Detection of Market Regimes and Optimization of Trading Strategies

- Ran Cao

Project Goals

Factor Relationship Exploration

Apply Machine Learning techniques to capture the underlying relationship between economic cycles, market regimes and the dynamics of financial assets

Portfolio Management

Explore how technique based insights can be used for factor investing and asset allocation

Market Regimes

To define market regimes from the asset returns and factor returns datasets

Trading Strategy

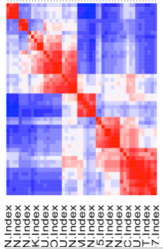
To find a trading strategy based for each regime

Obtain a high portfolio return with a low risk

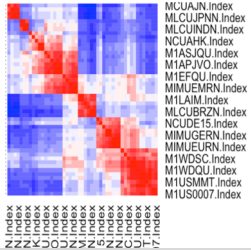
Market Regimes Using Corr Matrix & Clustering

Measure 1: Find the correlation matrix between a pair of window

corr matrix @ **1st** window

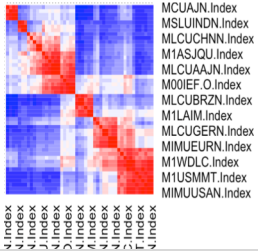


corr matrix @ **2nd** window



...

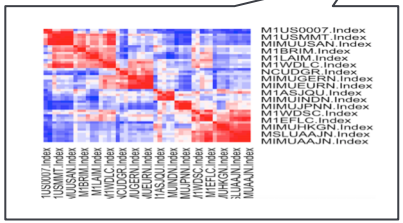
corr matrix @ **last** window



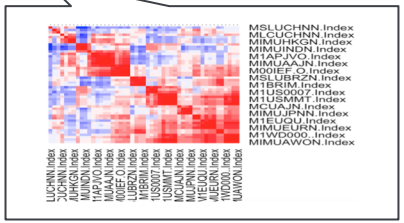
Red: positively correlated **Blue**: negatively correlated **White**: no correlations

Step 2: Calculate Difference Score between a pair of matrices

$$\zeta(t_1, t_2) \equiv \left\langle \left| C_{ij}(t_1) - C_{ij}(t_2) \right| \right\rangle_{ij}$$



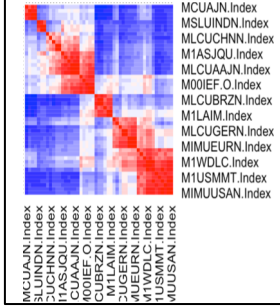
corr matrix @ 1st window



corr matrix @ 2nd window

Measure 2: Find the largest 5 eigenvalues of each correlation matrix

corr matrix @ **last** window



Explains >80% information

Largest 5 Eigenvalues:

30.70
21.51
7.14
4.54
3.40

In vector format:

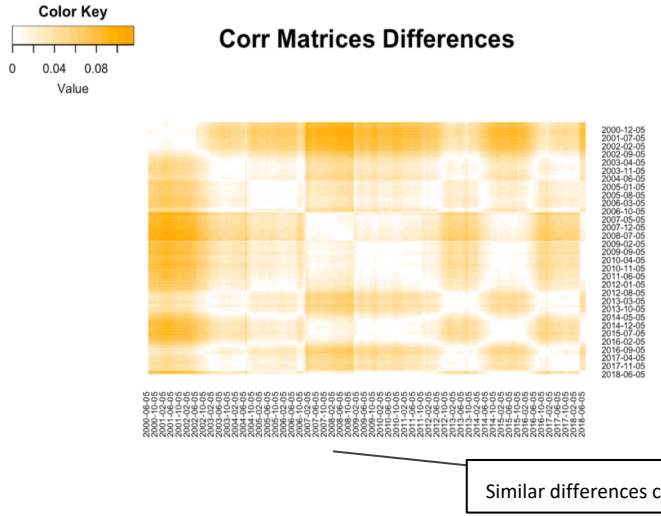
$$\begin{pmatrix} 30.70 \\ 21.51 \\ 7.14 \\ 4.54 \\ 3.40 \end{pmatrix}$$

