Principal Component Analysis
Application to Yield Curves

Richard Diamond

Introduction

The Principal Component Analysis (PCA) is approach to determine dominant factors that drive a multivariate system. This is achieved by analysing the information about changes in random variables by decomposing their covariance matrix. The decomposition achieves three purposes:

- Decorrelation: factors are independent and their covariance is zero.
- Dimension reduction: a few (typically) three factors describe well above 90 % of variance of the system (remember the goodness of fit statistic, R^2).
- Data structure revealed: the analysis operates under the assumption of no hidden factors.

These advantages often come at the cost of loosing the information and difficulty to find the attribution. In quantitative finance, some see the Principal Component Analysis as a calibration exercise (i.e., factors and their loadings change and tell nothing), but the systematic application of the analysis allows to see its predictive powers. For example, rotation of factors reveals the change in behaviour of market participants or cyclicality.

In this note we touch on PCA application to the forward curve analysis on order to provide some context. The evolution of the rate for a particular maturity (tenor) will be a risk factor. Dominance of a few factors means that, for example, changes in 1Y, 5Y or 25Y rates are likely to move the entire curve.[1]

[1] Rates with similar tenors, i.e., 4Y and 5Y move in a highly correlated way.

Eigenvectors and Eigenvalues

Orthogonal decomposition applies to the covariance matrix of changes (differences or log-differences). Eigenvectors are orthogonal to one another representing directions of minimal variance for the original data 'cloud'.

- The eigenvalues λ_i are simply the variances of movement along the eigendirections.
- The eigenvectors $\boldsymbol{e^{(i)}}$ are orthogonal projections of the changes in data in directions of minimum variance, defined to the nearest sign (see Inverted Signs section below).[2] Being orthogonal means that their covariance is zero: $\text{Cov}[\boldsymbol{e^{(i)}, e^{(j)}}] = 0$. This property is the strength as well as weakness of the analysis.

*Orthogonalisation allows making projections by adding up principal components in linear fashion:

$$\boldsymbol{\hat{f}_{t+\tau} = f_t + e^{(1)} X_1 + e^{(2)} X_2 + e^{(3)} X_3}$$

Where one standard deviation for each random variable (factor) X_i is $\sqrt{\lambda_i}$.

$$\mathbf{d\bar{f}(t, \tau_j) = \cdots + \sum_{i=1}^{k} \sqrt{\lambda_i} e_j^{(i)} dX_i} \quad \text{HJM projection}$$

Covariance Matrix

$\boldsymbol{\Sigma}$ is the covariance matrix of all available maturities $d\bar{f}(t, \tau_j)$. Because of its symmetry, it can be decomposed (factorised) according to *the spectral theorem*:

$$\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} = \mathbf{V \Lambda V'}$$

- $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues ranked and positive $\lambda_1 > \cdots > \lambda_n \geq 0$ (off-diagonal elements are zeroes).

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

- $\mathbf{V}$ is a vectorised matrix with eigenvectors $Vec(\mathbf{e^{(1)}}\ \mathbf{e^{(2)}}\ \ldots\ \mathbf{e^{(n)}})$ as columns.

Key properties:

- Multiplying $\mathbf{VV' = I}$ gives an identity matrix – this applies to *normalised* eigenvectors.
- Projection does not need the original matrix $\mathbf{\Sigma e^{(i)} = \lambda_i e^{(i)}}$.
- Confirmation of zero covariance: $\text{Cov}[\boldsymbol{e^{(i)\prime} X,\ e^{(j)\prime} X}] = \boldsymbol{e^{(i)\prime} V \Lambda V\, e^{(j)\prime}} = [\boldsymbol{\Lambda}]_{ij} = 0$.

*The fact that covariance of eigenvectors is zero makes it difficult (if not impossible) to create analysis (strategies) that are based on simultaneous changes in several eigenmodes. This is a problem in the light of the empirical phenomena of eigenvector rotation, for example.

[2] In fact, the eigenvector represents a 2D plane (an eigenmode) onto which a projection is made.

Significance of a factor

We would like to identify a few factors k that drive the evolution of the whole curve. This is easy to do, just take eigenvectors with the largest corresponding eigenvalues. Given that **eigenvalue λ_i is variance** of the movements in each eigendirection, the first factor explains

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_N}$$

The goodness of fit statistic for the k-factor model is

$$R^2 = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

Factor Attribution

PCA operates under the assumption that all factors can be made explicit, i.e., there are no hidden factor(s) U represented as $X_1(U), X_2(U)$, etc. It follows that eigenvectors have context-specific and unique interpretation.

Uncertain Attribution: if principal components have unclear attributions (i.e., they do not match their roles as parallel shift, skew, and 'twist' of the curve), then you might have to reduce the time period. For example, if you mix the data from the regime of the higher rates and their volatility (pre 2008) and low rates (2010- mid 2013) the input does not make for a good model and the eigenvectors would not behave well – there is no consistent data structure to reveal by PCA and the model becomes a 'garbage in, garbage out' story:

