EARLY WARNING SYSTEM

- Rajaa Bakir
- Margherita Basilico
- Carlotta Cacciatori
- Tiziano Cinciripini
- Marco Lepore

Introduction:

Financial assets in the medium to long term tend to generate positive risk premiums supported by basic economic principles, such as:

-For equities, the positive return on capital invested in business activities;

-For bonds, the remuneration for deprivation of money (and consumption) and for the risk of default.

Financial markets are periodically subject to crises where asset returns do not follow these principles.

Thus, investors' appetites for risk rise and fall over time and so do the prices of financial assets.

In risk-on situations, investors have a high risk appetite and bid up the prices of risky assets in the market (equities, corporate bonds, riskier government bonds, most commodities, etc);

In risk-off situations, investors become more risk-averse and sell risky assets, sending their prices lower, and have the tendency to gravitate toward lower-risk investments (cash, short term bonds of "safe" countries, sometimes gold, etc).

This is a **high systemic risk** situation.

Problems and goals:

- Financial markets tend to crash. Market crises correspond to "risk-off" situations, in which risk premia and financial assets exhibit anomalous behavior.
- Financial institutions would like to be able to find out promptly if they are in a risk-on or risk-off situation.
- The aim is to detect crises before mostdamage has been made and to reduce false alarms.

Data:

The dataset consists in weekly time series, from the beginning of 2000 to almost today from Bloomberg, made of 874 observations classified as normal and 237 as abnormal. It contains:

- Key equity indices;
- Bond indices (Global, Corporate IG/HY, Inflation-linked, Municipals, Mortgages);
- Short/medium/long term interest rates;
- Key exchange rates;

- Commodities Leading indicators (Economic surprise, Baltic Dry Index);
- VIX (option implied volatility;
- a label that indicates if the situation is normal (0) or abnormal (1).

Methodology:

Feature selection → PCA

Feature selection → barplot

Normalization

Box-Cox transformation

Box Cox + corrections for the anomally good situations

Approach with Mahalanobis distance

Unsupervised classification

KNN classifier

KNN for anomaly detection

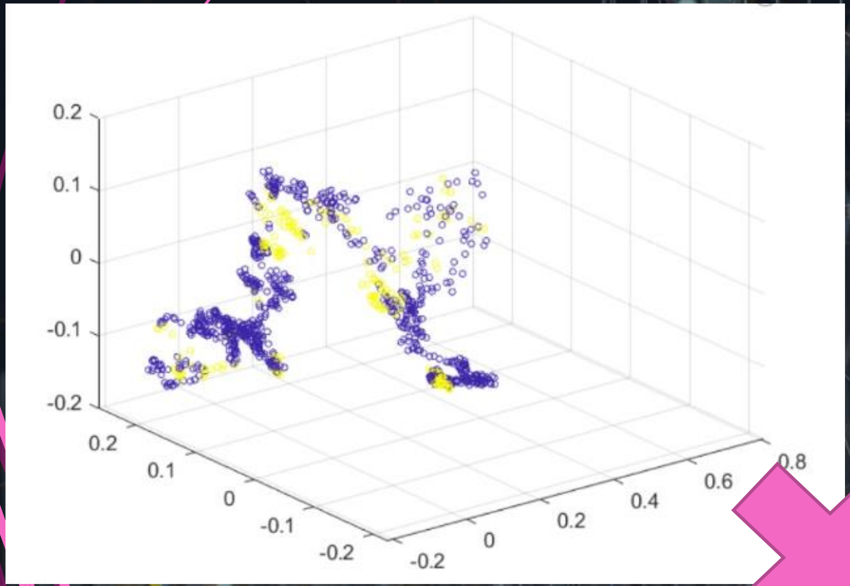
Copulae

Are ALL of these 42 features really significant in detecting risk-off situations?

Let's check with the **PCA**!

We could reduce the number of features from 42 to the first three or four, in which we have the elbow, components which alone explain more than 98% of variability.

We try to plot the three vectors of scores to find noticeable groups, but we can't identify any.



