PCA on different asset classes

Are there specific topics driving asset prices?

Sophie Grill, Sebastian Herzog, Alexei Volodin,
Arina Suhodolova, Dinara Zainullina

19.12.2023

High Level Roadmap

The research problem can be divided into the following **three steps:**
(1) Data Preparation & Brainstorming
(2) Statistical analysis & macro-economic interpretation
(3) Visualization & Next Steps

High Level Roadmap (Part 1)

I. Data Preparation & Brainstorming

1. Data Preparation

a. Define theoretically suitable Data

Economic and Capital Markets related time series across all asset classes, especially Factor-based Time-Series

b. Identify Data Sources

Bloomberg, Yahoo Finance, Kenneth R. French Library on Stock Return Factors, etc.

c. Data Preparation & Quality Assurance

2. Statistical Methods Brainstorming

Pro & Contra of e.g. Hidden Markov-Chains, Principal Component Analysis, Bayesian Nets, Neural Nets, etc.

Devise Long-List and most promising short list of suitable methods

High Level Roadmap (Part 2)

In progress

II. Statistical Analysis & Macro-economic Interpretation

a. Application of short-listed models to data, identify issues & solutions and come up with macro-economic interpretation of results

b. Time-Series Regression of Principal Components onto macro-economic/ Factor-Portfolios

III. Visualization of Theme Evolution through time

Outcome

- Fully integrated R Code (Data Input, Data Quality Checks, Statistical Analysis, Output)
- Sensitivity Assessment: Which asset classes are more heavily influenced by the identified topics, which are defensive safe havens?
- Interactive visualization dashboard / web application (e.g. R Shiny, Power BI) of "Driving Topics" through time (incl. conditional correlations)

1. Data Preparation

a) Define Theoretically Suitable Data

Chosen time series across multiple asset classes: commodity prices, bond indices, spread indices, equity indices, FX rates, as well as macro data such as CPI rates, unemployment rates, real GDP (%)

Time horizon: last 20 years

Frequency: daily

b) Identify Data Sources

Bloomberg and Kenneth R. French Library

c) Data Preparation and Quality Assurance

2. Statistical Methods Brainstorming

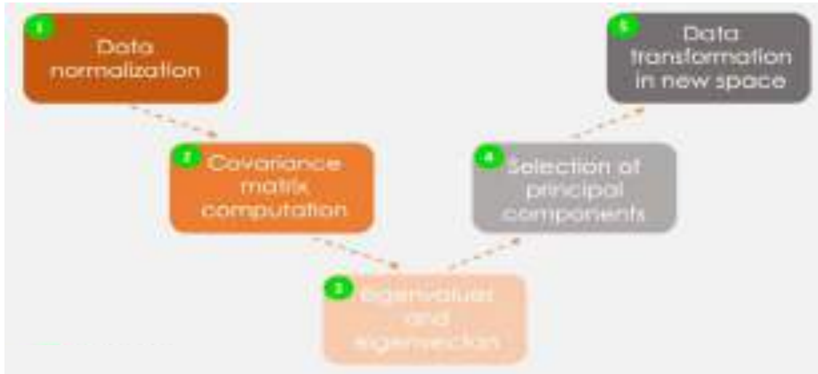
- Principal Component Analysis as a chosen statistical method for analysing high-dimensional data and capturing the most important information from it (principal components/ potential ”drivers” of asset prices)

- This is done by transforming the original data into a lower-dimensional space while collating highly correlated variables together

- Main advantages: Dimensionality Reduction, Multicollinearity Mitigation, Pattern Recognition

