MINISTRY OF ECONOMY
DEPARTMENT OF STATISTICS MALAYSIA

COMPARATIVE ANALYSIS OF TIME SERIES CLUSTERING: DYNAMIC TIME WARPING AND EUCLIDEAN DISTANCE MEASURE IN PRICE INDEX OF COMPANIES IN STANDARD AND POOR (S&P) 500 INDEX

Presenter Name : Dr. Norli Anida Binti Abdullah

Cheong Kah Ken[1]; Dr Norli Anida Binti Abdullah[2]; Dr Nur Anisah Binti Mohamed @ Abdul Rahman[1]; Dr Arief Gusnanto[3]

[1] Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Malaysia
[2] Center for Foundation Studies in Science (PASUM), University of Malaya, Malaysia
[3] School of Mathematics, Faculty of Engineering and Physical Sciences, University of Leeds, United Kingdom (UK)

11Th MALAYSIA STATISTICS CONFERENCE
"Data and Artificial Intelligence: Empowering the Future"
19Th September 2024
Organized by:
MINISTRY OF ECONOMY DEPARTMENT OF STATISTICS MALAYSIA
BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS
MyStats 2024

Introduction

- <u>Time Series Clustering – Objectives:</u>
  - To investigate the stock price index data extracted from companies listed in the Standard and Poor (S&P) 500 index using k-means clustering with DTW distance for centroid updating
  - To compare the clustering results obtained with DTW distance against Euclidean Distance to understand the impact on clusters formation
  - To interpret the characteristics and temporal patterns within the cluster identified through DTW-based clustering within the companies in S&P 500 index

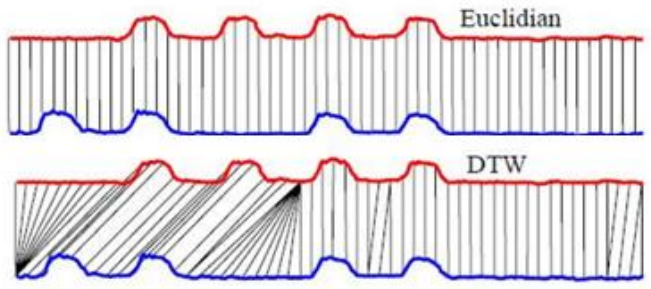
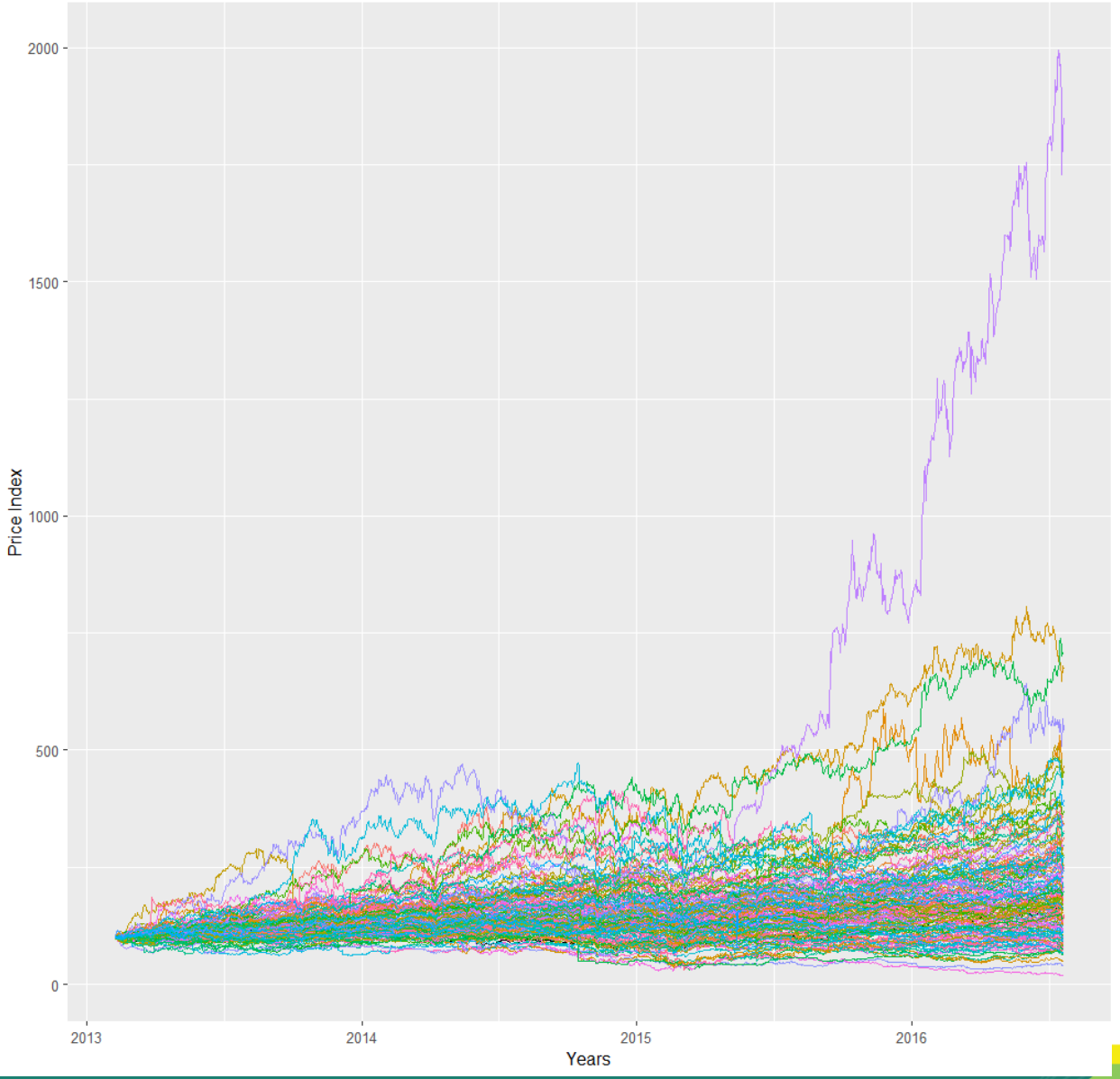