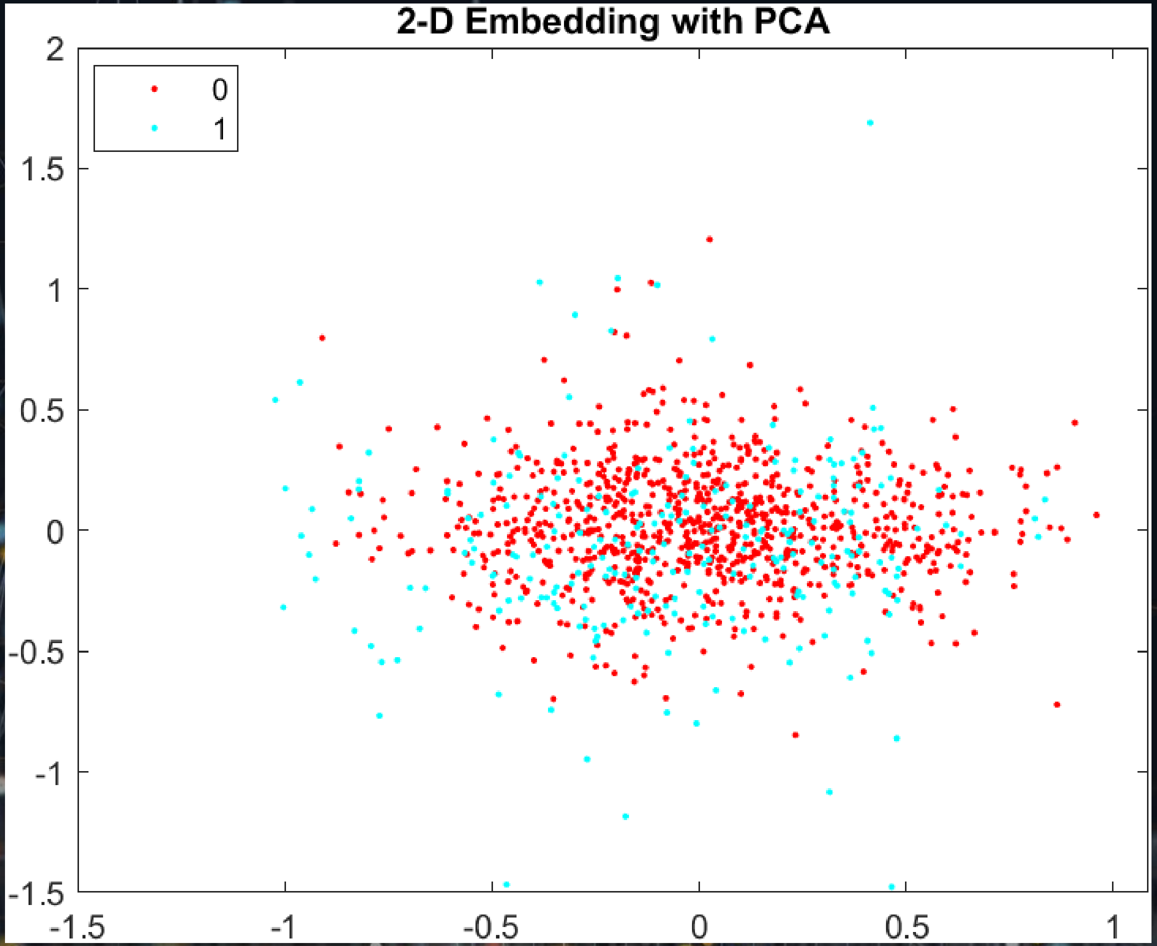
What about the differences, the log differences and stationary features?

Feature selection → PCA

WE USE:
- THE DIFFERENCES OR THE LOG DIFFERENCES (IN THE CASE THE VARIABLES WERE ALL POSITIVE)
- THE FEATURES THEMSELVES (WHEN THEY WERE ALREADY STATIONARY).

The first two principal components do not give any information about the type of the observation. In red we have the normal observations, in blue the anomalous ones. It seems that the blue dots are more towards the periphery of the sort of ellypsis and more on the bottom/left part but the results are not so evident.

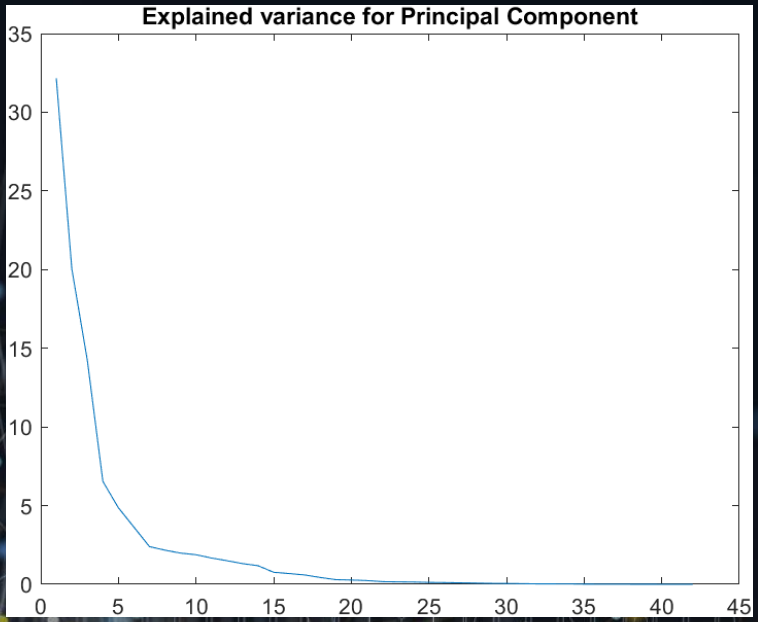
We now know that the ECSURPUS and the GTITL2YR may have an important role in the classification (and that we could also infer something about creating an algorithm that does not classify the very good situations as "risk off". The very good situation may be those on the right/top part of the graph of the PCA scores (all red points).