Step 2: Calculate Difference Score between a pair of vectors

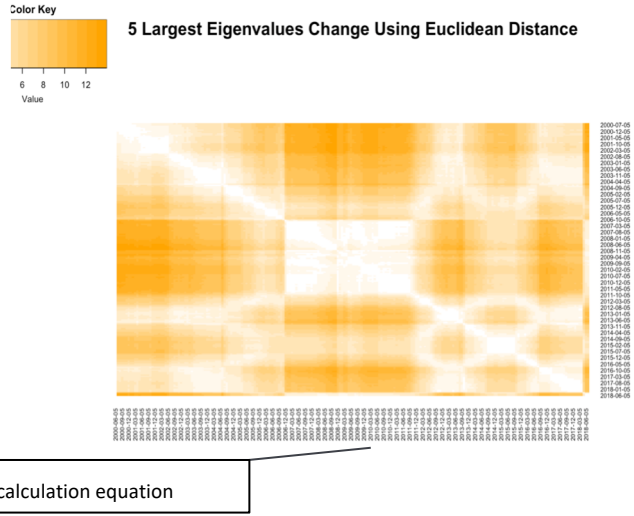
- Euclidean Distance $\sqrt{\sum |x_i - y_i|^2}$
- Cosine Distance $\frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}$

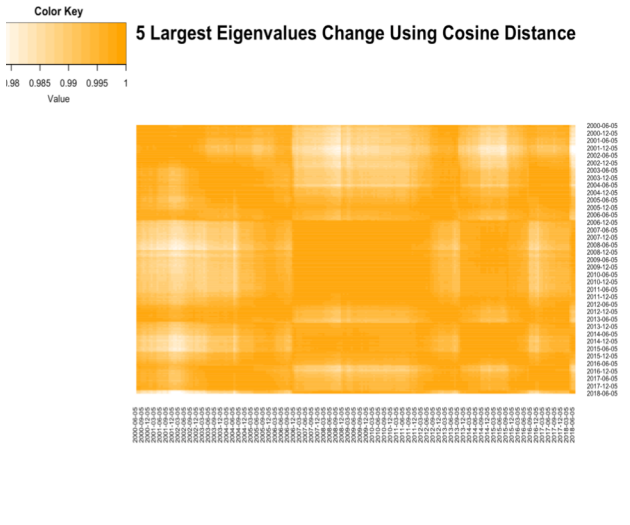
Two Measures Heat Map Comparison

Correlation Matrix Differences



5 Largest Eigenvalue Differences





Similar differences calculation equation

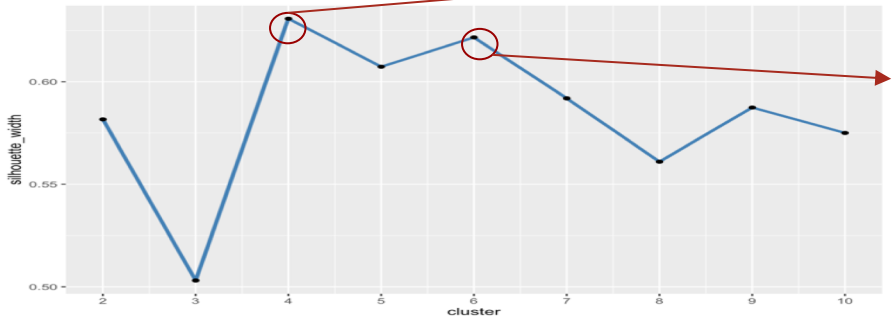
- Their scales are different
  - 5 Largest Eigenvalues Change Using Cosine Distance - **tiny**
  - Correlation matrix differences- **small**
  - 5 Largest Eigenvalues Change Using Euclidean Distance - **big**
- These three graphs look different, though the “Correlation Matrix Differences” graph and the “5 Largest Eigenvalues Change Using Euclidean Distance” graph look similar

PAM Method Clustering

PAM method

- PAM stands for “partition around medoids”. The algorithm is intended to find a sequence of objects called medoids that are centrally located in clusters.
- Input: Dissimilarity matrix
  - correlation matrix differences
  - 5 Largest Eigenvalues Cosine Distance
  - 5 Largest Eigenvalues Euclidean Distance