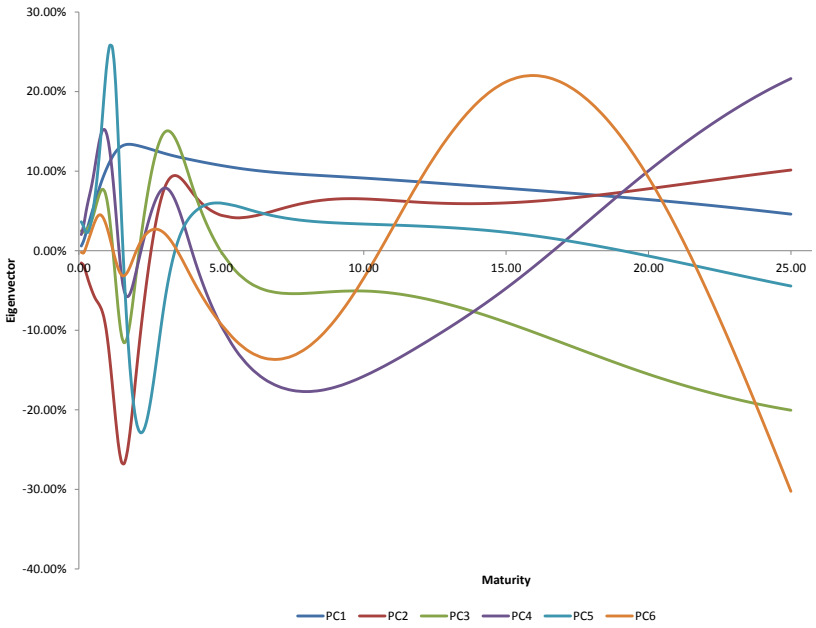

Figure 1: Principal Components with Uncertain Attribution

The illustration should help you to identify such situation: notice how the parallel shift that is typical for PC1 dominates through to the second eigendimension, while higher principal components replicate each other serving as a noise. Avoid the 'regime-switching' period (e.g., mid 2008–mid 2009 in your sample), otherwise the task becomes of research difficulty to study the relationship between curve regimes (low/high rates, normal/inverted curve) and internal structure that covariance suggests.

High vs. low rates regime: HJM lecture example has been done for high interest rate regime (pre-2008), under which it was fine to use PCA on rate differences, but since after 2008 we are in the low rates regime and it is best to conduct the PCA on log-differences in forward rates: $\log f(t+\Delta t) - \log f(t)$ for each column (that has time series for a particular tenor).

Numerical Methods for PCA

Understanding and implementation of numerical methods require advanced matrix algebra.

- The simplest method to estimate eigenvalues and eigenvectors is the Power Method. It relies on successful identification of largest eigenvalue first.
- More advanced is the family of QR decomposition methods, some of which allow computing eigenvalues and eigenvectors in one go. However, these methods require preliminary advanced transformations of a matrix into a three-diagonal matrix (Housesholder Transform).
- Jacobi Transformation method balances analytical tractability with slow convergence (compensated by modern computing power) therefore, the method can be implemented as an exercise.

  The idea is to get the diagonal matrix of eigenvalues by eliminating the largest off-diagonal element on each iteration (rotation). That 'orthogonal rotation' is implemented by premultiplying matrix that which we want to decompose by a matrix with trigonometric functions. The concept of the angle of rotation comes in. Eigenvectors are estimated through keeping the history of rotations.

  The number of rotations (Jacobi Transformations) can be high. Each new rotation destroys zero result for the particular off-diagonal element obtained on the previous step. Therefore, the convergence of the Jacobi Method is commonly compared to other numerical methods.

Jacobi Transformation represents a balance between being tractable and computationally efficient. Please see CQF HJM Lecture Solutions for the description of steps and properties of the Jacobi Transformation. Power method to calculate eigenvalues by one, starting with the largest, is also simple to present (see Chapter 37.13 in Volume 2 of PWOQF). Other matrix decomposition methods (including non-spectral) can suit the task and work much faster, in particular, see Cholesky decomposition applicable if the matrix is positive definite.

Issues with PCA and its Numerical Methods

The same *orthogonal basis* can be represented by several combinations of eigenvectors that are linearly dependent. Therefore, numerical methods of matrix decomposition can produce seemingly different results. A few checks must be made:

- All eigenvalues $\lambda_1 > \cdots > \lambda_n >= 0$ *must be positive and ranked.*
- Some decomposition methods or their implementations, particularly QR reduction, do not *standardise eigenvectors.* L2 norm (sum of elements by absolute values) of a standartised eigenvector is equal to one – dropping superscript (i) for each eigenvector $\sqrt{\sum_{k=1} \mathbf{e}_k^2} = 1.$

  Standardisation can be applied to the covariance matrix (i.e., before the decomposition). One advice is to standartise the matrix around a constant, such as the realized sample standard deviation over a given two-year period. Because the actual standard deviation of a Normal random variable (such as, residual) is not equal to one. By standartising the covariance matrix around a small constant you are not constraining estimates, such as portfolio allocations obtained from the matrix. Standardisation makes the variables expressed to the similar order and thus, any multivariate estimation will be more stable numerically.
- *Eigenvectors are only defined to the nearest sign* $\mathbf{e}^{(i)} \equiv . - \mathbf{e}^{(i)}$ where dot denotes the elementwise operation (same as MATLAB). In the practice of PCA application, it is common to observe two eigenvectors 'exchanging' signs. HJM implementation implications:
  - Inverted sign major components, such as PC2 vs. PC3 leads to the opposite attribution (e.g., negative vs. positive skew for a curve).
  - HJM drift calculation is invariant to the sign because

    $$\bar{\nu}_i(t,\tau) \times \int_0^\tau \bar{\nu}_i(t,s)ds = -\bar{\nu}_i(t,\tau) \times - \int_0^\tau \bar{\nu}_i(t,s)ds$$

    An eigenvector's sign matters for the diffusion term $\bar{\nu}_i(t,\tau)dX_i$, but the random Normal variable that can be positive or negative. The HJM model's drift is incrementally increasing and for longer maturities, should be much larger than diffusion.