Portfolio Diversification Based on Clustering Analysis

Young Sim, Meng Lu, Kit Qian, Shanshan Ding, Jenny Gray

“The objective of the project is to optimize stock **portfolio diversification** in order to reduce volatility and increase capital preservation by using clustering methods”

Data Source

- Comprised of the daily close price of the 470 common stocks in S&P 500 in the 5-year data
- Date range: Feb 7, 2013 - Feb 6, 2018
- Split into training period followed by test period
  - Training Period 1: Feb 2013 – Feb 2017
  - Training Period 2: Feb 2016 – Feb 2017
  - Test Period: Feb 2017 – Feb 2018

Sharpe Ratio

- Measures how much a portfolio outperforms the risk-free rate of return on a risk-adjusted basis.
- Formula: excess return divided by standard deviation
- Assumed zero risk-free rates in all cases of the study for simplicity
- Log Return = $\log(1+R_i)$ is used for normalization purposes

Clustering

K-MEANS

- Grouping by distance to the nearest cluster center
- Assume same density and equal weights on all directions

AGGLOMERATIVE

- Start by N (number of data points) clusters
- Similar clusters are merged until the specified clusters are left
- Reflects hierarchy via dendrograms

DBSCAN

- Consider density of data points
- No need to specify number of clusters
- Identify noise points

Algorithm Selection

K-MEANS

Only Euclidean distance of the features is allowed in Scikit-learn

AGGLOMERATIVE

Precomputed Correlation-based distance can be used to measure the distance

DBSCAN

Clusters are of very different sizes (no freedom of cluster numbers)

Algorithm Selection

N_Cluster

Set to 30 to match the number of stocks in DJI as benchmark for portfolio evaluation

Linkage

- Single
- Complete
- Average

Portfolio Construction & Evaluation

DIVERSIFIED PORTFOLIO

Selecting the stock with highest Sharpe ratio from each cluster with equal weight

1

Sharpe ratio

2

Annualized portfolio volatility

Benchmarks

Sharpe Ratio

- Dow Jones Index
- 30 randomly selected stocks

Volatility

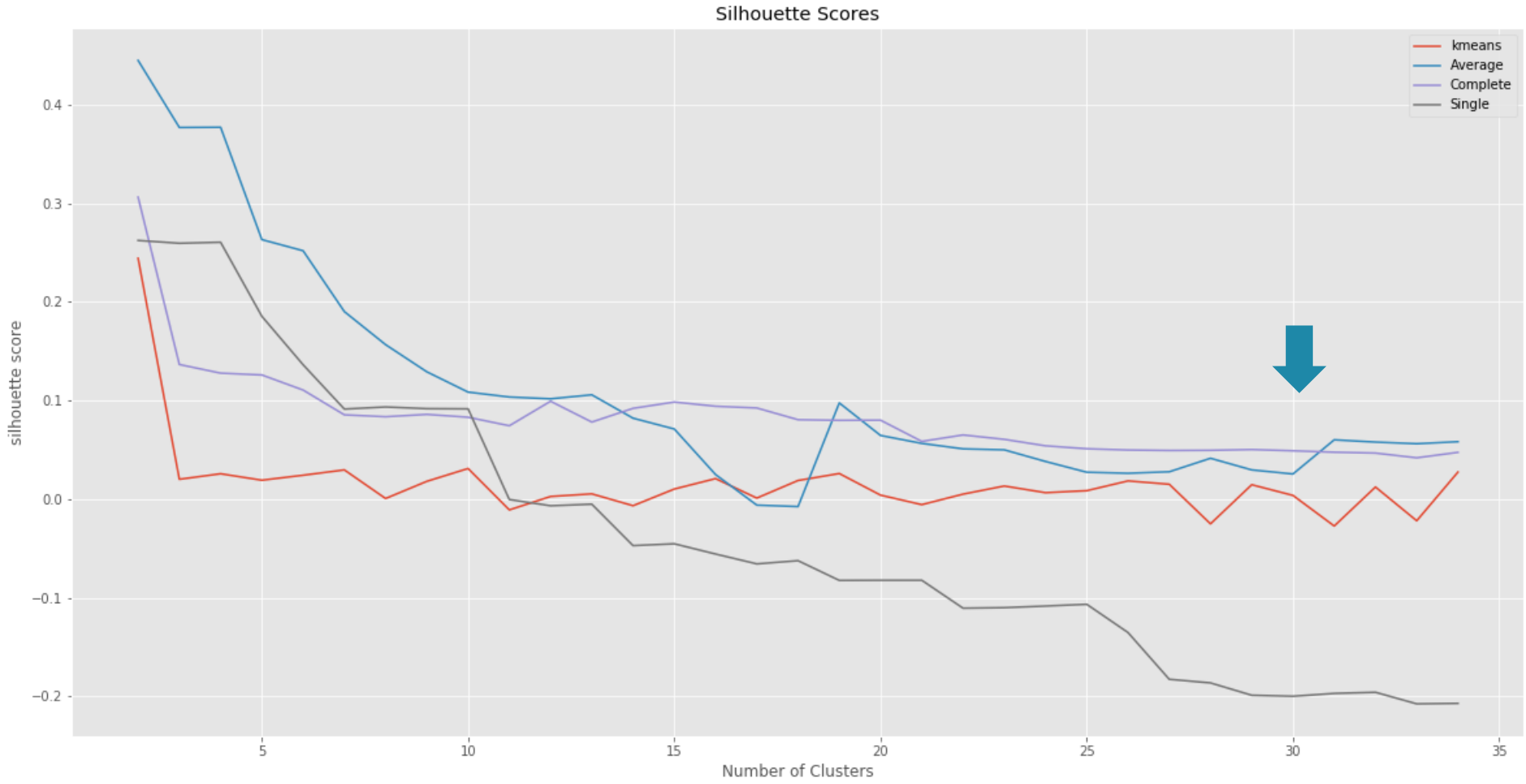
- Market(equally weighted S&P500 stocks)
- Random Selection
- Cluster