- Possible obstacles to be addressed: Interpretability, Sensitivity to Outliers

PCA in 5 Steps



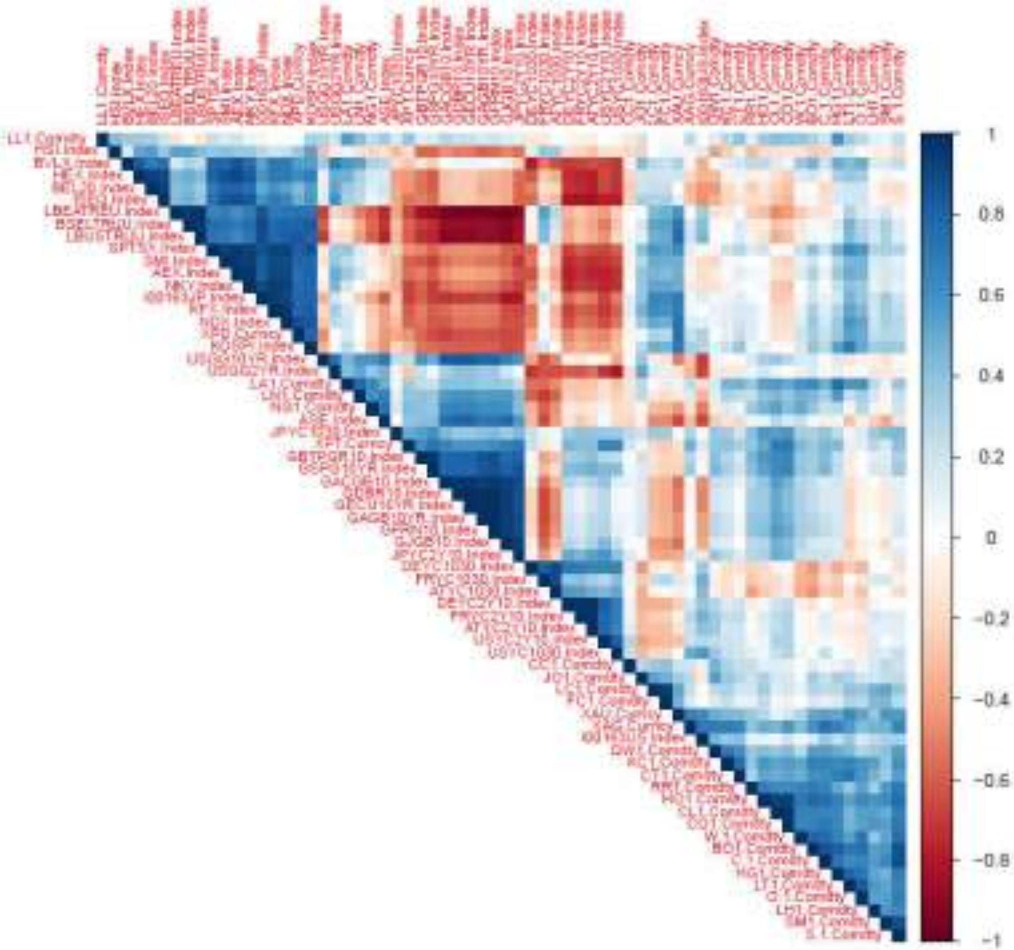
- **Step 1 - Data normalization**
  - Created log returns and normalized them
  - Attributes them on same level, no bias
- **Step 2 - Covariance matrix**
  - symmetric matrix, each element (i, j)
  - corr. to the covariance between variables i/j.
- **Step 3 - Eigenvectors and eigenvalues**
  - **Eigenvector** represents direction. An **eigenvalue** is a number representing the amount of variance present in the data for a given direction. Each eigenvector has its corresponding eigenvalue.
- **Step 4 - Selection of principal components**
  - Data variables determine the pairs of eigenvectors and eigenvalues. In our data are 76 columns (excluding macro data), hence 76*5721 pairs. Not all the pairs are relevant. So, the eigenvector with the highest eigenvalue corresponds to the first principal component.
- **Step 5 - Data transformation in new dimensional space**
  - re-orienting the original data onto a new subspace defined by the principal components This reorientation is done by multiplying the original data by the previously computed eigenvectors.