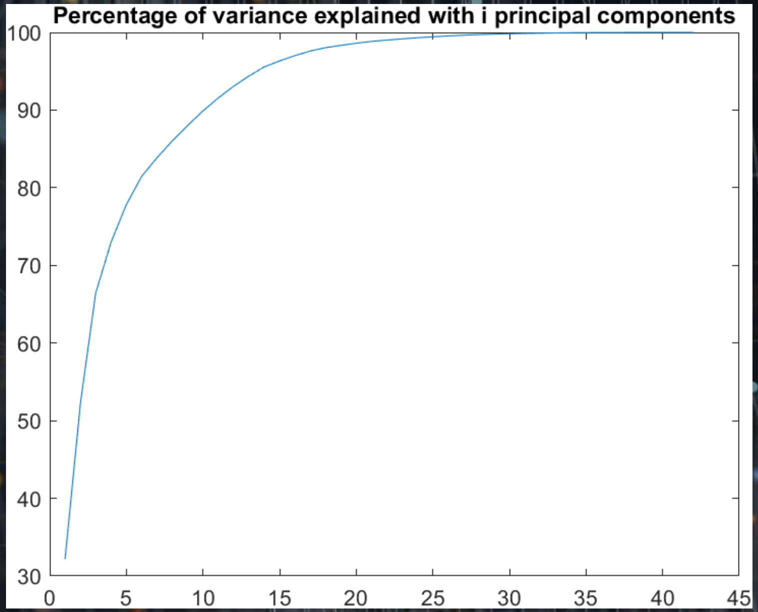


Feature selection → PCA

The explained variance is high even until about 15 components. We see an elbow in the explained percentage of variance at about 7 principal components, but if we look at the scores we cannot explain well what the first principal components represent.

We decided not to further use the PCA → difficulty to have full interpretability.

