afa FÖRSÄKRING

Interest rate principal component analysis

(Co)advisor: Andreas Lagerås, Afa Försäkring (andreas@math.su.se)

Principal component analysis (PCA) of the covariance matrix of a yield curve, i.e. of interest rates for different times to maturity, typically shows that a large proportion of the total variance is explained by the first three eigenvectors. The three eigenvectors can be interpreted as a shift of the yield curve, a change of slope and a change in curvature, respectively. Note that the eigenvectors as seen as functions of time to maturity change sign 0, 1 and 2 times, repectively, when moving from the shortest to the longest time to maturity.

This project concerns in what sense this is a general phenomenon and how it can be interpreted probabilistically.

Let $X = (X_1, \dots, X_n)$ be an n-dimensional random vector with covariance matrix C whose eigendecomposition is

$$C = VDV'$$

with D a diagonal matrix with the eigenvalues in falling magnitude on the diagonal and V a matrix whose columns are the corresponing eigenvectors $v_1, \dots, v_n$.

Let $v_{(i)}(t)$ be a piecewise linear interpolation of the eigenvector v_i:

$$v_{(i)}(t) \coloneqq (k+1-t)v_{ki} + (t-k)v_{k+1,i}, \text{ for } k \le t \le k+1$$

The *nodes* of $v_{(i)}$ are the roots of the equation $v_{(i)}(t) = 0$ arranged in increasing order. We also say that two sequnces $s_1 < \cdots < s_m$ and $t_1 < \cdots < t_{m+1}$ are *interlacing* if $t_k < s_k < t_{k+1}$ for all $k = 1, \dots, m$.

We are interested in the case where $v_{(i)}$ has $i-1$ nodes for $i = 1, \dots, n$ and the nodes of $v_{(i)}$ and $v_{(i+1)}$ are interlacing for $i = 1, \dots, n-1$, since this is seen in data, at least for the eigenvectors corresponding to the largest eigenvalues. *What kind of covariance matrix C produces this pattern?*

There is a considerable litterature in Linear Algebra that concerns this question. A sufficient condition is that C is *totally positive*, i.e. that all its minors (determinants of square submatrices) are positive.

The focus for this project is not necessarily to produce novel results, but to give probabilitic interpretations, e.g. what kind of dependence and conditional independence produces this pattern?