Financial Data for PCA

COMMODITIES		BOND INDICES		EQUITY INDICES		GVT BOND SPREADS	
Ticker	Description	Ticker	Description	Ticker	Description	Ticker	Description
~~CL1 Comdty~~	~~Crude Oil WTI Futures~~	~~GJGB10 Index~~	~~Japanese GVT 10Y~~	~~AEX Index~~	~~Amsterdam Exchange Index~~	USYC2Y10 Index	US 2Y/10Y
~~CO1 Comdty~~	~~Brent Crude Oil Futures~~	~~GAGB10YR Index~~	~~Austria GVT 10Y~~	ASE Index	Athens Stock Exchange General Index	DEYC2Y10 Index	Germany 2Y/10Y
HG1 Comdty	Copper Futures	~~I00163JP Index~~	~~Bloomberg Asian Pacific (mixed) - IG~~	BEL20 Index	Euronext Brussels Index	JPYC1030 Index	Japan 10Y/30Y
LA1 Comdty	Aluminium Futures	GDBR10 Index	Germany GVT 10Y	BVLX Index	PSI General Index Portugal	~~JPYC2Y10 Index~~	~~Japan 2Y/10Y~~
LL1 Comdty	Lumber Futures	USGG10YR Index	US GVT 10Y	~~HEX Index~~	~~Helsinki Stock Exchange Index~~	USYC1030 Index	US 10Y/30Y
LN1 Comdty	Leen Hog Futures	USGG2YR Index	US GVT 2Y	~~ISEQ Index~~	~~Irish Stock Exchange Quotient~~	~~DEYC1030 Index~~	~~Germany 10Y/30Y~~
LT1 Comdty	Gasoline Futures	~~GFRN10 Index~~	~~France GVT 10Y~~	~~KFX Index~~	~~Copenhagen Stock Exchange~~	FRYC1030 Index	France 10Y/30Y
XAG Curncy	Silver Spot Price	GBTPGR10 Index	Italy GVT 10Y	~~NDX Index~~	~~NASDAQ 100~~	~~FRYC2Y10 Index~~	~~France 2Y/10Y~~
XAU Curncy	Gold Spot Price	~~GACGB10 Index~~	~~Australia GVT 10Y~~	~~SPTSX Index~~	~~S&P/TSX Composite Index, Canada~~	~~ATYC2Y10 Index~~	~~Austria 2Y/10Y~~
XPD Curncy	Palladium Spot Price	~~GSPG10YR Index~~	~~Spain GVT 10Y~~	~~SMI Index~~	~~Swiss Market Index~~	~~ATYC1030 Index~~	~~Austria 10Y/30Y~~
XPT Curncy	Platinum Spot Price	~~GECU10YR Index~~	~~EURO GVT 10Y~~	NKY Index	Nikkei 225, Japan		
HO1 Comdty	Heating Oil Futures	~~BEELTRUU Index~~	~~Bloomberg Emerging Markets (mixed) IG & HY~~	KOSPI Index	Korea Composite Stock Price Index		
NG1 Comdty	Natural Gas Futures	I00163US Index		HSI Index	Hang Seng Index, Hong Kong		
~~BO1 Comdty~~	~~Soybean Oil Futures~~	~~LBUSTRUU Index~~	~~Bloomberg US Agg. (mixed) - IG~~				
W 1 Comdty	Wheat Futures	~~LBEATREU Index~~	~~Bloomberg Euro Agg. (mixed) - IG~~				
SM1 Comdty	Soybean Meal Futures						
~~S 1 Comdty~~	~~Soybean Futures~~						
RR1 Comdty	Rough Rice Futures						
QW1 Comdty	Sugar Futures						
O 1 Comdty	Oats Futures						
LH1 Comdty	Live Hog Futures						
LC1 Comdty	Live Cattle Futures						
KC1 Comdty	Coffe C Futures						
JO1 Comdty	Orange Juice Futures						
~~FC1 Comdty~~	~~Feeder Cattle Futures~~						
CT1 Comdty	Cotton No. 2 Futures						
CC1 Comdty	Cocoa Futures						
C 1 Comdty	Corn Futures						

Financial Data – Correlation Matrix

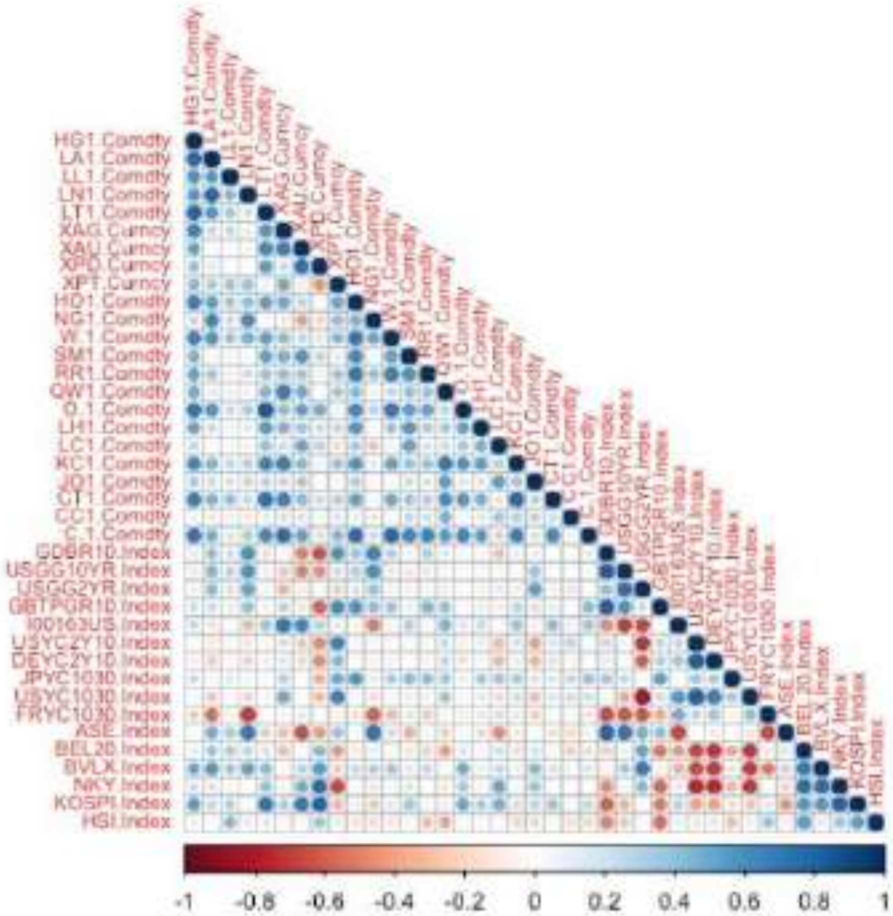
- Correlation matrix before handling multicollinarity: 66 variables