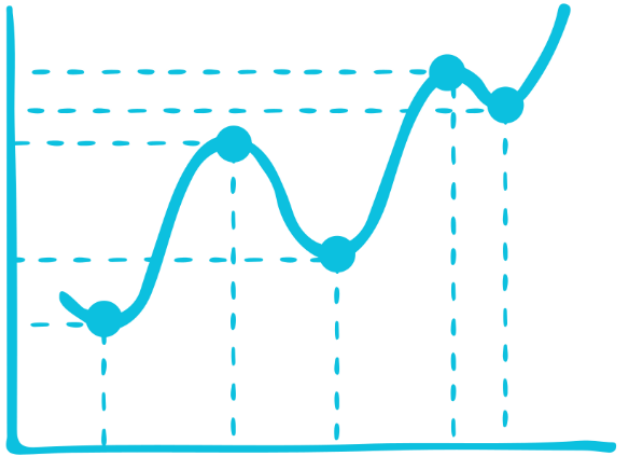
Illustration of Difference Between Euclidean and DTW Distance



Methodology

Methodology

- Data Source
  - S&P 500 Stock Prices (8 February 2013 – 8 February 2018)
    - 505 Company Tickers
    - 7 Variables
    - At Most 1259 Observations
- Data Preprocessing
  - Data Transformation
  - Data Smoothing



Price Index at time t,

$$I_t = \frac{P_t}{P_{t-1}} \times 100, t = 2,3, \ldots$$

Where $I_1 = 100$

P_t is the stock's closing price at time t

Simple Moving Average of t *days*,

$$\overline{X_t} = \frac{1}{n}\sum_{t=-n+1}^{n} X_t$$

Where X_t is the Stock's Price Index at time t

Methodology

- Clustering Process
  1. Apply K-means Clustering to S&P 500 data using
  - DTW-Average,
  - DTW-Median and
  - Euclidean distance

- Implementation of Clustering Models
  - K-Means Clustering by Dynamic Time Warping (DTW) Distance Measure using Averaged Time Series for Centroid Update
  - General Steps for K-Means Clustering Using DTW:
    1. Initialization
       - Randomly Choose K Centroids
    2. Assignment
       - Assign Data to Nearest Centroid, Forming K Clusters
    3. Centroid Updating
       - Recalculate Centroids
       - Minimizes Cumulative DTW Distance
    4. Repeat
       - Repeat Step 2 and 3 Until Minimal Change in Centroids

$$DTW(A,B) = min\left(\sqrt{\sum_{i,j} d(a_i,b_j)^2}\right)$$

Where A and B are Series A and B respectively
$d(a_i,b_j)$ are Distance between Points "a_i" from Series A and Points "b_j" from Series B

