

# Portfolio Diversification Based on Clustering Analysis

---

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



“ The objective of the project is to optimize financial portfolio diversification in order to reduce volatility and risk & losses, 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 the first 4 years of data (learning period) and the remaining one year of data (testing period)

# 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 =  $\log_i(1+R)$  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



## AGGLOMERATIVE

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 control group for portfolio evaluation



## Linkage

- Single
- Average
- Complete

# Portfolio Construction & Evaluation

## *DIVERSIFIED PORTFOLIO*

---

selecting the stock with highest **Sharpe** ratio from **each cluster** with **equal weight**

1

Sharpe ratio  
in comparison  
to Dow Jones  
Index as  
control group

2

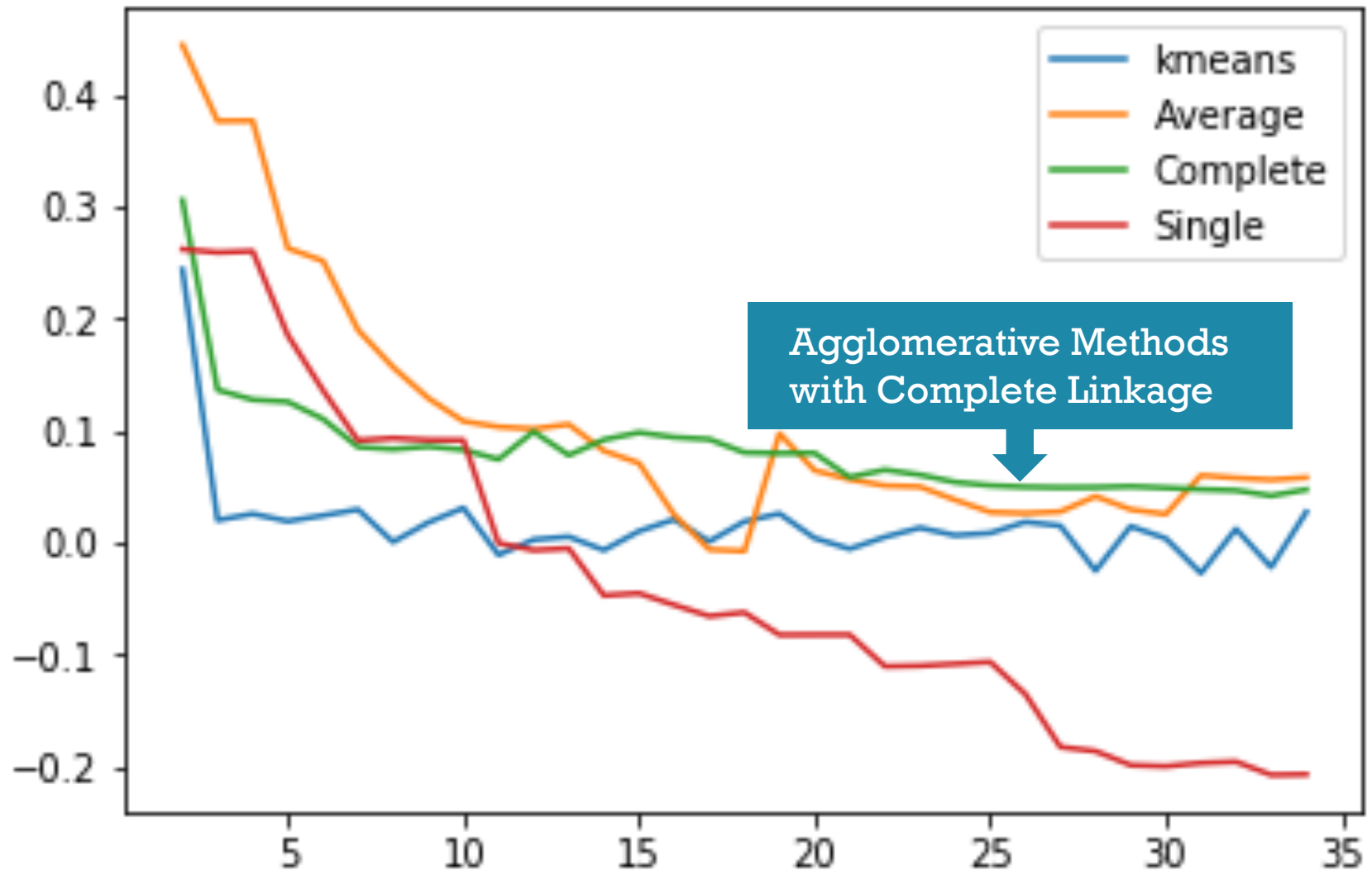
Portfolio  
volatility in  
comparison to  
single stock  
and market  
volatility



