

# **SOLUTION MANUEL for Handling Big Data**

**prepared by: Prof. Dr. Maxim Ulrich, Spring 2022**

# 1 Forecasting using PCA

Suppose your aim is to predict the ECB yield curve, consisting of all euro-area bonds. The yield curve of interest consists of maturities  $[3/12, 6/12, 1, 2, 3, 5, 7, 10, 20, 30]$  years. You got access to the last  $T > 0$  observations of the yield curve.

The yields that you got access to are called spot rates. These are zero-coupon bond yields.

**Task 1: Write down an algorithm to extract all principal components of the yield curve. Also, highlight how to get the most important principal component.**

Declare a  $T \times 10$  matrix  $X$  of zero coupon yields:

$$\underbrace{X}_{T \times 10} := \left[ \underbrace{y^{(3/12)}}, \dots, y^{(30)} \right]_{T \times 1} \quad : 10 \text{ time-series of ZCB yields}$$

Define  $X^{(dm)}$  as a column-wise demeaned  $X$

$$X^{dm} := \left[ y^{(3/12)} - \bar{y}^{(3/12)} \times I_{T \times 1}, \dots, y^{(30)} - \bar{y}^{(30)} \times I_{T \times 1} \right] \quad : 10 \text{ time-series of demeaned ZCB yields,}$$

where  $\bar{y}^{(j)}, j \in [3/12, \dots, 30]$  is the sample mean of ZCB  $j$ 's yield.

We want to eigenvalue decompose  $\Sigma$ , the covariance matrix of ZCB yields,

$$\Sigma = \frac{1}{T} X^{dm'} X^{dm}$$

such that the resulting covariance matrix of principal components is diagonal. As the covariance matrix is symmetric it holds

$$\Sigma \equiv E \Lambda E' \quad : \text{Eigenvalue decomposition of } \Sigma,$$

where  $E$  is the  $10 \times 10$  matrix of eigenvectors (one eigenvector per column),  $\Lambda$  is the diagonal matrix of eigenvalues and  $E' \equiv E^{-1}$ .

Notice, we sort  $E$  and  $\Lambda$  in descending order. This means we sort such that column one of  $E$  coincides with the eigenvector for the largest eigenvalue while the first element of  $\Lambda$  is the largest eigenvalue, while the last column of  $E$  contains the eigenvector for the smallest eigenvalue and the last element of  $\Lambda$  coincides with the smallest eigenvalue.

The  $T \times 10$  matrix of (ordered) principal components (from most important to least important) coincides with  $PC(X)$ :

$$PC(X) := X^{dm} \times E.$$

The most important principal component,  $PC1(X)$ , with dimension  $T \times 1$  coincides with

$$PC1(X) := [PC(X)]_{[:,0]}$$

where as in Python, we start counting at 0.

### **Task 2: How much variance does the first principal component explain?**

The variance share of each principal component  $i$  coincides with the ratio of respective eigenvalue divided by the sum of all eigenvalues:

$$\frac{[\Lambda]_{[i,i]}}{\sum_{m=0}^9 [\Lambda]_{[m,m]}}; \forall i \in [0, \dots, 9].$$

The variance explained by the most important principal component does therefore coincide with

$$\frac{[\Lambda]_{[0,0]}}{\sum_{m=0}^9 [\Lambda]_{[m,m]}}.$$

### **Task 3: Forecasting AR(p) model**

Assume the first principal component explains 92% of the variance of each yield-to-maturity. You aim to predict this principal component forward using an AR(p) model.

The BIC test reveals 2 to be the optimal lag length.

Rely on fitting an AR(2) model to the first principal component to compute its forecasts for  $T + k, k > 0$ . Once you have the forecast, rotate from the principal component space back to the yield space. This says, compute forecasts of future zero-coupon bond yield with maturities  $[3/12, 6/12, 1, 2, 3, 5, 7, 10, 20, 30]$  where forecasts are only based on the forecast of the first principal component. Hence, write down an algorithm of how to map the AR(2) forecast of the principal component to a forecast of future zero-bond yields.

Let  $PC$  with  $\dim(PC) = T \times 10$  be the descendigly sorted martrix of principal components. It holds from eigenvalue calculus

$$\underbrace{PC}_{T \times 10} = \underbrace{(y - \text{mean}(y))}_{T \times 10} \times \underbrace{E}_{10 \times 10}$$

which rotates correlated demeaned yields into the space of orthogonal principal components.

An inverse PCA transform allows to rotate from the space of orthogonal principal components to the space of correalted demeaned yields. Notice,  $E' \equiv E^{-1}$ :

$$\underbrace{(y - \text{mean}(y))}_{T \times 10} = \underbrace{PC}_{T \times 10} \times E'.$$

An inverse PCA transform with only the most important principal component  $PC1$  reads:

$$\underbrace{(\tilde{y} - \text{mean}(\tilde{y}))}_{T \times 10} = \underbrace{PC1}_{T \times 1} \times \underbrace{[E']_{[0, :]}}_{1 \times 10}.$$

where  $\tilde{y}$  differs from  $y$  because  $y$  is made up of all principal components while  $\tilde{y}$  is made of of only  $PC1$ .

Hence, if we have AR(2)-based forecasts for  $PC1$ , denoted as  $E_t[PC1_{t+k}]$  for  $k > 0$ , then we have  $k$ -period ahead forecasts for  $\tilde{y}$  by means of

$$E_t[\tilde{y}_{t+k}^{(\tau)}] = mean(\tilde{y}^{(\tau)}) + E_t[PC1_{t+k}] \times \underbrace{[E']_{[0, :]}}_{1 \times 10}, \quad \tau \in [3/12, 6/12, 1, 2, 3, 5, 7, 10, 20, 30], \quad k > 0.$$