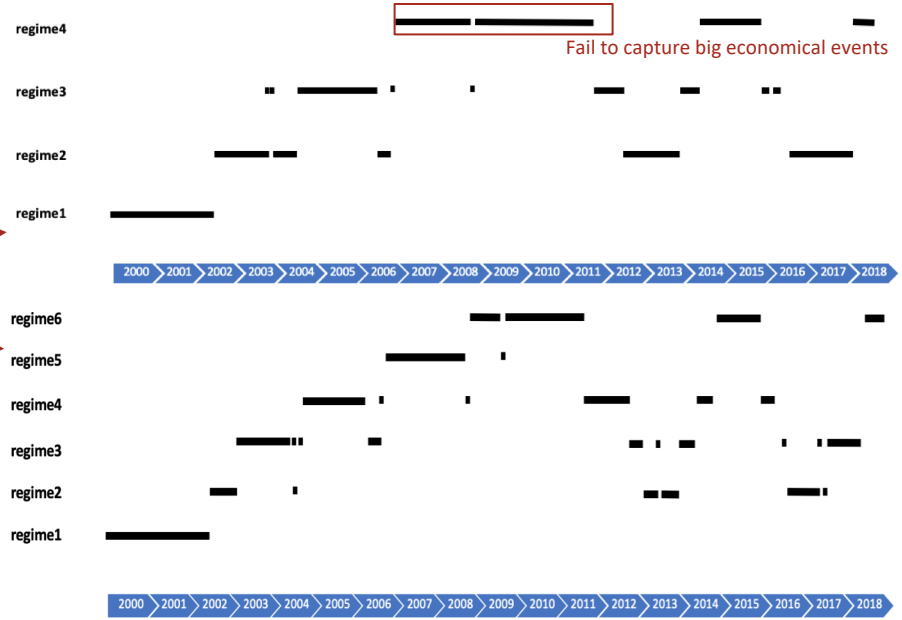
Note: use 2 yrs data first because it’s preferred



Eg: Silhouette graph of Corr Matrix Diff

Silhouette analysis

- reflects how well points fit in their respective clusters.
- number of optimal clusters are 4 and 6



Optimization: A New Clustering Method Construction

Use windows as observations, and assets & factors as features

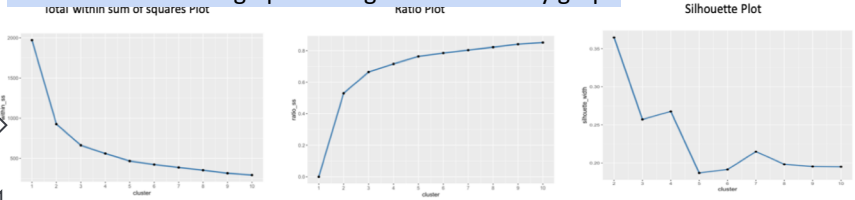
1. Within each window, find the sum of each asset&factor returns

Window1:		
Row1	(factor1, factor2, ..., factor49, asset1, asset2, … , asset 51)	
Row2	(factor1, factor2, ..., factor49, asset1, asset2, … , asset 51)	
	...	
Row515	(factor1, factor2, ..., factor49, asset1, asset2, … , asset 51)	
colSum	**sum factor1, … sum factor49, sum asset1, … sum asset51**	

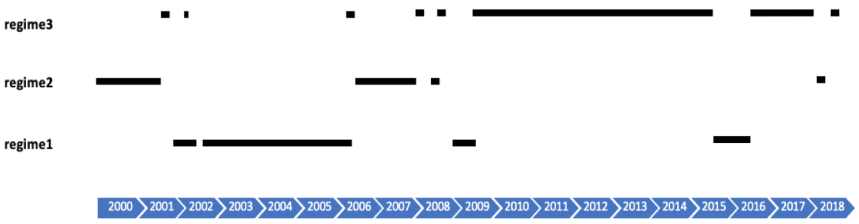
2. Cluster on observations (windows)

Window1 (sum factor1,..., sum factor49, sum asset1, … , sum asset 51)
Window2 (sum factor1,..., sum factor49, sum asset1, … , sum asset 51)
....
Last Window (sum factor1,...,sum factor49,sum asset1, … ,sum asset 51)

