CRAM 1 - Programming Lab

Problem Set 5: Statistical Risk Factors and Principal Component Analysis

WS 2017/18

Prof. Dr. Maxim Ulrich
M.Sc. Elmar Jakobs

Goals

This set of exercises has four goals:

- First, using a yield curve to describe the fixed income market.
- Second, how to perform a principal component analysis.
- Third, how to determine the importance of your principal components.
- Fourth, how to use the level, slope and curvature factor to forecast the yield curve.

Submission

The solution to the problem set has to be submitted in form of two files:

- *ps5.py*: Containing the code to reproduce the quantitative results.
- A power point presentation answering the qualitative questions ($\sim$ 1-2 slide per question). Please upload it as a pdf file.

Submit your solution via the Praktomat https://praktomat.cs.kit.edu/cram_2017_WS latest 11:59 a.m. on Wednesday, 22/11/2017. Note, please name your Python file 'ps5.py' to allow for evaluation by the Praktomat and set the status along your code to SOLN.

Statistical Risk Factors and Principal Component Analysis

Question 1a: Treasury Yield Curves

Read-in the provided txt-file with the daily Treasury yields and filter out the ‘NaN’ values. Then, plot the yield curves for the following dates:

- 31st of July 2001,
- 2nd of July 2007 and
- 31st of October 2008.

Describe the three yield curves. How can you explain the different shapes?

Wikipedia: ‘In finance, the yield curve is a curve showing several yields or interest rates (**y-axis**) across different contract lengths (2 month, 2 year, 20 year, etc. ...) (**x-axis**) for a similar debt contract.’

Question 1b: PCA Intuition

The ‘Principal Component Analysis’ is an unsupervised learning tool which can be used in finance to extract statistical risk factors.

Briefly summarize the methodology and state the main goals of the PCA.

Name two specific applications of the PCA in risk and asset management.

Hint: See Chapter 7.5 of the lecture notes.

Question 1c: Class 'PCA'

Create a class 'PCA' which performs a principal component analysis for given panel data:

- Given:

$$\underbrace{X}_{T\times K} = \left[\underbrace{y^{(1)}}_{T\times 1}, ..., y^{(K)}\right] \qquad \text{K time-series of yields.}$$

- First, we demean X and call the resulting $T \times K$ matrix X^{dm}, i.e.

$$X^{dm} = \left[y^{(1)} - \bar{y}_1 \times I_{T\times 1}, ..., y^{(K)} - \bar{y}_K \times I_{T\times 1}\right].$$

- Second, get the $K \times K$ covariance matrix of all (demeaned) Treasury yields:

$$\begin{aligned}\Sigma_X &:= E[(X - E[X])'(X - E[X])] \\ &= \frac{1}{T}\, X^{dm'}\, X^{dm}.\end{aligned}$$

- Third, perform an Eigenvalue decomposition of the covariance matrix:

$$\Sigma_X \equiv E_X\ \Lambda_X\ E_X'.$$

- Fourth, calculate the covariance matrix of the transformed data:

$$
\begin{aligned}
\tilde{\Sigma}_X &:= \frac{1}{T}\,(\tilde{X}^{dm})'\,\tilde{X}^{dm} \\
&= \frac{1}{T}\,\left(\underbrace{X^{dm}}_{T\times K} \times \underbrace{P}_{K\times K}\right)'\,\underbrace{X^{dm}}_{T\times K} \times \underbrace{P}_{K\times K} \\
&= P'\,\Sigma_X\,P,
\end{aligned}
$$

with $P \overbrace{\equiv}^{PCA} E_X$.

- Fifth, perform the following rotation to obtain the transformed data:

$$
\underbrace{\tilde{X}^{dm}}_{T\times K} := \underbrace{X^{dm}}_{T\times K} \times \underbrace{P}_{K\times K}.
$$

Question 1d: Principal Components of Yields

Use the class ‘PCA’ to determine the principal components (PCs) of your panel of yields.

Plot the first PC and compare it with the original yields. Also look at the correlation matrix between PC #1 and the yields.

Then, recover the orginal yield (of Period 1) based on all principal components. Use the following relationship:

$$\underbrace{X^{dm}}_{T\times K} := \underbrace{\tilde{X}^{dm}}_{T\times K} \times \underbrace{P^{-1}}_{K\times K}.$$

Question 1e: Importance of the PCs

Use the sorted Eigenvalues from the previous exercise to determine the importance of your principal components.

E.g. for PC #1:

$$\frac{\lambda_1}{\sum_{i=1}^{K} \lambda_i}.$$

Therefore, complete the implementation for the function 'explained_variance' of your class 'PCA' and apply it to the set of PCs from the previous Question.

How many principal components would you choose to model the cross-section of yields?

Question 1f: Forecasting the Yield Curve

Research has shown that the first three principal components (level, slope and curvature) of a panel of yields are sufficient to describe and forecast the dynamic of the whole cross-section.

Therefore, fit an AR(1) to each of the three PCs and make a forecast for the principal component based on the last known value (set α to zero). Then, obtain the yields based on the forecasted PCs and plot the expected yield curve.

How would you adjust your bond portfolio?