Economic Regimes Classification

Yijia Yu

08/05/2020

Introduction

1. No single asset allocation is resilient to all economic regimes. Different economic regimes call for different asset allocations.
2. Economic regimes or scenarios (recession, expansion, etc.) can be defined in terms of 4 key economic factors: economic growth, inflationary expectations, monetary policy, and unemployment rate.
3. Adopting a regime-based asset allocation policy within an overall strategic portfolio may significantly enhance the portfolio's efficiency.
4. Developing and implementing a regime-based allocation policy doesn't require perfect forecasting skills, since anticipating the direction of regime change matters more than predicting the absolute rate of change in the 4 factors.

Introduction

In this study, I used two unsupervised learning methods, principle component analysis (PCA) and k-means clustering to classify the historical data from 1960 to 2020 into different economic regimes.

- PCA: deriving a low-dimensional set of features from a large set of variables.
- K-means: partitioning a data set into k distinct, non-overlapping clusters.

Economic Indices

There are 13 indices reflecting US economic market performance used in this study. Those indices are listed in the table below, in which you will see the name of each index and what it represent for.

Index

Description

EHGDUS Index

US Real GDP (QoQ, %, SAAR)

CPI YOY Index

US CPI (inflation) Urban Consumer YoY NSA

CPI CHNG Index

US CPI (inflation) Urban Consumer MoM SA

EHUPUS Index

US Unemployment Rate (%)

IP CHNG Index

US Industrial Production MoM SA

NHSPATOT Index

Private Housing Units Permits Total SAAR (thousands)

NFP TCH Index

US Employment on Nonfarm Payrolls Total (SA, Net Monthly Change, thousands)

TMNOCHNG Index

US Manufacturing New Orders Total MoM SA

LEI TOTL Index

Conference Board US Leading Economic Indicator

PITL YOY Index

US Personal Income YoY SA

CICRTOT Index

Federal Resrve Consumer Credit Total Net Change SA

USCABAL Index

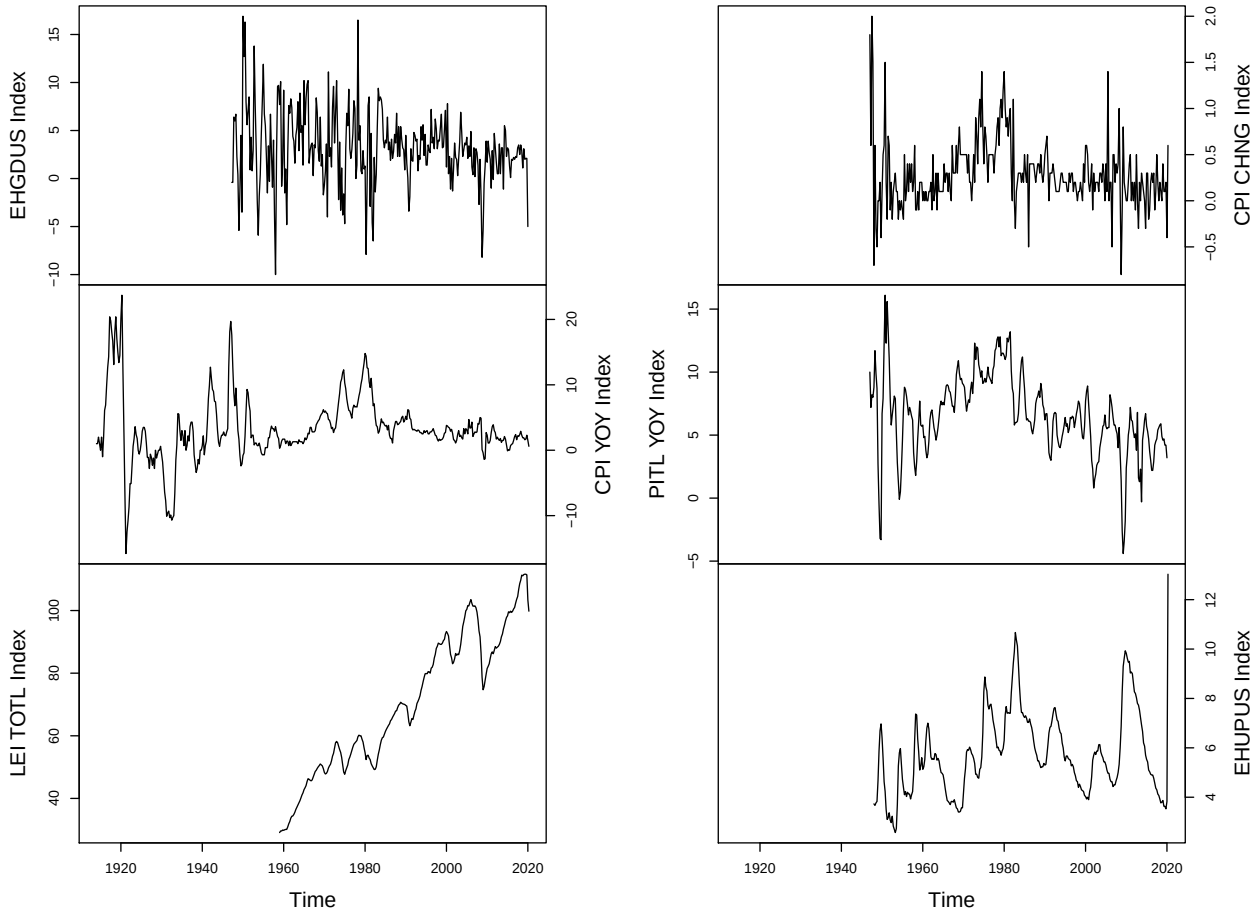
US Nominal Account Balance (Billions USD)