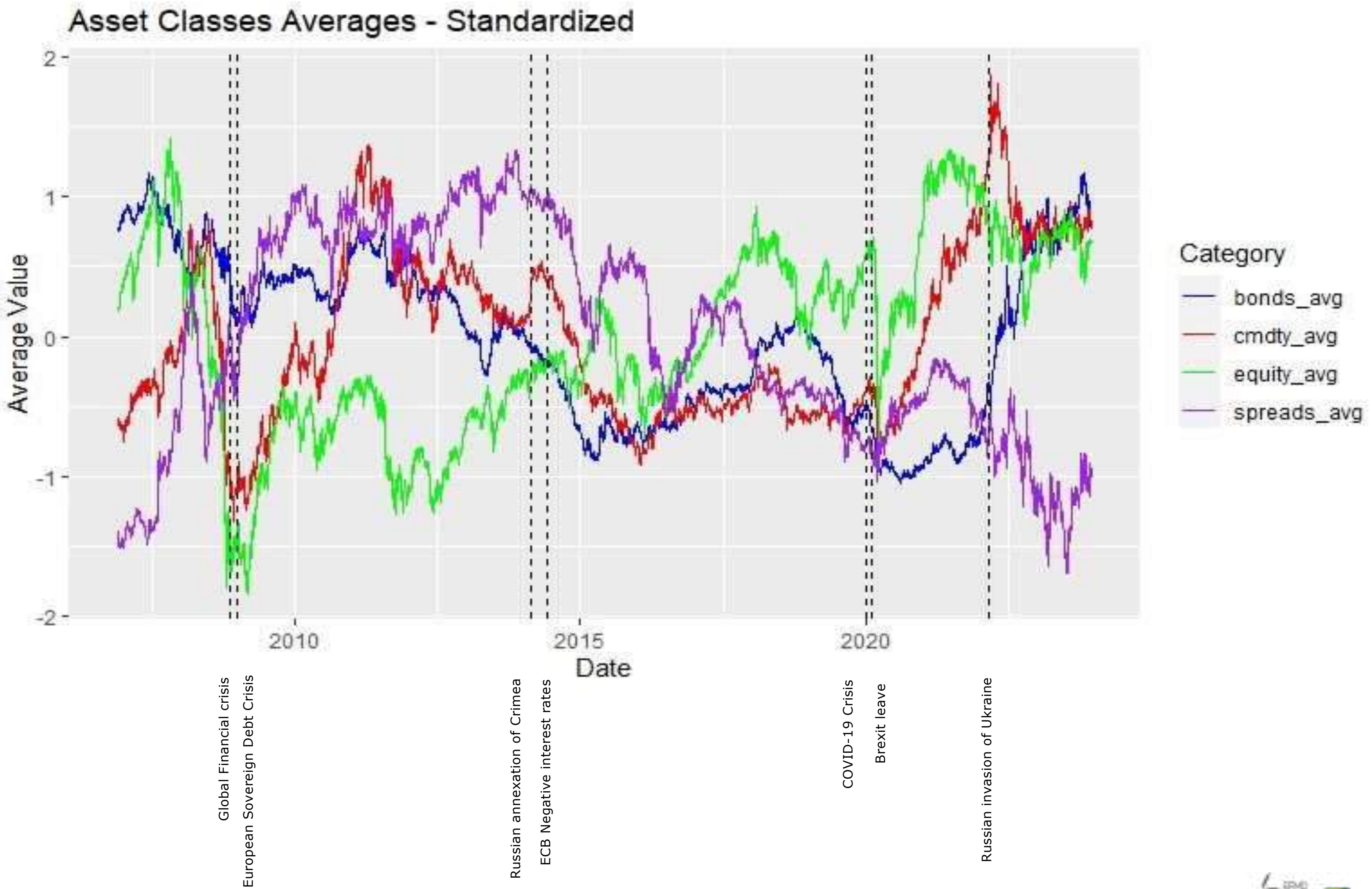
Financial Data – Correlation Matrix

Deletion of highly correlated/redundant variables

- Correlation threshold: 0.85
- Remaining variables: 39



Time Series Plot (normalized)



KMO test

Kaiser-Meyer-Olkin (KMO) Test measures sampling adequacy and assesses the suitability of data for factor analysis.

- **Purpose:**
  - Determines if the partial correlations among variables are small.
  - Indicates if factor analysis is likely to be informative.
- **How KMO Works:**
  - Calculates proportion of variance among variables that might be common variance.
  - Values range from 0 to 1.
- **Computation of KMO:**
  - Uses an anti-image correlation matrix.
  - Compares the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients.
- **KMO for Individual Items:**
  - Can also be computed for individual variables (KMO Measure of Sampling Adequacy).
  - Helps identify which variables to drop if overall KMO is low.
- **Significance of KMO:**
  - High KMO values indicate a high potential for factor analysis to yield distinct and reliable factors.
  - Low KMO suggests that factor analysis may not be appropriate.

