PCA on different asset classes

Are there specific topics driving asset prices?

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

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

22.04.2024

High Level Roadmap

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

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)

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)

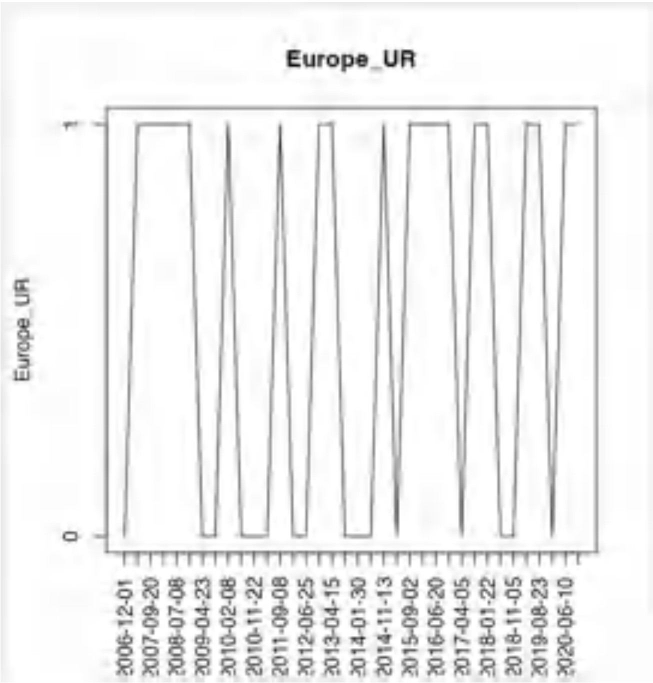
Next milestone: Final presentation

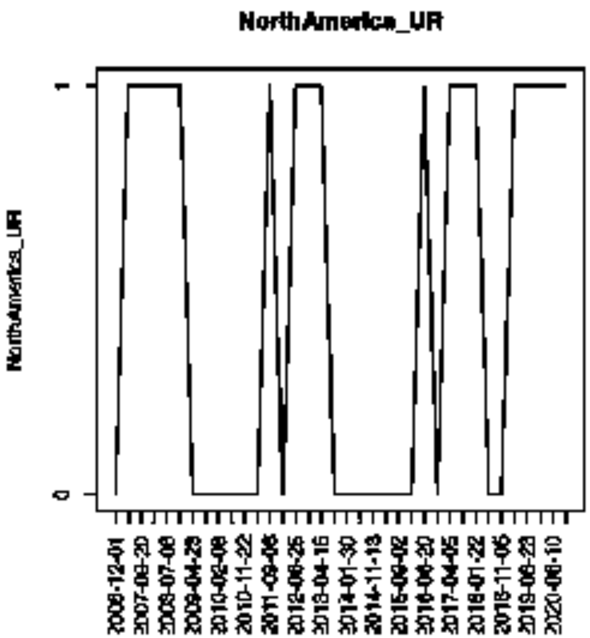
Added synthetic macro indices

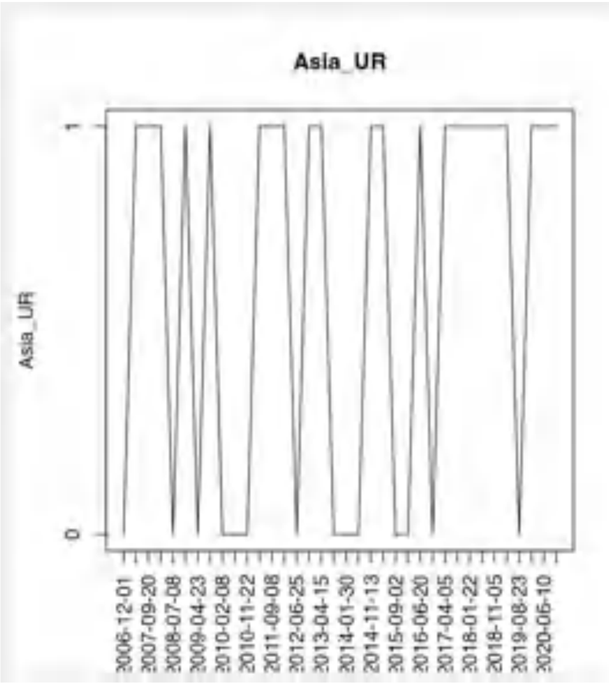
- **Europe CPI**
  - CPEXEMUY.Index, ITCPNICY.Index, FRCPEECY.Index, SPIPCYOY.Index
  - Eurozone, Italy, France, Spain
- **Asia UR**
  - EHUPCN.Index, EHUPJP.Index, EHUPHK.Index, EHUPASAY.Index