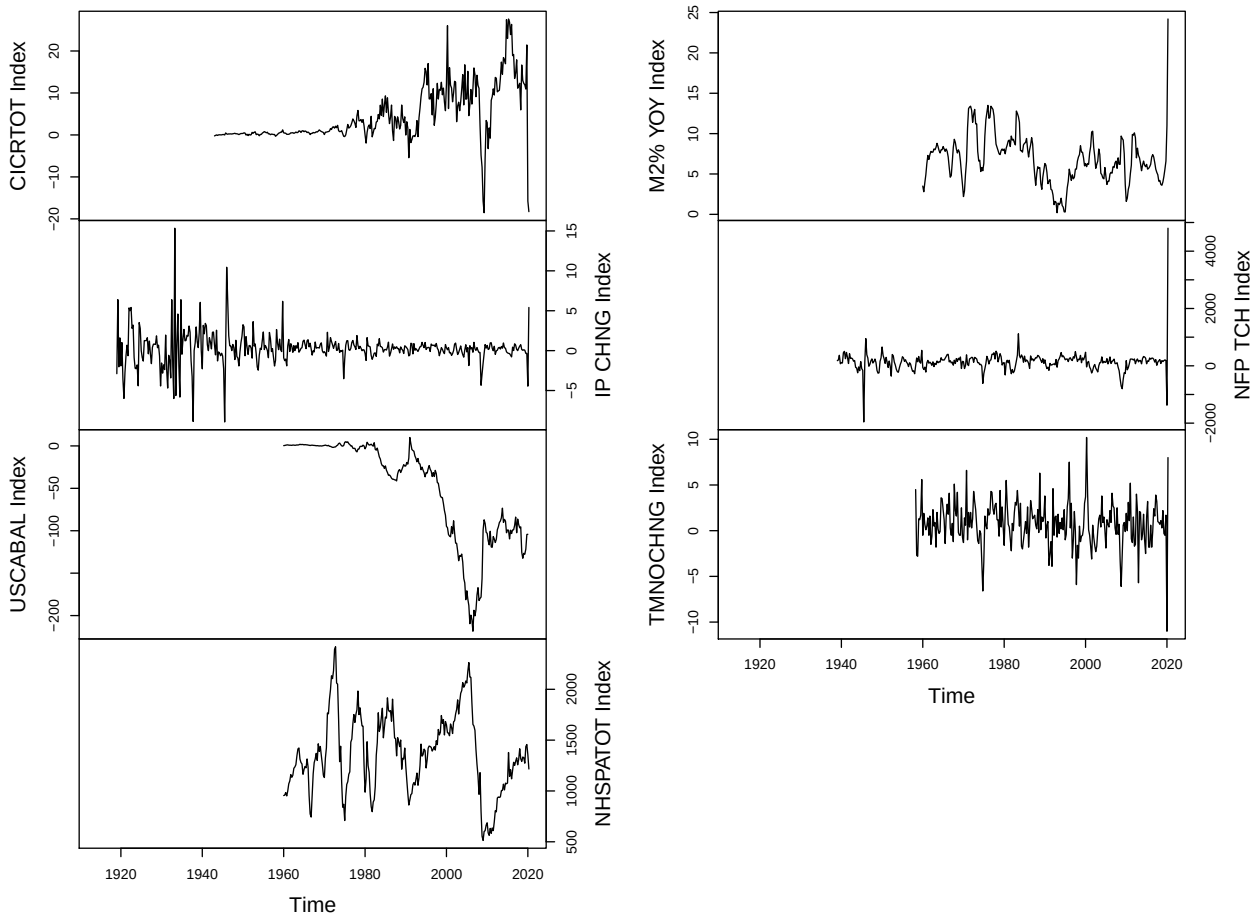
M2% YOY Index

Federal Reserve Money Supply M2 YoY % Change

Preparing Data

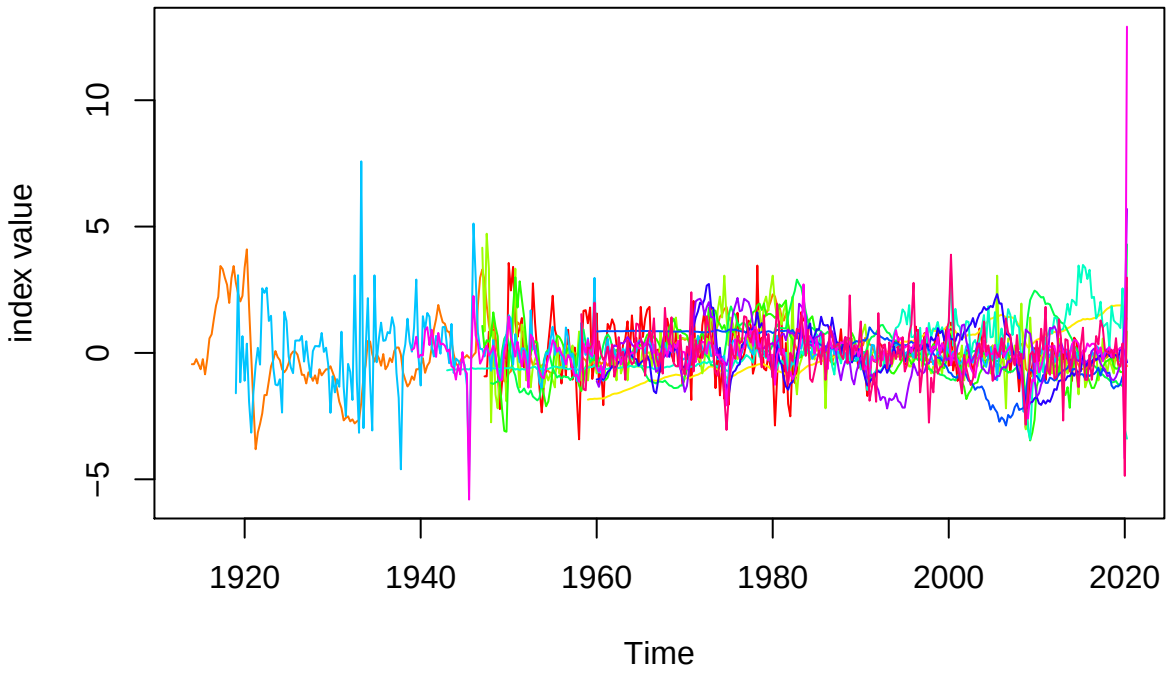
Since the original data are mixed-frequency time series. To solve this problem, the frequency of all indices used in this study was chosen as quarterly. Besides, to solve the inconsistency of time length, i.e. different starting time of all indices, I used the subset series of the longest common length which starts in 1960 Q1 and ends in 2020 Q1.


EHGDUS Index
CPI YOY Index
LEI TOTL Index
PITL YOY Index
CPI CHNG Index
EHUPUS Index
Time
Time



Preparing Data

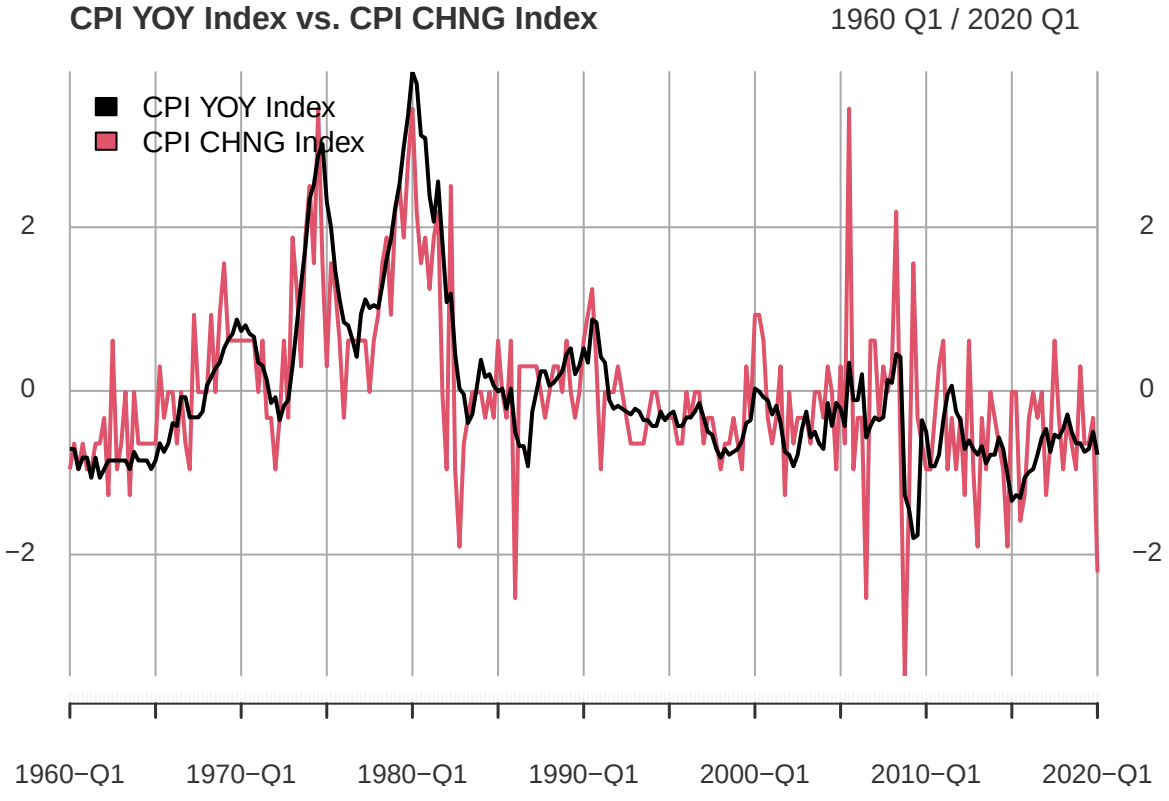
All index series are standardized, that is, centering each series (makes its mean equal to 0) and multiplying a constant to each series to make its variances equal to 1. By the data standardization, we may ignore the absolute value of a single series, but focus on the internal relationship within those economic indices.