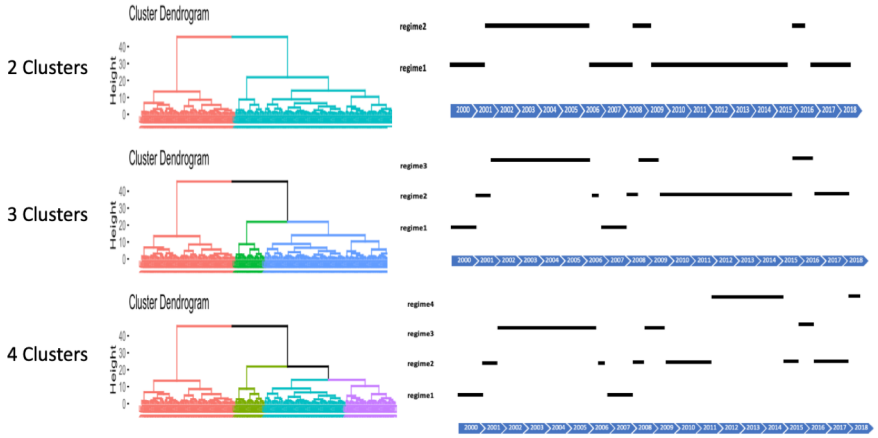
K-means: determining optimal regime numbers by graph



Optimal: 3 Clusters



Hierarchical Clustering



Optimization & Regime Selection

➔ **20** total market regime combinations:
4 time period (0.5 year, 1 year, 1.5 year, 2year) * 3 difference measures (corr matrix, 5 Largest Eigenvalues Cosine & Euclidean)
+ 8 (new method - k-means & hierarchical clustering for 0.5 year, 1 year, 1.5 year, 2year)

Which one is the best?

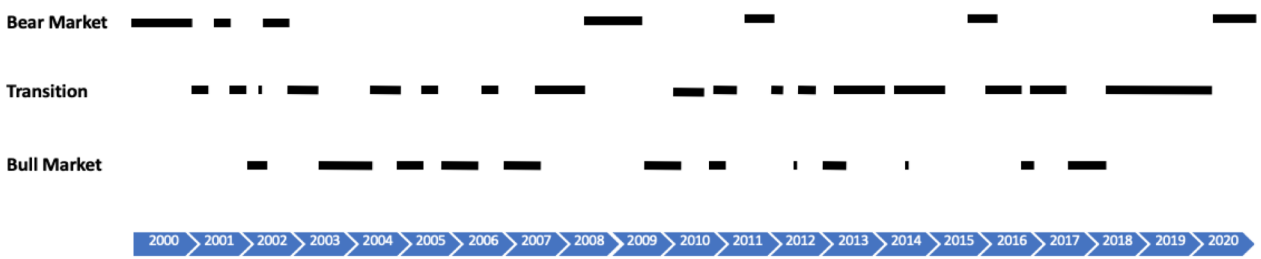
Criteria 1: Define Market Regime by Economic Cycle

- Regime should capture bull market (2002-2007 & 2009-2020) and major bear market (2000-2002, 2007-2009, 2020)

Criteria 2: Define Market Regime by volatility

- Step1: find the volatility of each assets under the same regime
- Step2: Within one regime, average all assets' volatilities (each regime has one averaged volatility value)
- **Bull markets**: slow and steady, low volatility **Bear markets**: assets lose value and prices become volatile

Preferred - Optimization method, Half year, 3 Clusters



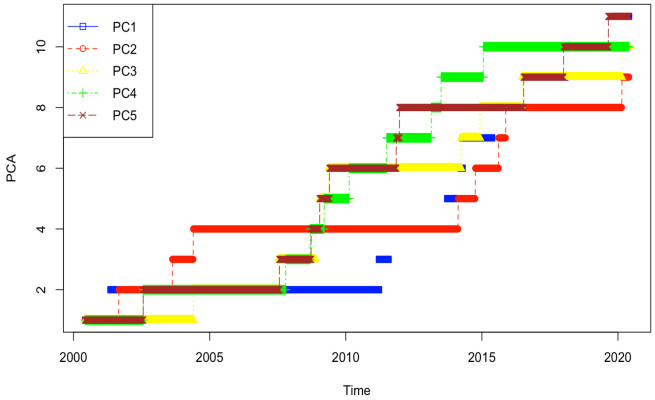
Why it's good?

