(d)}$ is the daily return, $r^{(w)}$ is the weekly average return (last 5 days), and $r^{(m)}$ is the monthly average return (last 22 days).

Network Discovery via Lasso Regression. We model the relationship between the set of commodities and the set of stocks as a **Bipartite Graph** where edges represent significant predictive power. Since the number of potential predictors is large, we employ **Lasso Regression (L_1 regularization)** to enforce sparsity (Tibshirani, 1996).

For each stock $i \in S$, we solve the following optimization problem:

$$\min_{\beta_i} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \sum_{j \in C} \sum_{k \in \{d,w,m\}} \beta_{i,j}^{(k)} X_{j,t}^{(k)} \right)^2 + \lambda ||\beta_i||_1 \right\}$$

where λ is the regularization parameter.

- **Graph Construction:** We draw a directed edge from Commodity Node j to Stock Node i if and only if $\beta_{i,j} \neq 0$.
- **Interpretation:** This allows us to empirically determine if a stock like *Chevron* is driven more by *Daily Oil Shocks* or *Monthly Natural Gas Trends*.

3 Exploratory Extension: Daily Order Flow Imbalance

While this project primarily utilizes daily CRSP data, we aim to incorporate insights from market microstructure, specifically the **Order Flow Imbalance (OFI)** metrics detailed in Cont et al. (2023). Since high-frequency LOBSTER data is outside the current scope, we will experiment with a **Daily Imbalance Proxy** to capture the "net buying pressure" concept:

$$OFI_{i,t}^{daily} \approx \text{Volume}_{i,t} \times \frac{Close_{i,t} - Open_{i,t}}{High_{i,t} - Low_{i,t}}$$

We will test if including this proxy as an additional feature in our Lasso regression improves the detection of lead-lag relationships during high-volatility regimes.

4 Case Study: The 2026 Venezuela Regime Shift

We will apply this Bipartite Lasso framework to the **Venezuela Regime Shift (Jan 2026)**. We hypothesize that:

1. The network topology will shift drastically: the number of active edges (non-zero β) connected to heavy-crude proxies will spike.
2. The Lasso model will select "Daily" features over "Monthly" features during the shock, indicating a shift to high-frequency information processing.

5 Expected Deliverables

1. **Network Visualization:** A bipartite graph visualization showing which commodities drive which stocks.
2. **Strategy Backtest:** A Long-Short strategy based on the residuals $\epsilon_{i,t} = r_{i,t} - \hat{r}_{i,t}$, evaluated via Sharpe Ratio.
3. **Stability Analysis:** A study of how the Lasso λ parameter affects the stability of the detected edges.

References

Cartea, Á., Cucuringu, M., & Jin, Q. (2023). Detecting Lead-Lag Relationships in Stock Returns and Portfolio Strategies. *SSRN*.

Cont, R., Cucuringu, M., & Zhang, C. (2023). Cross-impact of order flow imbalance in equity markets. *Quantitative Finance*.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.