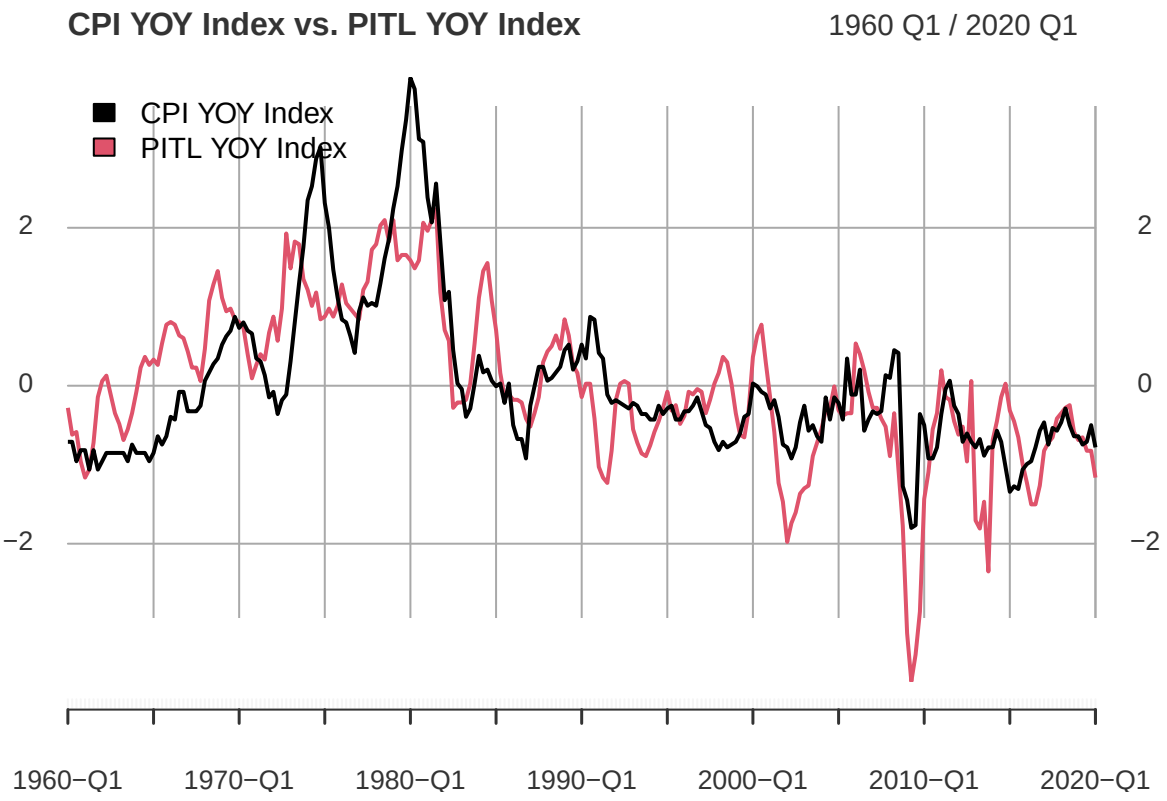
Data Correlation

Some of economic indices are closely correlated with each other.

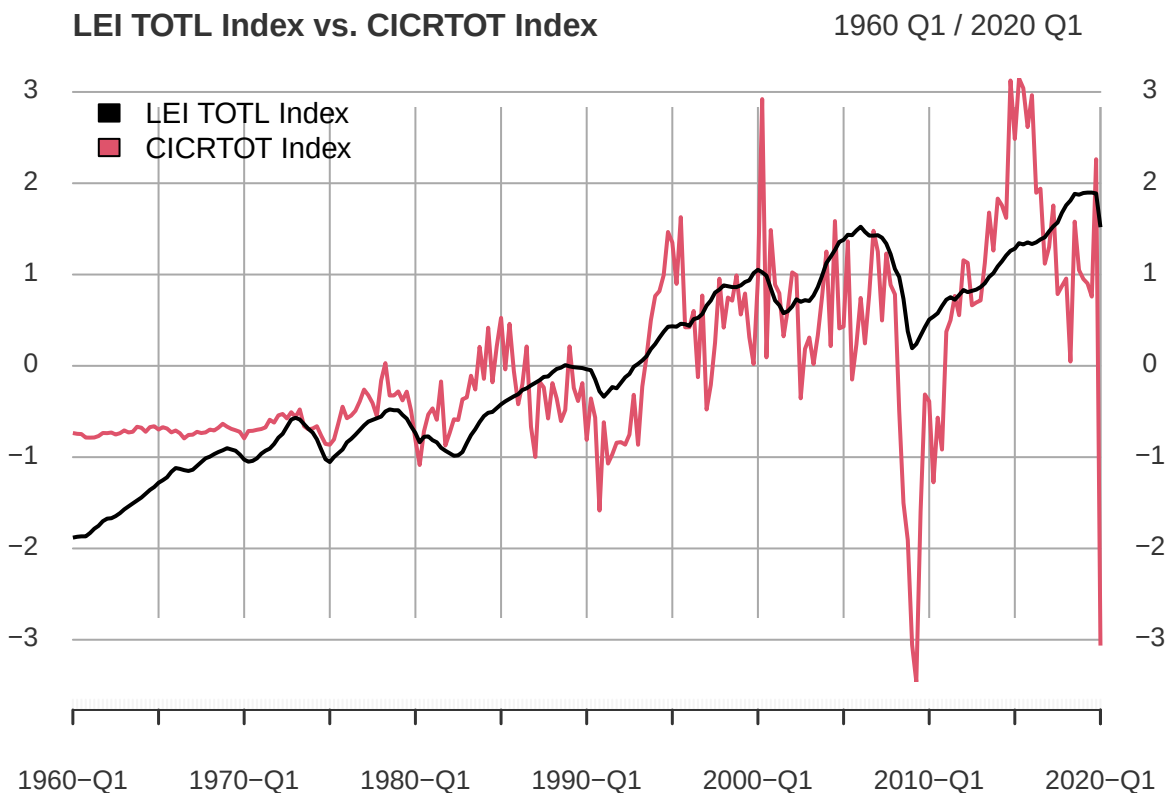
```
## The correlation between US CPI (inflation) Urban Consumer YoY NSA and
##  US CPI (inflation) Urban Consumer MoM SA is 0.719 .
```



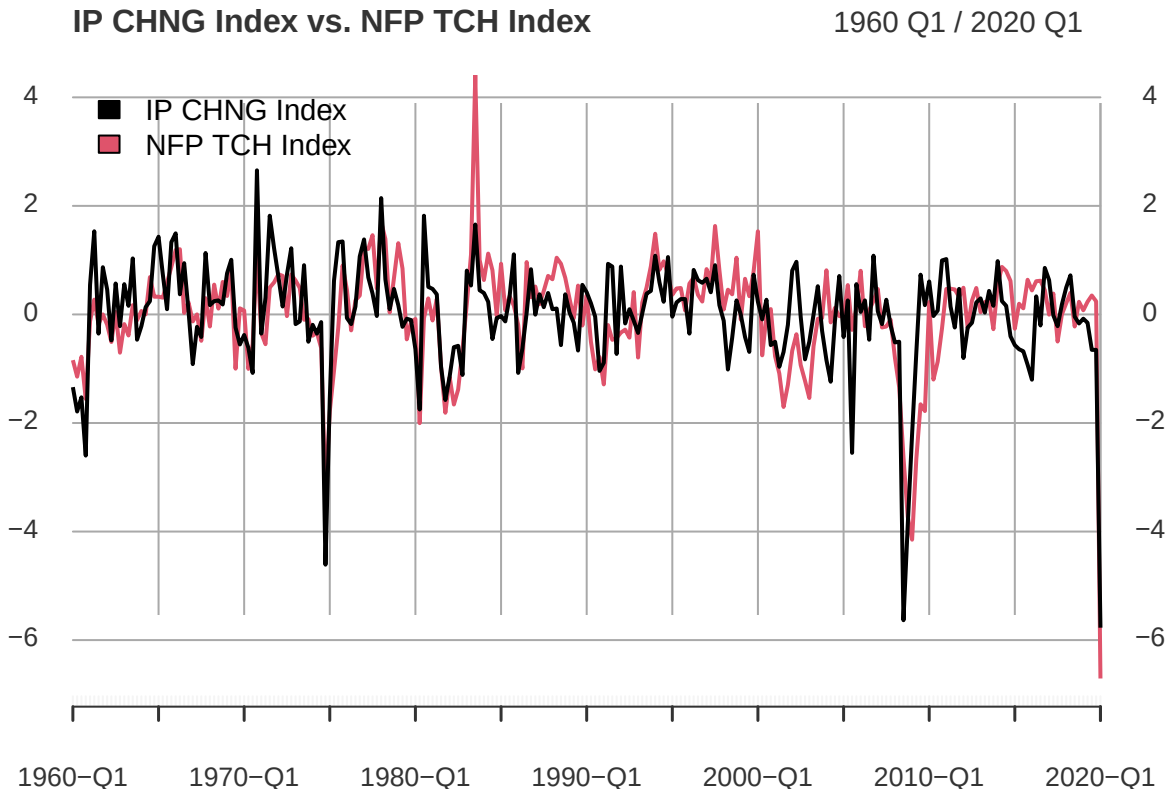
```
## The correlation between US CPI (inflation) Urban Consumer YoY NSA and
##  US Personal Income YoY SA is 0.709 .
```




```
## The correlation between Conference Board US Leading Economic Indicator and
##  Federal Resrve Consumer Credit Total Net Change SA is 0.662 .
```




```
## The correlation between US Industrial Production MoM SA and
##  US Employment on Nonfarm Payrolls Total (SA, Net Monthly Change, thousands) is 0.681 .
```