# Results

Portfolio	DJI	K means	Single	Average	Complete
Sharpe ratio	0.046057	0.058261	0.058983	0.069390	0.087760

# Results



# Results

Portfolio	DJI	Portfolio 1 (Same stocks)	Portfolio 2 (Same clusters)
Sharpe ratio	0.150174	0.085972	0.210182

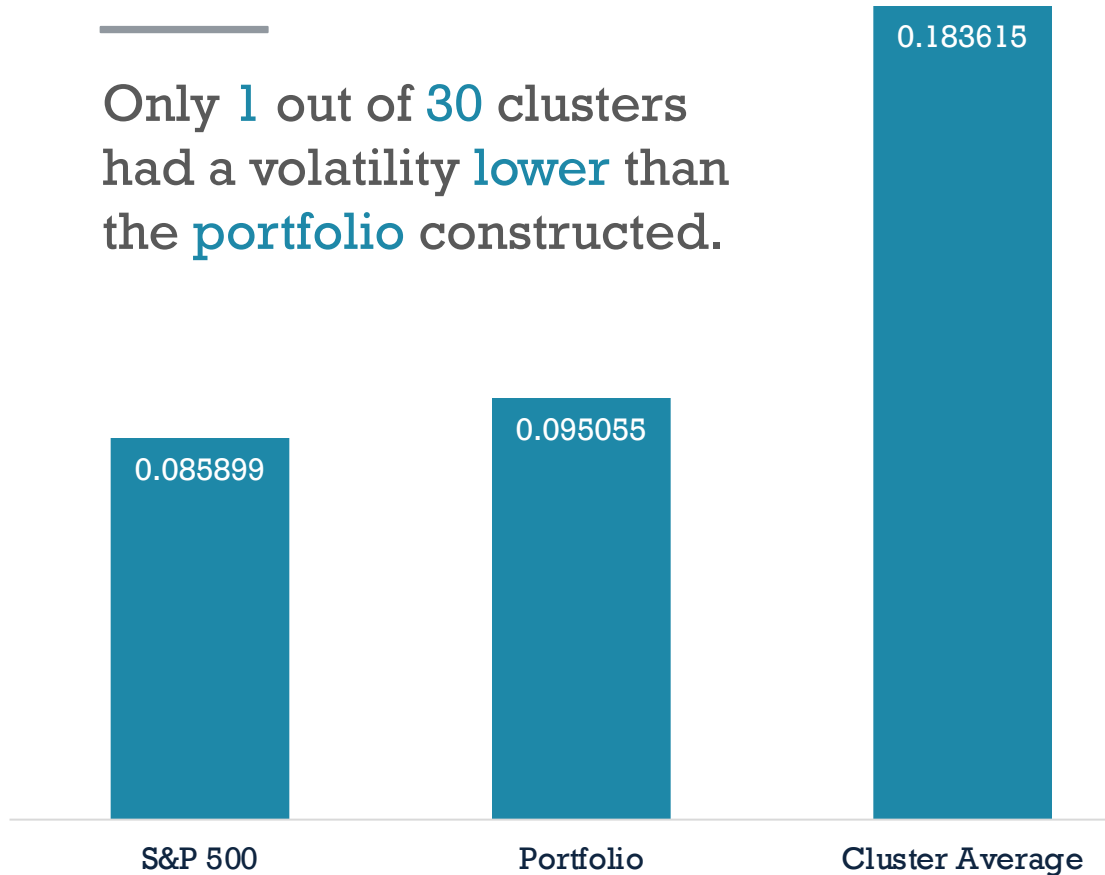
# Results

	VOLATILITY
S&P 500	0.085899
Portfolio	0.095055
Cluster Average	0.183615

0.2  
0.18  
0.16  
0.14  
0.12  
0.1  
0.08  
0.06  
0.04  
0.02  
0

## ***VOLATILITY***

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



# Conclusion

## *Further Research:*

1. Other ratios can be used as alternative of Sharpe Ratio: Sortino Ratio, Value-at-Risk
2. Improve on the risk-return of the portfolio by:  
    assigning weights on each stock of the portfolio  
    based on an optimization problem with  
    maximizing Sharpe ratio as a constraint
3. Use time series analysis to fit a model that describes the changes in stock prices throughout the duration



THANK  
YOU

---

*Q&A*