Results: Training Period 1

- Training period 1: Feb 7, 2013 - Feb 6, 2017
- Test period: Feb 6, 2017 - Feb 6, 2018

Benchmark /Portfolio	DJI	Random	K means	Single	Complete	Average
Sharpe Ratio	0.150174	0.102065	0.122447	0.11101	0.085972	0.089122

Results continued

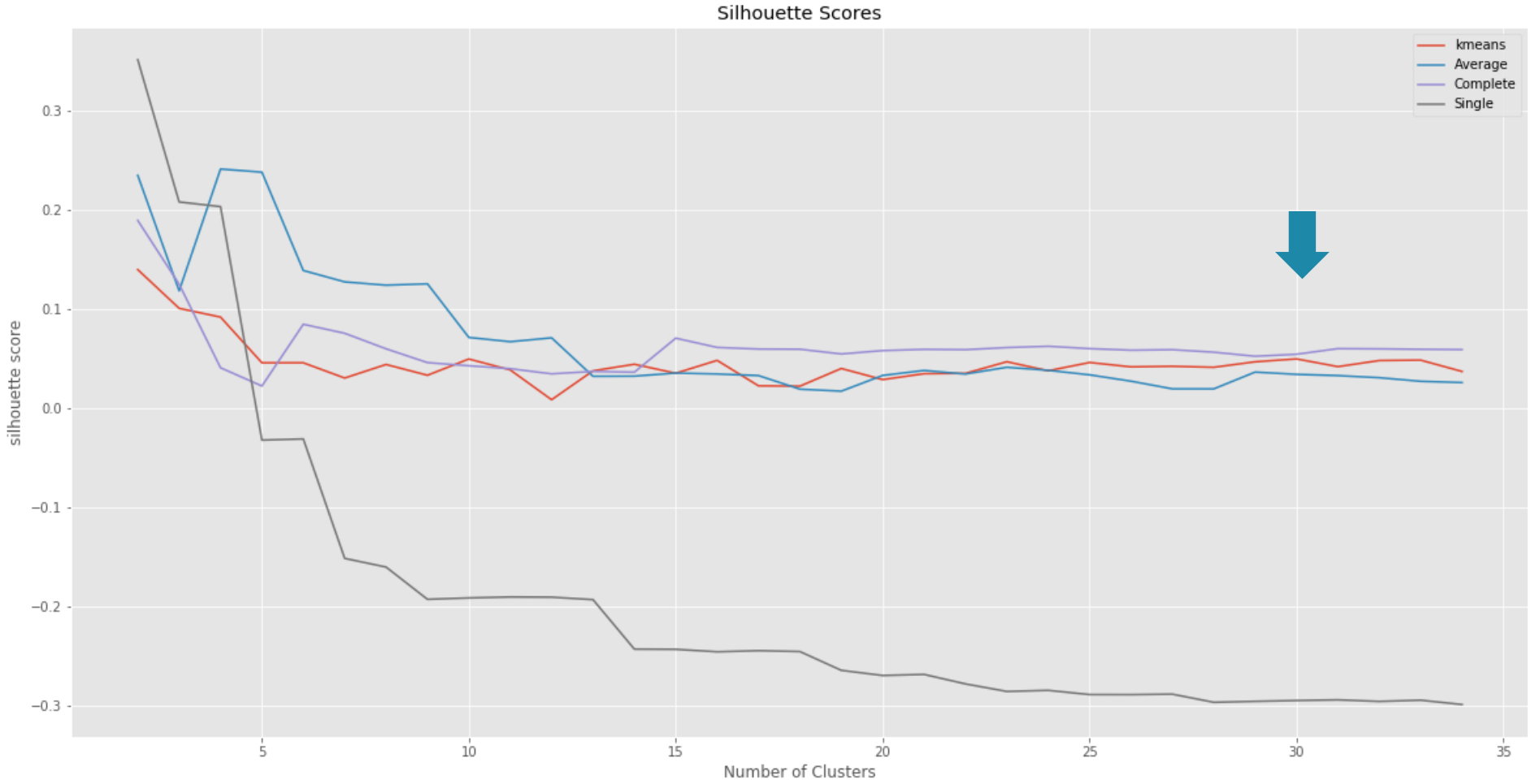


Results: Training Period 2

- Training period 2: Feb 7, 2016 - Feb 6, 2017
- Test period: Feb 6, 2017 - Feb 6, 2018
- Stock prices from 5 years ago may not necessarily be consistent with the prices today.