Labelling

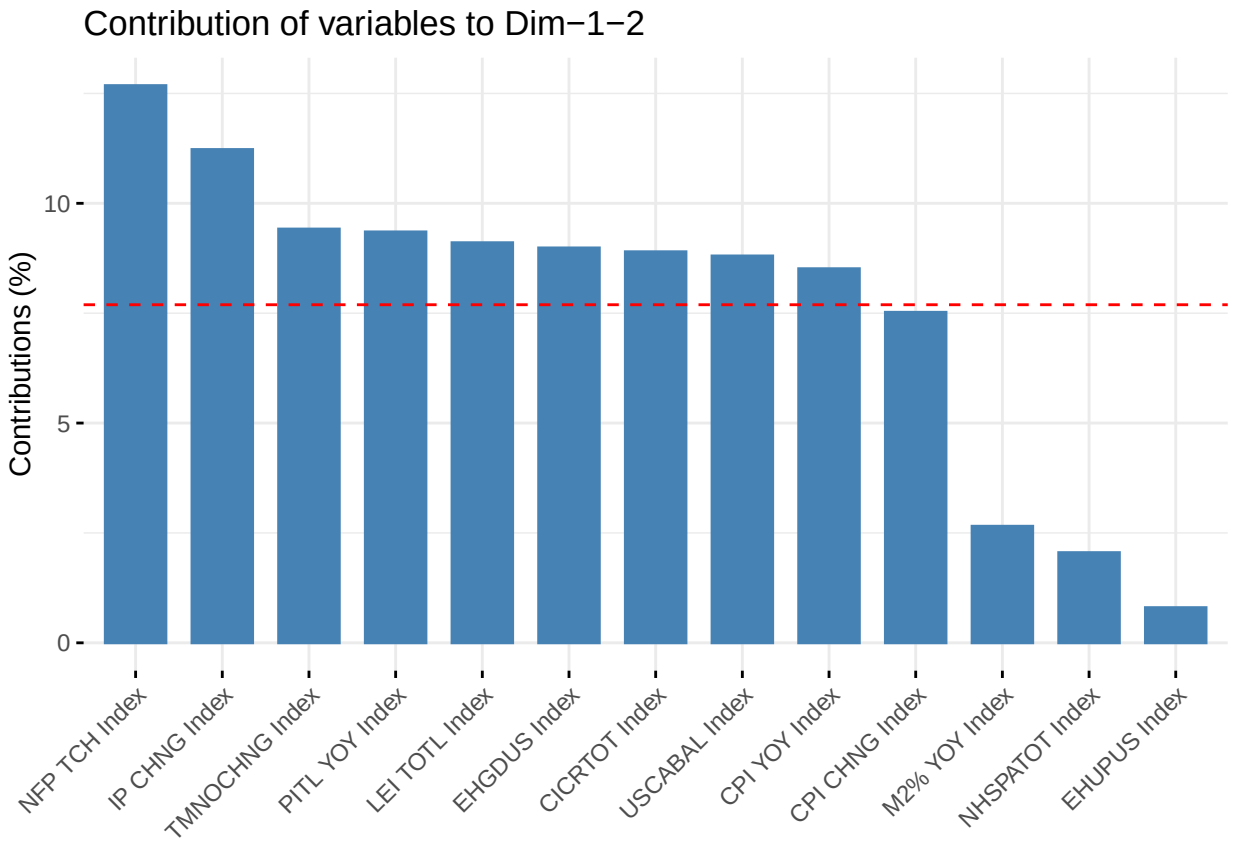
According to List of recessions in the US, all quarters involved in the listed US financial recession periods were labeled as “recession”.

According to List of expansions in the US, all quarters involved in the listed US financial expansion periods were labeled as “expansion”.

Principle Component Analysis (PCA)

- Variable contributions to PC

The plot below shows the contributions of variables in accounting for the variability to the top 2 principal components, that is, the higher contribution (%) of one economic index in this graph, the greater necessity to including this index into our analysis.

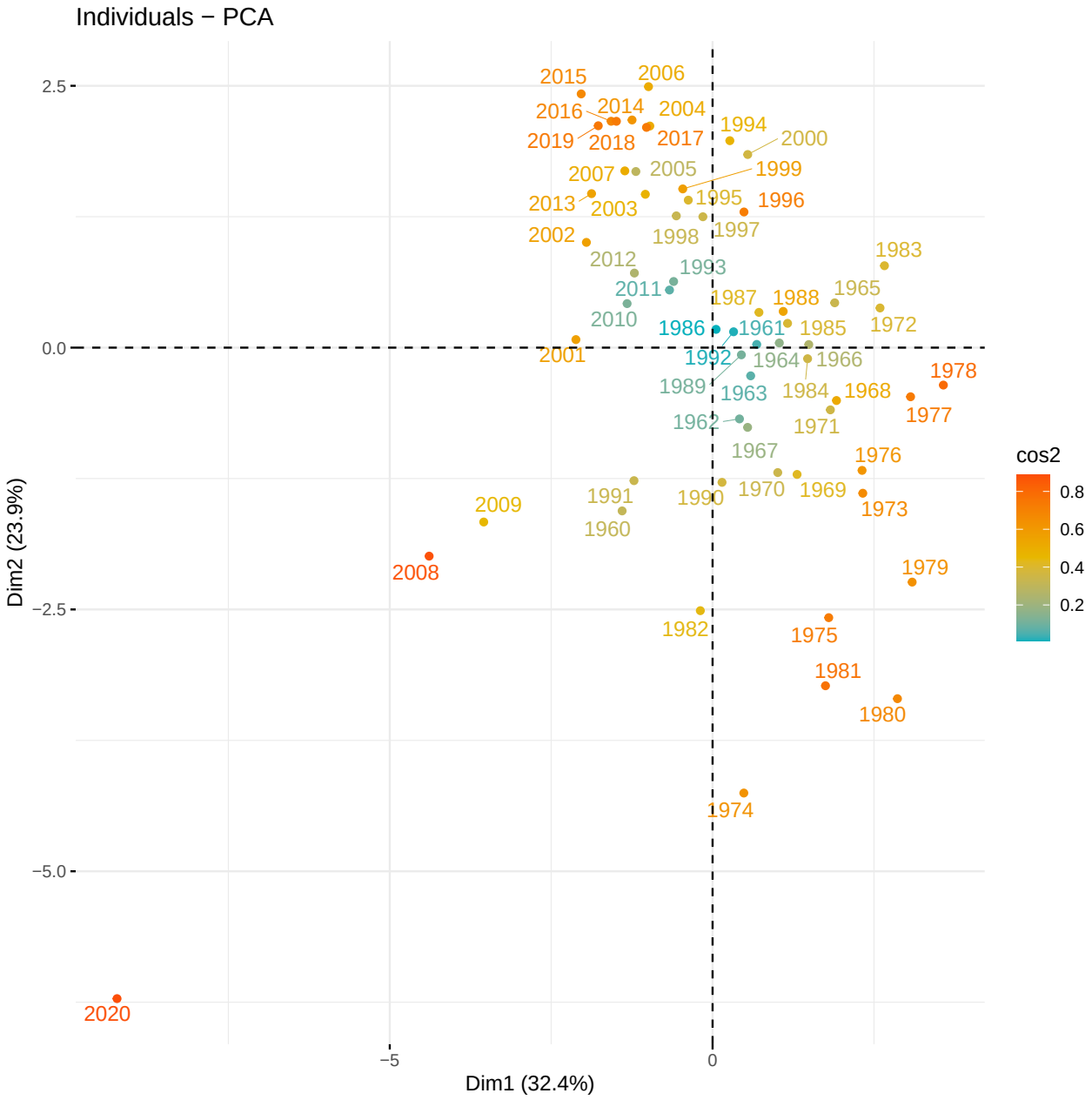


Some highly correlated indices result trivial variables which do not contribute variance to first two principle components.

