

STAT 5244 – Unsupervised Learning

Homework 3

Name: Chuyang Su UNI: cs4570

1 Graphical Models.

1.1 Data Processing.

The log return transformation was applied to the daily closing prices. The daily log return r_t for a stock price P_t was calculated as:

$$r_t = \ln(P_t) - \ln(P_{t-1})$$

This dataset of log returns, spanning 1,228 trading days, was used for all subsequent graphical model fitting.

1.1.1 Descriptive Statistics

The table below summarizes the descriptive statistics for the daily log returns.

Table 1: Descriptive Statistics of Daily Log Returns (Jan 2021 – Present)

| | AAPL | AMZN | BAC | CVX | GOOGL | JNJ | JPM | KO | META | MSFT | NVDA | PFE | PG | WMT | XOM |
|---------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------------|---------|--------------|---------|---------|---------|---------|
| Count | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 | 1228.00 |
| Mean ($\times 10^{-3}$) | 0.63 | 0.27 | 0.53 | 0.64 | 1.02 | 0.33 | 0.81 | 0.38 | 0.65 | 0.66 | 2.13 | -0.12 | 0.18 | 0.68 | 1.01 |
| Std ($\times 10^{-2}$) | 1.76 | 2.23 | 1.72 | 1.60 | 1.96 | 1.05 | 1.53 | 1.00 | 2.78 | 1.63 | 3.29 | 1.59 | 1.09 | 1.32 | 1.71 |
| Min | -0.097 | -0.151 | -0.117 | -0.086 | -0.100 | -0.079 | -0.078 | -0.072 | -0.306 | -0.080 | -0.186 | -0.070 | -0.064 | -0.121 | -0.082 |
| Max | 0.143 | 0.127 | 0.081 | 0.085 | 0.097 | 0.060 | 0.109 | 0.046 | 0.209 | 0.097 | 0.218 | 0.103 | 0.042 | 0.091 | 0.062 |

The data clearly demonstrates the risk-return trade-off. The semiconductor stock **NVDA** shows the highest average daily return ($\sim 0.213\%$) but also the highest volatility (Standard Deviation: 3.29%) and largest maximum single-day return ($\sim 21.8\%$). Conversely, consumer staples stocks like **KO** (Coca-Cola) and **PG** (P&G) exhibit the lowest standard deviations ($\sim 1.0\%$), indicating high stability but lower returns. The largest single-day drop belongs to **META** (former FB) at -30.6% .

1.1.2 Time-Series Exploration

The cumulative returns plot (Figure 1) illustrates the differential performance across sectors over the analysis period.

1.1.3 Correlation Analysis

The correlation heatmap (Figure 2) reveals strong clustering of dependence among stocks within the same sector, which confirms the pervasive influence of systematic market risk. Additionally, Figure 3 shows the within-sector correlation analysis, where stocks in the same industry (e.g., Tech, Energy) exhibit notably higher correlations than cross-sector pairs.

Key Observations from the Heatmap:

- **Strong Correlation (0.6+):** High-tech stocks (AAPL, MSFT, AMZN, GOOGL, NVDA) are tightly coupled (e.g., MSFT-AMZN at 0.66, MSFT-GOOGL at 0.65). Financials (JPM-BAC at 0.82) and Energy stocks (CVX-XOM at 0.86) exhibit the highest correlations, reflecting their singular dependence on industry-specific factors (e.g., oil price, interest rates).
- **Weak/Low Correlation (0.0-0.3):** Healthcare stocks (JNJ, PFE) show low correlation with most other stocks (e.g., JNJ vs. Tech stocks often below 0.2), confirming their defensive, counter-cyclical nature.
- **Negative Correlation:** A notable weak negative correlation exists between the pharmaceutical stock JNJ and the high-growth technology stock NVDA (~ -0.09), suggesting an interesting divergence in their underlying risk drivers.

This preliminary analysis confirms the existence of strong, sector-specific dependencies, which the Graphical Lasso will aim to distill into a network of conditional dependencies.