Benchmark /Portfolio	DJI	Random	K means	Single	Complete	Average
Sharpe Ratio	0.150174	0.102065	0.094728	0.075378	0.115475	0.125963

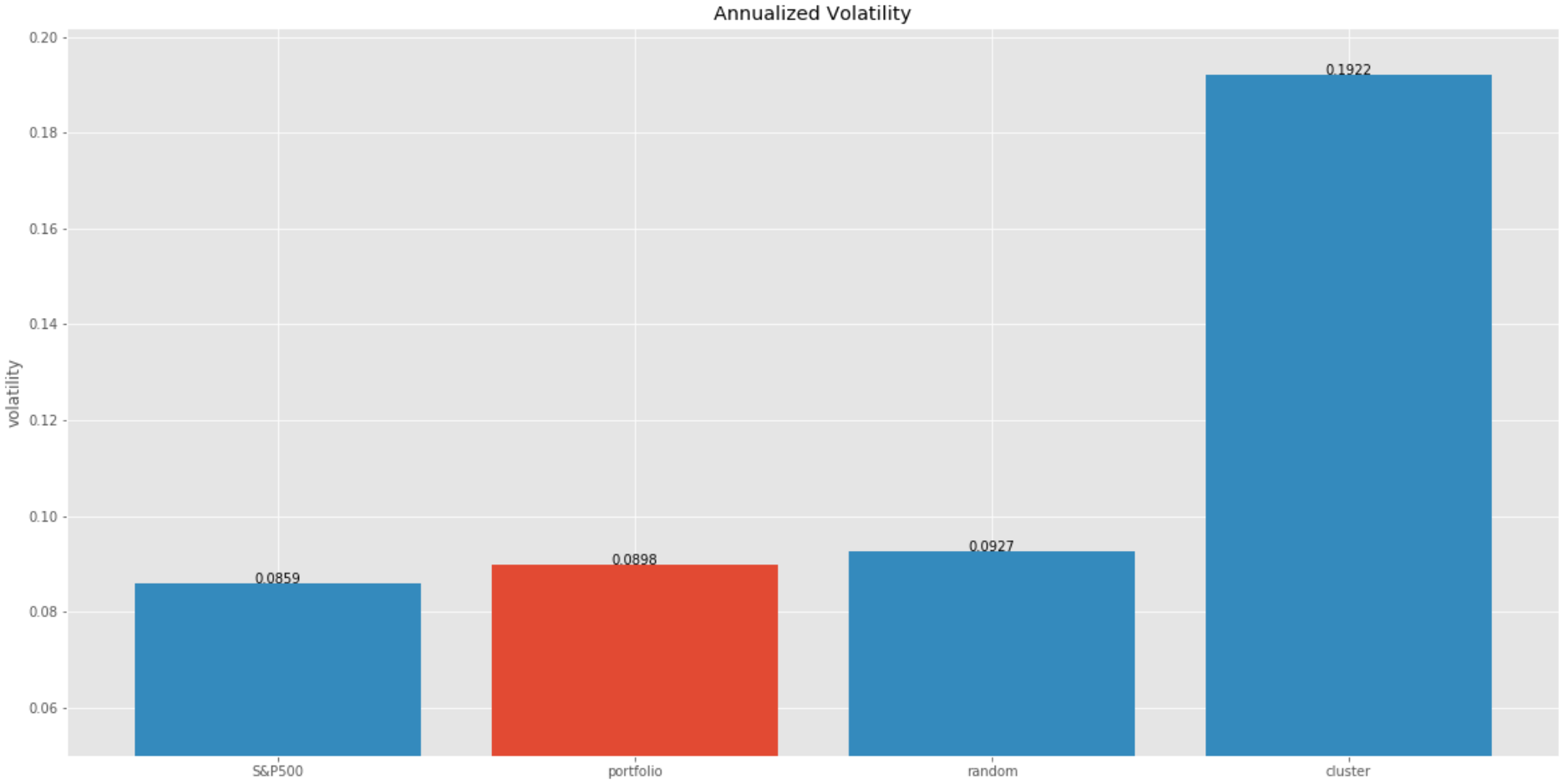
Results continued



Results continued

VOLATILITY

Only 1 out of 30 clusters had a volatility lower than the portfolio constructed.



Further Research

1. Other metrics can be used as alternatives of Sharpe Ratio: Sortino Ratio, Value-at-Risk, etc.
2. Adjust length of training period and number of clusters
3. Change/extend pool of stocks
4. Implement distance thresholds
5. Improve on the risk-return of the portfolio by: assigning weights on each stock of the portfolio based on an optimization problem to maximize Sharpe ratio
6. Add/utilize industry sector information for each company
7. Use time series analysis to describe the changes in stock prices throughout the duration

THANK YOU

Q&A