Graph of individuals

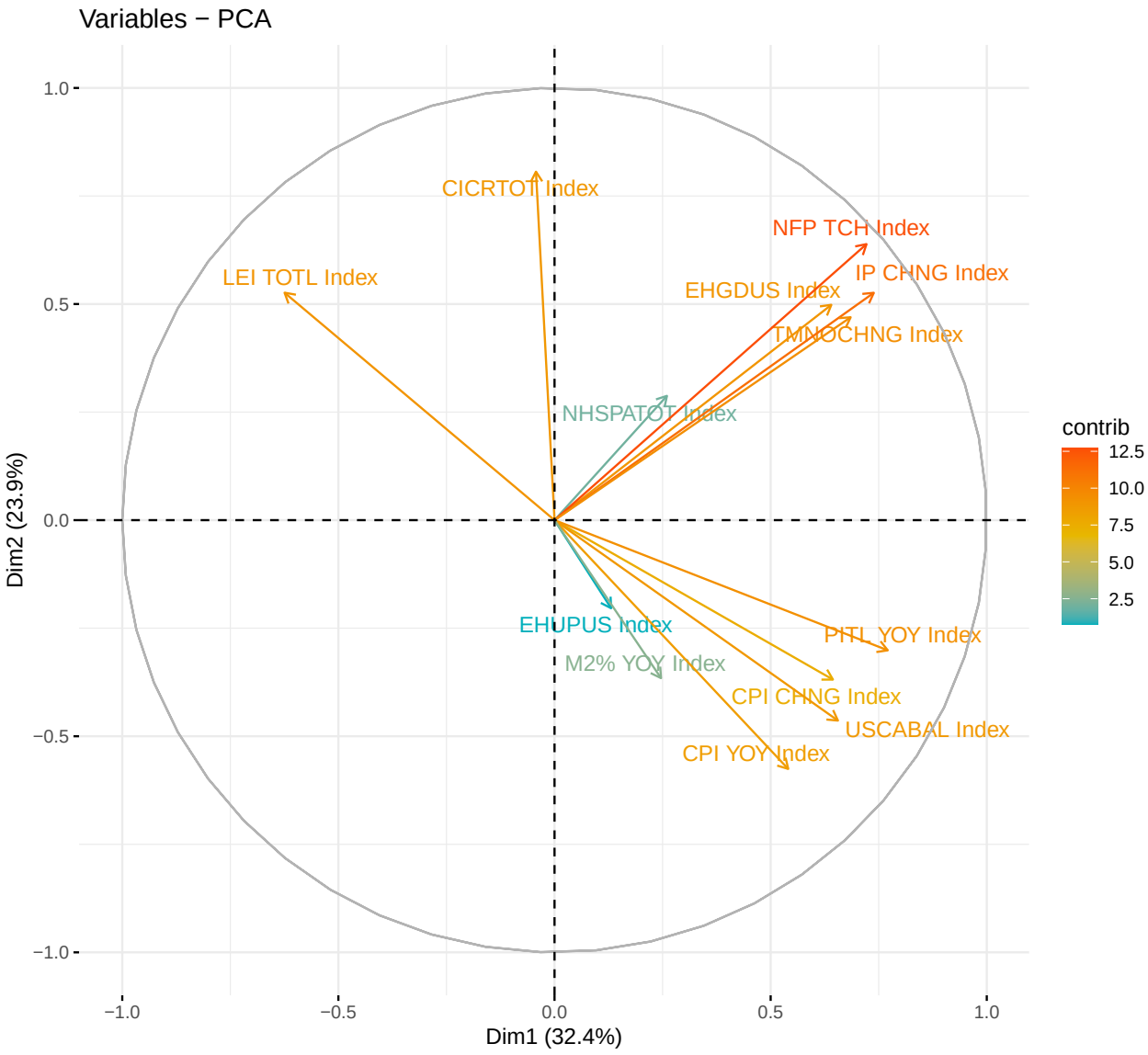
This plot displays the annually data which is projected onto the plane given by the first two PC score vectors.

Individuals with a similar profile are grouped together. Points with a higher cos2 (square cosine, the quality of representation of the PC) are far away from the center.



Graph of variables

Positive correlated variables point to the same side of the plot. Negative correlated variables point to opposite sides of the graph.



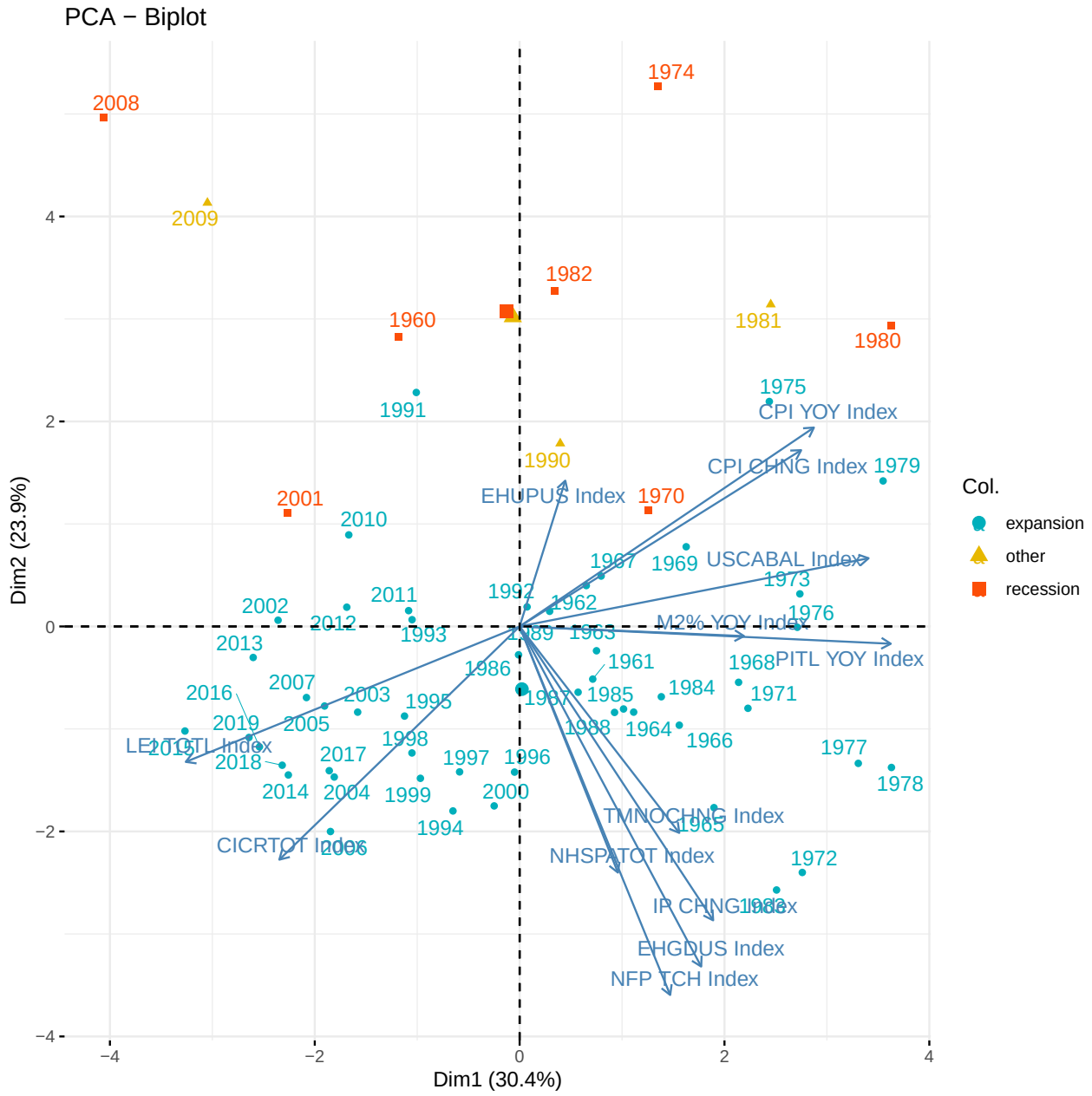
Biplot of both individuals and variables

Yearly data was labeled as recession/expansion if there are more than two quarters in the year labeled as recession/expansion. If there are exactly two quarters of recession and two quarters of expansion in a year, then it will be grouped in 'other'.