KMO results

- Use of **Kaiser-Meyer-Olkin (KMO)** for different rolling windows
  - Optimal window size: **423 days**
  - Threshold: 0.9
    - From: Kaiser, Henry F. 1974. "An Index of Factorial Simplicity." Psychometrika 39 (1): 31–36.

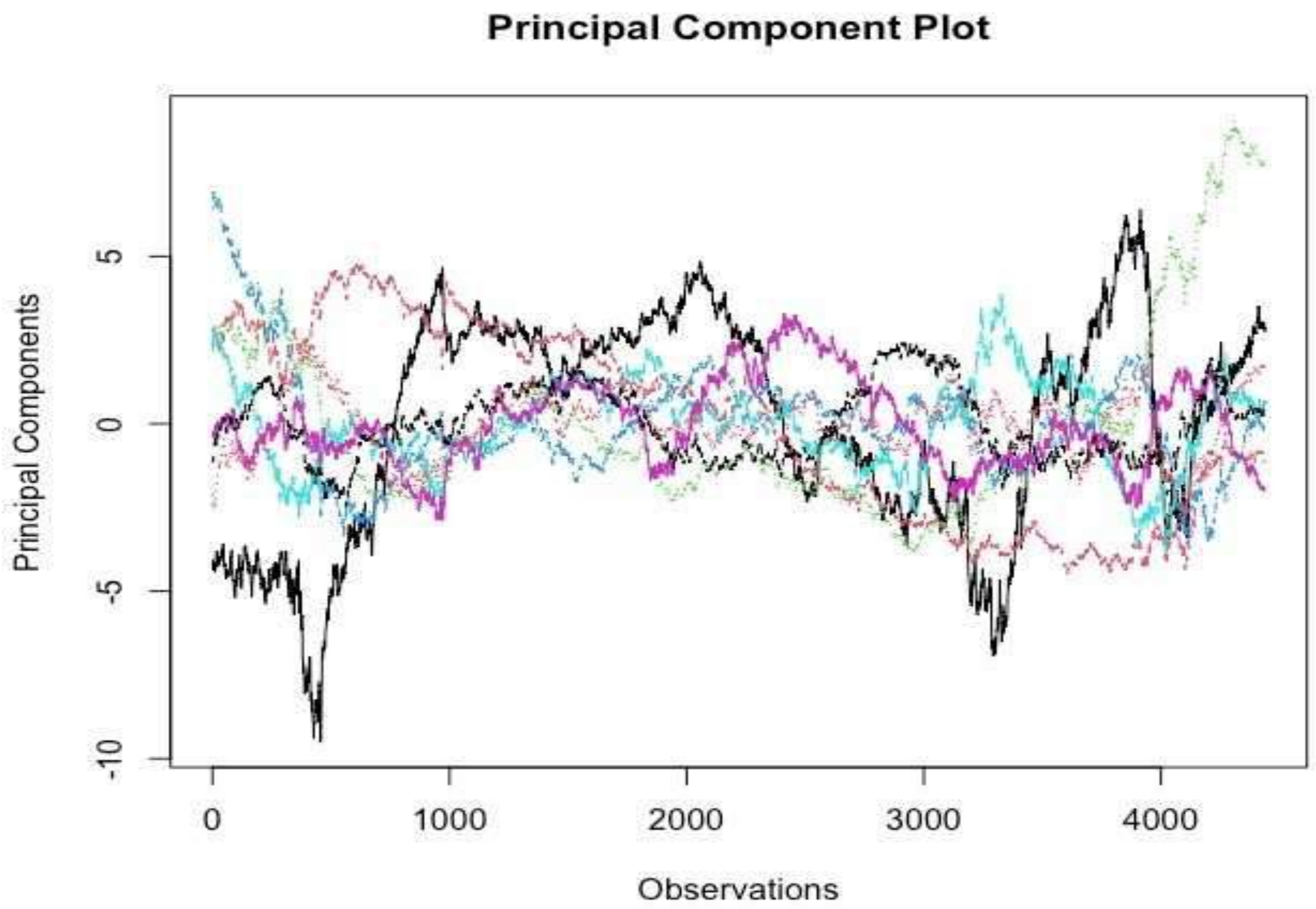
```
> kmo_result
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = df_temp)
Overall MSA =  0.9
MSA for each item =
    HG1.Comdty     LA1.Comdty     LL1.Comdty     LN1.Comdty     LT1.Comdty     XAG.Curncy     XAU.Curncy
          0.87           0.95           0.84           0.93           0.87           0.88           0.89
    XPD.Curncy     XPT.Curncy     HO1.Comdty     NG1.Comdty     W.1.Comdty     SM1.Comdty     RR1.Comdty
          0.96           0.78           0.94           0.92           0.95           0.64           0.75
    QW1.Comdty     O.1.Comdty     LH1.Comdty     LC1.Comdty     KC1.Comdty     JO1.Comdty     CT1.Comdty
          0.95           0.88           0.91           0.95           0.93           0.94           0.96
    CC1.Comdty     C.1.Comdty   GDBR10.Index USGG10YR.Index  USGG2YR.Index GBTPGR10.Index I00163US.Index
          0.94           0.92           0.91           0.86           0.88           0.91           0.93
USYC2Y10.Index DEYC2Y10.Index JPYC1030.Index USYC1030.Index FRYC1030.Index      ASE.Index    BEL20.Index
          0.81           0.92           0.80           0.64           0.67           0.93           0.86
    BVLX.Index      NKY.Index    KOSPI.Index      HSI.Index
          0.62           0.92           0.83           0.88
```