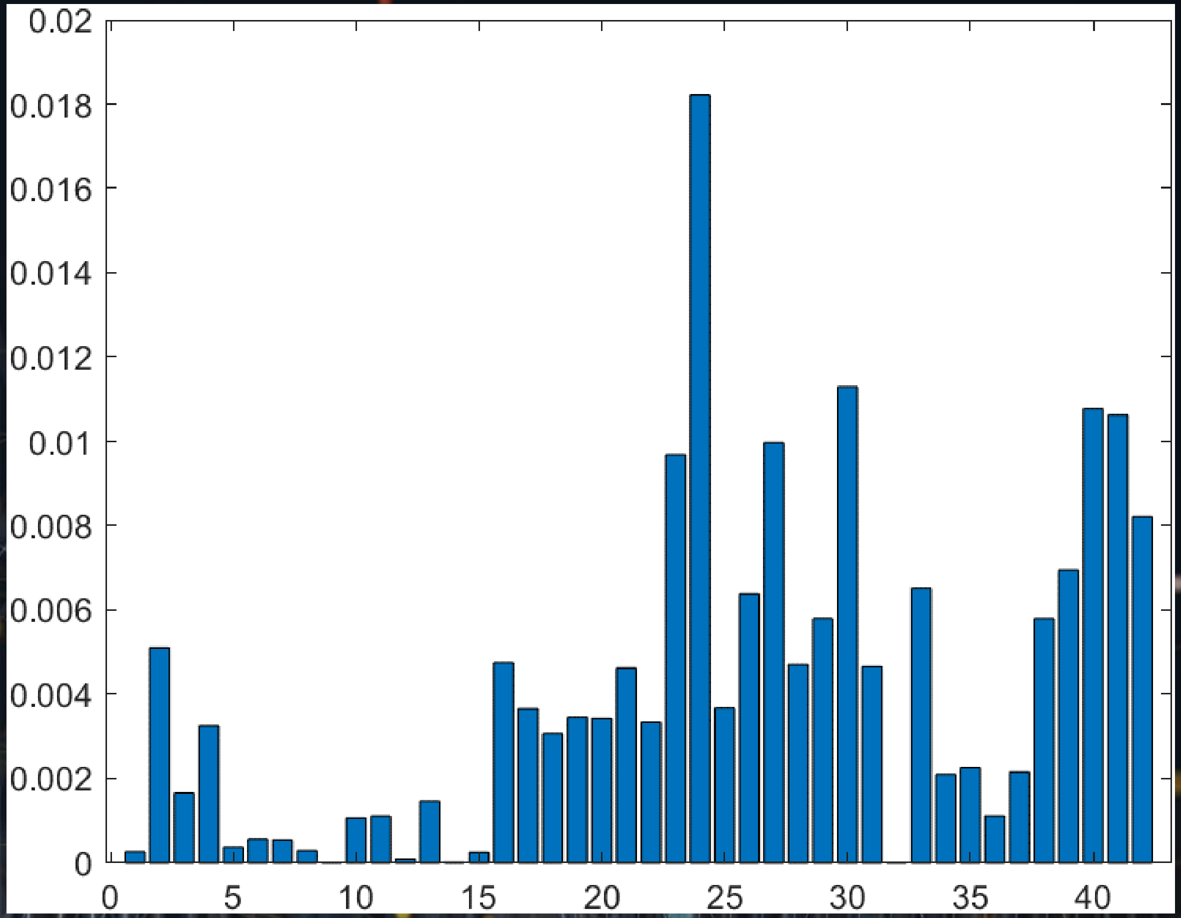




Alternative feature selection → Barplot

We tried to plot for every feature the mean of the not anomalous observations and the mean of the anomalous observations with a bar plot.

The barplot does not give any information. We now take the entire vector of the differences.



As we notice, the most marked difference is in the 24° feature, which is the **VIX**.

Normalization

We try to normalize the data and then we computed the probability for the observations in the cross validation set and in the test set to belong to a normal distribution.

We set mean and covariance matrix equal to the sample mean vector and the sample covariance matrix of the training set.

The observations are not normally distributed if we look by single feature.

Box-Cox transformation

After a Box-Cox transformation we tried the same classificator.

This is an example of NOVELTY detection, since in the training set we don't have the outliers.

The values of F1 and of the recall are good and there aren't many false negative, but there are some

false positive

We do not want this since this may lead to a **LOSS**

Box Cox
+ correction for the anomally good situations

We now consider also the situations in which the values indicate a big improvement.

With PCA we saw that the **ECSURPUS** and the **GTITL2YR** had a good impact on the variance of the dataset.

A similar reasoning may be done for the **VIX** referring to the difference in the mean of the two groups.

We set three thresholds respectively for the **ECSURPUS**, for the **GTITL2YR** and for the **VIX**, then we select only two for computational issues.

Approach with Mahalanobis distance

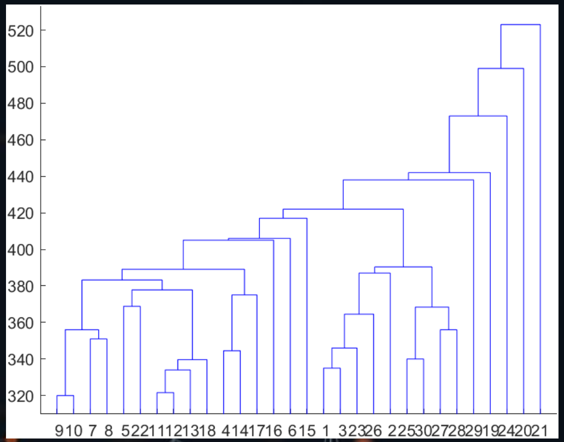
Now we do NOT normalize the data, but we classify the points in the CV set and in the Test set as anomalous if their **Mahalanobis** distance from the distribution of the training set is bigger than a threshold.

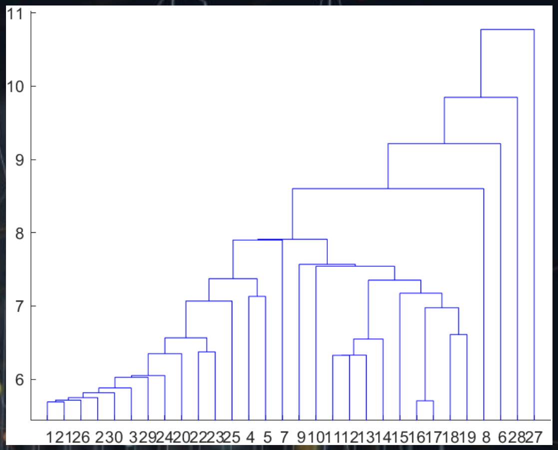
The Mahalanobis distance is unitless, scale-invariant, and takes into account the correlations of the dataset. That's why we don't need to rescale or normalize the variables.