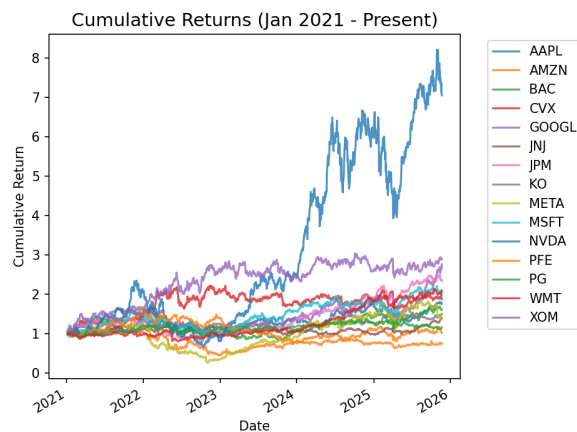


Figure 1: Cumulative Log Returns of Selected Stocks (Jan 2021 - Present)

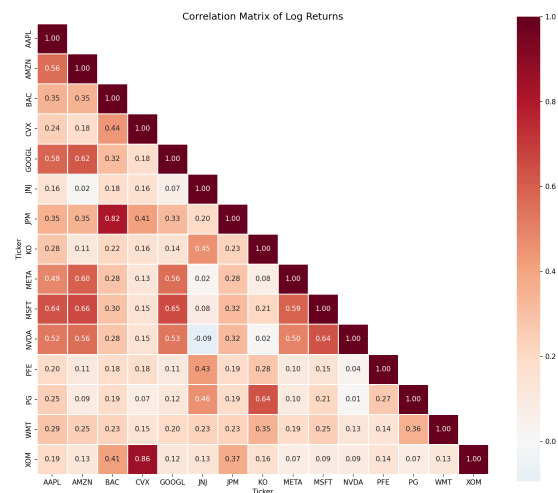


Figure 2: Correlation Heatmap of Daily Log Returns

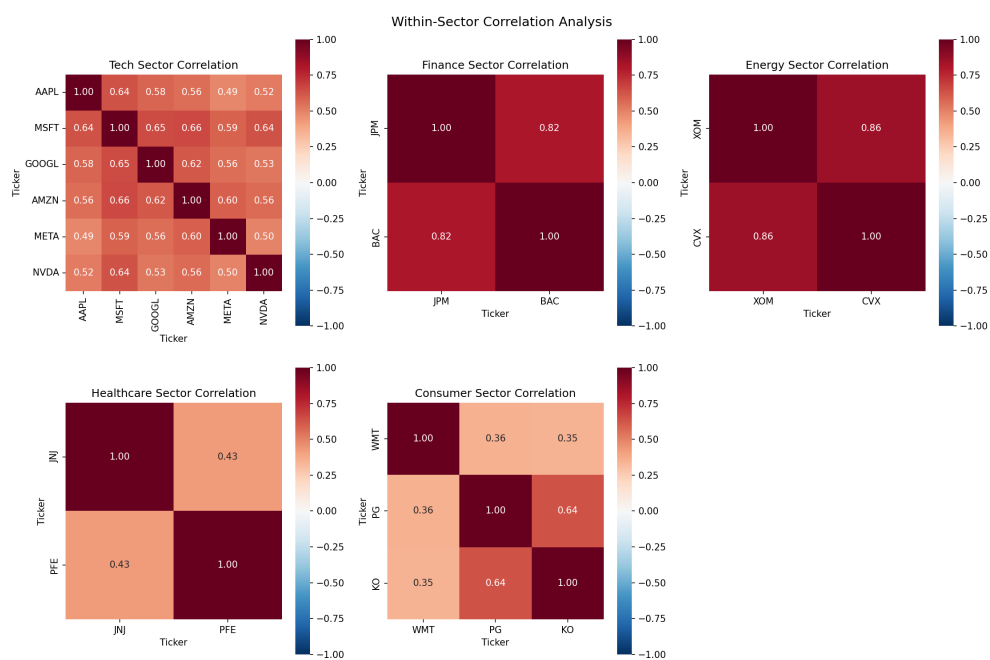


Figure 3: Sector Internal and Cross-Sector Correlation Overview