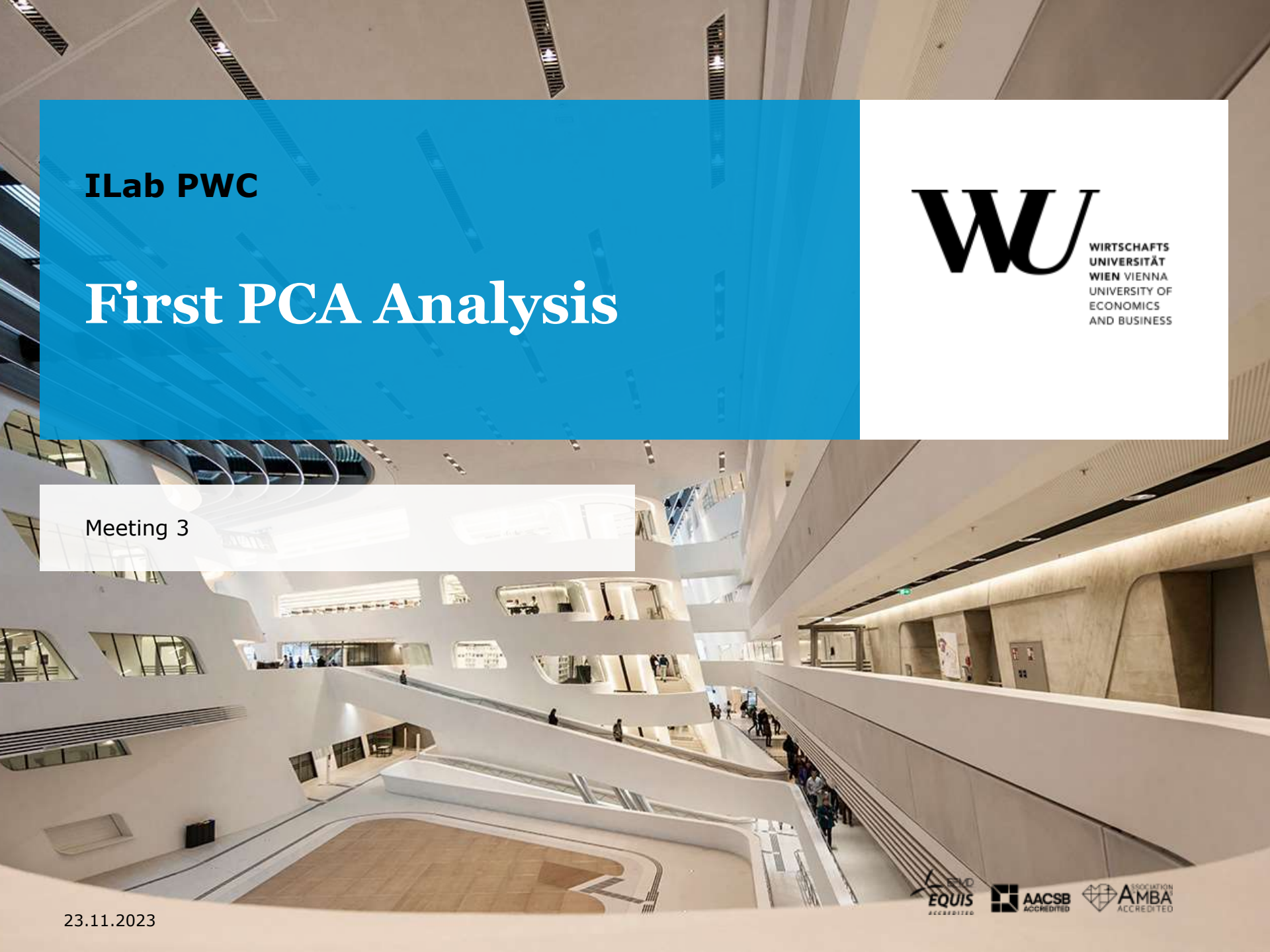


ILab PWC

First PCA Analysis

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Meeting 3

23.11.2023

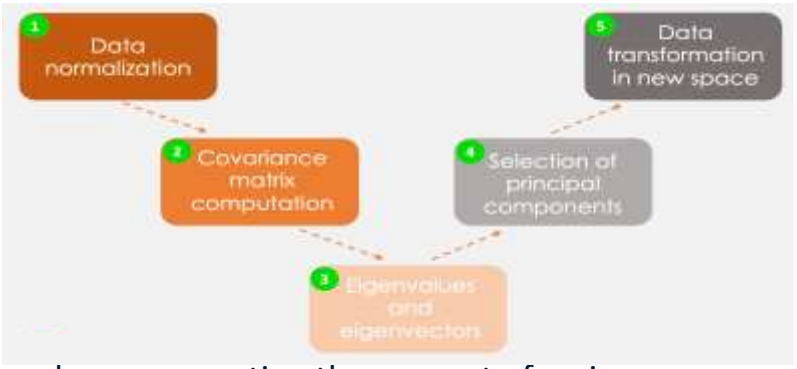
Principal component analysis (PCA)

- PCA is a statistical approach that can be used to analyse high-dimensional data and capture the most important information from it.
- This is done by transforming the original data into a lower-dimensional space while collating highly correlated variables together.

Data

- We gathered the following quarterly log price returns: A, AA , AAA, BAA, CAC40, DAX, FTSE100, IBEX35 and one log yield change: GER10Y
- Time Horizon is the last 20 Years.
- All log returns have been normalized.

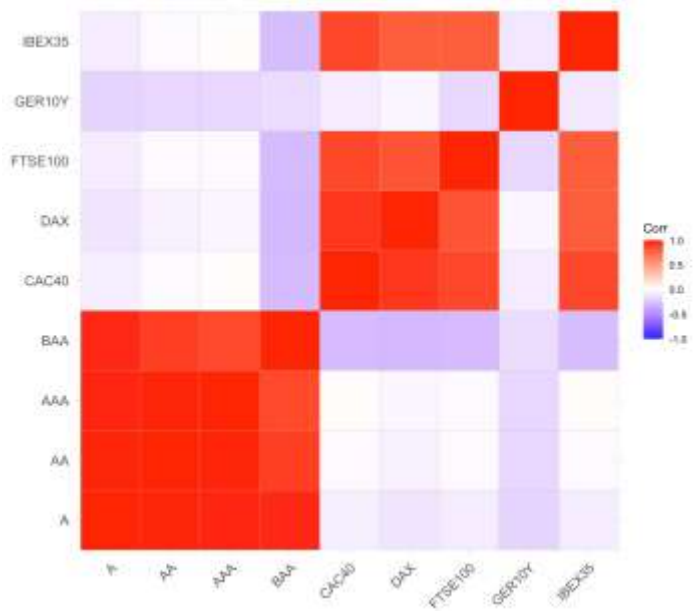
PCA in 5 Steps



- **Step 1 - Data normalization**
  - Created log returns and normalized them (Ger10Y)
  - Attributes then on same level, no bias
- **Step 2 - Covariance matrix**
  - symmetric matrix, each element (i, j) corresponds to
  - the covariance between variables i and 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 9 columns, hence 9*80 paires. 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.

→It is important to remember that this transformation does not modify the original data itself but instead provides a new perspective to better represent the data.

Descriptive Statistics



- Fixed Income Prices and Equity Indices are highly correlated among each other which is not a surprise.
- However, no great interdependence.
- Only High Yield relatively negatively correlated to equity.