- Economic life cycle: captured bear & bull
- volatility
  Bear: 0.01741643 high
  Bull: 0.0101847 low
  Transition:0.01045842

Alternative Approach

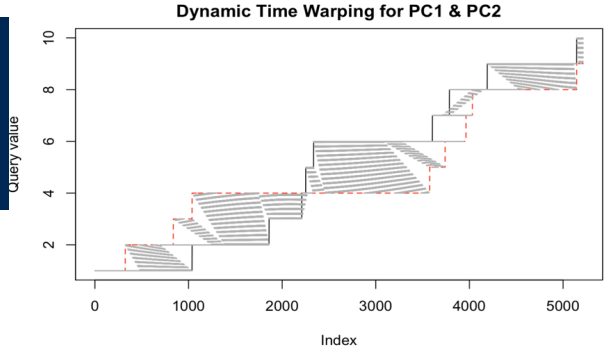
Dynamic Time Warping with 10 Changepoints

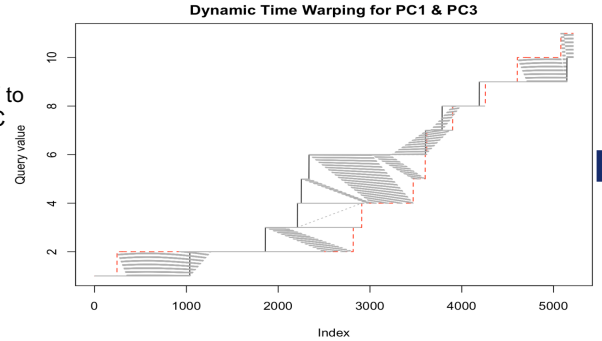
Plot all five PCs into the same plot

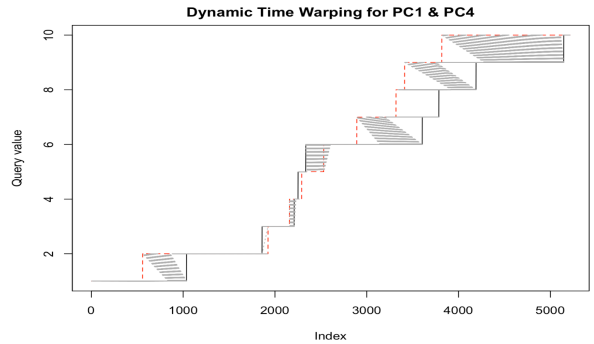


Dynamic time warping is one of the algorithms for measuring similarity between two temporal sequences – Use it to measure the similarity between two aligned PCs

Apply DTW to a pair of PC







Compute the area between the warping function and the no-warping path & List the areas in below:

	PC1	PC2	PC3	PC4	PC5
PC1					
PC2	3.75M				
PC3	2.93M	4.10M			
PC4	2.47M	4.48M	3.31M		
PC5	1.51M	4.06M	2.91M	1.00M	

Findings:
PC4 & PC5 has highest similarity. Then PC1 is also similar to PC5.

Later part, when clustering, consider priority:
PC5>PC4>PC1

Is our Model Robust? Are There Any Potential Limitations?

How quickly can we tell a regime change?

Proposed ways:

1. Define characteristics of each regime
   - bear market: high volatility, assets lose values (negative asset returns)
   - bull market: low volatility, assets slowly gain values
   - transition period: time in between of bear market and bull market

2. Find the probability of switching from one to the next: implement Monte Carlo Simulation

3. What's the early sign of moving to the end of the regime, or move to the new regime
   Will dig into the assets data at the end/beginning of the regime to study their traits

Limitation? Changepoint detection method is sensitive to noise, especially in the time series data.

- Our input is not the raw data, it's PCA. PCA reduces the noise in the dataset as we focus on the "main component", which successfully helped our model denoise.

Trading Strategy Validation on All Market Regimes for 20 Years

The cumulative return of all market regimes has an **uptrend** → overall trading strategy is profitable

The cumulative return of all market regimes for 20 years is **310%.**

Is this return good?

Compared with naïve approach: equaled weighted as our benchmark

Equaled weighted portfolio:

Allocation: Invest all factors (49 factors) for the entire 20 years

Result: 1.652992 < 3.1 our strategy