  - China, Japan, Hong Kong, South East Asia
- **Europe UR**
  - EHUPEU.Index, EHUPES.Index, EHUPIT.Index, EHUPGB.Index, EHUPFR.Index
  - EU, Spain, Italy, Great Britain, France
- **NorthAmerica UR**
  - EHUPUS.Index, EHUPMX.Index, EHUPCA.Index
  - US, Mexico, Canada
- **Developed GPD**
  - EHGDDEY.Index, EHGDUSY.Index, EHGDEUY.Index, EHGDFRY.Index, EHGDESY.Index, EHGDITY.Index, EHGDGBY.Index, EHGDMXY.Index, EHGDNOY.Index, EHGDEEUY.Index
  - Germany, US, EU, France, Spain, Italy, Great Britain, Mexico, Norway, Eastern Europe
- **Asia_GPD**
  - EHGDCNY.Index ,EHGDHKY.Index, EHGDASPY.Index
  - China, Hong Kong, Asian Pacific

Significance Test Macro Variables - Unemployment Rate

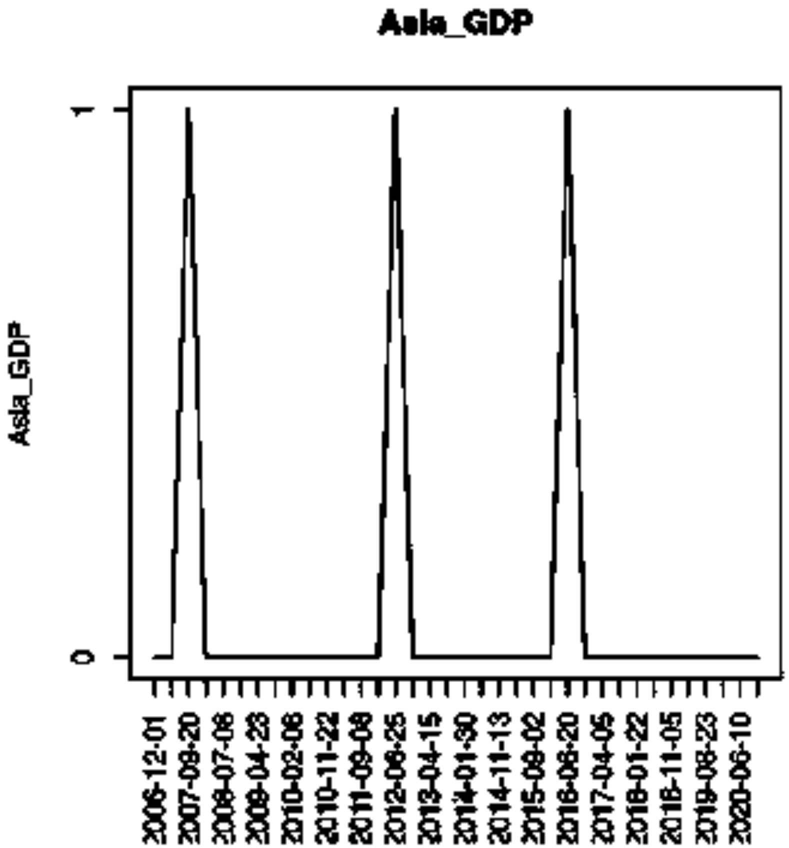
WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