Applying PCA

```
> summary(data.pca)
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4     Comp.5
Standard deviation     1.3718675 0.6067002 0.084690125 0.074143005 0.054761363
Proportion of Variance 0.8304466 0.1624186 0.003164849 0.002425648 0.001323232
Cumulative Proportion  0.8304466 0.9928651 0.996029989 0.998455637 0.999778868
                           Comp.6       Comp.7       Comp.8       Comp.9
Standard deviation     0.0179960562 1.310913e-02 2.331889e-03 1.220501e-08
Proportion of Variance 0.0001429032 7.582895e-05 2.399403e-06 6.573004e-17
Cumulative Proportion  0.9999217716 9.999976e-01 1.000000e+00 1.000000e+00
```

- **Nine** principal components have been generated (Comp.1 to Comp.9), which also correspond to the number of variables in the data.
- In the **Cumulative Proportion** section, the first principal component explains almost 83% of the total variance. This implies that almost two-thirds of the data in the set of 9 variables can be represented by just the first principal component. The second one explains 16%

Loading Matrix for PC 1 & 2

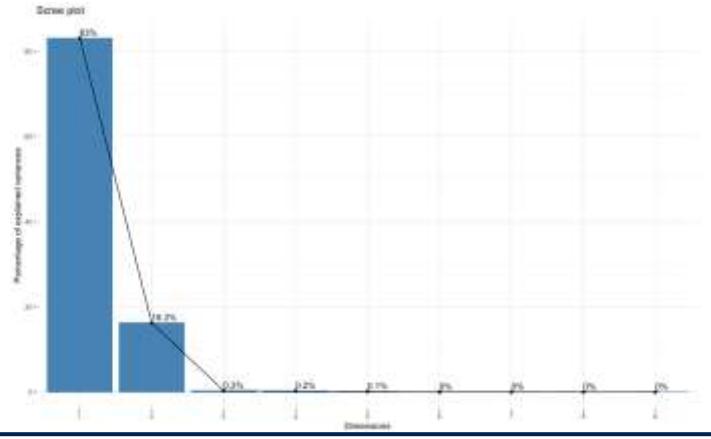
It's great to have the first two components, but what do they really mean?

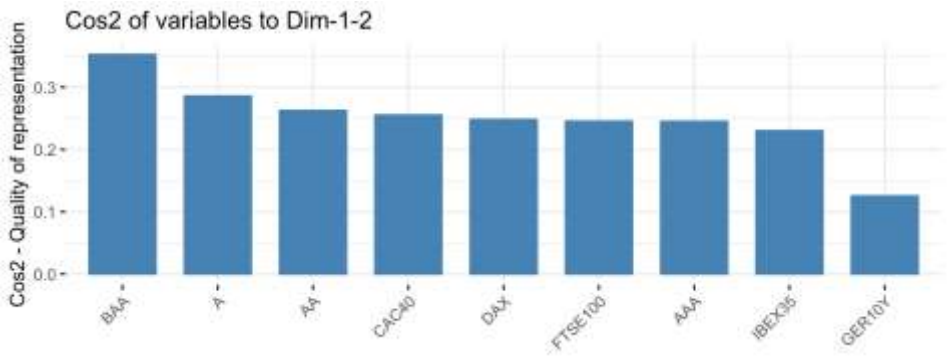
	Comp.1	Comp.2
A	0.36979515	0.2779464
AA	0.34977309	0.2964470
AAA	0.33501760	0.3027773
BAA	0.42526358	0.1820070
CAC40	-0.34155025	0.3115903
DAX	-0.34296810	0.2692544
FTSE100	-0.32849570	0.3383866
GER10Y	-0.04060333	-0.5763094
IBEX35	-0.32246346	0.3063331

- The loading matrix shows that the first principal component has high positive values for Fixed Income prices.

Visualization of the principal components

- **Scree Plot**
  - This plot shows the eigenvalues in a downward curve, from highest to lowest. The first two components can be considered to be the most significant since they contain almost 99% of the total information of the data.

