

Stock Trend Evolution

Taylor Agarwal
Henk Quelle
Cooper Ryan

Project Description

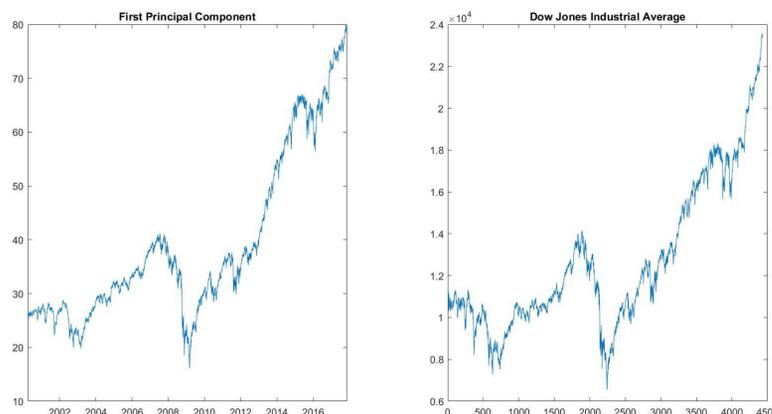
- We constructed a **novel tool** to form stock groupings based on the weights garnered from Principal Component Analysis
- Papers showed how to use **PCA** and how it relates to **stocks** [1] but none really had the same idea of using it to group them into **portfolios**
- We made a programmable tool that knows which parts of the **market** move together
- This information would allow for better investments and an increased level of diverse investing
- We were able to successfully group seemingly unrelated **stocks** that had almost identical fluctuations based on the found **eigenstocks**
- We were able to successfully model the Dow Jones Industrial Average, which demonstrates the validity of our approach
- We were able to successfully observe the effects of **catastrophes** on grouped stocks

Scientific Challenges

- Computing power of MATLAB does not allow for large enough matrices to group the whole **market**, had to develop a workaround for this
- There was not a lot of specific information on people doing this before
- Cleaning the **stock** data for use in the code
- Locating a sufficiently comprehensive data set

Potential Applications

- Grouping stocks based on their highest weighted **eigenstock** is a new approach to the portfolio method
 - The groupings based off their highest **eigenstock** weight will in theory give us **stocks** that vary in nearly the same manner as the original set of **stocks**
- Knowing which **stocks** vary together allow investors to monitor their **stock** investments and predict upcoming changes to **stock** prices
- Investors can use this to diversify investments into groups with dissimilar **eigenstocks** to decrease risk of financial losses



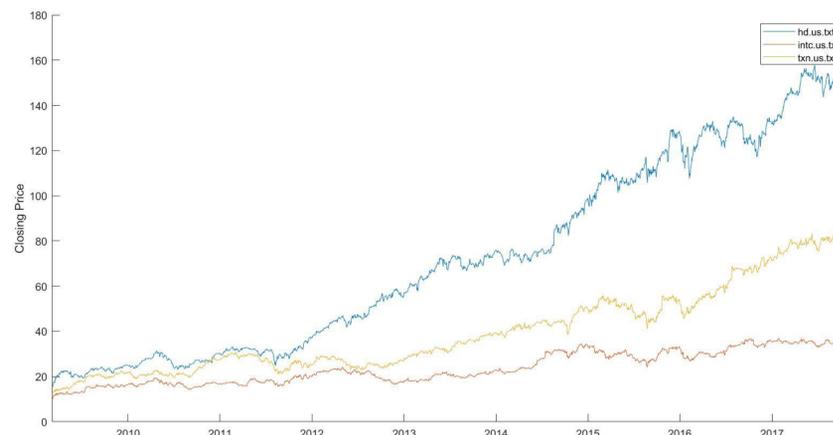
A side by side view of the first Principal Component found in PCA and the Dow Jones industrial average.