DTW Distance

Take Averaged Time Series Within the Clusters

Methodology

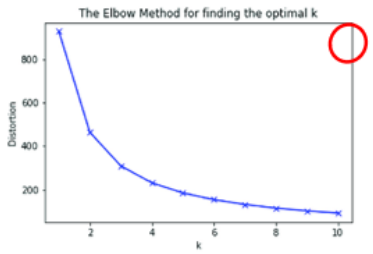
- <u>Clustering Process (continue)</u>

  2. Determine the optimal number of clusters using Elbow Plot
  3. Evaluate Performance using Rand Index (RI) and Davis-Bouldin Index (DBI)
  4. Results Comparison
  5. Properties Elucidation

- Elbow Plot

$$Within\ Cluster\ Sum\ of\ Squares\ (WCSS) = \sum_{T_i\ in\ Cluster\ 1} DTW\ (T_i, C_1)^2 + \sum_{T_i\ in\ Cluster\ 2} DTW\ (T_i, C_2)^2 + \cdots + \sum_{T_i\ in\ Cluster\ k} DTW(T_i, C_k)^2$$

Where T_i is the ith Series price index in the Particular Cluster
C_k is the Centroid of kth Cluster



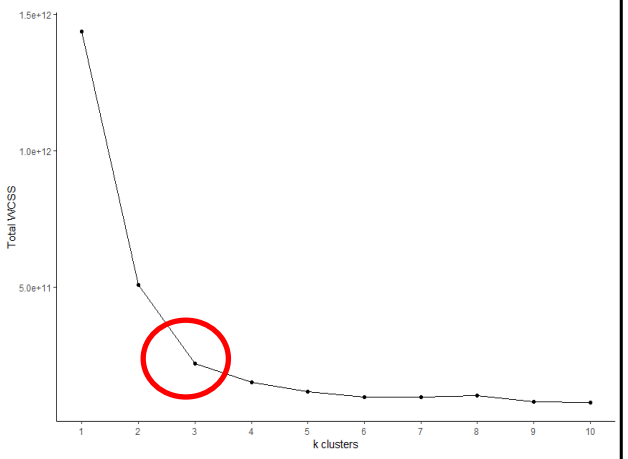
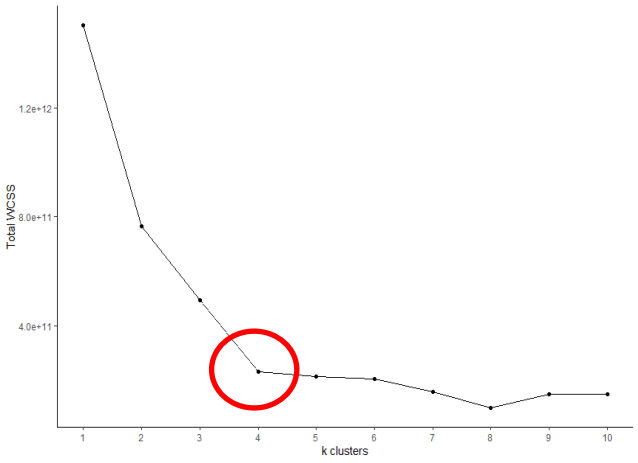
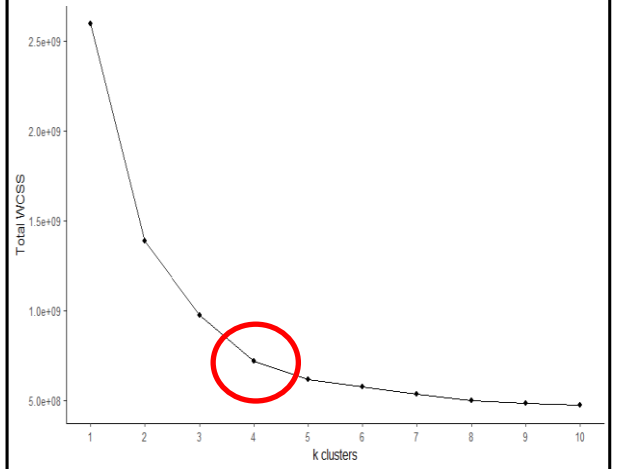
$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j}\left(\frac{S_i + S_j}{DTW(c_i, c_j)}\right)$$

Where k is the Number of Clusters
S_i is the Total Intracluster Distance within Cluster i
$DTW(c_i, c_j)$ is the Distance between Centroid i and Centroid j

Results and Discussions

Results and Discussions

- Clustering Results and Performance Evaluation

Elbow Plot			
Optimal No. of Clusters	3	4	4
Method	DTW (Average)	DTW (Median)	Euclidean
Rand Index	1.000	0.825	0.964
Davis-Bouldin Index	0.461	1.228	0.873