PCA Results – full event window
Summary

- **39** principal components have been generated (Comp.1 to Comp.39)
- In the **Cumulative Proportion** section, we see that the first 3 principal components explain almost 70% of the variability.

```
> summary(full_pca)
Importance of components:
                          PC1    PC2    PC3     PC4    PC5     PC6     PC7     PC8     PC9    PC10   PC11
Standard deviation     3.2348 2.8415 2.7786 1.69726 1.3423 1.28342 1.06322 0.94449 0.82457 0.71268 0.6369
Proportion of Variance 0.2683 0.2070 0.1980 0.07386 0.0462 0.04224 0.02899 0.02287 0.01743 0.01302 0.0104
Cumulative Proportion  0.2683 0.4753 0.6733 0.74716 0.7934 0.83559 0.86458 0.88745 0.90488 0.91791 0.9283
```

Plotting the first 8 Principal Components



PCA Results – Rolling event window

```
> print(results_table)
  Window     PC1      PC2      PC3      PC4      PC5
1    423 72.5698 5.624469 4.292571 3.197415 2.235493
```

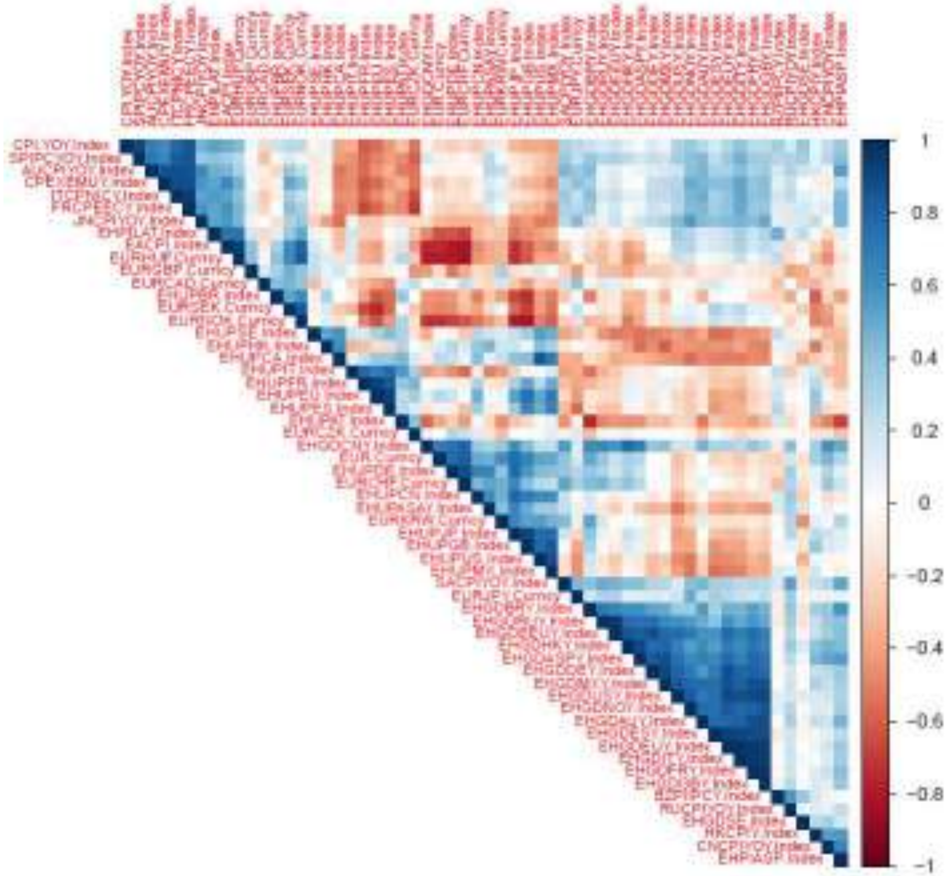
- Variance explained by each Principal Component

