S&P 500 Analysis Final Project

Presented by Group 8 - Maanav (O) Amy (O) Winta (D) Shane (C) Taku (I) Walt (I)





= Introduction

Who thinks a disruption in GPU supply can ripple through the entire tech industry?







When a certain sector has a change in mean stock price, how does that influence or predict the mean value of another sector's stocks?

Why?

Many stock analyses focus on performances of individual stocks. However, for the average investor, diversification is key. People like us should invest broadly in index funds that cover multiple sectors. Thus, it's more important to focus on sector interactions at a macro scale.





\equiv Data Description

Type: Working with a time series of sector indices.

Key definitions:

- Sector a category that groups companies with similar business activities (e.g., Technology, Healthcare, Financial Services). Total of 11 sectors.
- Adjusted close prices stock's price at day's end, adjusted for actions like dividends, inflation, and splits to better reflect performance
- Mean sector prices mean of the adjusted close prices of all stocks within a given sector on a particular day.

Period: Data from 2010-2024

Data sources: YahooFinance, through a Kaggle dataset provided by LARXEL, and our own web scraper.







\equiv Methods

- Initial data cleaning/prep:
 - Original columns: Date, Symbol, Sector, Adjusted Close
 - Performed a time series transformation of the data to quantify the relationships between sector returns.
 - Summarize raw prices into sectors
 - Spread each entry into a row (date) with columns of sectors and values of mean returns on that date.
 - Final columns: **Date, ...Health, Technology, Utilities, etc...**

Main method:

 Linear modeling on time series data to measure the relationships between sector returns.







What:

Linear regression with a penalty that increases with the absolute value of all coefficients. The penalty encourages coefficients to shrink to zero.

Why:

LASSO is valuable for datasets with high multicollinearity (variables are correlated to each other) which is the case with stock sectors.

Our initial 11 models:

Response (y): Mean sector price of a given sector Predictors (X): Date + mean sector price of other sectors



Attempt1

We will be using the technology sector just as a comparable example of the changes in model fit throughout the process.

- Overall fit looks decent, but residual plots are far more important in checking fit.
- The residual plot displays striping by year our data is still too dependent on time and needs to be adjusted or split.





Attempt 2

- Used daily returns (% change compared to yesterday) as response to remove time as predictor.
- Used non-log response for better interpretability.





8

Results

-NOT a correlation matrix -Just uses corr plot for easier interpretation

How to Read:

Row: Response (Each row is a model predicting it) **Columns: Predictors** Cell: Coefficient in that model Black: Coefficient of O

More symmetry in the graph = more equal influence between the pairs

Mostly positive coefficients A good economy helps everyone - sectors are not competitors

Basic Materials	
	Basic Materials
-0.03	mmunication Services
0.12	Consumer Cyclical
0.05	Consumer Defensive
0.23	Energy
0.05	Financial Services
0.10	Healthcare
0.20	Industrials
0.03	Real Estate
0.05	Technology
-0.02	Utilities

Commur

Communication Services	Consumer Cyclical	Consumer Defensive	Energy	Financial Services	Healthcare	Industrials	Real Estate	Technology	Utilities	
	0.19	0.09	0.09	0.08	0.11	0.47	0.02	0.04		0.51
	0.34	0.13	-0.02	0.03	0.12		0.05	0.39	-0.13	- 0.45
0.15		0.10	-0.02	0.10	0.05	0.29	0.08	0.13	-0.03	- 0.38
0.05	0.10			0.07	0.11	0.02	0.03		0.28	- 0.32
0.03	-0.09	-0.03		0.51	-0.11	0.45	-0.12	0.02	0.13	- 0.25
0.01	0.09	0.08	0.11		0.09	0.43	0.05	0.05	0.06	- 0.19
0.07	0.07	0.16	-0.03	0.13		0.13	0.10	0.19		- 0.13
	0.20	0.02	0.08	0.32	0.07			0.09	0.02	- 0.06
0.04	0.15	0.06	-0.05	0.09	0.14			0.12	0.46	- 0
0.27	0.20	-0.02	0.01	0.09	0.23	0.23	0.11		-0.12	- 0.07
0.09	-0.06	0.48	0.06	0.11	0.00	0.05	0.41	-0.12		-0.07
										-0.13

9

\equiv Conclusion

Using engineered time series data and LASSO regression, we identified potential relationships between sectors' mean prices and whether those relationships were mutually influential.

Understanding broad market dynamics is critical for regular investors who are seeking to diversify, not invest in particular stocks. It is also easier to translate into daily life, as you can intuitively relate world events to sector performance.

None of this is actual financial advice though. Just invest in the SP500 and don't touch it until you







Thank You