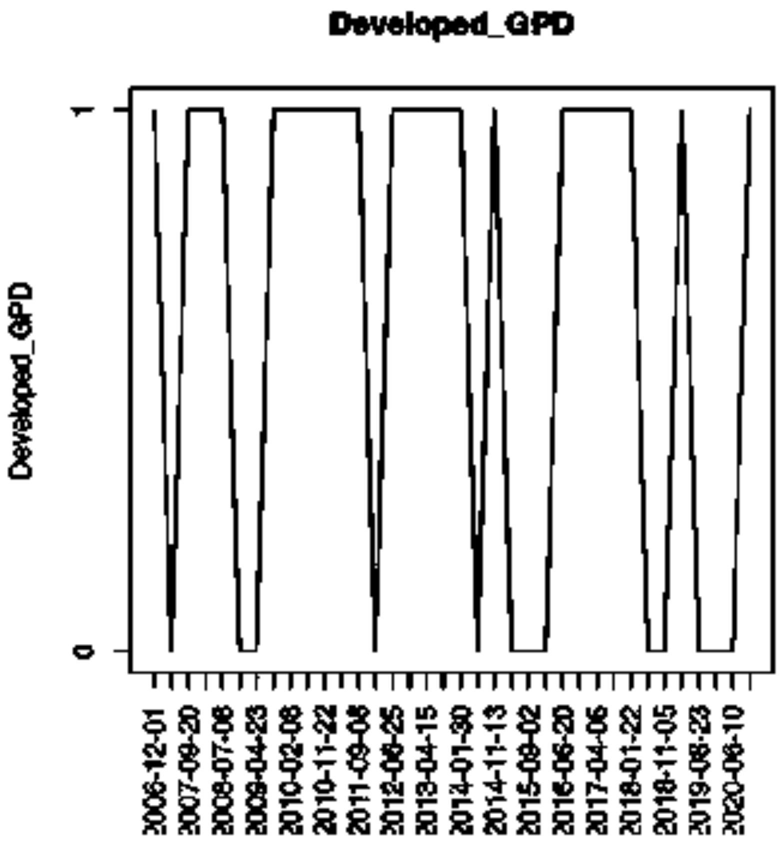




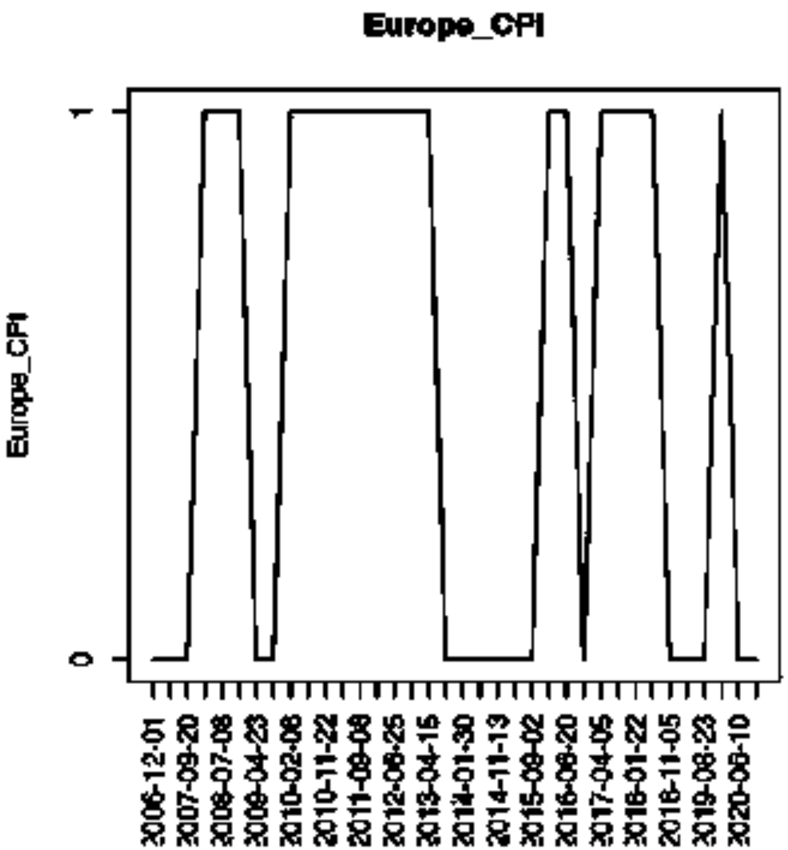


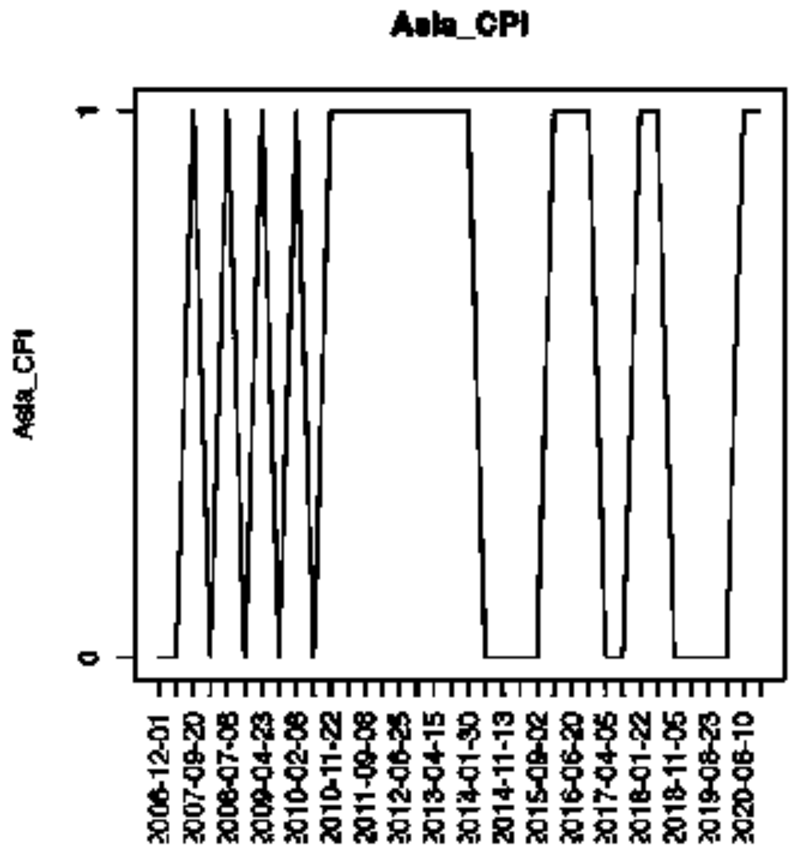
Significance Test Macro Variables - GDP



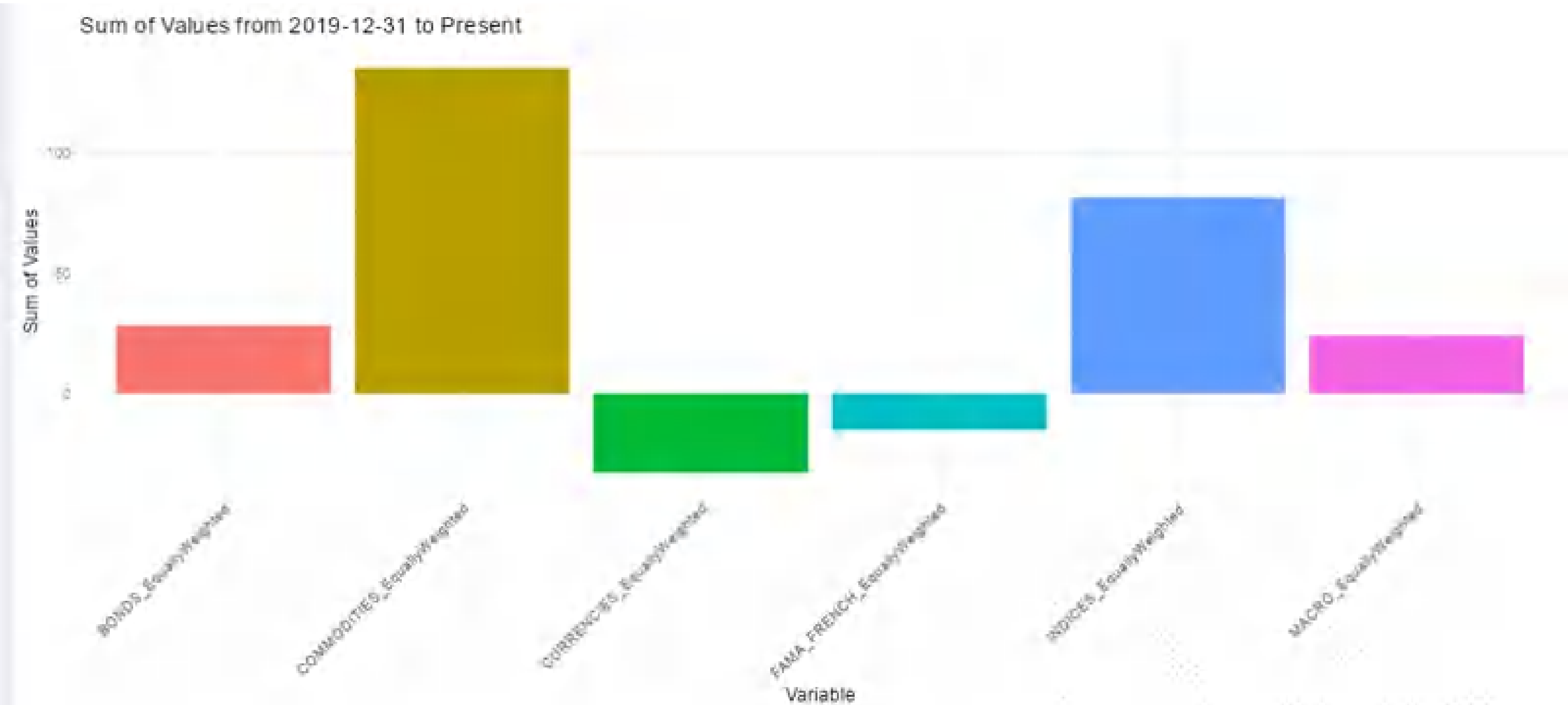


Significance Test Macro Variables - CPI





Loadings Timeframe 2019-2023 (Covid Pandemic)



Changes in the Shiny App

- No external computation anymore. All internal based on two data inputs: index and macro.
- Bug fixed in Tab Partial Regression
- Completed Shared Dataframes: PCA on Index-Tab & Significant Variables-Tab
- New tab: Significant Variables
- Renaming tabs and adjusting Paragraphs.

The runtime is only mildly affected. No delay in loading the app. Computational results are ready after 10 second depending on your CPU.

PCA vs. Other statistical techniques

Principal Component Analysis Explained

Why PCA?

Factor Analysis (FA):

Another dimensionality reduction technique similar to PCA, often used when the underlying factors influencing the data are assumed to be correlated. Like PCA, FA aims to identify latent variables that explain the correlations among observed variables.

Advantages of PCA:

In our case, we started with a variable X and wanted to explore the underlying patterns across financial time series data. PCA's orthogonality property (components are uncorrelated) are advantageous for this goal. It simplifies the interpretation of the identified components by ensuring that each captures unique variation in the data, without overlap due to correlations among factors.