Methodology

- We located a data set containing 7195 **stocks** dating from 1960 to 2017, roughly 10^8 data points with opening and closing prices for the day [5]
- **PCA** consists of approximating a data matrix A with a matrix of eigenvectors V that account for a large portion of the variance in the data, then finding W and μ such that $A \approx VW + \mu$
- Using subsets of the stock data and **PCA** [3] in MATLAB, we identified **eigenstocks** that account for variance in the subset
- We then formed **portfolios** from the subset based on the maximum projection weight for each **stock** after projecting onto the **eigenstocks**
- In one analysis, we aimed to explore how the **eigenstocks** were related to the overall trends of the **stock market**
- In another analysis, we aimed to explore the formed **portfolios** to see interesting relationships between seemingly unrelated stocks
- In our last analysis, we used the subset as a training set, then projected a **stock** outside the subset that experienced a **catastrophe** onto the **eigenstocks**, and observed it in the **portfolio** of its nearest neighbor to see if they experienced similar declines in value

Results

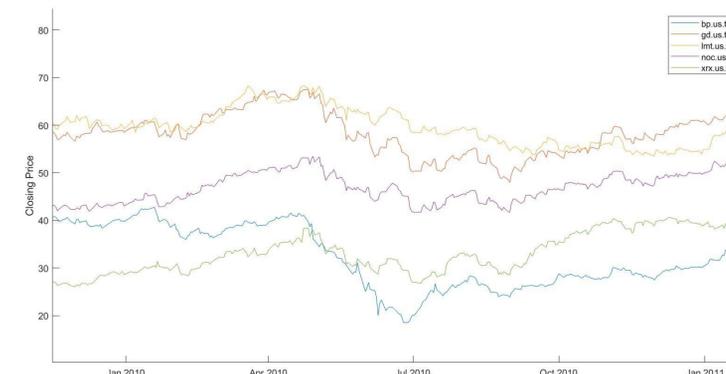
- Our tool successfully forms groupings of **stocks** based on any subsection of the **market**, while following user-defined timeframes and **variance** thresholds
- Using 100 Fortune 500 [2] **stocks** we were able to recreate an almost identical model of the Dow Jones using the first principal component
- We found that despite seemingly isolated incidents, such as the BP oil spill, other **stocks** such as Goodyear and Xerox trended with BP
- Our portfolios found dozens of seemingly unrelated **stocks** with similarly timed increases and decreases in value



It is evident that Home Depot, Intel, and Texas Instruments have different sizes but have similar rises and falls when referencing this specific example.

Glossary of Technical Terms

- Stock:** A fraction of ownership in a particular company that can be purchased and sold at ever-changing prices
- Market:** Where stocks are traded and sold. Here, this refers to the NYSE and the NASDAQ
- PCA:** Principal Component Analysis
- Eigenstock:** A created stock whose movements cover a section of the entire market's variance
- Portfolio:** A grouping of stocks designed by financial advisors
- Variance:** A statistical measure of how much a particular data set differs from its mean
- Covariance:** A statistical measure of how two data sets are related to each other
- Catastrophe:** An event that had major consequences on the price of a particular stock



A zoomed in look at the period of time around the BP Oil Spill. Notice that the grouped stocks all fall at the same time

References

1. Giorgia Pasini, Principal Component Analysis For Stock Portfolio Management, International Journal of Pure and Applied Mathematics, Volume 115, 153 – 168 (2017).
2. Fortune Media. "FORTUNE." Fortune, Fortune, 3 Apr. 2020
3. Sheng Zhang and Matthew Turk (2008) Eigenfaces. Scholarpedia, 3(9):4244.
4. The Ocean Portal Team. "Gulf Oil Spill." *Smithsonian Ocean*, Apr. 2018
5. Marjanovic, Boris. Huge Stock Market Dataset. Kaggle, 2017

Acknowledgments

This project was mentored by Dr. Melinda Lanius, whose help and guidance is greatly appreciated