Macroeconomic Data for Regression

CPI Index (monthly)		UNEMPLOYMENT RATE (quarterly)		REAL GDP % (yearly)		CURRENCY (daily)
Ticker	Country	Ticker	Country	Ticker	Country	Ticker
CPI YOY Index	US	~~EHUPUS Index~~	~~US~~	EHGDDEY Index	Germany	EURJPY Curncy
CNCPIYOY Index	China	EHUPCN Index	China	~~EHGDUSY Index~~	~~US~~	EURCHF Curncy
JNCPIYOY Index	Japan	~~EHUPJP Index~~	~~Japan~~	~~EHGDEUY Index~~	~~EU~~	EURGBP Curncy
BZPIIPCY Index	Brazil	~~EHUPEU Index~~	~~EU~~	~~EHGDCNY Index~~	~~China~~	EURSEK Curncy
~~EACPI Index~~	~~East Africa~~	~~EHUPDE Index~~	~~Germany~~	EHGDSE Index	Sweden	~~EURNOK Curncy~~
HKCPIY Index	Hong Kong	EHUPAT Index	Austria	~~EHGDFRY Index~~	~~France~~	EURCAD Curncy
AUCPIYOY Index	Australia	~~EHUPES Index~~	~~Spain~~	~~EHGDESY Index~~	~~Spain~~	EURCZK Curncy
~~ITCPNICY Index~~	~~Italy~~	EHUPIT Index	Italy	~~EHGDITY Index~~	~~Italy~~	~~EURHUF Curncy~~
~~FRCPEECY Index~~	~~France~~	EHUPHK Index	Hong Kong	~~EHGDGBY Index~~	~~Great Britain~~	EURKRW Curncy
EHPILAT Index	Latin America	EHUPBR Index	Brazil	~~EHGDHKY Index~~	~~Hong Kong~~	~~EUR Curncy~~
EHPIASP Index	Asian Pacific	~~EHUPGB Index~~	~~Great Britain~~	~~EHGDASPY Index~~	~~Asian Pacific~~	
~~CPEXEMUY Index~~	~~EUROZONE~~	EHUPSE Index	Sweden	~~EHGDBRY Index~~	~~Brazil~~	
SACPIYOY Index	South Africa	~~EHUPFR Index~~	~~France~~	~~EHGDMXY Index~~	~~Mexico~~	
RUCPIYOY Index	Russia	~~EHUPMX Index~~	~~Mexico~~	~~EHGDRUY Index~~	~~Russia~~	
~~SPIPCYOY Index~~	~~Spain~~	EHUPCA Index	Canado	EHGDAUY Index	Australia	
		EHUPASAY Index	South East Asian	~~EHGDEEUY Index~~	~~Eastern Europe~~	
				~~EHGDNOY Index~~	~~Norway~~	

Financial Data – Correlation Matrix

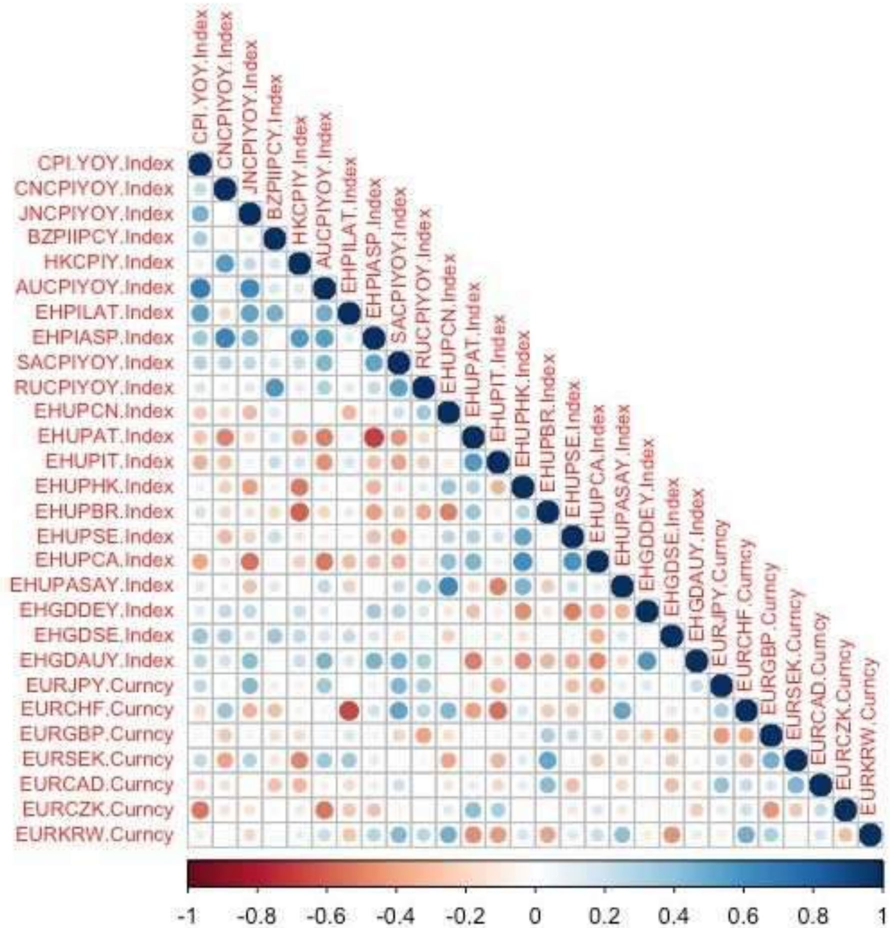
- Correlation matrix before handling multicollinarity: 58 variables