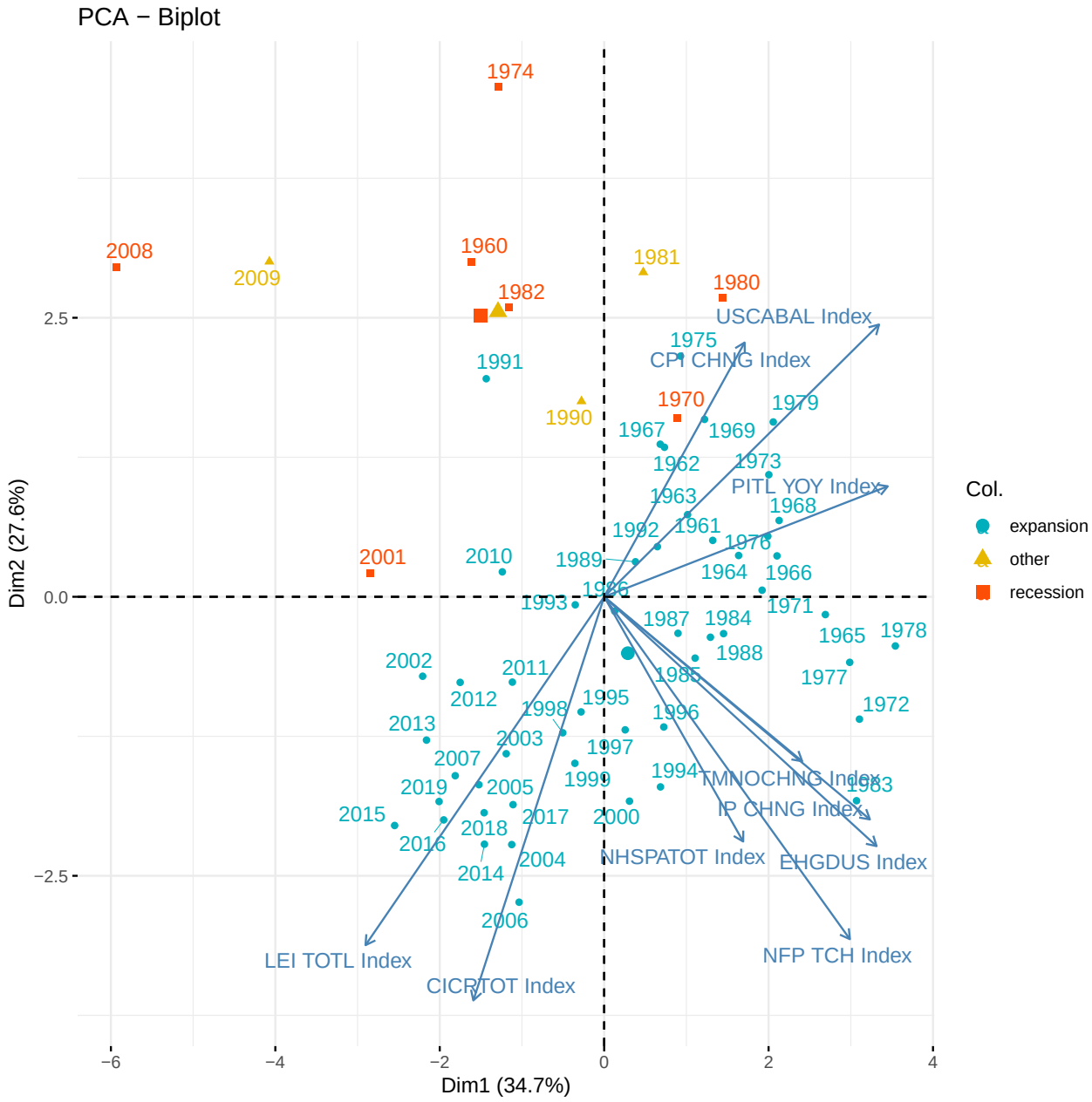
The score of 2020 is very unusual, far away from scores of all previous data.

Removing 2020

After removing 2020, the data become more dispersed in the panel. But some highly correlated variables are clustering together, giving bad classification result.

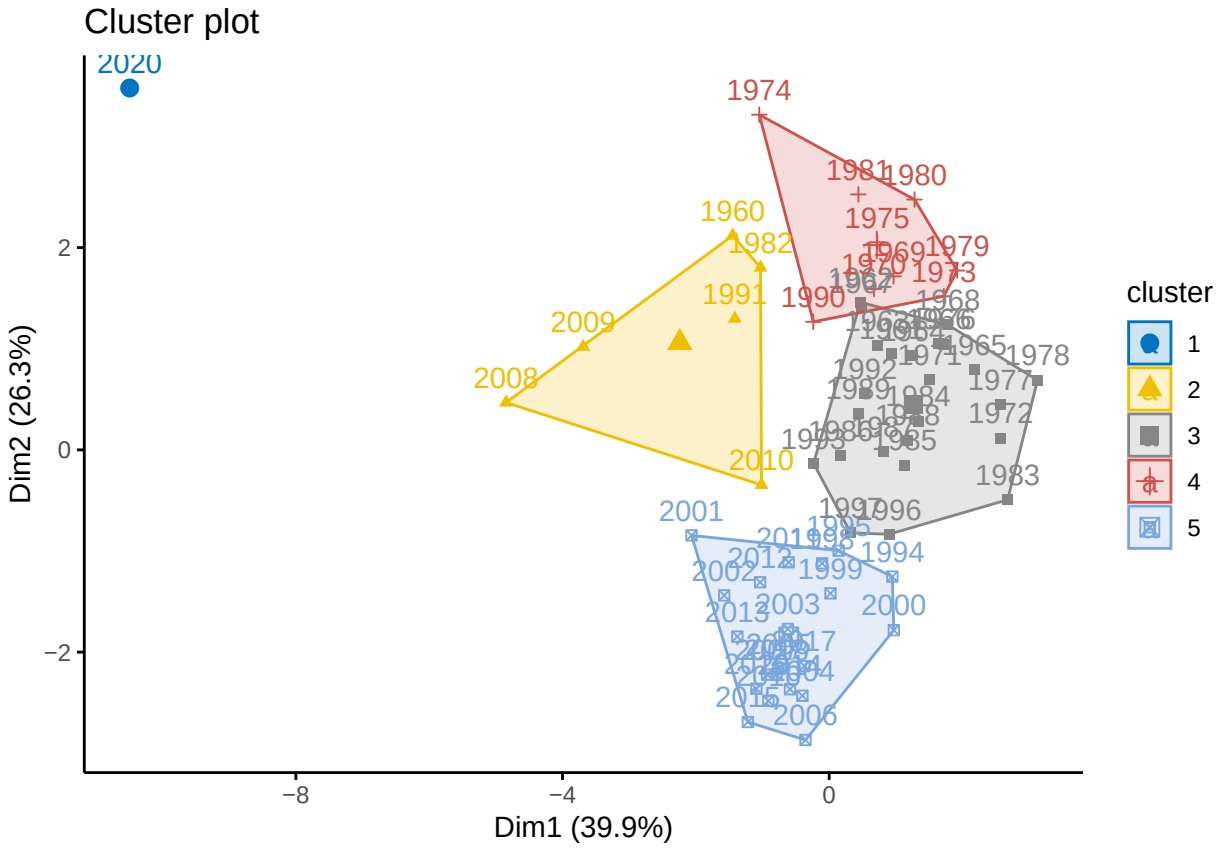


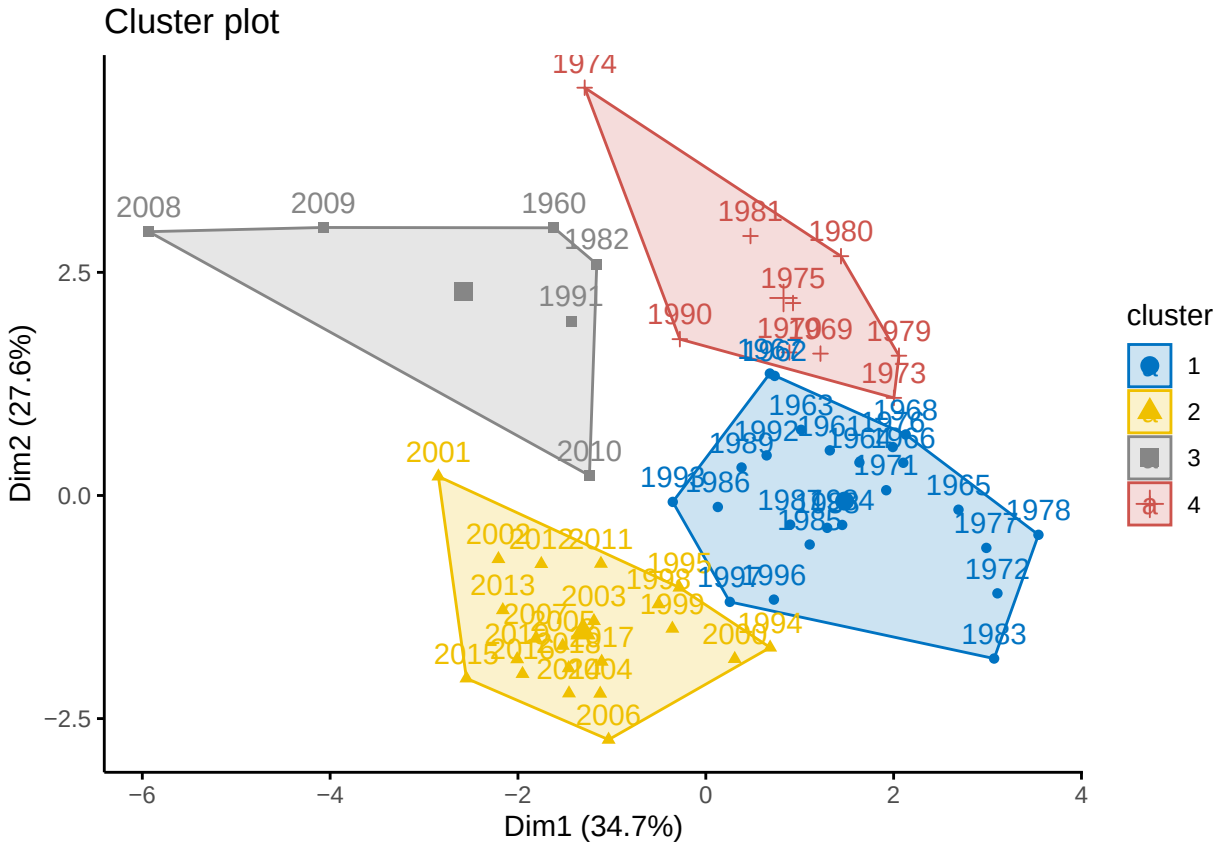
After removing the least 3 contributing variables, the results become more reasonable.



