Slides next offer **PCA Tutorial**

for **Data Analytics on Yield Curve**.

We start with a case for multi-factor modelling of the yield curve, and review the PCA side of implementation.

Case for a multi-factor model

A single-factor model for the short rate $r(t)$ can't hope to capture the richness of yield curve movements.

Consider *a spread option*. Its payoff is the difference between rates at two different tenors, e.g., $(L_{6M} - L_{3M})$.

- **If movements of two rates are not correlated (going up/down in sync), there is an extra source of risk**, ie, another factor.

  Instrument is sensitive to more than one factor.

  Cannot hedge with a single bond.

Bucket risks vs. Curve movement

Consider two 'natural' bucket risks, some short rate $f_{0.08Y}$ and a long-term rate f_{7Y},

the common risk methodology will study the CVA or derivative price *wrt* change at a single bucket.

However, if rates at $0.08, 7$ tenors move in the opposite direction, they represent another kind of curve movement, <u>one systematic factor</u>:

- steepening or flattening of the curve.

$$Corr[\Delta f_{3Y}, \Delta f_{5Y}], Corr[\Delta f_{3Y}, \Delta f_{10Y}], ... \gg 0$$

Interest rate *changes* are well-correlated across distant tenors (except *wrt* the short end).

But the covariance of changes $\boldsymbol{\Sigma(\Delta f_j, \Delta f_{j+h})}$ can be explained with a few independent factors.

There follows a possibility to represent the change about tenor τ_j as a linear decomposition of 'orthogonal' (independent) changes:

$$\Delta f_j = \text{PC}_1 + \text{PC}_2 + ... + \text{PC}_k$$

The linear components come from the **Principal Component Analysis**.

- Systematic factors that describe movement of a curve as a whole.
- Factor attribution is well-established for yield curve analysis.

HJM SDE (9), each **volatility function** is equal to the scaled principal component *i*.

$$\bar{\nu}(\tau) = \text{Std Dev} \times \text{Eigenvector}_\tau \qquad \text{or} \qquad \sqrt{\lambda_i}\, \boldsymbol{e}_{\tau}^{(i)}$$

volatility structure matches the data and calibration is fully numerical.

Data Preparation

If we have time series of each rate going back a few years, we can calculate covariances between **changes** in the rates.

- Inst. forward rates from BOE Yield Curve Statistics (use BLC)
- $\tau = 0.5$ increment for $0.08Y$ ... $25Y$ gives 50 columns.
- Jan 2002 – Jan 2007 regime. Consider regimes since then.

Government Liability Curve (GLC) bootstrapped from repo agreements, spot bonds (Gilts) and bond futures.

Bank Liability Curve built from short sterling futures, and FRAs. It is more suitable for pricing IR derivatives.

Covariance Matrix – Changes Percentage Rates

To estimate the covariance matrix $\mathbf{\Sigma}$,

1. Compute **daily differences** in fwd rate at each tenor, columnwise. Subtract the mean, if not a small quantity.

   ```
   B3-B2 - AVERAGE(B), B4-B3 - AVERAGE(B), ...
   5.7680 - 5.7733, 5.7757 - 5.7680
   ```

2. $\mathbf{\Sigma} = \frac{1}{N}\boldsymbol{XX'}$ where $\boldsymbol{X}$ relates to the dataset of Differences, see the tab in *HJM PCA.xls*.

3. We annualise covariance, and correct for the fact that we used percentages, eg, 5.768 not 0.05768 when computing differences,

$$\times \frac{252}{100 \times 100}.$$

Matrix Decomposition

$\mathbf{\Sigma}$ is the covariance matrix of fwd rate **changes**. Such symmetric matrix can be decomposed according to *the spectral theorem*:

$$\mathbf{\Sigma} = \mathbf{V\Lambda V'}$$

- $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues $\lambda_1 > \cdots > \lambda_n > 0$ positive and usually ranked in software output (Matlab, R).

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

- $\mathbf{V}$ is a vectorised matrix of eigenvectors $vec(\mathbf{e}^{(1)}\ \mathbf{e}^{(2)}\ \ldots\ \mathbf{e}^{(n)})$.

$$\Delta f(\tau_j) = \sqrt{\lambda_1}\boldsymbol{e}^{(1)}_{\tau_j} + \sqrt{\lambda_2}\boldsymbol{e}^{(2)}_{\tau_j} + \sqrt{\lambda_3}\boldsymbol{e}^{(3)}_{\tau_j} + \ldots \quad \text{in rows}$$

Using PCA output, we express a system of HJM SDEs in matrix form,

$$d\boldsymbol{f}(t,T) = \boldsymbol{M}(t,T)dt + \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\,d\boldsymbol{X} \tag{12}$$

where $d\boldsymbol{X}$ is a multi-dimensional Brownian Motion representing k independent factors.

Independence is achieved by decomposition of covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)' = \boldsymbol{A}\boldsymbol{A}' \quad \text{Cholesky decomposition}$$

The covariance matrix is estimated from <u>**changes in forward rates**</u>

$$\boldsymbol{\Sigma} = \mathbb{C}ov[\Delta f(\tau_j),\, \Delta f(\tau_{j+h})].$$

Eigenvectors

Volatility functions of τ

- For each *column* eigenvector $\mathbf{e}^{(\mathbf{i})}$, the first entry is the movement of one-month rate ($\tau = 0.08$), the second entry is of the six-month rate ($\tau = 0.5$) and so on.

$$\overline{\nu}_i(t^*, \tau) = \sqrt{\lambda_i}\, \boldsymbol{e}^{(\boldsymbol{i})}_{\boldsymbol{\tau}} \tag{13}$$

  To obtain a volatility function, it is naturally convenient to fit a column eigenvector to tenor τ. Eigenvector $\boldsymbol{e}^{(i)}_{\tau}$ has values at

$$\tau = 0.08Y,\ 0.5Y,\ 1Y,\ \ldots,\ 25Y$$

- Instead of picking numbers from the matrix of eigenvectors $\mathbf{V}$, **we use the fitted volatility functions**.