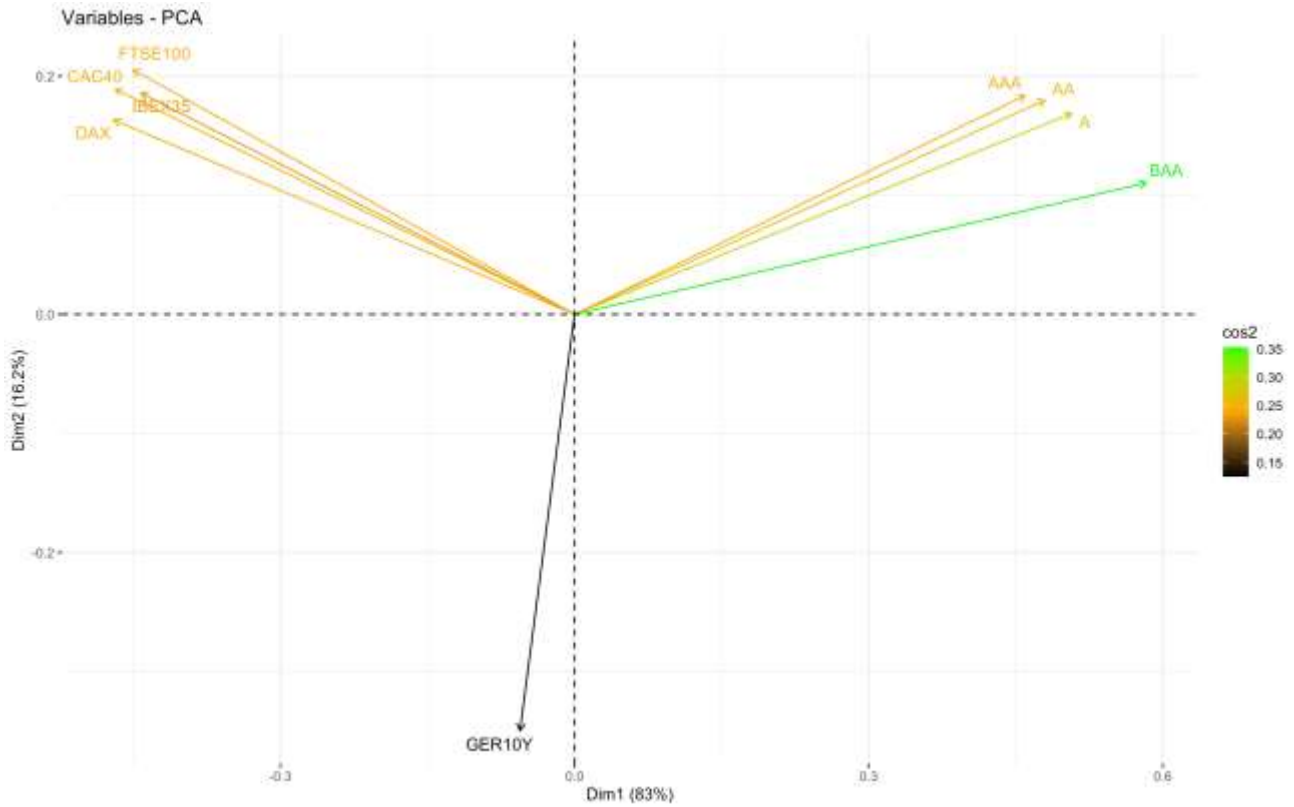




- **Contribution of each variable**
  - Determines how much each variable is represented in a given component. Such a quality of representation is called the Cos2 and corresponds to the square cosine
  - A high value, on the other hand, means a good representation of the variable on that component.

Biplot of the attributes

With the biplot, it is possible to visualize the similarities and dissimilarities between the samples, and further shows the impact of each attribute on each of the principal components.



1. All the variables that are grouped together are positively correlated to each other

2. The higher the distance between the variable and the origin, the better represented that variable is.

3. Variables that are negatively correlated are displayed to the opposite sides of the biplot's origin.