Alternative way: K-means

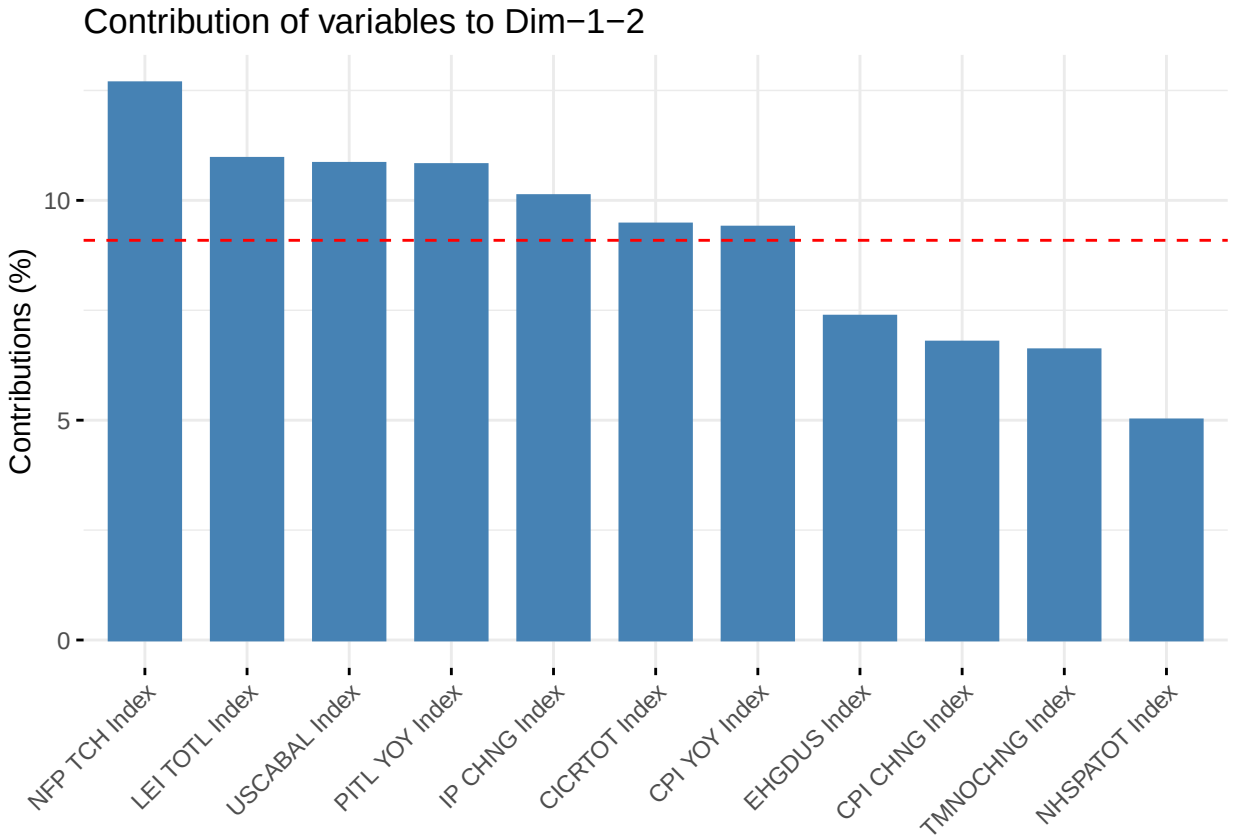
K-means clustering (MacQueen 1967) is one of the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of k groups (i.e. k clusters), where k represents the number of groups pre-specified by the analyst. It classifies objects in multiple groups (i.e., clusters), such that objects within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas objects from different clusters are as dissimilar as possible (i.e., low inter-class similarity).