Results and Discussions

- Clustering Results (DTW-Average)

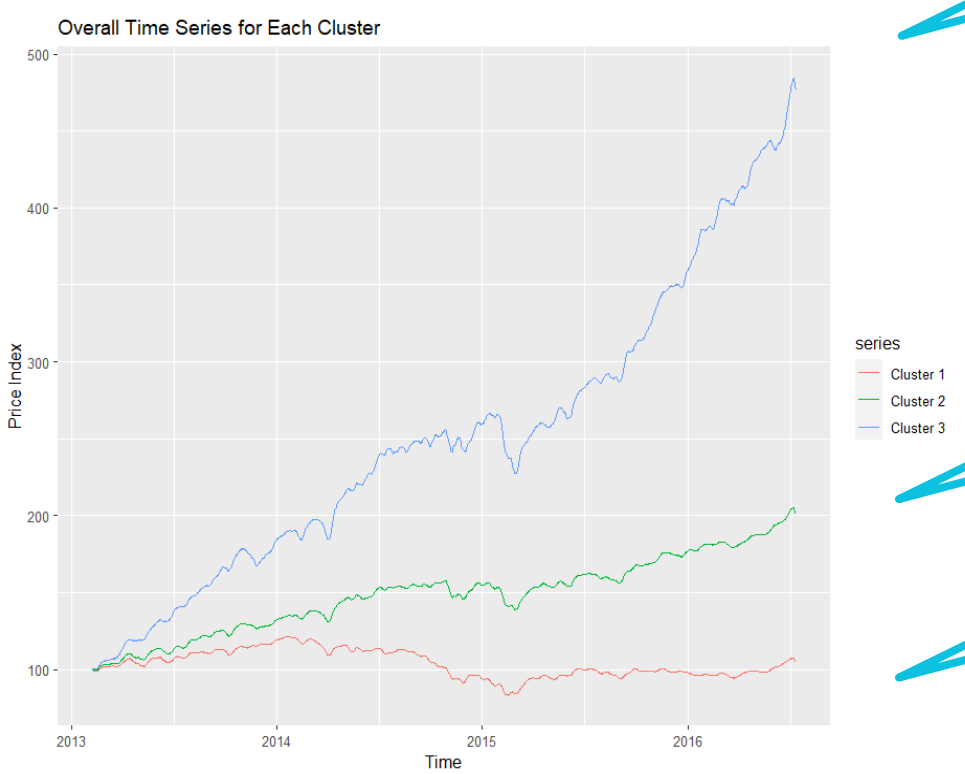


Illustration of Centroids for Each Clusters

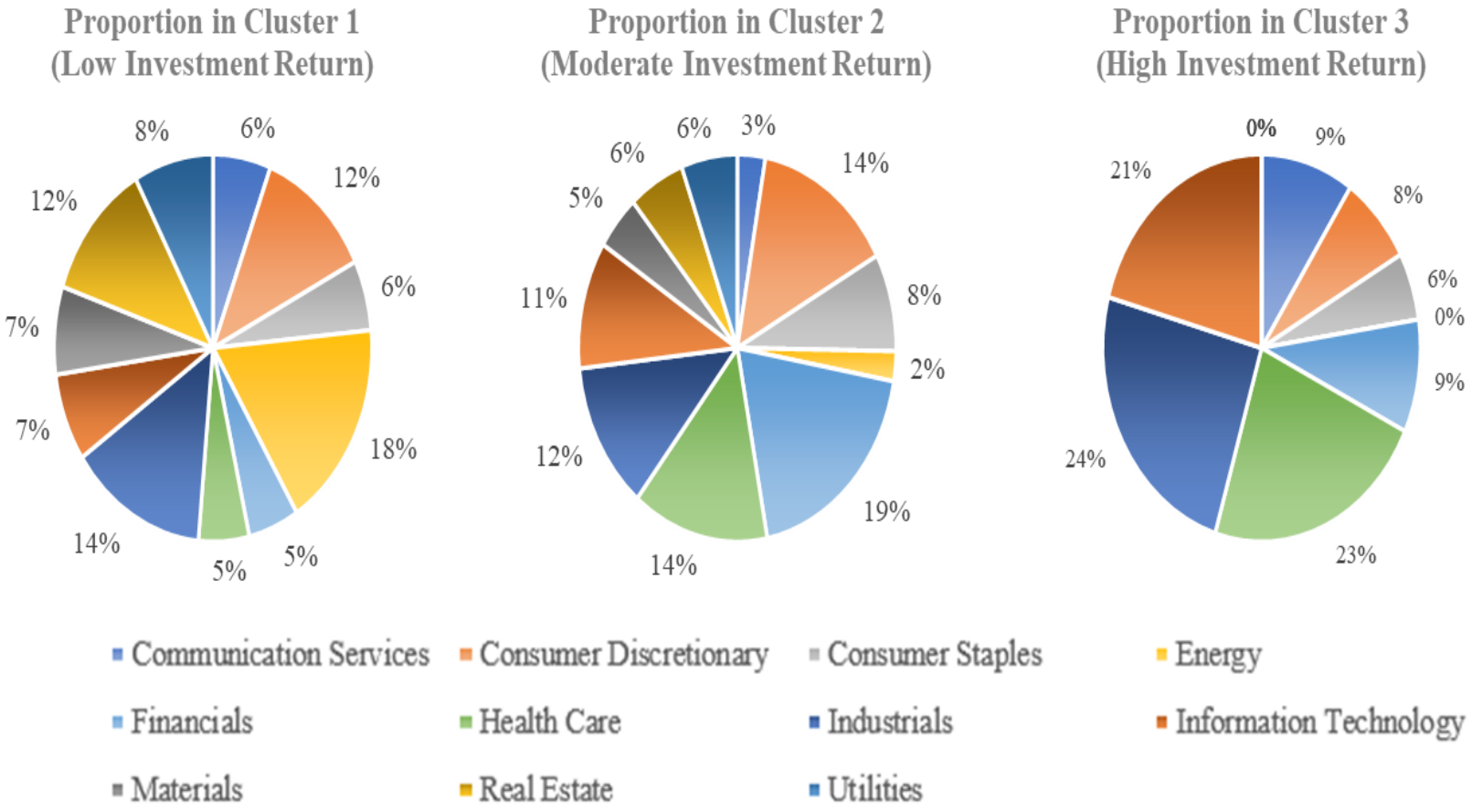
High Investment Return

Moderate Investment Return

Low Investment Return

Results and Discussions

- Sectorial Breakdown of Clusters

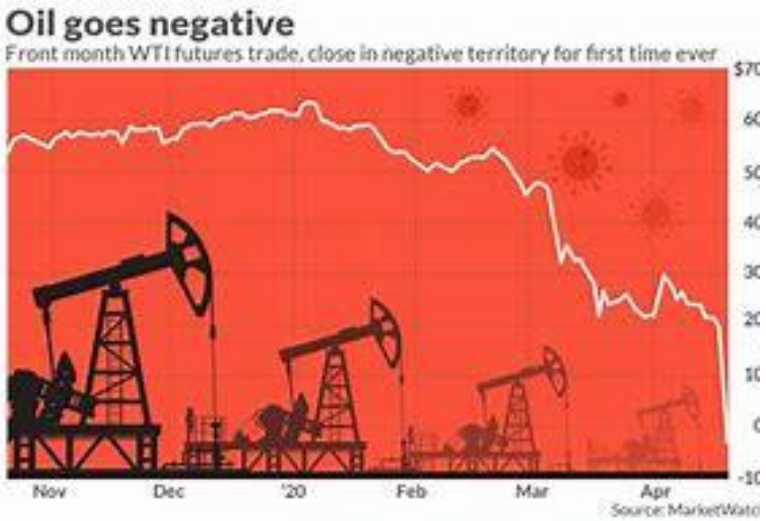


Results and Discussions

- <u>High Investment Return Cluster</u>
  - Health Care Sector
    - Health Insurance Expansion
    - Technological Breakthrough
  - Information Technology (IT) Sector
    - Gaming Boom
    - Growth of Cloud Computing
  - Industrials Sector
    - Aerospace and Defense Sub-Sectors

Results and Discussions

- Low Investment Return Cluster
  - Energy Sector
    - Oil Price Crash in 2014
    - Oversupply in the market
    - Decision of Organization of the Petroleum Exporting Countries (OPEC)



Conclusion

Conclusion

- Key Findings:
  - DTW Captured Patterns more Effective than Euclidean Distance Measures
  - Companies are Clustered into 3 Main Groups
  - Clustering Result is Meaningful
  - Healthcare, Industrials and IT Companies Provide High Investment Return
- Research Significance:
  - Underscore the Potential of DTW for Financial Analysis
- Future Work Suggestion:
  - Algorithm Optimization (Reduce Computational Time)

Thank you