Additionally, PCA is computationally less complex compared to FA, which can be beneficial when dealing with large datasets or when computational resources are limited.

Cluster Analysis:

A method used to identify groups, or clusters, of similar observations within a dataset. It could be considered to identify distinct groups of financial assets or time periods based on their characteristics.

Advantages of PCA:

While Cluster Analysis is effective for identifying groups within data, it may not directly address our objective of uncovering underlying drivers or themes across time series. PCA, on the other hand, focuses on capturing patterns of variation across variables, making it more suitable for identifying common themes or factors that drive the behavior of our financial data.

PCA's ability to condense information into principal components allows for a more concise representation of the underlying structure in the data, facilitating interpretation and subsequent regression analysis with macroeconomic indicators.

Multiple Regression Analysis:

Multiple Regression Analysis could be used directly on our original dataset to model the relationship between financial variables and macroeconomic indicators, without dimensionality reduction.

Advantages of PCA:

While Multiple Regression Analysis provides direct estimates of the relationships between individual variables and outcomes, it may encounter issues such as multicollinearity when dealing with highly correlated predictors, as is often the case with financial time series data.

PCA addresses multicollinearity by creating orthogonal components, allowing us to capture the most significant sources of variation without the need to explicitly specify relationships between variables. This can lead to more stable regression results and enhanced interpretability, particularly when assessing how financial data load on macroeconomic indicators across different time frames.

In summary, PCA was chosen as the preferred method for our analysis due to its ability to handle high-dimensional datasets, identify underlying drivers or themes across financial time series, mitigate multicollinearity issues in regression analysis, and facilitate interpretation of relationships across different time frames. Compared to alternative methods like Factor Analysis, Cluster Analysis, or Multiple Regression Analysis, PCA offered a more suitable approach tailored to our specific objectives and dataset characteristics.