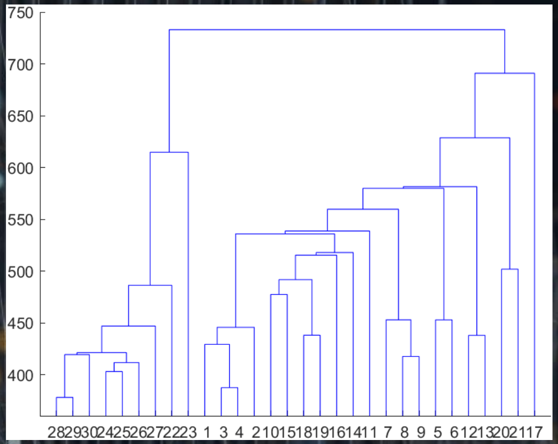
UNSUPERVISED CLASSIFICATION -hierarchiacal clustering on the whole dataset

- We tried an approach with hierachical clustering based on different distance metrics and with the original dataset. The idea is to start with each observation in its own cluster and then to combine the two closest clusters until we have one single cluster with all the observations. Then we compare it with the vector Y.

We were not satisfied with the outcome and especially with the comparison with the real labels.







«Sort of KNN»

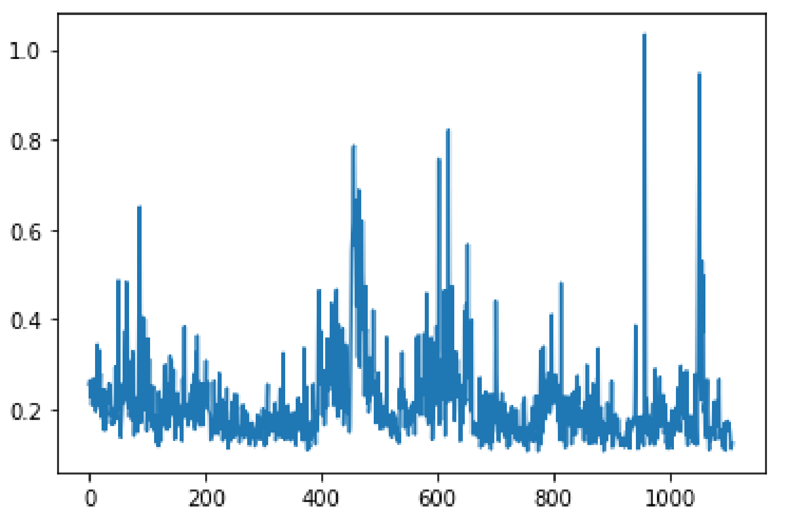
- **First**,we compute the mean of the features within the two groups.
- **Then**, we compute the distances between each unit.
- **At last**, we see whether the unit is classified into the group with the closer mean. We have decided to try with euclidean distance.

- We have to discard this approach since we have a high number of fp and fn, units that actually belong to a group different from that expected by this 'algorithm'.

KNN for anomaly detection

We used the stationary Data, but we performed anomaly detection.

The idea is to compute the k nearest neighbors for every observation, then to plot the mean of k-distances of each observation.



We want to visually determine a cutoff over which the observations are classified as anomalies.

The number of false negatives, false positives, the recall and the F1 are not good.

We discard this model !

COPULAE

We carried out a factorization of marginals and correlation, which are thus independently modeled.

We extract the sample of uniform random variables from cumulative distribution function and we fit a Gaussian copula.

The results seem good

BUT

the amount of false positive predicted on the test set is basically the same amount of not abnormal observation in it, so it basically gives the same result as a trivial classifier that assigns to everything the label 1; even worst, since it predicts also some false negative, that at least a trivial classifier doesn't.

We therefore discard the model !

FINAL RESULTS

We opted for the Box Cox transformation with the correction for the anomally good situations. The final results are not outstanding but we tried to have a good trade off between the false positives (→ risk of losing money due to the fear in the market because we think there is a crisis but we are not right) and the false negatives (→ risk of not detecting early a crisis).

```
tp = 107
fp = 50
fn = 11
prec = 0.6815
rec = 0.9068
F1 = 0.7782
```

REFERENCES:

- https://towardsdatascience.com/k-nearest-neighbors-knn-for-anomaly-detection-fdf8ee160d13
- https://www.machinelearningplus.com/statistics/mahalanobis-distance/

THANK YOU!