Polynomial Fitting

- The fitting is done by a single **cubic spline** *wrt* tenor τ

$$\overline{\nu}(t,\tau) = \beta_0 + \beta_1\tau + \beta_2\tau^2 + \beta_3\tau^3 \quad \forall\,\tau_j$$

  A spline is a piecewise-defined smooth polynomial function.

  In general, we can fit exactly using a piecewise polynomial – here, an improvement can be made by enquiring into fitting and methods behind functions like *polyfit()* in Matlab and *nls()* in R.

  Fitting recipes are domain-specific for yield curve, implied vol. under LMM/LVM/SABR. Principal components are polynomials, eg, the best PC4 fit requires τ^4.

Despite using *LINEST()* to calculate β, here we are **not** conducting any regression analysis.

In our PCA application to Pound Sterling curve, three factors explain 93.33% of movement (variation) in the yield curve.

Tenor	λ	Cum. R^2
1Y	0.002027	71.31%
25Y	0.000463	87.58%
6Y	0.000164	93.33%

But how did we choose these $k = 3$ eigenvectors to be our volatility functions?

$$\boldsymbol{e}^{(1Y)},\ \boldsymbol{e}^{(6Y)},\ \boldsymbol{e}^{(25Y)}$$

By the largest corresponding eigenvalue.

Factor Significance

Eigenvalue λ_i **is variance** of the movements of a curve in each eigendirection. For example, the first factor explains

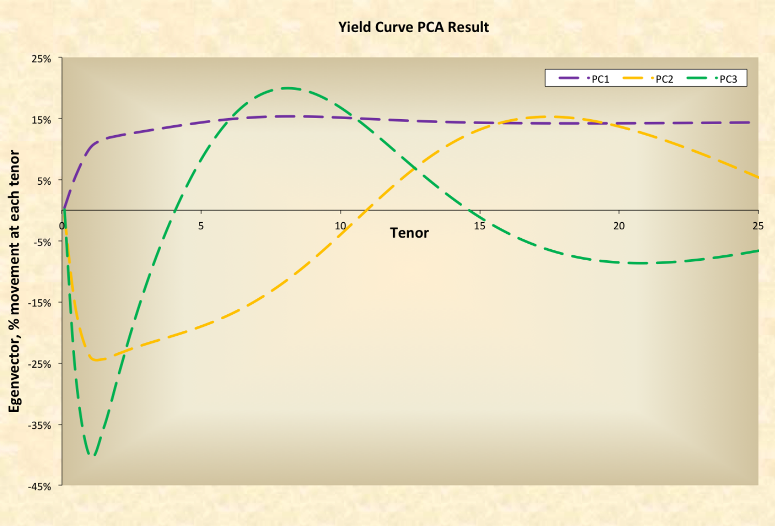
$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_N}$$

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

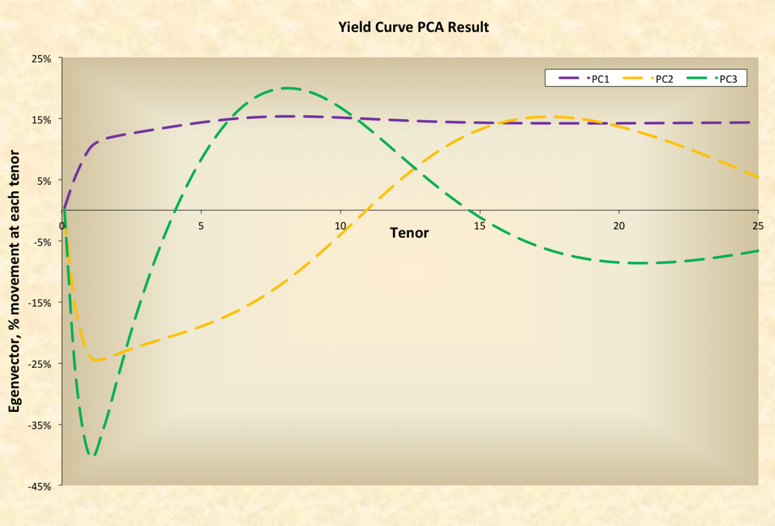
$$\text{Cum. } R^2 = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

By choosing the largest-impact factors we reduce an N-dimensional model to the three-factor model. Each factor represents systemic movement by the curve.

PCA Result: Three Largest Factors



PCA Result: Three Largest Factors



Factor Attribution

- **Parallel shift in overall level of rates** is the largest principal component of forward curve movement, common to all tenors.
- **Steepening/flattening of the curve** is the second important component (i.e., change of *skew* across the term structure)

  Inverted curve (backwardation for commodities term structure) would have a different shape for the PC2.
- **Bending about specific maturity points** is the third component to curve movement that mostly affects *curvature* (convexity).

Disclaimer. These are commonly accepted attributions but please read next.

PCA Factors and the Curve

Changes in the rate at certain tenor 1Y, 6Y, 25Y can be particularly sensitive to a systematic factor **BUT** causality is not proven, and eigenvectors tend to rotate particularly for PC2 vs. PC3 and above.

Short end of the curve has low correlation of changes with the longer tenors. Short end is sensitive to and often represents PC1.

In the current regime of low interest rates (to 2017) and flattened curve, PC2, PC3 components might have no attribution and rotate often. PCA is a limited tool for periods of rapid shifts in interest rates.

END OF TUTORIAL