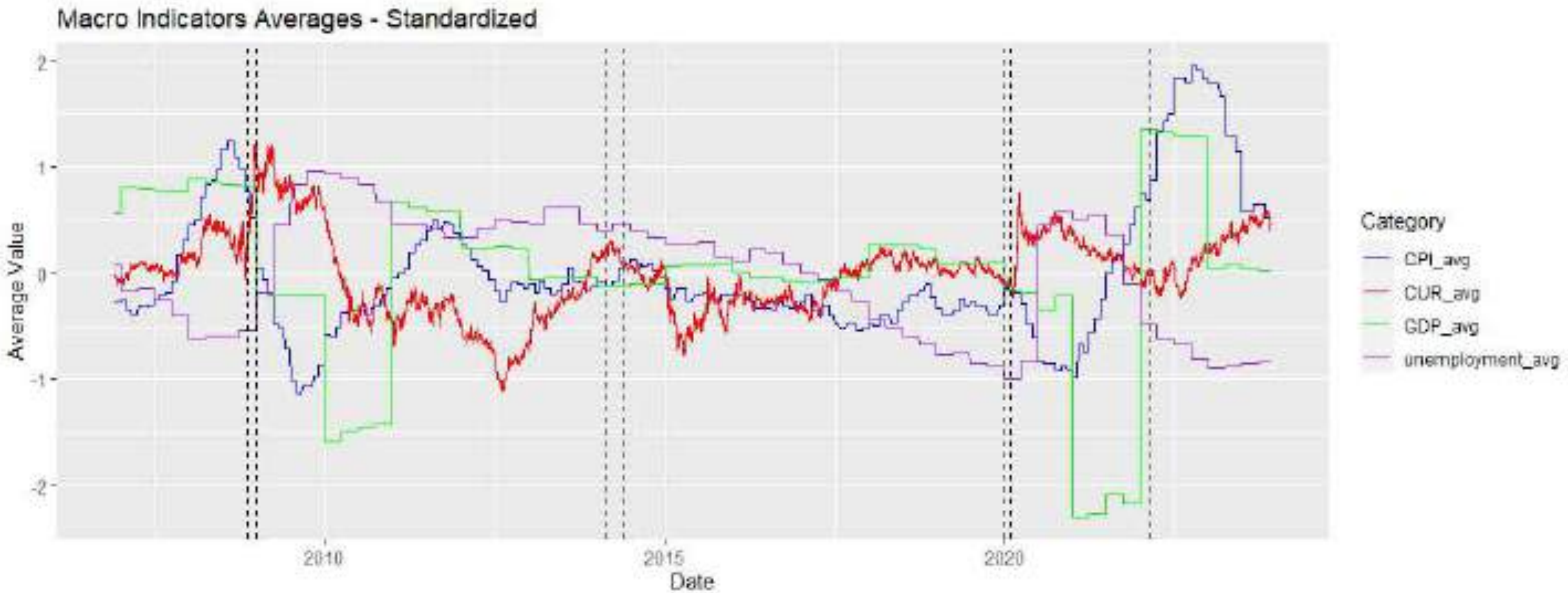
Financial Data – Correlation Matrix

Deletion of highly correlated/redundant variables

- Correlation threshold: 0.75
- Remaining variables: 28



Time Series Plot (normalized)



Global Financial crisis
European Sovereign Debt Crisis
Russian annexation of Crimea
ECB Negative interest rates
COVID-19 Crisis
Brexit leave
Russian invasion of Ukraine

Open Topics

- Need of advice how to conduct the regression (managing of ~4000 PCA)

Timeline

26.01.

- Model results (all issues of 12.01. addressed)
- discussion of economic interpretation expectations

03.02.-03.03.

- *Spring Break*

08.03.

- **First economic interpretation of the results**

22.03.

- Addressing remaining model and interpretation issues
- Discussion of dashboard visualization expectations

23.03.-07.04.

- *Easter Break*

19.04.

- **Final Model and Dashboard Visualisation (tool to be selected)**

xx.xx. (tbd)

- **Oral Presentation/Conclusion of Industry lab**

03.05.

- **Delivery of Final Report**