Cluster plot
Dim2 (26.3%)
Dim1 (39.9%)
cluster
1
2
3
4
5

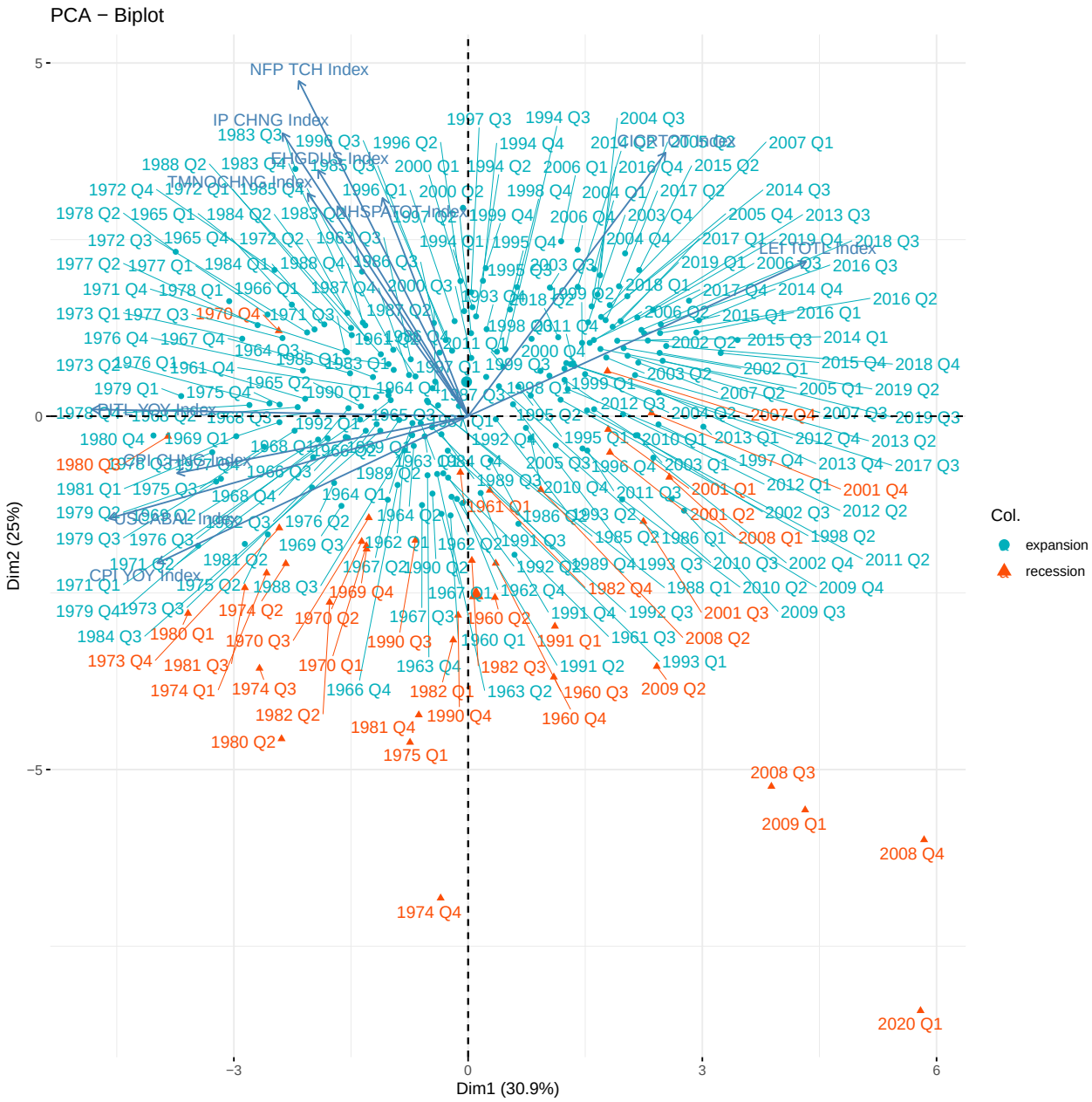

Cluster plot
Dim2 (27.6%)
Dim1 (34.7%)
cluster
1
2
3
4

Quarterly Data Analysis

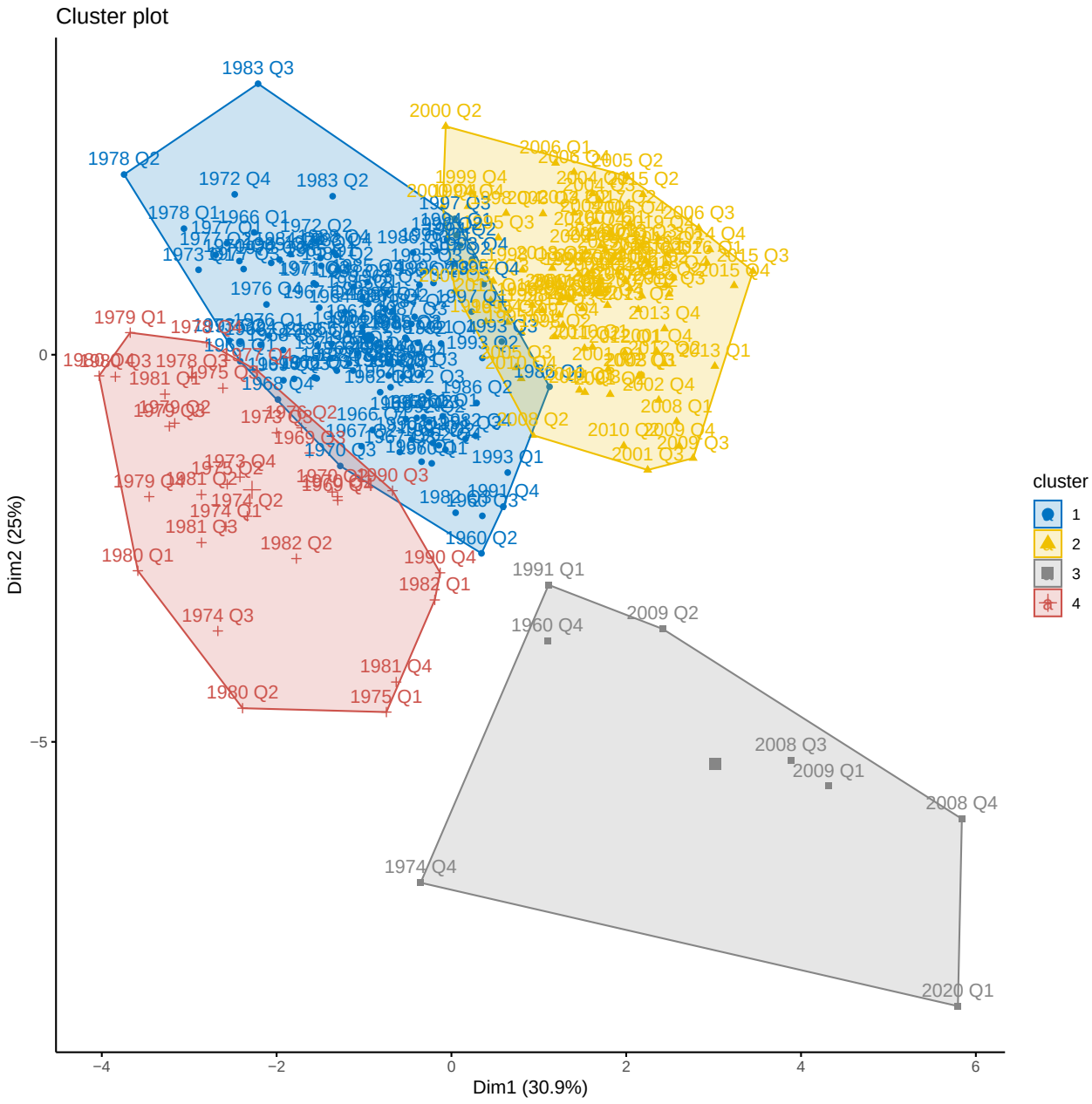
For the quarterly data, we can keep 11 economic indices since even the contribution of the least important index is above 5%



Biplot of quarterly data

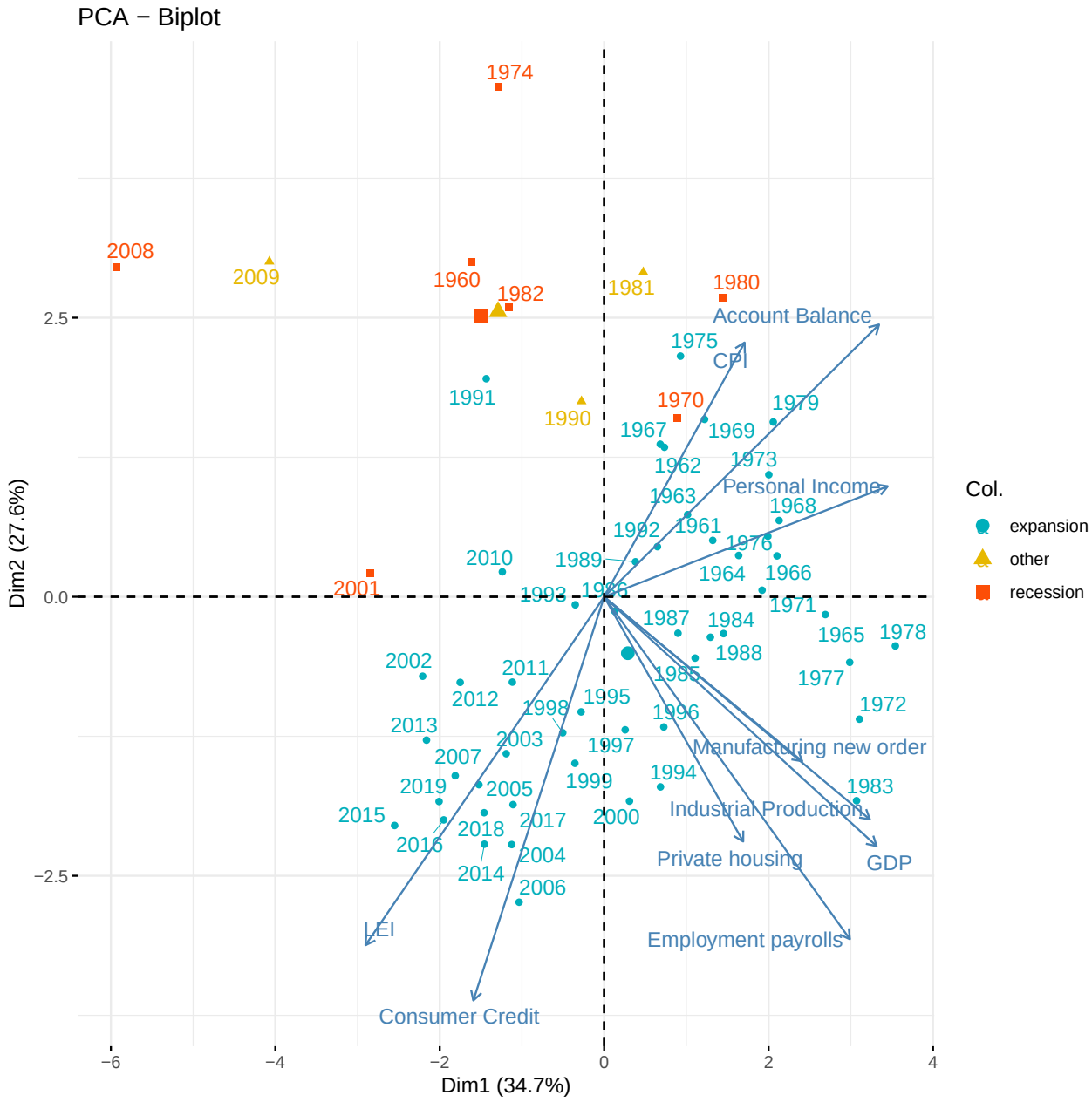


K-means clustering



Conclusion

1. 2020 (Covid-19 recession) is a new class of economic regime, the scale of which is unprecedented.
    - The first quarter of 2020 is relatively closed with 2009 Q1, 2008 Q3-Q4, 1974 Q4



2. PCA divided the data into four economic regimes:
   - recession : 2008, 2009, 2001, 1982, 1974, 1960...
   - strong economic expansion: 1983-1989
   - stagnation (slow economic growth): 2011-2019
   - moderate recovery: 1961-1969
3. The results of K-means is similar to PCA, however, PCA is a better approach since it measured the scales for each observation in all PC loading vectors.

Future Work

- Find more representative economic indicators other than the 13 economic indices used in the study, while continue removing redundant variables.

- Assign different asset allocation strategy to each economic regime, and backtest the portfolio allocation strategy.
- In terms of long-term forecasting, we can consider modeling the long-term trend on each economic indicator. Then the long-term economic regime can be predicted once we know the direction of economic indicators.
- More advanced machine learning approaches can be applied. e.g. Kernel PCA; Sparse PCA, Independent Component Analysis, Multidimensional Scaling and so on